IFS

Alison Andrew
Sonya Krutikova
Gabriela Smarrelli
Hemlata Verma

Working paper

22/41

Gender norms, violence and adolescent girls' trajectories: evidence from a field experiment in India

UKRI Economic and Social Research Council

Gender Norms, Violence and Adolescent Girls' Trajectories: Evidence from a Field Experiment in India[*][†]

Alison Andrew[‡]   Sonya Krutikova[§]   Gabriela Smarrelli[¶]   Hemlata Verma[‖]

Abstract

Striking gender gaps persist in fundamental aspects of human welfare. In India, the setting of this paper, these gaps are particularly large. Interventions often target adolescent girls with the aim of empowering them to make choices that go against the status quo – to remain in school longer or marry later, for example. This approach may inadvertently expose girls, who are often marginalized within their communities, to new risks if it encourages them to violate prevailing gender norms. In this study, we design an experiment to compare the effectiveness of targeting only adolescent girls with an approach that additionally engages with the enforcers of gender norms in the wider community. We find that both arms of the trial led to a reduction in school dropout and early marriage. We see large improvements in girls' mental health but only in the arm which engages with the wider community. Improvements in mental health can be explained by community engagement causing gender norms to become more progressive and causing a reduction in the severity of sanctions that girls face for breaking norms. Both adolescent girls and their mothers perceived these shifts in norms and sanctions. Our results demonstrate that in settings where unequal outcomes are sustained through restrictive gender norms, change in the attitudes and behavior of the enforcers of these norms is critical for achieving meaningful improvements in womens well-being.

1   Introduction

Striking gender gaps persist in some of the most fundamental aspects of human welfare, including health, education, the division of paid and unpaid work, consumption, and agency in decision-making. Many of these gaps are especially large in resource-poor contexts and can be exacerbated by poverty (Jayachandran 2015). However, it is becoming increasingly clear that economic development alone is not enough to close these gender gaps (Duflo 2012). India, the setting for this work, is a particularly stark example: despite rapid economic growth, there has been little progress on many dimensions of gender equality and even a marked deterioration in some, such as labor force participation. In light of this evidence, more and more attention is being paid to the role of "norms" – shared informal rules of appropriate behavior – in sustaining gender-unequal outcomes (Jayachandran 2021).

In this paper, we consider whether it is possible to improve women's welfare and make progress towards gender equity *despite* prevailing norms or whether real change requires a shift in restrictive norms and how they are enforced. This is a particularly important distinction in contexts, such as India, where inequalities in political voice and power mean that those who are most influential in creating and enforcing restrictive gender norms (community leaders, men, and older women) are a different social group from those whose behavior these norms control (adolescent girls and younger women) (Heller and Rao 2015).

Recent evidence shows that programs which encourage adolescent girls to question restrictive norms, and act to improve their outcomes in spite of these, *can* be effective, especially at improving girls' participation in education (Buchmann et al. 2017; Edmonds et al. 2021). However, these studies do not address the question of whether such approaches achieve improvements in women's well-being. Women and girls often face backlash from their families and communities when their behavior conflicts with accepted ideas of what is appropriate (Aizer 2010). Indeed, such sanctions may well be key to how norms are enforced and sustained. Policies that *only* target women, even if effective at changing particular outcomes, therefore risk reducing women's well-being if they expose women to an increased risk of violent sanctions.

We study a randomized experiment in rural Rajasthan, a context with highly restrictive gender norms. We compare two approaches to improving adolescent girls' education, delaying their marriage, and promoting well-being. The first approach engages only adolescent girls; it encourages them to change how they internalize prevailing norms and gives them strategies and motivation to go against these norms if they choose to. The second additionally seeks to change the prevailing norms, and the severity with which norms are enforced, by involving the wider community and, in particular, the community leaders who have the greatest influence in setting and enforcing gender norms. Critically, in addition to being the first study to compare the efficacy of these two approaches in improving adolescent girls' education and marriage outcomes, we consider the implications of these two approaches for the girls' mental health which we view as a key determinant of girls' well-being.

We start by setting out a simple conceptual framework that highlights the channels through which

prevailing norm structures in our study setting may affect girls' schooling choices and mental health. To motivate this framework, we use the rich data from our baseline to show that this is a setting in which strict rules exist over most dimensions of girls' lives and these rules are enforced through clear sanctions – including violence, harassment, shaming, and disapproval. Furthermore, we show that, for many girls, attending school requires actions such as traveling by bus, walking long distances, and getting home late, which directly conflict with existing norms. Importantly, we show high levels of private agreement with these restrictive norms suggesting that this is not a situation of "pluralistic ignorance" in which people misperceive others' attitudes (Bursztyn et al. 2018). Rather, our data suggest that norm change in this context requires community members to change their private beliefs (Jayachandran 2021).

Building on these insights, our conceptual framework captures the idea that both choices and girls' mental health will be shaped by norms through three channels: (i) girls' attitudes, internalized norms and motivation; (ii) externally set norms; and (iii) the severity of sanctions against girls who break the rules. It highlights that restrictive norms impose two types of costs on adolescent girls: the costs of incurring sanctions for those who violate the norms and the costs of forgoing schooling in order to avoid such sanctions. Our model implies a monotonically positive relationship between the progressivity of gender norms and educational attendance and a convex, possibly U-shaped, relationship between the progressivity of gender norms and girls' mental health. The rationale for this U-shaped relationship is that girls take into account both their immediate well-being and longer-run returns to education when making education decisions. This implies that the marginal girl who enters education in response to a small increase in the progressivity of gender norms may do so at the cost to her immediate well-being if the longer-run returns to education compensate for this. These costs arise because attending education requires violating norms and thus incurring harmful sanctions. We show robust correlational patterns in our data between education, mental health, gender norms, and the perceived likelihood of violent sanctions that are consistent with these key implications.

