Reshaping Adolescents' Gender Attitudes: Evidence from a School-Based Experiment in India*

Diva Dhar
University of Oxford

Tarun Jain
Indian Institute of Management Ahmedabad

Seema Jayachandran
Northwestern University

July 15, 2020

Abstract

Societal norms about gender roles contribute to the economic disadvantages facing women in many developing countries. This paper evaluates a school-based intervention in India that engaged adolescents in classroom discussions about gender equality for two and a half years with the goal of eroding their support for restrictive gender norms. Using a randomized controlled trial, we find that the program made attitudes 0.18 standard deviations more supportive of gender equality, or, equivalently, converted 16% of participants' regressive views. In addition, self-reported behavior became more aligned with progressive gender norms, particularly among boys. The effects observed in the short run were still present two years after the program had ended.

*Contact information: diva.dhar@bsg.ox.ac.uk, tj9d@virginia.edu and seema@northwestern.edu. We thank our partners, Breakthrough and the Government of Haryana, for collaborating on the project. We thank Sachet Bangia, Maaike Bijker, Srijana Chandrasekharan, Rachna Nag Chowdhuri, Alejandro Favela, Vrinda Kapoor, Vrinda Kapur, Lydia Kim, Akhila Kovvuri, Saumya Mathur, Suanna Oh, Priyanka Sarda, Ananta Seth, Niki Shrestha, Anantika Singh, and Rachita Vig for excellent research assistance and research management, the J-PAL survey staff for collecting the data, and Alice Eagly for helpful suggestions. An earlier version of this paper was submitted for pre-publication re-analysis to the Abdul Latif Jameel Poverty Action Lab (J-PAL), where a code replication exercise was conducted. We thank Andreas de Barros for conducting this re-analysis. We are also grateful to the J-PAL Post-Primary Education Initiative, International Initiative for Impact Evaluation, and Bill and Melinda Gates Foundation for being the primary funders of the data collection. We also thank the International Growth Centre and Northwestern Global Poverty Research Lab for additional small grants. The trial is registered in the AEA RCT Registry (AEARCTR-0000072) and received institutional review board approval from Northwestern University and the Institute for Financial Management and Research.

1 Introduction

Gender inequality exists in every society but is especially severe in many developing countries: Women and girls have fewer educational opportunities, less autonomy over marriage and fertility, and more restrictions on labor force participation (Duflo, 2012; Jayachandran, 2015). Many of these gender gaps are stagnant despite economic progress, suggesting that economic development on its own is unlikely to close them.

Recent work has drawn economists' attention to the importance of cultural norms in perpetuating gender gaps (Alesina et al., 2013; Bertrand et al., 2015). Studies suggest that even centuries-old norms might be amenable to change. For example, reserved seats for female politicians reduce gender-biased attitudes in India (Beaman et al., 2009), and television programming can change fertility preferences (Jensen and Oster, 2009; La Ferrara et al., 2012). These findings are part of a broader literature on how people's preferences are shaped (Becker and Mulligan, 1997; Bowles, 1998).

In this paper, we focus on a direct attempt to reshape gender attitudes through discussion and persuasion. We evaluate a school-based program in the state of Haryana, India, for seventh to tenth graders. The program centered around classroom discussions about gender equality in secondary schools, with a 45-minute session held every three weeks for two and a half school years. The sessions taught facts and endorsed gender equality and, as importantly, prompted students to reflect on their own and society's views. Discussion topics included gender stereotypes, gender roles at home, girls' education, women's employment outside the home, and harassment. A few sessions taught communication skills to help students convey their views to others and be able to, say, persuade their parents to permit them to marry at a later age. The program's messaging combined a human-rights case for gender equity with pragmatic reasons to value women, such as their economic contributions.

The intervention was designed and implemented by Breakthrough, a non-profit organization with extensive experience in gender-equality programming. The Government of Haryana allowed Breakthrough to conduct the classes in schools as part of the regular school day. The government was interested in eroding some of the restrictive gender norms in its society, and schools offer governments a powerful platform to shape the next generation's views. The reason for targeting secondary school students specifically is that adolescence is a critical time in the development of morality and formation of identity, when people are young enough to still have malleable attitudes but mature enough to reflect on complex

moral questions (Kohlberg, 1976; Markus and Nurius, 1986).[1]

As context, gender inequality is pronounced in India. While boys and girls start secondary school at the same rate, only 0.80 girls enroll in tertiary schooling for every boy (World Bank, 2011). Early marriage and childbearing are common, and many women have limited agency and physical mobility and face barriers to labor force participation (Field et al., 2010; Klasen and Pieters, 2015; Afridi et al., 2018; Calvi, 2020). India's sex ratio among children age 0 to 6 years is 1.09 boys per girl, reflective of the widespread practice of sex-selective abortion; Haryana's sex ratio of 1.20 is the most male-skewed among Indian states (Sen, 1990; Jha et al., 2006; Govt. of India, 2011).

We implement a randomized controlled trial across 314 government secondary schools in four districts of Haryana, with data collected from roughly 14,000 students. We measure effects a few months after the program ended and again two years later when the students are, on average, 17 years old.

We focus primarily on how the program changed participants' gender attitudes, that is, their views about what is right and wrong or desirable and undesirable, such as whether it is wrong for women to work outside the home and whether it would be good to have more women in politics. A second outcome is girls' aspirations for their education and career; for this outcome, we hypothesize an effect only on girls. Our third primary outcome is self-reported behaviors influenced by gender norms. It will only be once the participants become older that we can assess impacts on major outcomes they have control over, such as their employment and childbearing. Nonetheless, we view effects on self-reported day-to-day behaviors as adolescents to be informative about whether reshaping gender attitudes translates into behavior change. We focus on behaviors that adolescents likely have some say over, such as chores done at home and interaction with opposite-gender peers. Even for these actions, parents' views will be a moderating factor; a girl's parents might deny her request that her burden of household chores be reduced, for example. Another reason that attitude change might be insufficient for behavior change (even in the long run) is a desire to conform to social norms. A boy who now believes that he should help out with chores might feel that the social sanctions for doing so would be too costly.[2]

[1]There were also practical reasons for targeting adolescents rather than younger students. The government was less comfortable with younger children discussing possibly sensitive topics like sexual harassment. Also, the government was interested in effects on "hard" outcomes such as higher education and female employment; with younger participants, there would be a longer delay until these outcomes were realized.

[2]We examine perceptions of social norms as a secondary outcome.

We find that the intervention made gender attitudes more progressive by 0.18 standard deviations in the short run (three years after baseline). The measure of attitudes is an index that aggregates several survey responses about support for gender equality. The effect size is equivalent to 16% of the cases in which a student held a gender-regressive view being converted to support for gender equality. We find a similarly large effect on attitudes in the medium run (five years after baseline).

The intervention also affected self-reported gender-related behaviors. An index of gender-equal behaviors increased by 0.20 to 0.23 standard deviations, in both the short run and medium run. However, there was no effect on girls' educational and professional aspirations, which were quite high to begin with. We find a marginally significant effect on a revealed preference measure of girls' educational intentions included in the second endline, namely completing a college scholarship application, but no effect on boys' and girls' likelihood of signing a public petition to end the dowry system.

We examine heterogeneity of the treatment effects by two pre-specified characteristics, the parents' gender attitudes and the student's gender. We find no evidence of heterogeneity based on parents' attitudes, but we find important differences in treatment effects between boys and girls. At the first endline, the effect size on attitudes is similar for boys and girls, but by the second endline, the effect is larger for boys. In addition, behavior change is more pronounced for boys in both the short and medium run. For example, boys report doing more chores, but girls do not report doing fewer, and boys have a larger increase in how much they encourage their older sisters to pursue college. With chores, the smaller effect for girls can be explained by the asymmetry in the outcome; greater gender equality means that girls seek a lower burden, which might not be granted, while boys can voluntarily help out more. But the fact that we also see heterogeneity by gender in support given to sisters suggests a broader phenomenon of girls facing more external constraints on their behavior. This pattern highlights that, because behavior change requires not just a desire but also an ability to change, the very fact of boys' and men's greater power in society makes it important to include them in interventions aimed at increasing girls' and women's power.

A key concern with self-reported outcomes such as attitudes is social desirability bias. In our case, the specific concern is an experimenter demand effect, i.e., more social desirability bias in the treatment group. People who participated in a program that explicitly tried to increase their support for gender equality might disingenuously express more gender-progressive views to present themselves in a good light to the surveyors. To address this,

we make innovative use of the Marlowe-Crowne social desirability scale, a survey module developed by social psychologists that measures a person's propensity to give socially desirable answers (Crowne and Marlowe, 1960). The module, which we included on the baseline survey, asks respondents if they have several too-good-to-be-true traits such as never being jealous of another person's good fortune and always being a good listener.

We find that respondents with a high propensity for social desirability bias express more support for gender equality overall for the sample, but importantly, this pattern is not more true for the treatment group than the control group. In other words, the positive treatment effect on self-reported attitudes and behavior is similar in magnitude for respondents with a low and high propensity for social desirability bias.[3] We view this analysis as an important check on the validity of our results, and think that this approach could be useful in a wide array of studies in which experimenter demand is a concern. The method allows one to test for bias for all outcomes (e.g., attitudes about dowry, women's work, sexual harassment), so it complements techniques such as list experiments and revealed-preference measures, which often must focus on a narrower set of outcomes due to logistical costs.

Our study contributes to the literature on endogenous preferences, specifically on the formation of gender-related preferences.[4] Besides political quotas (Beaman et al., 2009) and television (Jensen and Oster, 2009; La Ferrara et al., 2012), other factors that have been shown to make attitudes more gender-progressive include mothers' employment (Fernandez et al., 2004), having daughters or sisters (Washington, 2008; Healy and Malhotra, 2013), serving with women in the military (Dahl et al., 2018), having teachers who hold weaker gender stereotypes (Carlana, 2019), and having female role models (Porter and Serra, 2019).

Unlike most of the studies above on endogenous gender preferences, our research examines an intervention that intentionally changed preferences. It thus also sits within the literature on persuasion, or communication expressly designed to change preferences or beliefs (DellaVigna and Gentzkow, 2010). Much of the persuasion literature focuses on ways to influence consumer or political preferences. Closer to our work are studies on attitude change about intimate partner violence (Gupta et al., 2013; Abramsky et al., 2014; Pulerwitz et al., 2015; Green et al., 2020), racial minorities (Donovan and Leivers, 1993), immigrants

[3]The fact that fade-out of the effect on attitudes differs for boys and girls offers further suggestive evidence that the results are not driven by experimenter demand. Experimenter demand effects that change over time differentially by gender are possible, but seem unlikely.

[4]Recent work has also studied the formation of preferences toward different castes and toward children from poorer families in India (Lowe, 2020; Rao, 2019), Muslims in the United Kingdom (Alrababa'h et al., 2019), and racial minorities in the US (Carrell et al., 2019), among other topics.

(Hopkins et al., 2019; Grigorieff et al., 2020), and women in STEM (Moss-Racusin et al., 2018), as well as studies that shift people's perceptions of social norms about gender or about ethnic discrimination and violence (Bursztyn et al., 2020; Aloud et al., 2020; Paluck, 2009). Our study is also related to Cantoni et al. (2017), which finds that Chinese students taught with textbooks designed to convey pro-Communist messages express more pro-government views and skepticism of free markets when they are adults.

We also add to the rapidly growing literature on educational/training interventions aimed at increasing girls' and women's agency and opportunities in developing countries. Related work includes Bandiera et al. (2020) on female empowerment and livelihood training in Uganda, Buchmann et al. (2018) on empowerment training and financial incentives to delay marriage in Bangladesh, Ashraf et al. (2020) on negotiation skills training for girls in Zambia, Edmonds et al. (2019) on life skills training for girls in India, and McKelway (2020) on self-efficacy training in India. While most of these interventions focus primarily on imparting human capital – either traditional skills or positive psychological traits – to women and girls, the central (though not exclusive) aim of the intervention we evaluate is to influence participants' preferences, specifically their attitudes about restrictive gender roles. Our study is also relatively unusual in the literature on women's empowerment in that it focuses on both boys and girls.

2 Description of intervention

The project emerged from the Government of Haryana's interest in testing policies to narrow gender gaps. Breakthrough, a gender human rights organization that had developed other social change campaigns, designed an intervention aimed at influencing adolescent boys' and girls' views related to gender norms, and implemented it in government schools, with the government's permission.

The participants in the program were the cohorts in grades 7 and 8 in the academic year 2014-15 when the program launched. It ran from April 2014 to October 2016, so one cohort participated in the program in grades 7, 8 and half of 9, and the other cohort participated in grades 8, 9 and half of 10. Grades 7 to 10 have high enrollment and low dropout in Haryana, so the program could reach a large share of the underlying age cohorts and have limited attrition due to school dropout (significant dropout occurs after grade 10) (DISE, 2011).

The objective of the program, which was named *Taaron ki Toli*, or Legion of Stars, was to create awareness of gender-based discrimination, change dominant gendered perceptions and

promote gender-equitable attitudes, raise girls' aspirations, and provide tools to participants to translate attitude change and greater aspirations into behavior change. By changing fundamental gender attitudes, the program aimed to ultimately influence a wide range of behaviors related to female education, mobility, work, marriage and fertility, for both female participants and male participants' female family members (e.g., their future wives).

The program emphasized both human rights and pragmatic (i.e., instrumental) reasons for giving girls and women more opportunities. For example, it conveyed that equal opportunity for education is a universal human right. The hypothesis is that this rights-based message would increase how much participants value girls having access to higher education; a boy would get disutility from seeing his sister denied the same chance to attend college that he has. As an example of a pragmatic argument for girls' education, the intervention informed participants that outcomes for children improve when their mother is more educated. The hypothesis is that this information causes belief updating; girls want to stay in school longer, and both boys and girls will want to educate their daughters down the road.

To ensure that the intervention would be widely accepted, Breakthrough engaged with multiple stakeholders at the state, district, and sub-district levels, gathering input from various education officials, school principals, and teachers as they developed the program. This helped them design a program that might ultimately be integrated into the standard school curriculum.[5]

The program featured classroom sessions, each 45 minutes long, led by a Breakthrough facilitator. The regular teacher was welcome to stay for the sessions or leave the classroom and have a break. There were a total of 27 sessions spread over two and half years.[6] Breakthrough hired 15 facilitators, 13 of whom were male, to cover the 150 treatment schools. The facilitator visited each school roughly once every three weeks. Other elements of the program included a one-time training for one teacher per school, optional youth clubs, and school-wide activities such as street theater performances held about once a year.[7]

[5]Scale-up options include hiring special-purpose teachers to lead the sessions, each of whom covers multiple schools; having regular school teachers deliver the lessons; incorporating some of the content into textbooks; or continuing to have NGOs implement the program. The Government of Punjab plans to roll out the program to all of its upper primary schools in 2021, to be taught by regular teachers as part of the "moral science" curriculum.