We use this framework to consider the mechanisms through which the two programs that we evaluate may operate. The first program, which we refer to as Girl Groups, consisted of group education and sports sessions for girls age 12–19. In these sessions, girls were encouraged to recognize restrictive gender norms, to understand their origin and arbitrariness, and to consider other, more gender-equitable, ideas of what constitutes appropriate behavior and success in life for women and girls. Therefore, the aim of the program was to influence girls' educational and marriage choices by changing their attitudes to these and altering their perspective on norms in ways that reduced the psychological cost of breaking them. Within our framework, we see this program as trying to influence channel (i) only – girls' attitudes, internalized norms, and motivation. In addition to the same group sessions with the girls, the aim of the second program, which we refer to as Girl Groups and Community Campaigns, was to also work through the other two channels (ii and iii) by changing externally set norms and the severity of sanctions against girls who break them. It did this by

targeting the broader community, with a particular focus on community leaders, through community events organized by the girls who participated in the Girl Groups. The discussions that took place during the events focused on issues affecting young women, advocating for more equitable treatment and thus seeking to relax the external constraints faced by girls, including stringency of norms and sanctions.

We use a three-armed cluster randomized controlled trial across 90 clusters and covering over 5,000 adolescent girls in rural Rajasthan to evaluate and compare the effects of these two programs. We follow two different cohorts of girls who were unmarried at baseline: those aged 12–14 at baseline, and those aged 15–17 at baseline. The girl groups were formed within each of these cohorts, and the specific content of the interventions varied slightly across cohorts. We find that, alone, the Girl Groups intervention led to a significant increase in the proportion of girls enrolled in and attending school or post-secondary education by 4.0 percentage points (hereafter p.p.; $p = 0.018$). This was driven by older girls for whom gender norms were most in conflict with pursuing education. Among these girls, the increase in educational attendance was 6.1 p.p. which corresponds to a 15% increase in educational attendance relative to the control group. This older group was also most at risk of early marriage during the study period. We find a significant 20% reduction in the likelihood of them being married at the time of follow-up, relative to the control group (effect size 3.6 p.p.; $p = 0.036$). This effect is even larger (5.3 p.p) once we include girls who were engaged or whose marriage had been fixed by the time of the endline. We find that this reduction in marriage rates is correlated with staying in education longer.

The addition of Community Campaigns did not change the effects on education and marriage: in both the Girl Groups and the Girl Groups and Community Campaigns programs, we see the same pattern of impacts on these outcomes, and we cannot reject the hypothesis that the size of these on each outcome is the same in the two arms of the trial. Strikingly, however, despite the significant improvements in education and marriage outcomes in both arms, we only see an improvement in the girls' mental health in the Girl Groups and Community Campaigns arm. We show that the Girl Groups program alone had no significant impact on the prevalence of symptoms of depression and anxiety among the targeted girls. In contrast, we see large improvements in both sets of symptoms with the addition of Community Campaigns, which is significantly different from the null effect in the Girl Groups arm. Impacts are especially large for the older girls; in this group, we see an improvement of 0.38 of a standard deviation in a combined mental health index relative to the control group.

Comparing the impacts of the two arms suggests that holding fixed girls' behavior, the Community Campaigns resulted in improvements in girls' mental health. Our conceptual framework suggests that these improvements may stem from changes in gender norms or changes in the likelihood of sanctions associated with breaking norms. We see clear empirical evidence that both mechanisms were at play. Girls and their mothers reported more progressive gender norms and girls perceived a lower likelihood of violent sanctions for breaking gender norms.

This study contributes to several literatures. The first of these is on how long-standing and restrictive norms change. To the best of our knowledge, this is the first study to isolate, quantitatively, the additional impact of engaging the wider community in an intervention that aims to improve the outcomes of adolescent girls. Dialogue and deliberation, particularly when involving members of the community with more power, are often cited as key to how community-wide norms are upheld and how they change (Heller and Rao 2015). Despite the fact that adolescent girls are often a marginalized group within their communities, much of the existing literature on improving the outcomes of adolescent girls in contexts with restrictive norms intervenes solely with this group (Adoho et al. 2014; Buehren et al. 2016; Bandiera et al. 2019; Buchmann et al. 2017; Bergstrom and Özler 2021). A few studies look at targeting young men and women together (Dhar et al. 2018; Edmonds et al. 2021) and men alone (Barker et al. 2017). None compare the efficacy of targeting different groups for creating norm change. The fact that we observe strong impacts on gender norms and sanctions from the community campaigns (reported by both girls and their mothers) supports models of norm change in which the process begins by those with new or different ideas seeking to persuade others of a need for a new norm through appealing to their empathy and altruism (Finnemore and Sikkink 1998).[1] It highlights how in order to be persuasive, marginalized groups may require the type of platform and status offered by the structure of the intervention.

We also contribute to work on the welfare impacts of restrictive norms and interventions aiming to help the subjects of these restrictive norms. Our framework makes clear that restrictive and enforced norms impose two distinct sets of costs on their subjects: the costs of incurred sanctions and the costs of forgone actions. We contribute to understanding both.

First, restrictive norms create a direct cost to the safety and well-being of those who go against the norm and who are thus the subject of sanctions. Existing evidence has shown that women may face higher risks of divorce (Bertrand et al. 2015) or violence (Krishnan et al. 2010; Guarnieri and Rainer 2018) for violating gender norms. Women's psychological well-being may suffer when they deviate from their families' wishes (Ashraf et al. 2014), perhaps out of fear of retaliation. This is closely related to the notion of backlash in which dominant parties (e.g. husbands or male community members) may use violence to reinforce their own dominance in response to improvements in the status of women (Aizer 2010). We add to this literature by showing that in our study context both young women and their mothers are acutely aware that breaking gendered norms of behavior puts young women at risk of sanctions, that the perceived likelihood of sanctions and the restrictiveness of gender norms are strongly related to girls' mental health, and that experimentally loosening the strictness of gender norms and the severity of their enforcements leads to large improvements in mental health.

Second, our framework highlights how the threat of sanctions prevents the subjects of restrictive norms from engaging in valuable activities. A large and growing literature has suggested that restrictive and enforced norms appear key to explaining, for example, gender inequalities in the labor

---

[1]These theories highlight that once a critical mass of people support a new norm, the process may accelerate.

market (Field et al. 2019; Jayachandran 2021), in entrepreneurship (Field et al. 2010; Field et al. 2016), in education (Beaman et al. 2012), and in social networks and political influence (Sanyal 2009; Kandpal and Baylis 2019). This is one of a handful of studies assessing whether group-based interventions designed to help young women thrive despite restrictive gender norms by targeting their enthusiasm and motivation for school, attitudes and internalized norms can impact behavior. Adolescent intervention "bundles" often include such group-based activities, but the impact of these on girls' outcomes cannot be evaluated separately from other intervention components that they are bundled with (Adoho et al. 2014; Buehren et al. 2016; Bandiera et al. 2019). We study a stand-alone girl group program which had not other components (such as providing girls with marketable skill, financial resources etc); we are thus able to attribute the positive effects we find on girls' behavior to the group activities. Recently, Edmonds et al. (2021), has shown that a life skills program administered to adolescent girls, also in rural Rajasthan, resulted in a significant reduction in school dropout and increased grade progression, mediated by an improvement in several life skills and gender attitudes. In the context of Bangladesh, Buchmann et al. (2017) find less promising results from a similar program. They found only marginal increases in schooling over the medium term and no impacts on early marriage. Our findings complement the existing evidence by suggesting that even working with girls alone, both educational and marriage outcomes can be shifted. However this is achieved by helping girls stay in education despite restrictive norms (by making the girls see greater value in education, for example) rather than changing these norms. We interpret this as further evidence that explicit engagement with more powerful members of the community is necessary to facilitate norm change.

The rest of the paper is structured as follows. We start (in Section 2) with a more detailed description of the study context, paying particular attention to characterizing the prevailing gender norms and attitudes. We then propose a simple conceptual framework to consider the channels through which the two programs, which we describe in detail, could affect girls' outcomes. In Section 3, we present the study design, a description of the data, and our empirical strategy. We present our results in Section 4, and discuss potential mechanisms in Section 5. We conclude in Section 6.

2 Context, Conceptual Framework, and Interventions

The setting for this paper is 125 rural villages in Dholpur district in the state of Rajasthan. Rural Rajasthan has some of the highest rates of early marriage and school dropout amongst young women in India. For example, while nationally around 35% of women have at least 10 years of education, only 13% of women in Dholpur have attained this level.[2] At 50%, female literacy is 35 percentage points lower in Dholpur than the national average. A high degree of son preference is evident in the

particularly skewed sex ratio: 845 girls to 1,000 boys in 2011, compared to 940 girls nationally. Early marriage is also pervasive: 40% of married women aged 20–24 had been married before the age of 18 here, compared to 25% nationally. These indicators motivated the Government of India to name Dholpur district a "priority district".

2.1 Prevailing Gender Norms and Their Enforcement

Our data offer us an insight into prevailing gender norms as perceived by adolescent girls and their families in our study. The data highlight just how pervasive norms are, the degree to which they reach the most mundane aspects of girls' daily lives, the degree to which they are accepted, and the perception of likelihood of sanctions for violating these norms. We consider these details as key for understanding how girls and their families make choices, and how these choices might be affected by programs such as the ones we evaluate in this paper.

Figure 1 presents the answers that both adolescent girls and their mothers gave during the study's baseline when asked about their own beliefs regarding the informal rules governing the behavior of girls and women. Specifically, girls and their mothers were asked the extent to which they agreed with a set of statements expressing strongly patriarchal rules. Panel A of Figure 1 shows that the great majority of mothers and girls agree that girls should primarily stay at home, only venturing into public spaces when absolutely necessary and, ideally, only when accompanied. It shows that such beliefs are especially strong in relation to girls who have begun their periods and that socializing with boys is particularly frowned upon. Panel B of Figure 1 goes on to show broad agreement that women should be responsible for homemaking and play a limited role in family decisions, even ones that greatly affect them.

Figure 2 further suggests that women and girls are not only aware of these norms but have also internalized and accepted the notion that those who break the norms should be punished, including violently; they expect adolescent girls to experience negative repercussions from breaking norms. For example, half of the mothers and girls in the sample believe that if a girl was to walk home from school with a male friend then he would act inappropriately towards her, while 59% of both mothers and girls believe girls will likely experience unwanted attention if they walked home alone.

Figure 3 makes clear that many of these norms are at odds with the reality of what is required to attend school. It shows that girls frequently have a long commute to school; even in upper primary, a fifth of girls commute for more than one hour, rising to three-fifths for girls in higher education. For upper secondary and beyond (beginning in 10th standard), commutes are frequently longer than 5 km and take place on public or shared transportation.[3] With the realities of delayed, overcrowded buses and the mixed-gender passengers on public transport, simply getting to and from school is likely to conflict with the gender norms we have documented. This suggests that norms relating to

---

[3]These averages likely underestimate the difficulty of the commute for the average girl in this sample because they are only for girls who have continued in education; it is probably the case that girls with the most difficult commutes are more likely to have dropped out of school.