[6]The program's total dosage was 20 hours, or about 23 hours adding in the school-wide assemblies. As comparisons, the negotiation program for girls in Zambia studied by Ashraf et al. (2020) consisted of six two-hour sessions (12 hours total); the safe space groups in Bangladesh evaluated by Buchmann et al. (2018) met for about 200 hours total over six months; and the empowerment and livelihood clubs in Uganda evaluated by Bandiera et al. (2020) were open five afternoons per week for two years (over 500 hours).

[7]The curriculum described was delivered to the two study cohorts, finishing in the middle of the 2016-7

The classroom sessions were discussion-based, with more student participation than is typical in secondary schools classes. Perhaps for this reason, and also because having a dynamic teaching style was one of the hiring criteria for facilitators, student reaction to the program was generally very positive. To complement the in-class material and encourage further reflection, the facilitators gave out some homework assignments such as writing stories and recording observations, and they encouraged students to talk to their family members about what they were learning. The sessions did not displace one specific subject like math or history; they crowded out a roughly even mix of material in other subjects. While the program could have harmed other learning by displacing instructional time, it is also possible that the discuss-and-debate style strengthened students' critical thinking and speaking skills, conferring value to them beyond the gender focus.

Discussion topics for the sessions included gender identity, values, aspirations, gender roles and stereotypes, and recognition and tolerance of discrimination. For example, one session focused on household chores. Students broke out into groups and listed whether males or females did various chores in their households. They then reconvened and discussed the answers. When the pattern emerged that women and girls did most of the chores, the facilitator asked why that was and whether it was fair. The class discussed why women cook at home, but men are cooks in restaurants, with the latter role earning more status in society. A few of the sessions aimed to impart skills such as public speaking, communication between the genders, and leadership, which could enable gender-equitable attitudes to translate into behavioral change. For instance, girls might be able to negotiate greater independence with their parents, leading to more freedom of movement in the short run and greater occupational choice in the long run. Through this curriculum, students explored gender identity and stereotypes, gained a better understanding of gender inequities and their consequences, understood their rights, and were encouraged to communicate and act on what they had learned. To map this to standard concepts used by economists, the intervention aimed to change students' preferences (i.e., their moral views on gender inequality); their factual beliefs (e.g., greater realization that restricting women's employment leaves money on the

school year. Breakthrough received additional funding and resumed activities in 59 of the 150 treatment schools in 2017-18, offering an extra module to the younger of our two study cohorts. This added 0.2 years of dosage on average (59/150 * 52% of sample in younger cohort * 1 year), or 2.1 additional sessions per participant. Breakthrough also initiated the full curriculum with new cohorts, 2 to 5 years younger than our study cohorts, in these schools. We do not expect spillovers from these younger children to our study participants to have added much dosage, especially since most of the older cohort had switched to a new school for grade 11 by then. We became aware of these additional activities in 2019.

table); and their skills (e.g., how to persuade their parents to let them go to college).

3 Study design and data

3.1 Experimental design

We designed and implemented a randomized evaluation of the gender attitude change program using a sample of 314 government schools across Sonipat, Panipat, Rohtak, and Jhajjar districts in the state of Haryana, India.[8] The randomization unit is the school.

The sample size of 314 schools was chosen to be able to measure the short- and medium-run effects of the program on gender attitudes, aspirations, and behavior, as well as long-term effects on educational attainment, occupational choice, marriage, and fertility that might emerge up to ten years after the program ended. There were 607 government-run secondary schools that offered grades 6 through 10 across the four districts. We first restricted attention to the 346 schools that officially enrolled at least 40 students in grades 6 and 7 combined, and then eliminated schools with low actual enrollment based on a preliminary visit. In cases where a village had more than one government secondary school, we chose at most one for the sample, to minimize the possibility of spillovers. Of the 314 schools, 59 enroll only girls, 40 enroll only boys, and the remaining 215 are co-ed. The schools have an average enrollment of about 80 students per grade (DISE, 2011).

We randomly selected 150 of the sample schools to be in the treatment group; the remaining 164 serve as control schools. Figure 1 shows the four study districts and the schools assigned to the treatment and control groups. The randomization was stratified by district, co-ed status of the school, school size, and distance to the district headquarters. Table 1 reports baseline characteristics of schools by treatment status. The first panel confirms that the two samples are balanced on various school characteristics such as urban/rural, number of male and female students, and number of teachers.

3.2 Enrollment of study participants and baseline data collection

The baseline survey was conducted between August 2013 and January 2014, covering 14,809 students. The data collection was conducted by the Abdul Latif Jameel Poverty

[8]These districts were the government's priority districts for addressing gender inequality based on their skewed sex ratios. The 2011 child sex ratio was 1.25 in Sonipat, 1.19 in Panipat, 1.22 in Rohtak, and 1.28 in Jhajjar. These districts do not necessarily have more male-biased preferences than the rest of Haryana, or the rest of north India, but their low fertility rate (they are near New Delhi) means that son preference translates into a higher rate of sex-selective abortions to ensure having a son (Jayachandran, 2017).

Action Lab, South Asia.

To select students for the sample, we visited the sample schools and distributed parental consent forms to all 6th and 7th graders who were present.[9] There was a high rate of parental consent; 84% of forms were returned. The rate of returned consent forms does not vary by gender or by village-level proxies for gender norms, namely the child sex ratio and female employment rate from the 2011 Census. Anecdotally, lack of consent was usually due to the student losing or forgetting the form.[10]

We randomly chose students to enroll from among those whose parent gave consent and who personally assented to participate, with a target of 45 students per school, stratified by gender and grade with a ratio of 3:2:2:2 for Female 6th:Male 6th:Female 7th:Male 7th. We included more girls because more girls than boys are enrolled in government schools, and we sampled more grade 6 girls than grade 7 girls because we expected lower attrition for younger grades.[11] To be in the study, the student also needed to be at school on the baseline survey day. The 35-minute-long baseline survey took place on the school premises.

We mistakenly omitted one school from the baseline survey. This school was randomized into the treatment group, and the intervention was conducted in it. We collected endline data in the school and include it in the analysis, imputing baseline variables with the gender-specific sample average for the district.[12]

To understand how parental attitudes influence program impacts, one parent of a random 40% subsample of the surveyed students participated in a short survey at the student's home ($N = 6,022$); we selected at random whether to interview the father or mother. Budget constraints are why we visited only 40% of households and surveyed one parent per household. We construct a gender attitude index for the parent based on nine attitude questions asked of the parents, which we use in heterogeneity analyses.

Table 1 summarizes baseline characteristics of the sample. The boys and girls were about 12 years old, on average. Religious and caste variables line up with the overall demographics

[9]These students would be in grades 7 and 8 in spring 2014 when the program began. The school year begins in April.

[10]Parents had to consent to their child participating in the study, but the program was added to treatment schools by the government, who did not offer parents a way to opt out. The classes were not held on a fixed day, so keeping a child home to avoid the classes would not have been straightforward. Anecdotally, parental complaints about the program to schools and the education department were negligible.

[11]Parents are more likely to send their sons than daughters to private schools. Because wealthier families use private schools, the boys in government schools are, thus, from poorer families than the girls, on average.

[12]We distributed consent forms to students present during a school visit just before the endline survey and then randomly chose sample students from among those with parental consent who assented to participate.

for these districts, as reported in the Census (Govt. of India, 2011); the participants are predominantly Hindu. Mothers' average age was 36 years and fathers', 41 years. There is a high illiteracy rate for mothers, reflecting the low level of female schooling in the parents' generation. A small proportion (29%) of mothers work outside the home, which is consistent with low female labor force participation rates in India. Baseline variables are balanced between the treatment and control groups. An F-test of joint significance fails to reject balance between the study arms.

The baseline survey also measured the respondent's propensity to give socially desirable responses. We administered the Marlowe-Crowne module, designed by social psychologists to measure a person's concern for social approval (Crowne and Marlowe, 1960). We used a 13-item validated version of the module (Reynolds, 1982). The questions, which are listed in the data appendix, ask the respondent whether he or she has certain almost saintly personality traits. While it is possible some people are saintly, the module uses traits that are too extreme for most people to hold (e.g., "I am never irritated by people who ask favors of me"). We use respondents' social desirability score (i.e., the count of very positive traits they say they have) to investigate whether the treatment effects we estimate are biased upward by experimenter demand effects.

3.3 Endline data collection

We conducted a first endline survey shortly after the program ended, or about three years after the baseline survey. We then conducted a second endline two years later.

Data collection for the first endline survey occurred between November 2016 and April 2017. We resurveyed 13,943 of the 14,809 students surveyed at baseline, which corresponds to an attrition rate of 5.8%. The endline sample also includes an additional 44 students from the sample school that we mistakenly did not survey at baseline, yielding a total sample for the first endline of 13,987 students.

Appendix Table 1 shows that sample attrition does not differ significantly between the treatment and control groups. In addition, attrition in the treatment versus control group is not differential by baseline variables, except for a small association with baseline behavior conditional on the other baseline outcomes.[13] Most respondents (76%) were surveyed at

[13]For the sample overall, girls' aspirations are negatively correlated with attrition; girls with high aspirations were more likely to still be enrolled in school, which made them easier to track. Appendix Table 2 details the reasons for attrition, which include permanent or long-term migration, death or poor health, refusal to participate by the student or parent, not being available at the time of their appointment, and our inability to track the respondent.

school. Several students had moved to a different school, either in the same or a different village, or dropped out of school. These students were surveyed at home (24% of respondents). If the student had moved to another village that was far from the survey districts, we conducted a truncated phone survey (0.1% of respondents).[14]

We conducted a second endline survey more than two years after the intervention had ended, from January to July 2019, when the students were finishing or had just finished grades 11 and 12 (if they had not repeated a grade). The survey was conducted in students' homes, rather than at schools; the highest grade offered in most of the sample schools was grade 10, so the participants were scattered across various schools (or had dropped out). We conducted truncated phone surveys for those who had moved (3% of respondents). The attrition rate is 7.9%, and attrition is not significantly correlated with treatment status or correlated with baseline characteristics differentially by treatment status. The sample size for the second endline is 13,685 individuals.

3.4 Primary outcomes: Attitudes, aspirations, and behavior

We pre-specified three primary outcomes for the first endline: gender attitudes, girls' aspirations, and self-reported gender-related behavior.[15]

We combine responses to 18 gender attitude questions into a variance-weighted index, following Anderson (2008). The specific variables and procedure for constructing the index were pre-specified (see the data appendix for more details). We measure gender attitudes through direct questions about female and male roles and rights (e.g., whether women should work outside the home, the appropriate age of marriage for girls). We also measure attitudes via questions about a vignette about investing in a son's or daughter's education.

Gender attitudes are balanced between the treatment and control group at baseline (see Table 1).[16] They are also quite regressive. For example, about 80% of boys and 60% of girls

[14]Appendix Table 1 shows that treatment status is not significantly correlated with the survey location for the endline. Appendix Table 3 summarizes participants' schooling status at endline: 82% were enrolled in the same school as baseline, 8% had dropped out of school, and the remainder had changed schools. The table also reports that 87% of the treatment group were aware of the program activities.

[15]The pre-analysis plan (PaP) for the first endline survey was posted to the AEA RCT Registry at the beginning of endline data collection in November 2016. It specified the primary outcomes and how they would be constructed, the secondary outcomes, heterogeneity analyses, and the procedure for choosing control variables. The PaP for the second endline was posted in December 2018 and specified similar information. The two PaPs and a four-page document listing and explaining the deviations we made are available at bit.ly/PaP4RAGA.

[16]The gender attitudes module at baseline was shorter than the endline module. The baseline attitude index aggregates nine variables.

believe that a woman's most important role is being a good homemaker. The pattern that girls are less likely than boys to endorse gender-discriminatory positions is seen for each of the attitude questions (see Appendix Table 4).

We measure girls' aspirations with a variance-weighted index combining 5 questions about educational and career aspirations. We also collected data on boys' aspirations and note that aspirations are quite similar for boys and girls at baseline.

Arguably the most challenging outcome to measure was self-reported behavior. We focused on behaviors that are influenced by gender attitudes and norms and that we expected adolescents to have some say over. We construct an index of 6 questions, asked of both boys and girls. Most of these questions are coded the same way for boys and girls (e.g., being supportive of sisters' career aspirations, level of interaction with the opposite gender), whereas household chores is coded so that, for girls, more gender-equal behavior is to do fewer chores whereas, for boys, it is to do more. We also included some questions applicable to only girls (e.g., mobility), which we do not include in the main index but examine in auxiliary analyses.

For the second endline, we again pre-specified gender attitudes, girls' aspirations, and behavior as primary outcomes. We construct the attitudes index identically between the first and second endlines, using the same questions and weights (which are based on the first endline data). This makes the magnitude of the effect directly comparable across the two waves. For aspirations and behavior, we updated the modules, as the relevant questions changed as the sample grew older, so we construct the variance-weighted indices independently for the two endlines.

We included two additional primary outcomes in the second endline. The first is a revealed-preference measure of girls' aspirations for or expectation of attending college. We set up a girls' scholarship program for college expenses and use as an outcome whether respondents filled out and mailed in the application, which we gave to them at the end of the survey visit. The second is a revealed-preference measure of attitudes, more specifically, willingness to publicly espouse a feminist position. We informed respondents about a petition to end the dowry system, with the names of signatories to be published in the local newspaper (through an advertisement we placed). Students were given a toll-free phone number to call to add their name to the petition. We use signing the petition as an outcome.

4 Empirical specification and results

The intervention is hypothesized to make participants' attitudes less discriminatory against females, raise girls' aspirations, and increase gender-equitable behavior. This section describes the estimation strategy used to test these hypotheses and then presents the results.

4.1 Specification

We estimate the following ordinary least squares regression, with one observation per student:

$$Y_{ij} = \beta_0 + \beta_1 Treated_j + \beta_2 Y_{ij}^0 + \beta_3 \mathbf{X}_{ij} + \epsilon_{ij} \quad (1)$$

Y_{ij} is the outcome variable measured at endline for student i in school j. $Treated_j$ is a binary variable that equals 1 if the school was assigned to the treatment group, and 0 otherwise. Thus, β_1 represents the average effect of the intervention on the outcome. The outcomes are constructed so that a higher value represents more gender progressiveness, so the hypothesis is $\beta_1 > 0$.