Figure 1: Gender Norms as Understood by Mothers and Girls
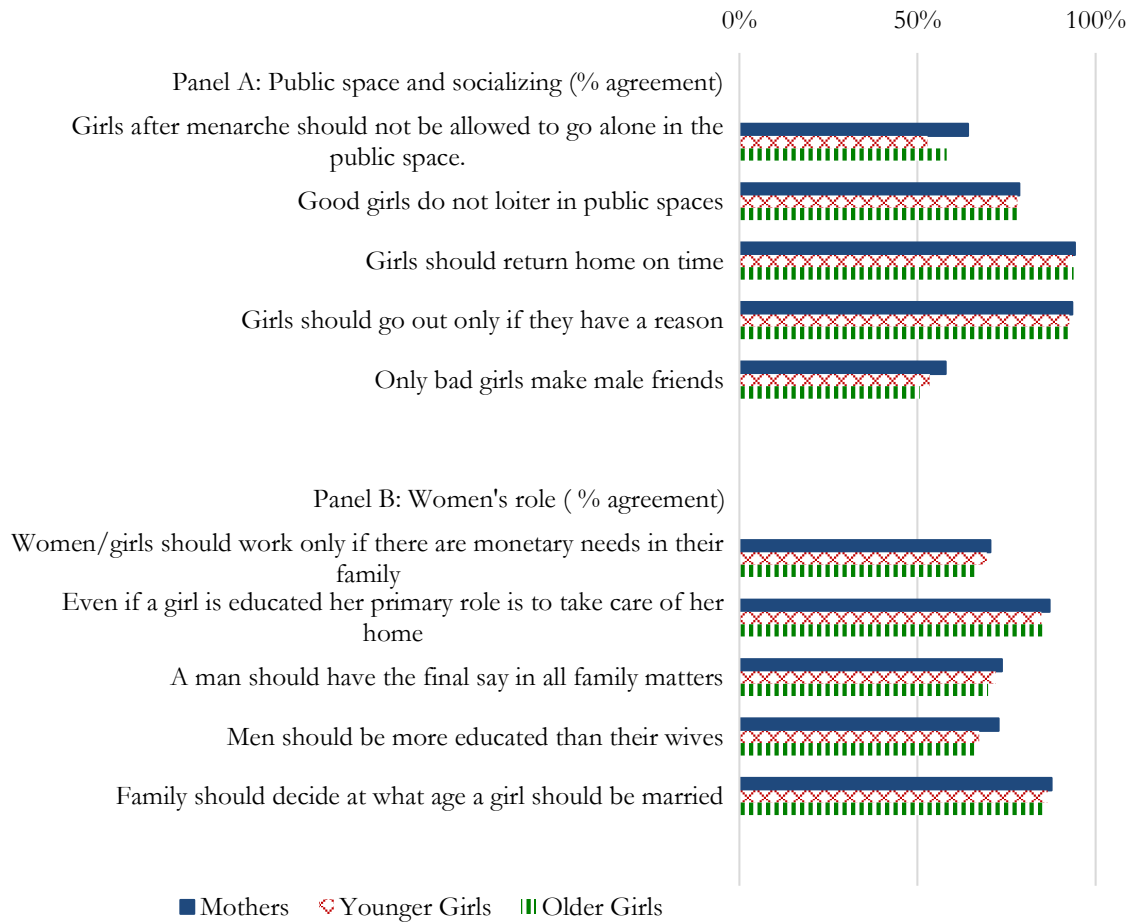
Notes: The figure plots the percentage of mothers, younger girls (12–14 at baseline), and older girls (15–17 at baseline) who reported that they either "strongly agree" or "agree" with each statement.

girls' everyday behaviors are of first-order importance for female educational outcomes. Indeed, our data suggest that in this context such norms may be substantially more important than aspirations or attitudes *directly* relating to girls' education and later careers - concepts that have received much attention in the literature (Beaman et al. 2012; Riley 2022).[4] Our data suggest that even though both girls and their mothers have high educational aspirations, girls tend to fall short of these aspirations as time passes (see Figure A.1 in Appendix A).

---

[4]Directly making girls' commutes easier and less in conflict with gender norms is another approach that has shown promise (Muralidharan and Prakash 2017).

Figure 2: Acceptance and Perceived Likelihood of Violence by Mothers and Girls

Panel A: Acceptance of violence (% agreement)

A family's honour lies in a girl's hand.

A daughter deserves to be beaten if she does not obey her parents.

If a girl is a victim of some sexual abuse, it is the fault of the girl.

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

There are times when a woman deserves to be beaten

Panel B: Likelihood of violence (% reporting violence likely)

If Rama arrived home later than agreed her parents would hit, slap, beat or kick her

If Rama walked home from school just with a male friend, he would act inappropriately towards her e.g. try to kiss her

If Rama walked to the market alone a stranger would hit her and steal her money

If Rama walked back from school alone a boy from a neighbouring village would follow her and call her names

If Rama has a mobile phone, her husband would demand to check her phone to see who she had been talking to

■ Mothers   ⨯ Younger Girls   ‖ Older Girls

Notes: Panel A plots the percentage of mothers, younger girls (12–14 at baseline), and older girls (15–17 at baseline) who reported that they either "strongly agree" or "agree" with each statement. Panel B plots the percentage who report that it is either "very likely" or "quite likely" that Rama will experience this mode of violence in each situation.