We control for Y_{ij}^0, the baseline analogue of the outcome. The vector $\mathbf{X}_{ij}$ comprises other control variables. In our basic specification, these are grade-gender and district-gender fixed effects and the school-gender mean of Y_{ij}^0. When the outcome is an index, we also include a missing flag for each component of the index.[17] We also estimate an enhanced specification that controls for additional baseline student, parent, and school characteristics chosen using LASSO following Belloni et al. (2014). Appendix Table 5 lists the extra control variables selected for each outcome and the larger set of variables from which the LASSO procedure chose them. We allow the error term, ϵ_{ij}, to be clustered at the school level.

We also test for heterogeneous treatment effects along pre-specified dimensions: parents' attitudes and student gender at the first endline, and just student gender at the second endline. Given the parsimonious set of primary outcomes and heterogeneity analyses, we do not adjust the statistical inference for multiple hypothesis testing.

4.2 Short-run results

We now present results from the first endline survey, which was conducted one to six months after the intervention ended.

[17] If an observation has a missing value for a component of the index, we construct the index using the remaining non-missing variables. When the baseline outcome Y_{ij}^0 is missing, we impute its value with the the district-gender mean; we include a flag for these imputed observations in the specification.

4.2.1 Effects on primary outcomes

Table 2 reports the short-run effects of the intervention on gender attitudes, girls' aspirations, and behavior using the basic specification.

We find that the intervention made gender attitudes more progressive: Column 1 shows that students in treatment schools have a 0.18 standard deviation higher attitude index than those in control schools ($p < 0.01$). Appendix Table 6 reports the effects for each of the 17 variables that comprise the index.

The coefficient is very similar (0.17) when the LASSO-selected extended controls are included, as shown in Appendix Table 7. Because our results are similar with or without the extended controls, subsequent tables only present results without the extended control variables. To account for the possibility that attrition is endogenous to treatment, we also estimate Lee bounds on the treatment effects (Lee, 2009). The attrition-adjusted lower bound on the point estimate is 0.16, as shown in Appendix Table 8.

One benchmark for the effect size is that endline gender attitudes are 0.50 standard deviations higher for girls than boys in the control group, as reported in Appendix Table 9. Thus, the treatment effect is 36% as large as the status quo gender gap in attitudes. Also, a one standard deviation increase in parent gender attitudes is associated with student gender attitudes being 0.05 standard deviations higher; the treatment effect is much larger than this.[18]

Another way to express the effect size is as a 'persuasion rate' (DellaVigna and Gentzkow, 2010). We find that the intervention converted 15.5% of gender-regressive views into support for gender equality. This persuasion rate is calculated by stacking all of the variables included in the attitudes index, which are coded as binary values. In the treatment group, on average 29.9% of views are gender-regressive, compared to 35.4% in the control group.

Turning to our second outcome, Table 2 shows that the program did not affect girls' aspirations. The average effect is 0.03 standard deviations and not significantly different from zero. Even though there likely will be a sizable gender gap in college completion and employment under the status quo (extrapolating from the cohorts ten to twenty years older than the sample), there is not a large gender gap in aspirations even without the intervention; girls' starting point of high aspirations might explain the lack of an effect on this outcome.

[18] In Dhar et al. (2018), we present an arguably better version of this parental attitudes benchmark, using indices constructed from the same set of questions for parents and students and collected at baseline for both groups. A one standard deviation increase in a parent's attitudes is then associated with a 0.11 standard deviation increase in the child's attitudes.

The third primary outcome is self-reported behaviors influenced by gender attitudes. As reported in column 3 of Table 2, behavior became more aligned with gender-progressive norms by 0.20 standard deviations ($p < 0.01$). The effect size for behavior is robust to including an extended set of control variables and has tight Lee bounds.

4.2.2 Assessing bias due to experimenter demand effects

A key concern is that participating in the program might have made salient what the socially desirable responses to our survey questions were without changing actual views. The treatment group participated in a program that made them more aware that many outsiders to their community regard support for gender equality as a laudable or 'correct' view to hold.

To investigate whether experimenter demand effects are upward biasing the estimated program impacts, we construct a social desirability score for each respondent. The score is based on responses on the Marlowe-Crowne module administered at baseline and measures a person's general tendency to give socially desirable answers. We test for heterogeneous treatment effects based on the social desirability score. The worrisome pattern would be if the treatment effects were driven by students with a high propensity to disingenuously give socially desirable answers and vanished for those with a low such tendency.

As shown in Table 3, the main effect of having a below-median social desirability score on gender attitudes is negative and significant, suggesting some upward shading of responses overall for the sample. We view this pattern as reassuring, a validation that the score seems to capture the propensity to shade responses. Importantly, there is no more of this shading up in the treatment group than the control group; the interaction term is small and insignificant. In other words, the treatment effect remains large and significant when we focus on the subsample with lower susceptibility to experimenter demand effects. We see a similar pattern for gender-related behavior. Interestingly, for girls' aspirations, where there is no average treatment effect, there appears to be a positive "effect" just for girls with a strong desire for social approval.

4.2.3 Heterogeneous treatment effects

There are many reasons the program could have had different effects on boys and girls. Girls might have felt more invested in the program's messages, or the ideas presented and discussed could have been more eye-opening for boys, for example. As reported in column 1 of Table 4, we find the effect of the program on attitudes is not statistically different for

boys and girls. (The point estimates suggest a perhaps somewhat smaller effect for girls.)

The fact that girls started out with more progressive attitudes than boys raises the question of whether the estimated heterogeneity by gender is entangled with heterogeneity by initial attitudes. As shown in Appendix Table 10, the patterns of gender heterogeneity are similar when we simultaneously allow for heterogeneity by baseline attitudes.[19]

For behavior, we find that the program had a significantly smaller impact for girls than boys (interaction coefficient of -0.13, $p < 0.01$), although the net effect for girls is still positive and significant. One interpretation of this finding is that boys and girls can adopt gender-equal attitudes with relatively equal ease, but girls face more constraints on translating their attitudes into behavior.

The second dimension of heterogeneity that we pre-specified was parental attitudes. Here we use the subsample for which we surveyed one of the parents at baseline, measuring his or her gender attitudes. A priori it is possible the program could have had either larger or smaller effects for students whose home environment is more conservative. Columns 3 to 5 of Table 4 show no significant differential effects by parents' gender attitudes. For both attitudes and behavior, the point estimates for the interaction coefficients are small compared to the main effects, while the point estimates are suggestive that the program may have raised aspirations among girls from more gender-progressive families. Overall, there is very little evidence that parental support for gender equality either facilitates or hinders the average success of the intervention.[20]

4.2.4 Effects on attitude and behavior subindices

In the appendix, we break down the primary outcomes into thematic sub-indices to show which specific attitudes and behaviors the intervention affected. Appendix Table 12 shows that the program had the strongest effects on attitudes about employment and gender roles, followed by education attitudes. The effect on gender-equitable fertility attitudes is statistically significant but small (0.02 standard deviations), which is possibly due to the intervention sessions having quite limited discussion about fertility. Appendix Table 13 examines behavior sub-indices. The intervention generated more interaction with the

[19]The gender heterogeneity analysis is also robust to correcting for the gap in economic status between girls and boys in government schools, which is due to boys' higher rate of attending private schools as discussed in footnote 11. See Appendix Table 11, which controls for several wealth proxies in parallel to gender.

[20]We also tested for heterogeneity based on several factors that seminar audiences frequently asked about. We find no heterogeneous effects by whether the school was co-ed, facilitator gender, siblings' sex composition, the sex ratio in the village, or the female employment rate in the village.

opposite sex, improved support for female relatives' ambitions, and made effort on household chores more gender-equal. Among girls, it also led to greater mobility (e.g., walking to school alone) but had no impact on decision-making power.

4.2.5 Effects on secondary outcomes

We examine perceptions of social norms as a secondary outcome. While the program only directly reached 100 to 200 adolescents per village, participants might have started regarding pro-equality views as more mainstream and, thus, updated their belief about how common those views were in their community.[21] We examine parallel questions about (1) personally holding a positive gender attitude, (2) believing one's community has a positive gender norm in that domain, and (3) personally holding the positive attitude and believing the community will not oppose you if you act on it. We thus assess both how perceived social norms change and how much participants view social norms as preventing them from acting on their progressive attitudes.

The intervention made personal attitudes about female employment more progressive by 13 percentage points (Table 5, column 1). It also increased by 5 percentage points the perception that others in the community hold that gender-progressive view (column 2), which is consistent with other findings that signals from institutions (Breakthrough in this case) can be effective in changing subjective perceptions of norms (Tankard and Paluck, 2016). In addition, the treatment group is more likely to hold a progressive attitude and believe society will be supportive (column 3); many of those who changed their attitudes do not expect the community norm to stand in their way of acting as they wish. However, the estimate in column 3 is smaller by 6 percentage points than the effect on simply holding that attitude in column 1: Some students whose own attitude changed think that restrictive cultural norms will hinder them from acting on it. Columns 4 to 6 show a similar pattern for norms about women leaving home to attend college.

We also examine school performance as a secondary outcome to rule out the concern that the program hurt achievement in subjects like math and Hindi by taking away some of their instructional time. We find no impact on school performance, as shown in Appendix Table 14. The administrative exam-score data we use are aggregated at the school-grade level, so we cannot estimate results separately by gender. However, girls generally perform better than boys in secondary school, so especially in light of their high aspirations to attend

[21] Recent evidence from India and Saudi Arabia suggests that people overestimate their community's opposition to female employment (Bernhardt et al., 2018; Bursztyn et al., 2020).

college, improving girls' school performance was not one of the expected impacts of the program.

Results for the three other pre-specified secondary outcomes (described in more detail in the data appendix) are reported in Table 6. First, we find that the program improved girls' self-esteem. Second, we find a small increase in awareness of gender discrimination; status-quo awareness is already high. Third, we find no impact on two implicit association tests (IATs), each of which was administered to about 3,000 respondents.[22] One IAT measured how respondents associate girls' and boys' faces with positive and negative words. The other, originally developed by Beaman et al. (2009), associates men's and women's faces with market labor or domestic work. Because of challenges administering the IATs at baseline (13% of responses were invalidated because the completion time was too fast or slow), we did not make the IAT a key focus for the endline.[23]

4.3 Medium-run results

We next discuss results from the second endline survey, conducted 2 to 2.5 years after the program ended.

4.3.1 Effects on primary outcomes

We continue to find a large and significant effect on gender attitudes in the second endline, as reported in column 1 of Table 7. The attitudes index is constructed identically to the first endline's index, so the 0.18 effect sizes in both endlines are directly comparable.[24] The coefficient is robust to adding extended controls to the regression (Appendix Table 7) and using Lee bounds (Appendix Table 8).

While experimenter demand effects are somewhat less of a concern in the second endline because the intervention is further in the past, it is still important to investigate this potential confound. Using the same approach as earlier, we find that having a low propensity to want to give socially desirable answers is not associated with having a smaller treatment effect.

[22] At baseline, we administered an IAT to 50% of the sample, and in the first endline we re-administered an IAT to this subsample, with half receiving each version of the IAT. Some endline IAT observations are missing due to technical problems with the laptops used.

[23] Another concern with IATs is that they measure not only personal attitudes but also awareness of shared cultural stereotypes. Thus, a program that discusses stereotypes could lead to a "worse" IAT score (Arkes and Tetlock, 2004).

[24] The control group mean of 0.003 in the second endline indicates that, absent the intervention, average attitudes changed very little between the endlines; the variable is normalized to have a mean of 0 in the control group for the first endline.

That is, the coefficient on the interaction of *Treated* and having a low social desirability score is small and statistically insignificant, as shown in Appendix Table 15.

We continue to find no effect on girls' stated aspirations (Table 7, column 2). It is possible an effect on this outcome could have emerged between the endlines, for example if girls' aspirations in the control group became more tempered over time.

Meanwhile, as in the short run, there is a sizable (0.23 standard deviations) and statistically significant effect on self-reported behavior, pooled for boys and girls. Note that we updated the elements in the behavior index between the two endlines, so the effect size is not as directly comparable over time as the effect on attitudes is.

We next examine treatment effects on two revealed-preference outcomes. The first measure is applying for a college scholarship (for girls). The theory of change for this outcome is that the program made girls' desire to attend college more intense (higher aspirations) or enabled them to persuade their parents to support their goal (expected behavior, conditional on aspirations). We find that the intervention led to a marginally significant increase in the application rate of 3.1 percentage points, or 8% (Table 7, column 4). When the extended set of controls is added, the coefficient remains relatively similar, but the p-value increases to 0.12 (see Appendix Table 7).

The intervention could have affected scholarship applications either by strengthening girls' resolve to go to college or by enabling them to secure their parents' support. That is, it could have changed their preferences or relaxed a constraint. We conduct some exploratory heterogeneity analysis (which was not pre-specified) to further probe this. First, we find that scholarship take-up is significantly higher for girls who had higher aspirations at baseline, as reported in Table 8, column 1. The total effect is close to four times as high among those with above-median aspirations compared to those with below-median aspirations (column 2). Second, scholarship applications are significantly higher among the 80% of girls who, at baseline, said they had talked with their parents about their education goals (column 3).[25] Thus, the program seems to have increased girls' intention to go to college by raising already-high aspirations and convincing already-engaged parents to support their daughter's goals, rather than by converting girls and families who started out lower on these dimensions.

Our final primary outcome is signing a public petition to end the dowry system. The intervention might have either made participants more opposed to the dowry system (attitude

[25]Aspirations and discussing goals with parents are positively correlated, so column 4 puts them together, and the larger effect seems to load somewhat more onto the discussing goals variable.

change) or reduced the cost to them of expressing their attitude (change in perceived social sanctions). In the control group, 15% of respondents called to add their names to the petition, and this rate is not significantly higher in the treatment group, as shown in Table 7, column 5.[26] This null result stands in contrast to the positive effect on self-reported opposition to dowry we observe. One interpretation is that the self-reported attitude is disingenuous. Another is that students worried about repercussions in their family or community from a public statement of their view, and the intervention did not lower this perceived cost (enough).

4.3.2 Heterogeneous effects by gender

We next examine heterogeneous effects by gender for the medium-run outcomes, with the results reported in Table 9. While there was no significant difference in attitude change between boys and girls in the short run, two years later there is. The medium-run effect size for boys is 0.25 (column 1), even larger than the short-run effect size: There is no fade-out for boys. The negative interaction term for females implies that the net effect for girls is 0.12 standard deviations. While this effect is still statistically significant ($p < 0.01$), it is about 75% of the short-run effect size for girls. We do not have a ready explanation for this gender gap in persistence — fade-out only among girls — but view the pattern as interesting and worthy of future research.

Self-reported behavior change continues to be larger for boys than girls in the medium run. Our interpretation of this pattern in the short run was that girls were more constrained by external factors in translating attitudes to behavior, but now another contributing factor is that attitude change is less sustained for them. We find no significant heterogeneity by gender in the treatment effect on signing the dowry petition.

4.3.3 Effects on secondary outcomes

We re-examine perceived social norms as a secondary outcome, with the results reported in the bottom panel of Table 5. We continue to find that the treatment group views the community as more supportive of female employment, but there is no longer an effect on perceived norms about female education. In addition, we continue to see a positive effect on

[26]Ideally the measure would have had a higher mean in the control group; the estimated effect size on self-reported attitudes corresponds to a large (42%) increase in the likelihood of signing. The requirement of needing to phone in may have been an impediment to signing. We chose not to have respondents sign in the presence of the surveyor, as that variant could still be susceptible to experimenter demand effects, which would defeat the purpose of adding this outcome.

girls' self-esteem. This result is reported in Table 10, along with results for the remaining secondary outcomes.

The third secondary outcome is an index of girls' education outcomes. It is based on school enrollment, studying a STEM subject, taking classes to obtain extra skills (e.g., computer, English), and taking after-school tutoring for exam preparation. We see a marginally significant increase of 0.06 standard deviations in the index (Table 10, column 2).

We also added a set of questions on aspirations related to age of marriage and son preference. We find a statistically significant change of 0.05 standard deviations in an index of these marriage and fertility aspirations, in the direction of more gender progressiveness (column 3).

The final two secondary outcomes are related to sexual harassment. We asked girls about the harassment they experienced, and find that the intervention increased their reported harassment (column 4). While we cannot disentangle whether this is due to actual experiences or reporting, we speculate that the intervention made girls more aware of harassment or destigmatized victimization, so they recognized and reported it at a higher rate. Finally, we measured boys' self-reported engagement in sexual harassment and assault. Because of concern about underreporting, we used an item count (or list) experiment, with one group receiving an extra item that asked them if the following statement is true: "In the past year, I have passed dirty comments about a girl; made dirty gestures in a girl's presence, or inappropriately touched or groped a girl." We do not find a significant effect of the intervention on this outcome.

5 Conclusion

There are a host of ways to intervene to improve women's and girls' well-being in contexts where cultural norms deny them equal opportunities. These include laws granting equal rights, affirmative action policies, subsidies to encourage investment in girls, programs that impart skills or target resources to women, and many others.

The approach studied in this paper is to try to directly reshape the gender attitudes that people hold. We examined whether an attitude-change intervention in schools in Haryana, India, could make adolescents' attitudes and, in turn, their behavior, less gender biased. The intervention had students discuss and think about gender norms and gender discrimination through class sessions held over two and a half school years.

We find that the intervention succeeded in making attitudes more supportive of gender

equality and coaxing more gender-equal behavior. Moreover, the impacts observed shortly after the program ended persisted: When we re-surveyed participants two years later, we continued to find strong effects. At the two-year mark, we also find suggestive evidence that the program strengthened girls' belief that they will get a college education.

North India has particularly strong gender discrimination, and gender norms are often highly dependent on the context. Thus, the exact programming would need to be adapted for other contexts, but this general approach of engaging adolescents, or even younger children, in school discussions could be a widely applicable way to change gender norms. While it might seem surprising that a series of class discussions changed views formed over many years, most adolescents had likely spent very little time thinking hard about gender inequality prior to the program. Interrogating one's personal prejudices and the norms that prevail in one's community might be especially powerful in schools, without voluntary, self-selected participation and when participants are young enough that their views are still quite pliable.

That said, perhaps a more indirect approach would be even more effective than explicit discussions about the importance of gender equality; assigned readings could feature empowered women, or a history assignment could ask students to discuss women's changing role in society. This aspect of the program as well as other ones, such as participants' age, the dosage, and the relative emphasis on moral versus pragmatic arguments for equality, could be varied and assessed to optimize programs like this.

One lesson from our results is the importance of including boys and men in programs aimed at altering gender norms. We find that attitude change translates into larger shifts in behavior for boys than girls. We speculate that the very problem the program aims to solve — that males have more power in society — means that they have more freedom to act on their gender-progressive views. For this reason, as we look ahead, we expect a larger increase in employment for male participants' wives than for female participants. Of course, men also face familial constraints on their behavior and pressure to conform to traditional norms, so it is an open question whether such an effect on female employment will materialize. In future work, we hope to measure whether it indeed does, as well as how the program affects other adult outcomes such as higher education, age of marriage, and childbearing.

References

Abramsky, T., K. Devries, L. Kiss, J. Nakuti, N. Kyegombe, E. Starmann, B. Cundill, L. Francisco, D. Kaye, T. Musuya, et al. (2014). Findings from the SASA! study: A cluster randomized controlled trial to assess the impact of a community mobilization intervention to prevent violence against women and reduce HIV risk in Kampala, Uganda. *BMC Medicine 12*(1), 122.

Afridi, F., T. Dinkelman, and K. Mahajan (2018). Why are fewer married women working in rural India? A decomposition analysis over two decades. *Journal of Population Economics 31*, 783–818.

Alesina, A., P. Giuliano, and N. Nunn (2013). On the origins of gender roles: Women and the plough. *Quarterly Journal of Economics 128*(2), 469–530.

Aloud, M. E., S. Al-Rashood, I. Ganguli, and B. Zafar (2020). Information and social norms: Experimental evidence on the labor market aspirations of Saudi women. NBER Working Paper No. 26693.

Alrababa'h, A., W. Marble, S. Mousa, and A. Siegel (2019). Can exposure to celebrities reduce prejudice? The effect of Mohamed Salah on Islamophobic behaviors and attitudes. Working paper No. 19-04, Immigration Policy Lab.

Anderson, M. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry preschool, and early training projects. *Journal of the American Statistical Association 103*(484), 1481–1495.

Arkes, H. and P. Tetlock (2004). Attributions of implicit prejudice, or "Would Jesse Jackson 'fail' the Implicit Association Test?". *Psychological Inquiry 15*(4), 257–278.

Ashraf, N., N. Bau, C. Low, and K. McGinn (2020). Negotiating a better future: How interpersonal skills facilitate inter-generational investment. *Quarterly Journal of Economics 135*(2), 1095–1151.

Bandiera, O., N. Buehren, R. Burgess, M. Goldstein, S. Gulesci, I. Rasul, and M. Sulaiman (2020). Women's empowerment in action: Evidence from a randomized control trial in Africa. *American Economic Journal: Applied Economics 12*(1), 210–259.

Beaman, L., R. Chattopadhyay, E. Duflo, R. Pande, and P. Topalova (2009). Powerful women: Does exposure reduce bias? *Quarterly Journal of Economics 124*(4), 1497–1540.

Becker, G. and C. Mulligan (1997). The endogenous determination of time preference. *Quarterly Journal of Economics 112*(3), 729–758.

Belloni, A., V. Chernozhukov, and C. Hansen (2014). Inference on treatment effects after selection among high-dimensional controls. *Review of Economic Studies 81*(2), 608–650.

Bernhardt, A., E. Field, R. Pande, N. Rigol, S. Schaner, and C. Troyer-Moore (2018). Male social status and women's work. *AEA Papers and Proceedings 108*, 363–67.

Bertrand, M., J. Pan, and E. Kamenica (2015). Gender identity and relative income within households. *Quarterly Journal of Economics 130*(2), 571–614.

Bowles, S. (1998). Endogenous preferences: The cultural consequences of markets and other economic institutions. *Journal of Economic Literature 36*(1), 75–111.

Buchmann, N., E. Field, R. Glennerster, S. Nazneen, S. Pimkina, and I. Sen (2018). Power vs. money: Alternative approaches to reducing child marriage in Bangladesh, a randomized control trial. Working paper, Duke University.

Bursztyn, L., A. Gonzalez, and D. Yanagizawa-Drott (2020). Misperceived social norms: Women working outside the home in Saudi Arabia. Working paper, University of Chicago.

Calvi, R. (2020). Why are older women missing in India? The age profile of bargaining power and poverty. *Journal of Political Economy 128*(7), 2453–2501.

Cantoni, D., Y. Chen, D. Yang, N. Yuchtman, and Y. Zhang (2017). Curriculum and ideology. *Journal of Political Economy 125*(2), 338–392.

Carlana, M. (2019). Implicit stereotypes: Evidence from teachers' gender bias. *Quarterly Journal of Economics 134*(3), 1163–1224.

Carrell, S., M. Hoekstra, and J. West (2019). The impact of college diversity on behavior toward minorities. *American Economic Journal: Economic Policy 11*(4), 159–82.

Crowne, D. and D. Marlowe (1960). A new scale of social desirability independent of psychopathology. *Journal of Consulting Psychology 24*(4), 349–354.

Dahl, G., A. Kotsadam, and D. Rooth (2018). Does integration change gender attitudes? The effect of randomly assigning women to traditionally male teams. NBER Working Paper No. 24351.

DellaVigna, S. and M. Gentzkow (2010). Persuasion: Empirical evidence. *Annual Review of Economics 2*(1), 643–669.

Dhar, D., T. Jain, and S. Jayachandran (2018). The intergenerational transmission of gender attitudes: Evidence from India. *Journal of Development Studies 15*(12), 2572–2592.

DISE (2011). District information system for education. `http://dise.in`.

Donovan, R. and S. Leivers (1993). Using paid advertising to modify racial stereotype beliefs. *Public Opinion Quarterly 57*(2), 205–218.

Duflo, E. (2012). Women empowerment and economic development. *Journal of Economic Literature 50*(4), 1051–1079.

Edmonds, E., B. Feigenberg, and J. Leight (2019). Can a girl influence her own schooling in early adolescence? Evidence from Rajasthan. Working paper, American University.

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

Field, E., S. Jayachandran, and R. Pande (2010). Do traditional institutions constrain female entrepreneurship? A field experiment on business training in India. *American Economic Review Papers and Proceedings 100*(2), 125–9.

Govt. of India (2011). Census of India. Available at `www.censusindia.gov.in/`.

Green, D., A. Wilke, and J. Cooper (2020). Countering violence against women by encouraging disclosure: A mass media experiment in rural Uganda. Working paper, Columbia University.

Grigorieff, A., C. Roth, and D. Ubfal (2020). Does information change attitudes toward immigrants? *Demography 57*, 1117–1143.

Gupta, J., K. Falb, H. Lehmann, D. Kpebo, Z. Xuan, M. Hossain, C. Zimmerman, C. Watts, and J. Annan (2013). Gender norms and economic empowerment intervention to reduce intimate partner violence against women in rural Côte d'Ivoire: A randomized controlled pilot study. *BMC International Health and Human Rights 13*(46).

Healy, A. and N. Malhotra (2013). Childhood socialization and political attitudes: Evidence from a natural experiment. *Journal of Politics 75*(4), 1023–1037.

Hopkins, D., J. Sides, and J. Citrin (2019). The muted consequences of correct information about immigration. *Journal of Politics 81*(1), 315–320.

Jayachandran, S. (2015). The root causes of gender inequality in developing countries. *Annual Review of Economics 7*, 63–88.

Jayachandran, S. (2017). Fertility decline and missing women. *American Economic Journal: Applied Economics 9*(1), 118–139.

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

Jha, P., R. Kumar, P. Vasa, N. Dhingra, D. Thiruchelvam, and R. Moineddin (2006). Low male-to-female sex ratio of children born in India: National survey of 1.1 million households. *The Lancet 367*, 211–218.

Klasen, S. and J. Pieters (2015). What explains the stagnation of female labor force participation in urban India? *World Bank Economic Review 29*(3), 449–478.

Kohlberg, L. (1976). Moral stages and moralization: The cognitive-developmental approach. *Moral development and behavior: Theory, research, and social issues*, 31–53.

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.

Lee, D. (2009). Training, wages, and sample selection: Estimating sharp bounds on treatment effects. *Review of Economic Studies 76*(3), 1071–1102.

Lowe, M. (2020). Types of contact: A field experiment on collaborative and adversarial caste integration. CESifo Working Paper No. 8089.

Markus, H. and P. Nurius (1986). Possible selves. *American Psychologist 41*(9), 954–969.

McKelway, M. (2020). Women's employment in India: Intra-household and intra-personal constraints. Working paper, Massachusetts Institute of Technology.

Moss-Racusin, C., E. Pietri, E. Hennes, J. Dovidio, V. Brescoll, G. Roussos, and J. Handelsman (2018). Reducing STEM gender bias with VIDS (video interventions for diversity in STEM). *Journal of Experimental Psychology: Applied 24*(2), 236.

Paluck, E. (2009). Reducing intergroup prejudice and conflict using the media: A field experiment in Rwanda. *Journal of Personality and Social Psychology 96*(3), 574–87.

Porter, C. and D. Serra (2019). Gender differences in the choice of major: The importance of female role models. *American Economic Journal: Applied Economics 12*(3), 226–254.

Pulerwitz, J., L. Hughes, M. Mehta, A. Kidanu, F. Verani, and S. Tewolde (2015). Changing gender norms and reducing intimate partner violence: Results from a quasi-experimental intervention study with young men in Ethiopia. *American Journal of Public Health 105*(1), 132–137.

Rao, G. (2019). Familiarity does not breed contempt: Generosity, discrimination, and diversity in Delhi schools. *American Economic Review 109*(3), 774–809.

Reynolds, W. (1982). Development of reliable and valid short forms of the Marlowe-Crowne social desirability scale. *Journal of Clinical Psychology 38*(1), 119–125.