Figure 3: Commuting to School

Commute 5kms +

Commute one hour +

Commute on shared public transport

■ Upper Primary   ◇ Lower Secondary   ‖ Upper Secondary   ■ Higher Education
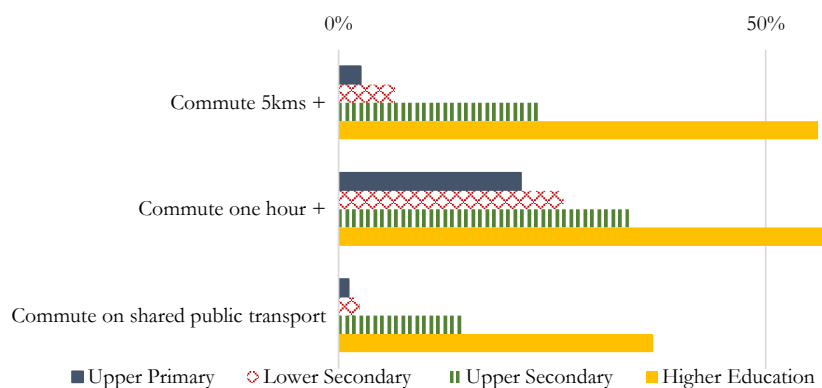
Notes: Figure plots the proportion of girls whose commute to their school or higher educational institution is: (i) 5 kms or more, (ii) takes them one hour or more in total during a typical a typical day, and (iii) on shared public transport (excluding a dedicated school bus). Sample is those girls attending any school or formal educational establishment at endline.

9

2.2 Interventions

2.2.1 Model 1: Girl Groups

In this context, the aim of the Girl Groups intervention was to encourage and enable girls to remain in education and delay marriage *despite* living in a context of strong external norms and constraints. The intervention was implemented by Pradan (Professional Assistance for Development Action), a non-governmental organization (NGO) which organizes women's self-help groups and livelihood programs locally but had no direct experience running programs for adolescent girls. Group activities in each community were facilitated by between one and three "peer mentors"- young women, aged 19–25, from the community. Mentors were paid Rs. 2000 (roughly USD 25) per month which corresponded roughly to 1.5 times the minimum wage for 2 days work a week. Each girl group consisted of around seven adolescent girls and aimed to meet twice a week for an hour – once for discussion and education sessions, and once to play sport. The groups were open to all girls living in the community between the ages of 12 and 19.[5]

The discussion and education sessions followed a curriculum designed for this program. This curriculum stressed the value of education for the girls' futures and encouraged them to think about the things they enjoyed about school. It also sought to lessen the grip of internalized gender norms by highlighting how restrictive and arbitrary these norms can be, as well as to correct misinformation that contributes to the perpetuation of norms. For example, the girls were encouraged to consider and question the idea that it is wrong for girls to travel to school with boys. They were also provided with accurate information relating to marriage, the female body, and sexual and reproductive health. The intention here was to dispel beliefs such as the belief (held by 70% of girls in our baseline sample) that menstrual blood is "dirty" which motivates the norm that girls' social contact should be limited during menstruation.

The curriculum comprised many types of activities including interactive games, role-play, and group discussions. As an illustration, one activity that groups worked on was the "power walk". This involved group members being given a hypothetical identity, for example a scheduled caste adolescent girl, or a dominant caste older man. The peer mentor of the group then read a series of statements such as "I can stay outside home after 7pm without any reason", and girls whose hypothetical identity corresponded to someone who can do that action took a step forward. After a series of statements, the group reflected about the relative position of different hypothetical identities, and how arbitrary characteristics create inequalities in individuals' power and freedoms.

Sessions were organized into three segments to introduce new topics in a progressive way, from more simple concepts to more complex and sensitive issues.[6,7] To ensure that the content was appropriately targeted, girls were divided into two groups based on their age: girls aged 12–14 at the

---

[5]There were between one and ten groups per community with an average of four. The number depended on the number of girls in the community who were interested in participating.

[6]The basic, intermediate, and advanced components comprised 11, 12, and 5–8 sessions, respectively.

[7]The discussion and education component consisted of a total of 31 sessions for girls aged 12–14 and 28 sessions for girls aged 15–19 years, each session lasting about 45–60 minutes.

program's start and girls aged 15–19.[8].

In addition to education sessions, there were also weekly sports sessions. These were intended as a way of bringing the girls closer together and providing them with new empowering experiences. Throughout most of the period during which the education sessions were running, the girls also met for weekly games of *kabaddi*, a team contact sport that is popular across India. The sports sessions culminated in a multi-week kabaddi tournament across the 60 treatment clusters. Whilst adolescent girls were familiar with the basic structure of kabaddi, most had not ever played it or had not played it since childhood. The sports sessions were novel in that they gave girls the opportunity to occupy a public space within the village, to play and to be physically active in public, and to hold the center of the communities' focus during the tournament.

Peer mentors who led the education and sports sessions were given extensive training on curriculum content and on how to keep the girls engaged, as well as ongoing practical support within their community from *sakhis*, older women who had worked with the implementing NGO Pradan on other projects. Sakhis were particularly crucial in reassuring families that the program was appropriate for adolescent girls to participate in.

2.2.2 Model 2: Girl Groups and Community Campaigns

The second intervention model combined the Girl Groups program with Community Campaigns. These sought to make norms at the community level more equitable and to reduce the severity with which girls were punished for breaching restrictive norms. So while the Girl Groups only model aimed to improve girls' outcomes *despite* restrictive norms, the Girl Groups and Community Campaigns arm of the trial aimed to change these norms as well.

The campaigns, which were called "call for action" events, were run by the girls attending the girl groups, their mentors and the sakhis. They were organised to coincide with the girls group completing a module of the curriculum. This roughly corresponded to an event after every two to three girl group sessions. Girls invited their parents, siblings, and other people they knew in the community to the events, while mentors and the mentors' supervisors reached out to the community leaders.