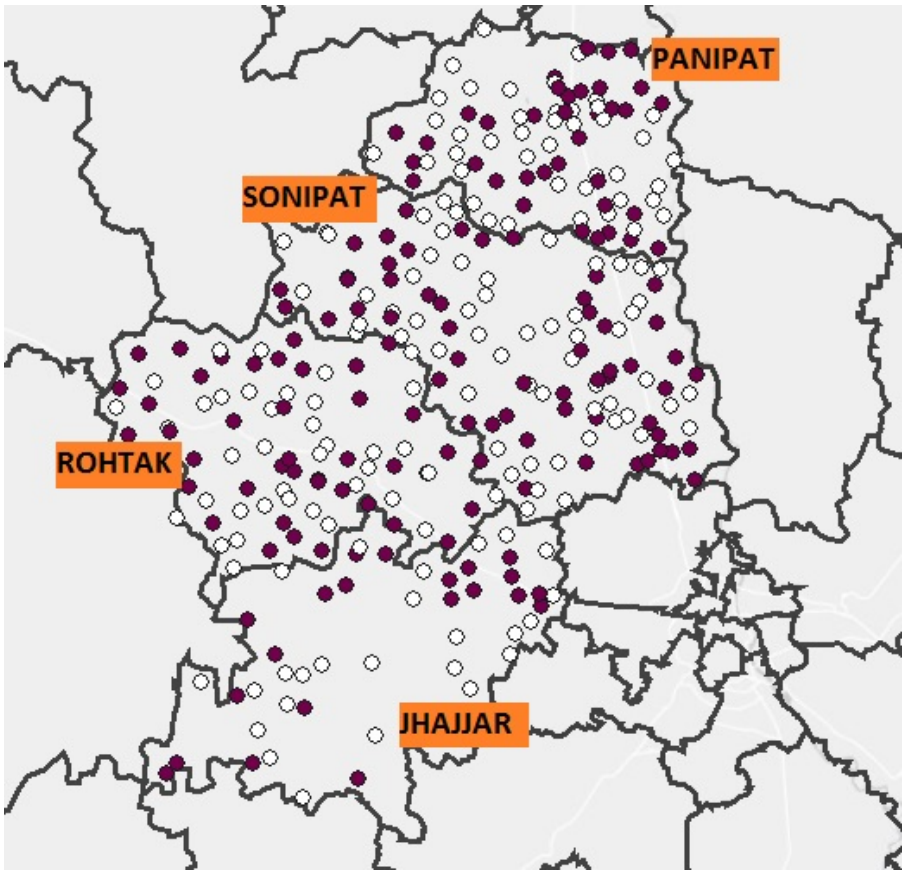
Sen, A. (1990). More than 100 million women are missing. *The New York Review of Books 37*(20).

Tankard, M. and E. Paluck (2016). Norm perception as a vehicle for social change. *Social Issues and Policy Review 10*(1), 181–211.

Washington, E. (2008). Female socialization: How daughters affect their legislator fathers' voting on women's issues. *American Economic Review 98*(1), 311–332.

World Bank (2011). *World Development Indicators*. Washington, D.C.: World Bank.

Figure 1: Map of treatment and control schools within the study districts



Notes: Schools in the treatment group are marked with dark dots, and schools in the control group are marked with white dots.

Table 1: Descriptive statistics: School and student characteristics at baseline

Variable	Treatment	Control	Standardized diff
Number of schools	149	164	
Urban	0.107	0.073	0.119
	[0.311]	[0.261]	
School is co-ed	0.698	0.677	0.045
	[0.461]	[0.469]	
Males in grades 6 and 7	66.427	65.270	0.028
	[45.948]	[35.963]	
Females in grades 6 and 7	75.125	74.212	0.015
	[60.081]	[58.344]	
Number of teachers	17.793	17.074	0.078
	[10.235]	[8.197]	
Number of students	7,051	7,758	
Student's age	11.833	11.854	-0.017
	[1.261]	[1.250]	
Female	0.566	0.544	0.044
	[0.496]	[0.498]	
Hindu	0.945	0.953	-0.036
	[0.227]	[0.211]	
Enrolled in grade 6	0.526	0.521	0.011
	[0.499]	[0.500]	
Scheduled caste	0.268	0.285	-0.039
	[0.443]	[0.451]	
Mother's age	35.462	35.572	-0.017
	[6.351]	[6.513]	
Father's age	40.497	40.611	-0.016
	[6.895]	[7.131]	
Mother is illiterate	0.370	0.374	-0.009
	[0.483]	[0.484]	
Mother works full-time	0.292	0.292	-0.002
	[0.455]	[0.455]	
Dwelling has flush toilet	0.155	0.130	0.070
	[0.362]	[0.337]	
Gender attitudes index	0.032	0.000	0.032
	[1.008]	[1.000]	
Girls' aspirations index	0.039	0.000	0.039
	[1.001]	[1.000]	
Behavior index	-0.012	0.000	-0.012
	[0.994]	[1.000]	

Notes: F-stat for joint significance of above baseline student variables is 0.980.

Table 2: Treatment effects on attitudes, aspirations, and behavior at Endline 1

	Gender attitudes index (1)	Girls' aspirations index (2)	Behavior index (3)
Treated	0.179***	0.027	0.195***
	[0.020]	[0.024]	[0.021]
Control group mean	0.000	0.000	0.000
Basic controls	Yes	Yes	Yes
Number of students	13,987	7,767	13,974

Notes: Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. All regressions control for the baseline analogue of the outcome, its school-gender mean, grade-gender and district-gender fixed effects (columns 1 and 3) or grade and district fixed effects (column 2), missing flags for each variable used to construct the outcome index, and a missing flag for the baseline outcome. Standard errors are clustered by school.

Table 3: Robustness check for social desirability bias (Endline 1)

	Gender attitudes index (1)	Girls' aspirations index (2)	Behavior index (3)
Treated	0.184***	0.056**	0.187***
	[0.022]	[0.028]	[0.024]
Low social desirability score	-0.083***	-0.027	-0.065***
	[0.021]	[0.027]	[0.020]
Treated × Low social desirability score	-0.007	-0.071*	0.022
	[0.032]	[0.042]	[0.028]
Treated + Treated × Low SD = 0	0.000	0.691	0.000
Control group mean	0.000	0.000	0.000
Basic controls	Yes	Yes	Yes
Number of students	13,987	7,767	13,974

Notes: Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. Social desirability (SD) score is a baseline measure of the student's propensity to give socially desirable answers. Low SD score refers to having a score that is below median for the sample. All regressions control for the baseline analogue of the outcome, its school-gender mean, grade-gender and district-gender fixed effects (columns 1 and 3) or grade and district fixed effects (column 2), missing flags for each variable used to construct the outcome index, and a missing flag for the baseline outcome. Standard errors are clustered by school.

Table 4: Heterogeneous effects by gender and baseline parent attitudes (Endline 1)

	Gender attitudes index (1)	Behavior index (2)	Gender attitudes index (3)	Girls' aspirations index (4)	Behavior index (5)
Treated	0.204***	0.266***	0.183***	0.032	0.184***
	[0.030]	[0.029]	[0.036]	[0.044]	[0.033]
Treated × Female	-0.044	-0.126***			
	[0.039]	[0.036]			
Treated × Above median baseline parent attitudes			-0.017	0.042	-0.009
			[0.045]	[0.049]	[0.033]
Treated + Treated × Female = 0	0.000	0.000			
Treated + Treated × Above median attitudes = 0			0.000	0.050	0.000
Control group mean	0.000	0.000	0.000	0.000	0.000
Basic controls	Yes	Yes	Yes	Yes	Yes
Number of students	13,987	13,974	5,718	3,231	5,717

Notes: Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. All regressions control for the baseline analogue of the outcome, its school-gender mean, grade-gender and district-gender fixed effects (columns 1, 2, 3, and 5) or grade and district fixed effects (column 4), missing flags for each variable used to construct the outcome index, and a missing flag for the baseline outcome. Columns 3–5 include the main effect of above-median baseline parent attitudes. Standard errors are clustered by school.

Table 5: Effect of intervention on perception of social norms (Endline 1 and 2)

	Social norms towards work			Social norms towards education		
	Student agrees that...			*Student agrees that...*		
	women should be allowed to work	community thinks women should be allowed to work	women should be allowed to work and thinks community will not oppose them	women should be allowed to study in college even if it is far away	community thinks women should be allowed to study in college even if it is far away	women should be allowed to study in college and thinks community will not oppose them
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Endline 1						
Treated	0.129***	0.052***	0.072***	0.084***	0.056***	0.067***
	[0.011]	[0.013]	[0.012]	[0.008]	[0.014]	[0.013]
Control group mean	0.690	0.435	0.464	0.853	0.592	0.637
Number of students	6862	6464	6409	7074	6752	6717
Panel B: Endline 2						
Treated	0.061***	0.034**	0.044***	0.028***	0.006	0.011
	[0.008]	[0.014]	[0.014]	[0.007]	[0.013]	[0.012]
Control group mean	0.864	0.612	0.647	0.911	0.675	0.713
Number of students	6736	6474	6447	6531	6295	6262

Notes: Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. All columns control for grade-gender and district-gender fixed effects. Each respondent was given either the set of questions on norms about work or norms about education, determined by randomization. Standard errors are clustered by school.

Table 6: Effects on other secondary outcomes (Endline 1)

	Girls' self-esteem	Awareness of gender-based discrimination	IAT: Associates girls with positive words	IAT: Associates women with market work
	(1)	(2)	(3)	(4)
Treated	0.100***	0.033**	0.004	-0.050
	[0.023]	[0.016]	[0.033]	[0.055]
Control group mean	0.000	0.000	0.000	0.000
Basic controls	Yes	Yes	Yes	Yes
Number of students	7,788	13,939	2,926	3,198

Notes: Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. All regressions control for grade-gender and district-gender fixed effects (columns 2–4) or grade and district fixed effects (column 1). All columns except column 2 also control for the baseline analogue of the outcome, its school-gender mean, and a missing flag for the baseline outcome. Columns 1 and 2 also control for missing flags for each variable used to construct the outcome index. Standard errors are clustered by school.

Table 7: Treatment effects on attitudes, aspirations, behavior, and revealed preferences at Endline 2

	Gender attitudes index (1)	Girls' aspirations index (2)	Behavior index (3)	Applied to scholarship (4)	Signed petition (5)
Treated	0.177***	-0.024	0.226***	0.031*	0.012
	[0.021]	[0.019]	[0.024]	[0.017]	[0.009]
Control group mean	0.003	0.000	0.000	0.408	0.150
Basic controls	Yes	Yes	Yes	Yes	Yes
Number of students	13,679	7,560	13,677	7,347	13,303

Notes: Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. All regressions control for grade-gender and district-gender fixed effects (columns 1, 3, and 5) or grade and district fixed effects (columns 2 and 4). Columns 1 to 3 also control for the baseline analogue of the outcome, its school-gender mean, missing flags for each variable used to construct the outcome index, and a missing flag for the baseline outcome. The outcomes in columns 4 and 5 are binary variables, not indices, and were not collected for the 3% of respondents who were surveyed by phone for the second endline (because these outcomes involved giving printed material to the respondent). Standard errors are clustered by school.

Table 8: Exploring mechanisms for the treatment effect on scholarship applications

	Applied to scholarship			
	(1)	(2)	(3)	(4)
Treated	0.029*	0.014	-0.023	-0.017
	[0.017]	[0.019]	[0.027]	[0.034]
Treated × BL aspirations index	0.022**			0.007
	[0.011]			[0.015]
Treated × Above-median BL aspirations		0.040		
		[0.024]		
Treated × Has discussed educ goals with parent			0.068**	0.059
			[0.028]	[0.037]
Treated + Treated × Above-median aspir. = 0		0.020		
Treated + Treated × Has discussed goals = 0			0.016	
Control group mean	0.408	0.408	0.408	0.408
Number of students	7,347	7,347	7,347	7,347

Notes: Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. All regressions include grade and gender fixed effects, the main effects for the baseline variables used in the interaction terms, and flags for whether they are missing. Standard errors are clustered by school.

Table 9: Heterogeneous effects by gender (Endline 2)

	Gender attitudes index	Behavior index	Signed petition
	(1)	(2)	(3)
Treated	0.248***	0.313***	0.003
	[0.031]	[0.040]	[0.010]
Treated × Female	-0.129***	-0.158***	0.016
	[0.039]	[0.043]	[0.016]
Treated + Treated × Female = 0	0.000	0.000	0.140
Control group mean	0.003	0.000	0.150
Basic controls	Yes	Yes	Yes
Number of students	13,679	13,677	13,303

Notes: Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. All regressions control for grade-gender and district-gender fixed effects. Columns 1 and 2 also control for the baseline analogue of the outcome, its school-gender mean, missing flags for each variable used to construct the outcome index, and a missing flag for the baseline outcome. Standard errors are clustered by school.

Table 10: Effects on secondary outcomes (Endline 2)

	Girls' self-esteem	Girls' education	Marriage and fertility aspirations	Girls' experienced harassment	Boys' perpetrated harassment (school-grade level)
	(1)	(2)	(3)	(4)	(5)
Treated	0.086***	0.058*	0.049**	0.063**	0.060
	[0.026]	[0.033]	[0.021]	[0.029]	[0.062]
Control group mean	0.000	0.000	0.000	0.000	-0.003
Basic controls	Yes	Yes	Yes	Yes	Yes
Number of observations	7,341	7,566	13,288	7,314	504

Notes: Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. The unit of observation is the student in columns 1–4 and the school-grade in column 5. Columns 1 to 4 control for grade-gender and district-gender fixed effects (column 3) or grade and district fixed effects (remaining columns) and missing flags for each variable used to construct the outcome index. Column 1 also controls for the baseline analogue of the outcome variable, its school-gender mean, and a missing flag for the baseline outcome. Column 5 controls for grade and district fixed effects. Standard errors are clustered by school. A higher value of the outcome in column 4 corresponds to higher reported harassment by girls. The outcome in column 5 is the proportion of boys who report engaging in sexual harassment, based on a list experiment.

ONLINE APPENDIX

Appendix Table 1: Determinants of sample attrition and endline survey location

	Endline 1			Endline 2	
	Attrited	Surveyed in school	Surveyed student in-person	Attrited	Surveyed student in-person
	(1)	(2)	(3)	(4)	(5)
Treated	0.004	-0.026	-0.000	0.010	-0.002
	[0.010]	[0.019]	[0.001]	[0.011]	[0.005]
Treated × Gender attitudes index	-0.005	-0.015*	0.001	-0.002	0.001
	[0.004]	[0.008]	[0.001]	[0.005]	[0.003]
Treated × Girls' aspirations index	0.003	0.016*	0.001	0.011	-0.002
	[0.005]	[0.009]	[0.001]	[0.007]	[0.004]
Treated × Behavior index	0.011**	-0.007	0.001	0.005	-0.002
	[0.005]	[0.010]	[0.001]	[0.006]	[0.004]
Treated × Female	0.006	0.030	-0.000	0.000	-0.001
	[0.013]	[0.025]	[0.002]	[0.015]	[0.008]
Gender attitudes index	-0.002	0.020***	-0.000	-0.004	0.001
	[0.003]	[0.006]	[0.001]	[0.003]	[0.002]
Girls' aspirations index	-0.010***	0.006	-0.000	-0.017***	0.004*
	[0.003]	[0.006]	[0.001]	[0.004]	[0.003]
Behavior index	-0.010***	0.013**	-0.000	-0.005	0.003
	[0.003]	[0.007]	[0.001]	[0.004]	[0.002]
Control group mean	0.056	0.757	0.999	0.074	0.973
Treatment group mean	0.062	0.752	0.998	0.084	0.971
Treatment = Control p-value	0.317	0.556	0.489	0.154	0.376
Basic controls	Yes	Yes	Yes	Yes	Yes
Number of students	14,809	13,987	13,987	14,809	13,685

Notes: Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. All columns control for grade-gender and district-gender fixed effects and missing flags for the baseline primary outcomes. Standard errors are clustered by school. The sample for columns 1 and 4 is the baseline sample, and the sample for columns 2, 3, and 5 are those who were successfully surveyed in the relevant endline round.

Appendix Table 2: Reasons for sample attrition

Panel A: Endline 1

	Female	Male	Total
Tracked and surveyed			
In school	6,241	4,312	10,553
At home	1,547	1,868	3,415
Over the phone	14	5	19
Tracked but could not be surveyed			
Student deceased or unwell	24	26	50
Student or parent refused assent	42	39	81
Not tracked			
Address not trackable	150	113	263
Student not found at home	15	31	46
Family in village but student moved	48	12	60
Family and student moved	113	102	215
Other	62	89	151

Panel B: Endline 2

	Female	Male	Total
Tracked and surveyed			
In-person	7,347	5,956	13,303
On phone	219	163	382
Parent survey*	182	116	298
Tracked but could not be surveyed			
Student deceased or unwell	43	22	65
Student or parent refused assent	189	134	323
Not tracked	276	206	484

Notes: The sample analyzed in this table are the 14,853 potential endline respondents (baseline respondents plus 44 students enrolled in the school with missing baseline data). * For Endline 2, we collected some data about students from their parents if the student was unavailable; these observations are considered to be in the "attrited" sample, as the main outcome variables are missing for these respondents.

Appendix Table 3: Descriptive statistics on school enrollment at endline

	Endline 1	Endline 2
Same school as at baseline	0.816	0.489
	[0.388]	[0.500]
Private school in same village/town as at baseline	0.036	0.053
	[0.186]	[0.225]
Govt school in different village/town than at baseline	0.036	0.092
	[0.186]	[0.290]
Private school in different village/town than at baseline	0.031	0.060
	[0.174]	[0.237]
Currently in formal schooling/college	0.919	0.717
	[0.273]	[0.451]
Dropped out of school and not pursing any other course	0.079	0.239
	[0.270]	[0.426]
Aware of program (treatment group only)	0.866	
	[0.341]	
Number of observations	13,987	13,685

Notes: Table reports variable means and standard deviations.

Appendix Table 4: Boys' and girls' baseline gender attitudes and aspirations

Variable	Boys	Girls
Gender attitudes index	-0.307 [1.030]	0.274 [0.902]
Disagree: A woman's most important role is being a good homemaker	0.201 [0.401]	0.403 [0.490]
Disagree: A man should have the final word about decisions in his home	0.334 [0.472]	0.496 [0.500]
Disagree: A woman should tolerate violence to keep her family together	0.606 [0.489]	0.665 [0.472]
Disagree: Wives should be less educated than their husbands	0.562 [0.496]	0.748 [0.434]
Disagree: Boys should get more opportunities/resources for education than girls	0.177 [0.381]	0.421 [0.494]
Agree: Men and women should get equal opportunities in all spheres of life	0.901 [0.299]	0.918 [0.274]
Agree: Girls should be allowed to study as far as they want	0.869 [0.337]	0.959 [0.199]
Agree: Daughters should have a similar right to inherited property as sons	0.823 [0.381]	0.874 [0.331]
Agree: It would be a good idea to elect a woman as the village Sarpanch	0.685 [0.465]	0.814 [0.389]
Aspirations index	0.117 [0.931]	-0.081 [1.057]
Student has discussed education goals with parent or adult relative	0.840 [0.367]	0.794 [0.405]
Student's highest desired level of education is above sample median	0.535 [0.499]	0.465 [0.499]
Student expects white collar job when he/she is 25 years old	0.772 [0.420]	0.717 [0.450]
Number of students	6,597	8,212

Notes: Table reports variable means and standard deviations.

Appendix Table 5: LASSO-selected extended set of control variables

Extended control variable	Gender attitudes index	Girls' aspirations index	Behavior index	Applied to scholarship (girls)	Supported petition
Student's age	X	X		X	
Student's grade at baseline					
Rural location					
Scheduled caste	X				
Scheduled tribe					
Muslim					
Number of female siblings					
Number of male siblings	X	X			
Student stays with parents					
Mother has completed 8th grade	X	X			
Mother works part-time					
Mother works full-time					
House is pukka		X			
Dwelling has electricity					
Dwelling has flush toilet		X			
House has a no-flush toilet					
Family owns the house					
Household owns radio or tape recorder					
Household owns TV					
Household owns refrigerator	X				
Household gets newspaper daily					
Household gets tap water					
Household owns water pump					
Self-efficacy index	X	X			
Social desirability score	X				
Parent's baseline gender attitudes index	X				
Number of guest teachers in the school					
Number of full-time teachers in the school					
Fraction of female teachers				X	
Availability of counsellor in the school					
Number of PTA meetings held in the last year					
School has a functional library					
School has functional toilets					
School has electricity					
School has access to computers					
School has access to internet					
School has sports field					
School has mid-day meals					
School has auditorium					
School has EduSat					
Bal Sabha sessions: number of times in a week					
Library sessions: number of times in a week					
School is co-educational					
Village-level adult female literacy rate	X				
Village-level adult male literacy rate			X		
Village-level female labor force participation					

Appendix Table 6: Treatment effects on individual gender attitudes (with Bonferroni correction)

	Endline 1	Endline 2
Disagree: Wives should be less educated than their husbands	0.041***	0.043***
Disagree: Boys should get more opportunities/resources for education	0.025	0.029
If HH head, would send both children or girl for education	0.066***	0.042***
Disagree: Woman's most important role is caring for home and children	0.108***	0.071***
Disagree: Men are better suited than women to work outside the house	0.081***	0.086***
Disagree: Marriage is more important for Pooja than her job	0.020	0.044**
Disagree: Being a teacher would be a more suitable job for Pooja	0.030	0.023
Agree: Women should be allowed to work outside home	0.127***	0.060***
Agree: Daughters should have a similar right to inherited property as sons	0.023***	0.008
Agree: It would be a good idea to elect a woman as the Sarpanch of your village	0.019***	0.015**
Disagree: A man should have the final word about decisions in his home	0.100***	0.080***
Disagree: A woman should tolerate violence in order to keep her family together	0.061***	0.038***
Disagree: Parents should maintain stricter control over daughters than sons	0.059***	0.050***
Has gender equal views on getting higher education for better marriage prospects	-0.016	0.019**
Sister/female cousins/friends should be married after age 19	0.065***	0.036***
Difference between boys and girls age to marry is less than control median	0.050***	0.041***
Disagree: Keep having children if no sons yet but not if no daughters	0.016	0.020***

Notes: Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$, using Bonferroni-adjusted p-values, i.e., the raw p-value is multiplied by 17. All regressions control for grade-gender and district-gender fixed effects, the baseline gender attitudes index and a flag for whether it is missing, and the school-gender average of the baseline gender attitudes index. Standard errors are clustered by school.

Appendix Table 7: Treatment effects using extended set of control variables

Panel A: Endline 1

	Gender attitudes index (1)	Girls' aspirations index (2)	Behavior index (3)
Treated	0.169***	0.022	0.189***
	[0.019]	[0.023]	[0.021]
Control group mean	0.000	0.000	0.000
Number of students	13987	7767	13974

Panel B: Endline 2

	Gender attitudes index (1)	Girls' aspirations index (2)	Behavior index (3)	Applied to scholarship (girls) (4)	Signed petition (5)
Treated	0.164***	-0.029	0.222***	0.027	0.012
	[0.020]	[0.019]	[0.024]	[0.017]	[0.009]
Control group mean	0.003	0.000	0.000	0.408	0.150
Number of students	13679	7560	13677	7347	13303

Notes: Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. All regressions include grade-gender and district-gender fixed effects (columns 1, 3, and 5) or grade and district fixed effects (columns 2 and 4), plus a set of additional controls selected via LASSO. Panel A and columns 1 to 3 of Panel B also control for the baseline analogue of the outcome, its school-gender mean, missing flags for each variable used to construct the outcome index, and a missing flag for the baseline outcome. Standard errors are clustered by school.

Appendix Table 8: Lee bounds on treatment effects

Panel A: Endline 1

	Gender attitudes index (1)	Girls' aspirations index (2)	Behavior index (3)
Treated	0.179***	0.027	0.195***
	[0.020]	[0.024]	[0.021]
Treated (Lower bound)	0.162***	0.026	0.182***
	[0.019]	[0.024]	[0.021]
Treated (Upper bound)	0.189***	0.029	0.206***
	[0.019]	[0.024]	[0.021]
Observations	13,987	7,767	13,974
Observations (Lee bounds)	13,944	7,765	13,928

Panel B: Endline 2

	Gender attitudes index (1)	Girls' aspirations index (2)	Behavior index (3)	Applied to scholarship (girls) (4)	Signed petition (5)
Treated	0.177***	-0.024	0.226***	0.031*	0.012
	[0.021]	[0.019]	[0.024]	[0.017]	[0.009]
Treated (Lower bound)	0.140***	-0.040**	0.192***	0.028	0.010
	[0.021]	[0.019]	[0.024]	[0.017]	[0.009]
Treated (Upper bound)	0.194***	-0.009	0.254***	0.037**	0.023***
	[0.021]	[0.019]	[0.024]	[0.017]	[0.009]
Observations	13,679	7,560	13,677	7,347	13,303
Observations (Lee bounds)	13,599	7,527	13,597	7,313	13,208

Notes: Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. All regressions include grade-gender and district-gender fixed effects (columns 1, 3, and 5) or grade and district fixed effects (columns 2 and 4). Panel A and columns 1 to 3 of Panel B also control for the baseline analogue of the outcome, its school-gender mean, missing flags for each variable used to construct the outcome index, and a missing flag for the baseline outcome. Standard errors are clustered by school. At both endlines, attrition is (insignificantly) higher in the treatment group than the control group, so the Lee bounds estimates trim control group observations to equalize the attrition rate, using either those with the highest or lowest value of the outcome variable.

Appendix Table 9: Correlates of primary outcomes in Endline 1 control group

	Gender attitudes index	
	(1)	(2)
Female	0.498***	
	[0.026]	
Baseline parent gender attitudes index		0.054***
		[0.018]
Control group mean	0.000	0.000
Basic controls	Yes	Yes
Number of students	7,326	3,003

Notes: Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. Sample consists of Endline 1 respondents in the control group. All regressions control for grade-gender and district-gender fixed effects (column 2) or grade and district fixed effects (column 1) and missing flags for each variable used to construct the outcome index. Standard errors are clustered by school.

Appendix Table 10: Heterogeneity by gender, controlling for heterogeneity by BL attitudes

	Endline 1		Endline 2		
	Gender attitudes index (1)	Behavior index (2)	Gender attitudes index (3)	Behavior index (4)	Signed petition (5)
Treated	0.210***	0.258***	0.246***	0.307***	0.003
	[0.030]	[0.028]	[0.031]	[0.040]	[0.010]
Treated × Female	-0.055	-0.113***	-0.125***	-0.148***	0.016
	[0.040]	[0.036]	[0.041]	[0.043]	[0.016]
Treated × Baseline gender attitudes	0.020	-0.026	-0.006	-0.020	0.000
	[0.017]	[0.016]	[0.019]	[0.018]	[0.006]
Control group mean	0.000	0.000	0.003	0.000	0.150
Basic controls	Yes	Yes	Yes	Yes	Yes
Number of students	13,987	13,974	13,679	13,677	13,303

Notes: Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. All regressions control for grade-gender and district-gender fixed effects. Columns 1 to 4 also control for the baseline analogue of the outcome, its school-gender mean, missing flags for each variable used to construct the outcome index, and a missing flag for the baseline outcome. Columns 2, 4, and 5 also control for the main effect of the baseline gender attitudes index and its missing flag. Standard errors are clustered by school.

Appendix Table 11: Heterogeneity by gender, controlling for heterogeneity by wealth proxies

	Endline 1		Endline 2		
	Gender attitudes index	Behavior index	Gender attitudes index	Behavior index	Signed petition
	(1)	(2)	(3)	(4)	(5)
Treated	0.172***	0.271***	0.158***	0.300***	-0.002
	[0.051]	[0.051]	[0.055]	[0.062]	[0.020]
Treated × Female	-0.043	-0.127***	-0.140***	-0.147***	0.017
	[0.039]	[0.036]	[0.041]	[0.043]	[0.016]
Treated × Father works full-time	0.021	-0.033	0.114**	0.071	-0.003
	[0.045]	[0.037]	[0.047]	[0.046]	[0.016]
Treated × House is pukka	0.015	0.001	-0.017	-0.003	0.009
	[0.037]	[0.032]	[0.043]	[0.038]	[0.016]
Treated × Dwelling has flush toilet	-0.044	0.025	-0.027	-0.003	0.019
	[0.043]	[0.042]	[0.050]	[0.051]	[0.020]
Treated × Household gets newspaper daily	0.059	-0.027	0.042	0.010	0.012
	[0.050]	[0.041]	[0.050]	[0.056]	[0.021]
Treated × Household owns some land	-0.001	0.075	0.061	-0.220***	-0.018
	[0.056]	[0.047]	[0.062]	[0.058]	[0.025]
Father works full-time	0.054*	0.016	0.019	-0.007	0.008
	[0.032]	[0.028]	[0.029]	[0.029]	[0.011]
House is pukka	0.044*	0.022	0.070**	-0.006	-0.004
	[0.023]	[0.023]	[0.031]	[0.026]	[0.009]
Dwelling has flush toilet	0.082***	0.043	0.084**	0.066**	0.008
	[0.029]	[0.029]	[0.037]	[0.032]	[0.015]
Household gets newspaper daily	0.059	0.076***	0.082**	0.083**	-0.003
	[0.038]	[0.029]	[0.037]	[0.038]	[0.015]
Household owns some land	0.093**	0.033	0.066	0.021	0.007
	[0.041]	[0.031]	[0.045]	[0.041]	[0.016]
Treated + Treated × Female = 0	0.015	0.003	0.761	0.006	0.509
Control group mean	0.000	0.000	0.003	0.000	0.150
Basic controls	Yes	Yes	Yes	Yes	Yes
Number of students	13,987	13,974	13,679	13,677	13,303

Notes: Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. All regressions control for grade-gender and district-gender fixed effects. Columns 1 to 4 also control for the baseline analogue of the outcome, its school-gender mean, missing flags for each variable used to construct the outcome index, and a missing flag for the baseline outcome. All regressions include missing flags for each wealth variable; missing values are imputed with the gender-district average. Standard errors are clustered by school.