During the events, the girls raised the issues that they had been considering in their groups. They used different communication mediums such as plays, songs, posters, and slogans. After the girls introduced the issue, the mentor, and sometimes more senior Pradan staff, facilitated a discussion between community members with the aim of building support for action towards positive change at the community level. Importantly, the floor was often given to community leaders who explicitly and publicly spoke against violence and harassment of girls, aiming to reinforce the message that violence and harassment of adolescent girls was unacceptable, regardless of the situation. Since the girl groups themselves crafted the agenda, they exact content of the campaigns across villages varied but issues of mobility, education, marriage, and girls' independence were frequent topics.

---

[8]Further details of the curriculum and implementation can be found in the implementation report (Andrew et al. 2018)

2.3 Conceptual Framework

We now sketch a simple conceptual framework of how norms may affect the choices of families and girls, drawing on features of the context that we highlighted in Section 2.1. We use this framework to consider how the two programs we evaluate may have affected behavior.

1. **Constraints.** We begin by drawing on the observation that what is required practically for girls to attend school, particularly after 10th standard, often conflicts with what prevailing norms dictate for girls' behavior. To capture this, we let y denote behaviors that are conducive to attending school.[9] For instance, this might capture actions such as traveling on the bus alone or sometimes arriving back late. Then let there be some minimum level of y, which we call $\bar{y}$, that a girl's behavior must surpass in order to practically be able to attend school. If S is an indicator equal to 1 if a girl attends school and 0 otherwise, then this gives us the constraint that: $S \leq \mathbf{1}(y \geq \bar{y})$.

2. **Preferences.** Girls and their families have preferences over schooling as well as over girls' behavior y. We distinguish conceptually between (1) girls' own immediate enjoyment (or dislike) of school and actions y, and (2) any other considerations that might not directly affect girls' current well-being but might still be important drivers of choices. For instance, we might expect that any economic or marriage market returns to education would enter here.[10] We let the immediate component of utility be $u(S, y) = u_S S + u_y y$ and use $v(S) = vS$ to capture other considerations that may make girls and/or their parents want to invest in a daughters' education.

3. **Norms and Punishments.** Finally, there exist norms, or rules, regarding girls' behavior that are set externally (i.e. not by the girls themselves). Let y^* be the level of girls' behavior y that a girls' community believes to be acceptable. If a girl's action breaks this norm (i.e. if $y > y^*$), then the girl will experience (or perceives that she will experience) a punishment proportional to the size of the deviation: $\max(0, y - y^*)P$. For simplicity, we assume that punishments are additively separable from preferences in driving girls' decisions and current well-being. We note that even within the same village, differences in the social and caste groups that families belong to might result in them facing different norms and differing levels of sanctions for breaking norms.

Bringing together these components, we assume that girls and their families choose S and y to maximize the sum of their own utility function net of the costs of deviating from the external norms y^* subject to the constraint that pursuing schooling necessarily involves choosing $y > \bar{y}$. In other

---

[9]Without loss of generality, we assume that y is bounded from below by 0.

[10]We anticipate that these preferences will be themselves shaped by girls' internalized norms and attitudes. For example, girls may believe that it is inappropriate for them to mix with boys after menarche because they have internalized community-wide norms and, therefore, have a preference against mixed-gender public transport.

words, they solve:

$$\max_{S,y} \quad \underbrace{u_S S + u_y y}_{\substack{\text{Girls' immediate} \\ \text{enjoyment/dislike} \\ \text{of } S \text{ and } y}} + \underbrace{vS}_{\substack{\text{Future benefits} \\ \text{from schooling}}} - \underbrace{\max(0, y - y^*)P}_{\substack{\text{Punishment for Deviation} \\ \text{from norm}}} \quad s.t. \quad \underbrace{S \le \mathbf{1}(y > \bar{y})}_{\substack{\text{School only possible} \\ \text{if } y > \bar{y}}} \quad (2.1)$$

While future expected benefits from schooling, v, will be important in shaping schooling choices, it seems likely that what is most relevant for girls' current well-being is their enjoyment or dislike of both school and the actions y they have to undertake to attend school in practice. We formalize this by suggesting that mental health takes the following form:

$$M(S, y, \bar{y}, y^*, P) = u_S S + u_y y - \max(0, y - y^*)P \quad (2.2)$$

In discussing the implications of our model, we make two assumptions on the parameter values that we consider to be sensible given our context.[11] First, we assume that $v > 0$, i.e. that girls and their families place a positive value on attending school over and above the short-term well-being benefits. We consider this to be highly likely given evidence of strong labor market and marriage market returns to education for girls in India (Andrew and Adams-Prassl 2021; Jensen 2012). Second, we assume $u_y < 0$, i.e. that girls have internalized restrictive gender norms and dislike y but recognise that these actions have instrumental value in allowing them to pursue education. Third, we assume that in most communities $\bar{y} > y^*$.

Implications We now draw out the implications of this model and check whether they are consistent with the patterns we see in the data for the control group. Appendix C details how these implications are derived. Figure 4(a) illustrates the prediction of our model for how the school attendance and mental health of a particular girl (with a specific $\bar{y}$) will vary with externally set gender norms. Figure 4(b) illustrate what our model would imply for average school attendance and average mental health in a population of girls who differ only in their external constraints to attending school ($\bar{y}$).[12]

1. **Monotonic relationship between gender norms and school attendance.** Our model predicts that school attendance will increase with the progressivity of community gender norms. For an individual girl, there is a threshold below which the cost of punishments from the broader community will deter her from attending school. This is shown in Figure 4(a). Averaged across girls with differing levels of constraints (i.e. different $\bar{y}$), this implies that overall school attendance will be monotonically increasing in the progressivity of gender norms (Figure 4(b)).