Appendix Table 12: Treatment effects on gender attitudes sub-indices

	Education attitudes	Employment attitudes	Attitudes towards female gender roles	Fertility attitudes
	(1)	(2)	(3)	(4)
Panel A: Endline 1				
Treated	0.129***	0.195***	0.193***	0.016***
	[0.021]	[0.019]	[0.020]	[0.006]
Control group mean	0.000	0.000	0.000	0.879
Number of students	13983	13984	13984	13508
Panel B: Endline 2				
Treated	0.125***	0.200***	0.169***	0.020***
	[0.022]	[0.022]	[0.021]	[0.005]
Control group mean	0.001	0.015	0.012	0.913
Number of students	13674	13677	13674	13373

Notes: Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. All regressions control for grade-gender and district-gender fixed effects and missing flags for each variable used to construct the outcome index. All columns except column 4 also include the baseline analogue of the outcome, its school-gender mean, and a missing flag for the baseline outcome. Standard errors are clustered by school.

Appendix Table 13: Treatment effects on behavior sub-indices

	Interaction with the opposite sex	Participation in HH chores	Supporting female relatives' ambitions	Girls' decision-making	Girls' mobility
	(1)	(2)	(3)	(4)	(5)
Panel A: Endline 1					
Treated	0.264***	0.040**	0.133***	0.017	0.027***
	[0.031]	[0.019]	[0.019]	[0.028]	[0.008]
Control group mean	0.000	0.000	0.000	0.000	0.908
Number of students	13504	13969	13927	7791	7510
Panel B: Endline 2					
Treated	0.265***	0.020	0.147***	-0.006	0.138***
	[0.030]	[0.018]	[0.019]	[0.030]	[0.032]
Control group mean	0.000	0.000	0.000	0.000	0.000
Number of students	13675	13668	13480	7544	7544

Notes: Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. All regressions control for grade-gender and district-gender fixed effects (columns 1–3) or grade and district fixed effects (columns 4–5) and missing flags for each variable used to construct the outcome index. All columns except column 3 also include the baseline analogue of the outcome, its school-gender mean, and a missing flag for the baseline outcome. Standard errors are clustered by school.

Appendix Table 14: Treatment effects on school performance (Endline 1)

Panel A: SCERT school data (2014-16)

	Proportion scoring >50 in...					
	Hindi	English	Math	Science	Social Science	All subjects
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	0.013	-0.007	0.012	-0.020	-0.012	-0.002
	[0.018]	[0.019]	[0.018]	[0.019]	[0.018]	[0.008]
Control group mean	0.547	0.429	0.348	0.506	0.458	0.320
Control SD	0.137	0.148	0.148	0.151	0.145	0.061
Number of schools	234	230	229	228	228	237

Panel B: 10th board exam data (2017)

	Proportion passing in...					
	Hindi	English	Math	Science	Social Science	All subjects
	(1)	(2)	(3)	(4)	(5)	(6)
Treated	0.001	-0.005	-0.010	-0.011	-0.021	-0.013
	[0.010]	[0.023]	[0.027]	[0.026]	[0.022]	[0.027]
Control group mean	0.924	0.667	0.751	0.735	0.733	0.552
Control SD	0.092	0.223	0.253	0.260	0.198	0.263
Number of schools	307	307	307	307	307	307

Notes: Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. Each observation is a school. We were able to match 237 and 307 sample schools with the SCERT and board exam datasets, respectively. The first panel uses data for both cohorts in our sample, from when each was in Grade 8. The second panel uses only the older cohort because the outcome is an exam taken in Grade 10 and the younger cohort was in Grade 9 at the time of these data. Some schools have missing observations in the SCERT dataset for certain subjects, so the sample size varies across columns within the first panel. All columns control for district fixed effects. Standard errors are heteroskedasticity-robust.

Appendix Table 15: Robustness check for social desirability bias (Endline 2)

	Gender attitudes index (1)	Girls' aspirations index (2)	Behavior index (3)	Applied to scholarship (4)	Signed petition (5)
Treated	0.183***	-0.010	0.212***	0.027	0.004
	[0.024]	[0.025]	[0.029]	[0.020]	[0.010]
Low social desirability score	-0.089***	-0.030	-0.058**	-0.026	-0.013
	[0.024]	[0.022]	[0.023]	[0.018]	[0.009]
Treated × Low social desirability score	-0.010	-0.035	0.033	0.013	0.019
	[0.034]	[0.035]	[0.032]	[0.024]	[0.013]
Treated + Treated × Low SD = 0	0.000	0.100	0.000	0.079	0.049
Control group mean	0.003	0.000	0.000	0.406	0.150
Basic controls	Yes	Yes	Yes	Yes	Yes
Number of students	13,679	7,560	13,677	7,347	13,303

Notes: Asterisks denote significance: * $p < .10$, ** $p < .05$, *** $p < .01$. Social desirability (SD) score is a baseline measure of the student's propensity to give socially desirable answers. Low SD score refers to having a below-median score among students. All regressions control for grade-gender and district-gender fixed effects (columns 1, 3, and 5) or grade and district fixed effects (columns 2 and 4). Columns 1 to 3 also control for the baseline analogue of the outcome, its school-gender mean, missing flags for each variable used to construct the outcome index, and a missing flag for the baseline outcome. Standard errors are clustered by school.

A Data Appendix

A.1 Sample selection and tracking

From the universe of 607 government secondary schools in the study districts, we focused on 346 schools with enrollment of at least 40 students in grades 6 and 7 combined based on DISE (2011) data.[27] If a boys' and a girls' school shared a building or were adjacent to each other, we considered them a single school. We conducted pilot activities in 6 of the schools. We removed 26 schools from the list for the following reasons (a) if there were multiple schools on the list in a village, we randomly selected one (b) based on initial visits, actual attendance was much lower than 40 students or (c) school officials did not agree to let us conduct surveys, despite our letter of approval from the Government of Haryana. The remaining 314 schools constitute the sample.

We distributed consent forms to 30,685 students, 84% of whom returned the form signed by their parent or guardian. Most students who did not have the signed consent form when the enumerators visited said they lost it or forgot their signed form at home. Providing consent is uncorrelated with student gender, which is suggestive that it was not driven by parental gender attitudes. The school-level consent rate is uncorrelated with village-level measures of gender equality. Our sample of students for each school was randomly selected from those returning the consent form who were present on the baseline survey day in their school and assented to participate.

For the baseline parent survey, if after multiple visits and follow-up phone calls, we could not interview the selected parent, we randomly chose a replacement household. We collected data for 2,546 fathers and 3,476 mothers. The completion rate was higher for mothers than for fathers because fathers were more often unavailable during daytime hours due to work. We did not survey parents at endline.

To reduce sample attrition, we conducted two tracking surveys to verify respondents' contact information between baseline and the first endline, in January to March 2015 (98.5% tracking rate) and February to June 2016 (93.8%). We also conducted a tracking survey between the first and second endlines, in February to July 2018 (96.4%). They were conducted through a combination of in-person visits and phone calls and verified the respondent's and parents' contact information and asked about intentions to move.

A.2 Primary outcomes

A.2.1 Procedure for index construction

Most of the outcomes variables are constructed by aggregating the responses to several individual questions into an index. The index is the weighted average value of the individual variables, with weights constructed by normalizing the variables to have the same standard deviation and then recovering the weights from the inverse covariance matrix, following the procedure of Anderson (2008). The steps involved in producing the final indices are as follows:

1. The individual variables are first converted to dummy variables. For questions that used a 5-point Likert scale, the binary variable was coded as 1 if the respondent answered "Strongly Agree" or "Agree" with a gender-progressive statement (or "Strongly Disagree" or "Disagree" with a gender-regressive statement), and 0 otherwise.

[27] Because it excludes schools with low enrollment, our sample has, on average, larger villages than the universe of villages with government secondary schools. In addition to government schools, there are 731 private unaided secondary schools in the four districts, which are disproportionately in urban areas.

2. For the purpose of constructing the weights (but not for the final outcome variable), we impute missing values with gender-district-treatment averages. This is done to enable us to invert the covariance matrix to calculate the weights.

3. Each individual variable is normalized by subtracting the overall sample mean and dividing by the control group standard deviation.

4. Weights are generated from the inverted covariance matrix of all the normalized and imputed variables in the respective index. For some index P consisting of variables a, b, c, and d, let the weights thus produced be wt_a, wt_b, wt_c, and wt_d.

5. If an observation has missing variables (which occurs, for example, because we asked some questions to only a random subset of respondents), we construct the index using only non-missing items. We weight the non-missing variables by their respective weights and normalize by the appropriate sum of weights. For example, if a, b, and c are non-missing, the sum $S = a \times wt_a + b \times wt_b + c \times wt_c$. Let W be the sum of weights for each variable, whether missing or not. So, $W = wt_a + wt_b + wt_c + wt_d$. Let N be the sum of weights of the non-missing variables. So, $N = wt_a + wt_b + wt_c$. Then the index is calculated as $S \times (W/N)$.

6. This weighted index is then re-scaled such that the control group mean is 0 and the standard deviation is 1.

7. In our regression specifications, we control for flags for each variable in the index, indicating whether it is missing.

8. For the gender attitudes index at endline 2 (and its sub-indices), we deviate from steps 4 to 6 above. The questions in the index were identical at endline 1 and endline 2, so to make these two outcome variables identical, we use the weights generated with the endline 1 data when we construct the endline 2 attitudes index. The final index is re-scaled by subtracting the endline 1 control group mean and dividing by the endline 1 control group standard deviation.

A.2.2 Gender attitudes index

The baseline attitude index aggregates the following 9 survey questions.

B1. A woman's most important role is being a good homemaker

B2. A man should have the final word about decisions in his home

B3. A woman should tolerate violence to keep her family together

B4. Wives should be less educated than their husbands

B5. Boys should get more opportunities/resources for education

B6. Men and women should get equal opportunities in all spheres of life

B7. Girls should be allowed to study as far as they want

B8. Daughters should have a similar right to inherited property as sons

B9. It would be a good idea to elect a woman as the village Sarpanch

The endline index aggregates 17 variables created from responses to 18 questions. Both the endlines use the questions listed here. We also divide the attitude questions into four mutually exclusive sub-indices for auxiliary analysis: gender equality in education, gender equality in employment, women's roles, and fertility preferences.

Education attitudes

E1. Wives should be less educated than their husbands

E2. Boys should be allowed to get more opportunities and resources for education than girls

E3. Education Vignette: If you were the head of the family, who would you have sent to the town for further studies?[28]

Employment attitudes

E4. A woman's most important role is to take care of her home, feeding kids and cook for her family

E5. Men are better suited than women to work outside of the house

E6. Work Vignette: Marriage is more important for Pooja than her job[29]

E7. Work Vignette: Being a teacher would be a more suitable job for Pooja

E8. Do you think women should be allowed to work outside home?

Women's role attitudes

E9. Daughters should have a similar right to inherited property as sons

E10. It would be a good idea to elect a woman as the village Sarpanch

E11. A man should have the final word about decisions in his home

E12. A woman should tolerate violence in order to keep her family together

E13. Parents should maintain stricter control over their daughters than their sons

E14. Girls [boys] should attain higher education so that they find better husbands [wives][30]

E15. At what age would you like your sister/female cousins/friends to get married *minus* At what age would you like your brother/male cousins/friends to get married?[31]

Fertility attitudes

E16. Suppose the first two children born to a husband and wife are both girls. Which of the following should they do? *minus* Suppose the first two children born to a husband and wife are both boys. Which of the following should they do?[32]

[28]This question was based on a vignette about a family deciding whether to send a son or daughter to further schooling. The variable was coded as 1 if the respondent said the daughter or both children, and 0 if they answered the son.

[29]Based on a vignette about a young woman named Pooja who wants to delay marriage to pursue a job as a police officer.

[30]The variable is coded as gender progressive if the respondent gave the same responses to the question about boys and the question about girls.

[31]We code two dummies from this, the first for saying that the age for girls should be > 19 and the other for stating a gap in the appropriate age between boys and girls that was larger than the control group median response.

[32]Coded as gender regressive if the respondent said "have no more children" after having two boys but not after having two girls, and gender progressive otherwise.

A.2.3 Aspirations index

We construct a gender aspirations index that measures educational and occupational aspirations for girls only. The questions used for the baseline aspirations index were as follows.

B1. Have you ever discussed your education goals with your parent or adult relative?

B2. What is the highest level of education you would like to complete if finances and opportunity of the school/college are available?

B3. What occupation do you expect to have when you are 25 years old?

The questions used for the aspirations index in the first endline were as follows.

E1. How many marks, according to you, will you score in the SSE 10th board examinations?

E2. Have you ever discussed your education goals with your parents or adult relatives?

E3. Suppose you were to get married right after school, would you want to continue your education after marriage?

E4. What is the highest level of education you would like to complete if finances and opportunity of the school/college are available?

E5. What occupation do you expect to have when you are 25 years old?[33]

The endline aspirations index is missing for a few observations because the respondent stopped the survey midway or refused to answer that module.
The questions used for the aspirations index in the second endline were as follows.

E2.1. How many marks, according to you, will you score in the SSE 12th board examinations?[34]

E2.2. Have you ever discussed your education goals with your parents or adult relatives?

E2.3. Suppose you were to get married right after school, would you want to continue your education after marriage?

E2.4. What is the highest level of education you would like to complete if finances and opportunity of the school/college are available? [35]

E2.5. What occupation do you expect to have when you are 25 years old?[36]

E2.6. Do you plan to go to college/pursue a vocational course/professional course/join civil services or army?

E2.7. What course would you like to pursue for higher studies?[37]

E2.8. I would like to have a job outside the home that I continue to pursue when I am married and have children.

[33]White collar occupations are coded as more progressive.

[34]Coded as 1 if the listed marks were greater than the gender-control group median and 0 otherwise. Question was only asked to students currently enrolled in grades 11 and 12.

[35]Coded as 1 if the level of education is greater than the gender-control group median and 0 otherwise.

[36]Coded as 1 if the respondent is able to report her expectations about having a job irrespective of the nature/type and 0 otherwise.

[37]Coded as 1 if the respondent is able to report any course irrespective of the nature/type and 0 otherwise.

A.2.4 Gender behavior index

We construct a gender behavior index that measures gender equitable behavior. Questions marked with # are coded with opposite signs for boys and girls. The questions used for the baseline behavior index were as follows.

B1. Are you comfortable talking to children of the opposite gender who are not related to you inside or outside school?

B2. In the past week, did you help with cooking/cleaning/washing clothes?#

The endline 1 behavior index was constructed using the following questions. Questions marked with # are coded with opposite signs for boys and girls. Questions marked with $*$ were also asked in the second endline.