2. **Convex, and possibly U-shaped, relationship between gender norms and current well-being.** Our model predicts that girls' mental health will have a convex, and possibly

---

[11]Appendix C details and justifies these assumptions in more depth.

[12]Heterogeneity in $\bar{y}$ is motivated by the idea that some girls live closer to or further away from school, or the fact that some girls can travel to school with a brother and others cannot.

13

Figure 4: Implications of Model for Relationship between Gender Norms, Education and Mental Health.

((a)) For an individual girl   ((b)) Averaged across girls with heterogeneous $\bar{y}$
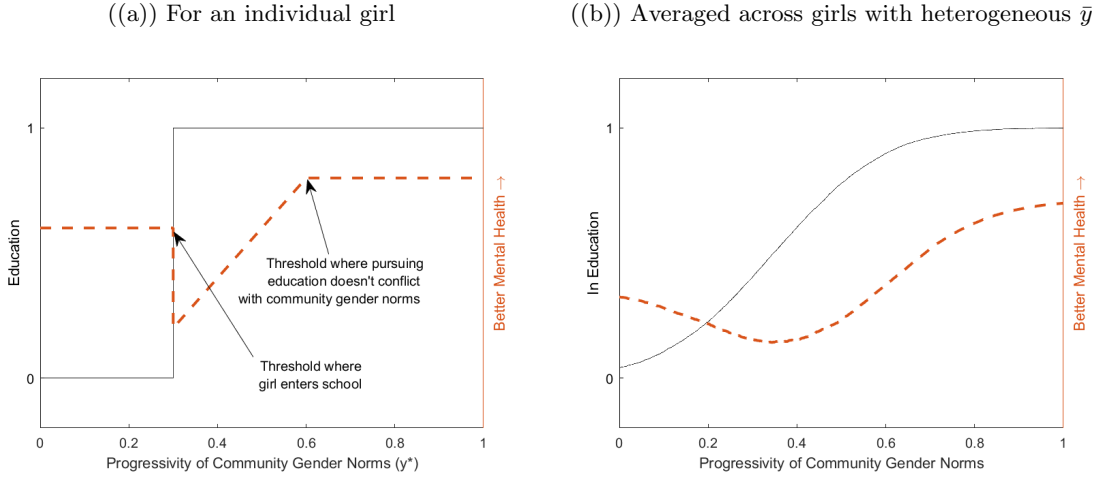
Notes: Black lines and left hand axes describe school attendance. Orange dashed lines and right hand axes describe mental health. Plot (a) plots school attendance and mental health for a particular girl facing a particular constraint $\bar{y}$. Plot (b) plots average school attendance and average mental health averaging across girls facing differing levels of $\bar{y}$. See Appendix 4 for a detailed explanation of how our model implies these relationships.

U-shaped, relationship with community gender norms. Girls' decision to pursue education is driven by considerations related to current well-being, *as well as* future returns. This implies that the marginal girl who enters school does so at a cost to her current well-being because of the sanctions and harassment she faces for violating the norms that she must violate in order to enter school. [13] We see this in Figure 4(a): as soon as gender norms pass the threshold where the girl starts going to school her well-being deteriorates. As norms continue to become more progressive, well-being starts to improve.

In 4(b), we plot what this means for average mental health across a population of girls who differ only in the external constraints on what is required to enter school ($\bar{y}$). This suggests that average mental health will follow convex, and possibly U-shaped relationship with gender norms. In communities with the most restrictive norms, few girls go to school and, therefore, few girls face sanctions for breaking gender norms. However, as norms become more progressive, more girls go to school but, on the margin, they face a mental health cost for doing so. Therefore, on average mental health will be higher in the former compared to the latter communities. As norms continue to become more progressive, girls who are going to school (and there are many of them) face fewer sanctions and average mental health improves .

The model suggests a similar pattern with respect to the severity of sanctions, P (see Appendix C).

In Figure 5, we check these predictions against the correlational patterns in our data. In particular, we analyze the correlations between factor measures of progressivity of girls' gender norms and of

---

[13]If current well-being was the only driver of girls schooling choices then at the margin school would have no impact on current well-being.

Figure 5: Descriptive Relationship between Gender Norms, Education and Mental Health in the Control Group