Interaction with the opposite sex

E1. Are you comfortable talking to children of the opposite gender who are not related to you inside and outside school?*

E2. Do you sit next to students of the opposite gender in class?*[38]

Participation in household chores

E3. In the past week, did you cook/clean/wash dishes?#*

E4. In the past month, have you missed school due to household based responsibilities?#*

Supporting female relatives' ambitions

E5. Do you discourage your sister/female cousin from working outside home?*

E6. Do you discourage your sister/female cousin from studying in college if it is far away?*

The following questions only pertain to girls and are not included in the main behavior index. These questions are included in the construction of behavior sub-indices.

Girls' decision-making

E7. I am able to talk to my parents about what work I would like to do in the future.*

The next 3 questions, E8 to E10, ask about decision-making using the following structure: "Who mostly makes decisions about the following, or if this is in the future for you, who do you expect will make this decision? Will you make the decision, make the decision jointly with parents or will parents make the decision for you?"[39]

E8. Whether or not you will continue in school past grade 10 (grade 12 in the second endline)*

E9. If you will work after you finish your studies*

E10. What type of work you will do after you finish your studies*

E11. How many days were you absent from school last week?[40]

Girls' mobility

[38] This question was not asked in single-sex schools.

[39] Coded as 1 if the respondent alone makes the decision and 0 otherwise.

[40] Coded as 1 if the respondent was not absent to school in the previous week and 0 otherwise.

E12. Are you allowed to travel to school alone or with friends?*

At endline 2, we excluded some behavior questions used at endline 1 (those not marked with * above) and added the following new questions to the index:

Interaction with the opposite sex

E2.1 Is friends with the opposite gender/has friends from the opposite gender.

E2.2 Plays with the opposite gender (who are not related to him/her) inside or outside of school.

E2.3 In the past one week, spoke with children (not related to him/her) of the opposite gender inside or outside of school.

The following new questions at endline 2 only pertain to girls and are not included in the behavior index. The questions are included in the 'girls' mobility' sub-index for the second endline.

Girls' mobility

E2.4. Has gone to the market within his/her village to buy personal items alone.

E2.5. Has attended community events without guardians present (either alone or with friends).

E2.6. In the past one week, has gone out of his/her house alone for any kind of purpose.

The endline behavior index is missing for a few observations because the respondent stopped the survey midway or refused to answer that module.

A.2.5 Revealed preferences measures - Endline 2 only

In the second endline, we included two revealed-preference measures. First, we offered girls an opportunity to apply for a financial scholarship to go toward college education or other post-secondary training. Second, we gave both girls and boys the opportunity to pledge support for a public petition to abolish the dowry system.

We set up a scholarship program that offered a Rs. 10,000 (150 USD) scholarship to each of 20 winners. At the end of the in-person endline survey, we informed girls about the scholarship and gave them the application form. To apply, they had to fill it out and mail it in by the stated deadline. The forms had a unique student ID, so we used the mailed-in applications to measure whether respondents applied. We randomly varied the degree of parental endorsement required on the application. Half of girls received a basic application on which they had to fill out basic information about themselves and the school and course they would like to pursue. The form also required a parental signature that stated that they understand the terms and conditions of the scholarship. In the second version of the form, there was an additional section that had to be filled by the parent or in consultation with the parent and had a weightier parental declaration that stated that they support their daughter's decision to attend college and apply for the scholarship. As pre-specified, we pool them in our analysis. (We do not see a difference in response rates between the two versions, in the control or treatment group.)

For the petition, at the end of the endline 2 survey, the enumerators informed respondents about a petition against dowry. We told respondent that names and villages of signatories would be printed in their local newspaper (and we then ran newspapers advertisements to do so). They were asked to call a toll-free number to register their support. We left a flyer with information on the petition text and the number to call. Due to a technical problem with the toll-free vendor, we lost 6 days' worth of data on potential calls from one phone carrier. Thus, we called those missed calls back to inquire if the respondents wanted to record their support.

A.3 Secondary outcomes

A.3.1 Social norms

The following questions were asked during both the endlines. Students were randomized to receive either Set 1 questions or Set 2 questions.

Set 1

E1. Do you think that women should be allowed to work outside home?

E2. Do you think that people in your village/community think that women should be allowed to work outside home?

E3. Do you think the community will oppose you since [if] you disagree with them (regarding women being allowed to work)?

Set 2

E1. Do you think that girls should be allowed to study in college even if it is far away?

E2. Do you think that people in your village/community think that girls should be allowed to study in college even if it is far away?

E3. Do you think the community will oppose you since [if] you disagree with them (regarding women being allowed to study in college)?

A.3.2 School performance - Endline 1 only

We examined academic outcomes to test if the intervention crowded out other academic instruction. We used overall pass rates and subject-wise test scores from two data sources:

- State Council of Educational Research and Training (SCERT): We were able to match 237 sample schools with the SCERT data. We have data for both cohorts in our sample, from when each was in Grade 8.
- Haryana Board of School Education: We were able to match 307 sample schools with the board exam dataset. We have data for only the older cohort of our sample because the outcome is 10th grade exam, and the younger cohort had not taken the exam at the time of this data collection.

A.3.3 Girls' self-esteem index

E1. On the whole, I am satisfied with myself

E2. I feel that I have a number of good qualities

E3. I am able to do things as well as most other people

A.3.4 Gender discrimination awareness index - Endline 1 only

E1. Do you know about female feticide and infanticide?

E2. Are female feticide and infanticide practiced in Haryana?

E3. According to you, what is the main reason for female feticide and infanticide?[41]

E4. In Haryana, are the number of girls less than the number of boys?

[41] Coded as 1 if any reason(s) given, 0 if respondent says "don't know" or doesn't answer.

A.3.5 Implicit association tests - Endline 1 only

We use two gender-related implicit association tests as secondary outcomes. Fifty percent of all student respondents took an IAT associating good-bad behavior characteristics to boys and girls during baseline. During endline, the same students were administered either the baseline IAT or a second IAT which asked them to associate gender stereotypical activities like factory work and ironing clothes to men and women. We use as outcomes the implicit association of girls with positive words for the first IAT, and of women with market work.

The IATs were created using Millisecond Software and administered on laptops. We aimed to collect each IAT for 25% of the sample but the usable sample size is smaller because, following guidelines for processing IAT data, we exclude observations that were completed very quickly or slowly (faster than 300 milliseconds or slower than 10 seconds on >10% of trials).

A.3.6 Observed classroom behavior - post-Endline 1 only

After analysis of our Endline 1 data was complete and we had presented our results to some audiences, based on feedback, we decided to collect objectively measured gender-related behaviors in 2017. We developed and conducted three activities in the co-ed schools in our sample. The three activities aimed to measure (1) girls' participation in classroom discussions (2) students' views about girls' knowledge, and (3) interaction with opposite-gender peers in the classroom. In activity (1), a surveyor facilitated a class discussion about "What changes do you want to see in your society?" Another surveyor took note of how many girls and boys made comments in the discussion. In activity (2), students were told about an inter-school competition based on a general knowledge quiz. The winning classroom in each district would get school bags for every student in the class. Students were asked to vote for three students in their class to represent them. The outcome is how many girls are elected for the quiz competition. For activity (3), students were asked to form groups of five for a poster-making activity about "Swachh Bharat Abhiyan" (India's Cleanliness Drive). The surveyor recorded how many of the groups were mixed-gender.

There were some major limitations of this exercise. First, our pilot activities were too limited to reveal to us that there is no gender gap in class participation in the status quo (i.e., in the control group), and students do not perceive girls' knowledge as lower than boys', making the first two activities ill-suited for testing for changes in gender roles and stereotypes. Second, we have low power to detect changes in the outcomes, partly because we only received permission from principals to conduct the exercises in 197 schools. Also, for our third outcome (co-ed poster-making teams), only 5% of self-formed groups in the control group were mixed-gender, so we only have power to detect a very large proportional increase in this outcome. With those caveats, we find no significant effect on these outcomes. Results available upon request.

A.3.7 Girls' education index - Endline 2 only

E2.1. Which school are you enrolled in?[42]

E2.2. What stream are you currently following?[43]

E2.3. In the past one year, have you enrolled for an English speaking, computer training, or vocational class?

E2.4. Do you take after-school/college tuitions?

[42]Coded as 1 if respondent is in any school or college, including open school, and 0 if dropped out or in vocational training.

[43]Coded as 1 if pursuing a science, commerce with math, or arts with math stream.

A.3.8 Marriage and fertility aspirations - Endline 2 only

E2.1. At what age do you want to marry?[44]

E2.2. At what age do you want to have your first child?[45]

E2.3. How many children do you want to have? How many of these children would you like to be boys, how many would you like to be girls?[46]

E2.4. Suppose your spouse and you are going to have N children, how many of them would you want to be boys?[47]

E2.5. If instead of X boys and $N - X$ girls, you could either have $X - 1$ boys and $N - X + 1$ girls OR $X + 1$ boys and $N - X - 1$ girls, which would you prefer?[48]

A.3.9 Girls' experience of sexual harassment/assault - Endline 2 only

The index is coded so that a higher value corresponds to more instances of harassment.

E2.1. In the past one year, have you ever been slapped, hit, or otherwise physically hurt by a boy in a way you did not want?

The following questions are coded as 1 if the incidence ever occurred, and 0 otherwise.

E2.2. How frequently have you been teased, whistled at, or called names by boys in school in a way you did not want?

E2.3. ...teased, whistled at, or called names by boys outside of school in your village/town in a way you did not want?

E2.4. ...touched or groped by boys in school in a way you did not want?

E2.5. ...touched or groped by boys in your village/town in a way you did not want?

A.3.10 Boys' engagement in sexual harassment/assault - Endline 2 only

We asked the boys in our sample about sexual harassment/assault, using list randomization. Half the boys in the sample, stratified by treatment, were given a list of 5 questions including the sensitive question. The other half were given a list of the 4 non-sensitive questions. They were asked how many of the statements were true without having to list which statements were true. We calculate the school-grade level differences between the mean true statements in the two sets as a measure of the proportion of boys who engage in harassment/assault. This outcome is analyzed at the school-grade level. Single-sex girls schools are thus excluded when we analyze this outcome.

The statements given to the respondents are as follows, with the sensitive item italicized:

E2.1. In the past year, I have made new friends.

[44] Coded as 1 if the age is greater than the gender-control group median and 0 otherwise.

[45] Coded as 1 if the age is greater than the gender-control group median and 0 otherwise.

[46] Coded as 1 if the number of girls desired is greater than or equal to the number of boys and 0 otherwise.

[47] The N was randomly generated integer between 1 and 5, inclusive. Question coded as 1 if the number of girls desired is greater than or equal to the number of boys and 0 otherwise.

[48] X is the number of boys that the respondents said they desire to have out of the randomly generated N number of children.The response options of the questions are as follows: (a) Prefer $X-1$ boys, and $N-X+1$ girl or (b) Prefer $X + 1$ boys, and $N - X - 1$ girls. The question was coded as gender progressive if the respondent chooses response option 1 and 0 otherwise.

E2.2. *In the past year, I have passed dirty comments about a girl; made dirty gestures in a girl's presence or inappropriately touched or groped a girl.*

E2.3. In the past year, I have gone on a vacation with my parents (to a relative's place etc.)

E2.4. In the past year, I have scolded my friend/cousin.

E2.5. In the past year, I have watched a program (sports, cultural etc.) on television.

A.4 Social desirability score

We use a 13-question short form of the Crowne and Marlowe (1960) module developed by Reynolds (1982). The following questions were asked at baseline with two answer choices: agree or disagree. The social desirability score sums how many of the responses are the socially desirable one. A low score means a lower tendency to give answers that have social desirability bias.

B1. It is sometimes hard for me to go on with my work if I am not encouraged

B2. I sometimes feel resentful when I don't get my way

B3. On a few occasions, I have given up doing something because I thought too little of my ability

B4. There have been times when I felt like rebelling against people in authority even though I knew they were right

B5. No matter who I'm talking to, I'm always a good listener

B6. There have been occasions when I took advantage of someone

B7. I'm always willing to admit it when I make a mistake

B8. I sometimes try to get even rather than forgive and forget

B9. I am always courteous, even to people who are disagreeable

B10. I have never been irked when people expressed ideas very different from my own

B11. There have times when I was quite jealous of the good fortune of others

B12. I am sometimes irritated by people who ask favors of me

B13. I have deliberately said something that hurt someone's feelings

A.5 Parent's gender attitude index

To understand how parental attitudes influence program impacts, one parent of a random 40% subsample of the surveyed students participated in a survey during baseline. The following questions were used to construct our parent's gender attitudes index at baseline.

B1. A woman's most important role is being a good homemaker

B2. A man should have the final word about decisions in his home

B3. A woman should tolerate violence to keep her family together

B4. Wives should be less educated than their husbands

B5. Boys should get more opportunities/resources for education

B6. Men and women should get equal opportunities in all spheres of life

B7. Girls should be allowed to study as far as they want

B8. Daughters should have a similar right to inherited property as sons

B9. It would be a good idea to elect a woman as the village Sarpanch

The heterogeneity analysis with the parent index is restricted to students whose parents were surveyed. We also use the parent attitude index as a possible control variable in our LASSO extended controls procedure. We impute missing values at the mean value for those students whose parent was not surveyed.

A.6 Extended control variables

The following variables are used in the LASSO procedure to select the extended controls.

Student-level variables from baseline survey

- Student's age
- Student lives with both his or her mother and father
- Number of sisters
- Number of brothers
- Mother has completed 8th grade
- Mother works part-time
- Mother works full-time
- Scheduled caste
- Scheduled tribe
- Muslim
- Asset variables: House is pukka, House is connected to electricity, Flush toilet, No flush toilet, Family owns the house, Household owns radio or tape recorder, Household owns TV/cable TV/satellite TV/dish TV, Household owns refrigerator, Household gets newspapers daily, Tap water, Household owns water pump
- Self-efficacy index[49]
- Social desirability score
- Parent's gender attitude index

School and village characteristics from other data sources

- Number of full-time teachers

[49]The self-efficacy index is constructed using the respondent's agreement with the following statements: (1) On the whole, I am satisfied with myself. (2) I enjoy learning. (3) I feel that I have a number of good qualities. (4) I am able to do things as well as most other people. (5) I help make my community a better place. (6) I am full of ideas. (7) I think about social problems. (8) I have parents who try to help me succeed. (9) Some people say that it is important to have definite opinions about lots of things, whereas other people think that it is better to remain neutral on most issues. I think it is better to have definite opinions.

- Number of guest (i.e., temporary) teachers
- Fraction female teachers
- Has school counselor
- PTA meetings
- Frequency of extracurricular activities
- School facilities (Indicators for has library, has toilets, has electricity, has computers, has internet, has sports field, has mid-day meals, has auditorium, has EduSat (satellite-based service delivering educational content)
- Co-ed school
- Rural location
- Village-level adult literacy rate by gender
- Village-level female labor force participation