((a)) Norms & Education

((b)) Norms & Mental Health

((c)) Violence & Education

((d)) Violence & Mental Health
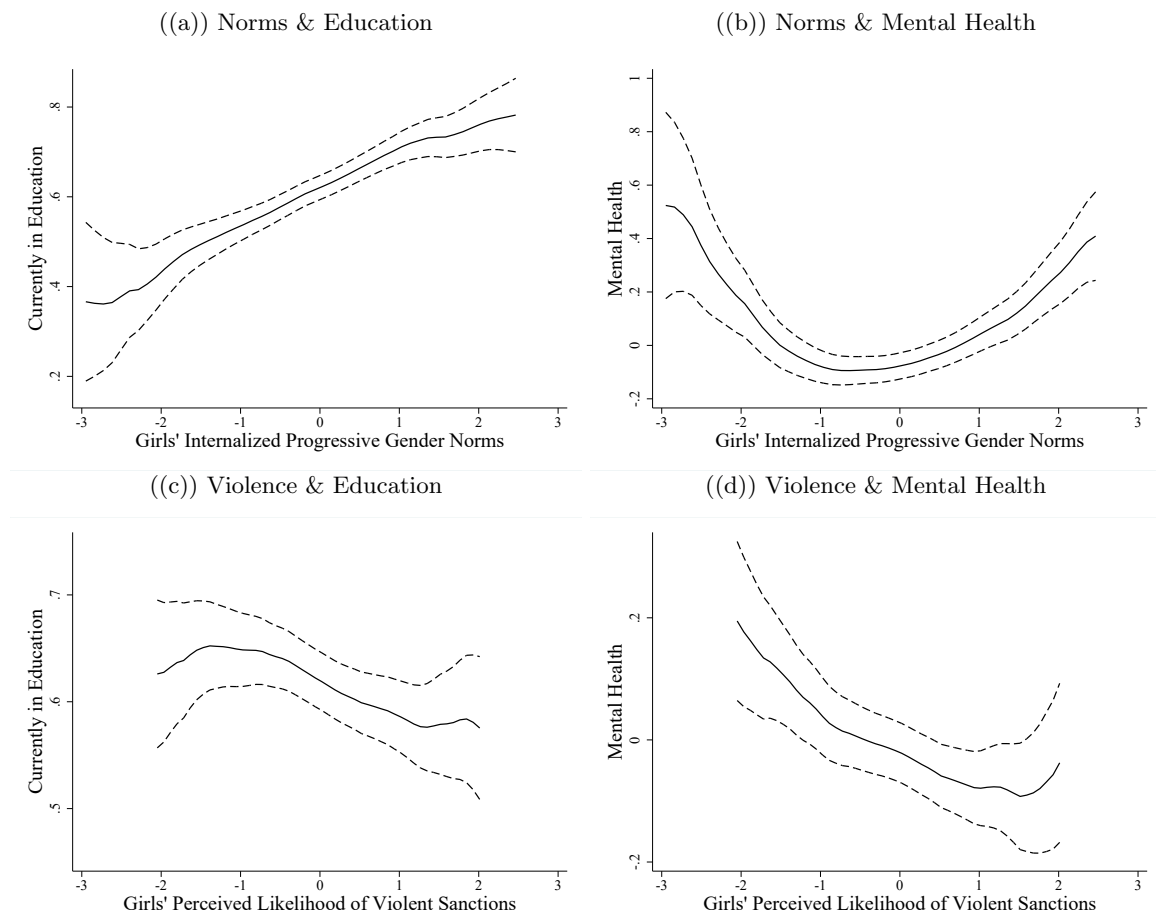
Notes: Solid lines plot relationships between girls' perceived gender norms, girls' perceived likelihood of facing sanctions for breaking norms, girls' school enrollment and girls' mental health in the control group using kernel-weighted local polynomial smoothing with an Epanechnikov and a bandwidth of 0.5. Dashed lines plot 95% confidence intervals. Construction of all measures is detailed in Section 3.3.

perceptions of the likelihood of violent sanction with, in turn, girls being enrolled in formal education and a factor measure of good mental health. The construction of all these measures is described in detail in 3.3. This analysis is purely descriptive and there could be many confounding factors. We present it as motivating evidence only although we note that the same patterns hold true separately for the older and younger sample (Figures A.3 and A.2) and when we use gender norms and likelihood of violent sanctions reported by the *mothers* of the girls the rather than girls themselves (Figure A.4). We do this analysis on the control group only to avoid treatment effects confounding the correlations.

Figure 5(a) confirms that we do indeed see that education is monotonically increasing in the progressivity of girls reported gender norms. Girls who report gender norms that are two standard deviations less progressive than average are in education at a rate of around 43% compared to a rate of 75% among girls who report norms to be two standard deviations above average.

In Figure 5(b) we turn to the relationship between gender norms and mental health and find evidence of the U-shaped relationship suggested by our model. Both girls who report particularly strict gender norms and girls who report particularly progressive norms have better mental health than those who perceive their community to have average gender norms. This is consistent with the story that as gender norms become more lenient, girls may be encouraged into school but, at the margin, this might come at a cost to their immediate mental health. As norms continue to become more progressive, more progressive norms begin to be associated with better mental health.

Figures 5(c) and 5(d) show that we see a similar set of patterns in relation to the perceived likelihood of violent sanctions for breaking norms. The severity of sanctions is negatively correlated with school attendance. And while there is a clear overall negative relationship between the severity of sanctions and girls' mental health, the shape appears convex.

2.4 Intervention Mechanisms

The framework we have set up is helpful for considering the mechanisms through which the two programs we evaluate might affect girls' outcomes. The first thing to note is that we see the distribution of constraints (i.e. the distribution of $\bar{y}$ which characterize the participation constraint that $S \leq \mathbf{1}(y \geq \bar{y})$) as being fixed across the control and the two program arms as neither explicitly aimed to relax practical constraints, such as, for example, making girls' commutes to school easier. We propose that the primary potential mechanisms associated with the two intervention arms are:

1. **Girl Groups.** We think of the first Girl Groups program as trying to improve outcomes through changing girls' own attitudes and preferences (i.e. through changing $u(S, y)$ and $v(S)$). It did so by seeking to strengthen their preference and enthusiasm for education and encouraging them to place less intrinsic value on the importance of abiding by restrictive gender norms.

2. **Community Campaigns.** Over and above the standalone effect of the Girl Groups program, we view the addition of the Community Campaigns program as directly targeting the restrictive

16

norms in the communities where the girls live (y^*), and the severity of the punishments for deviating from these norms (P). As discussed above, our framework suggests that the impact of changes in gender norms and the perceived risk of sanctions on girls' mental health will depend on where in the distribution of these factors any changes occur. Such changes could come at short-run costs to girls' mental health if they push some girls into school who then face violence or harassment as the result. However, if the changes primarily reduce the sanctions that girls who are already attending school face then our framework suggests they should lead to an improvement in mental health.

This design maps well to the conceptualization of norm change set out by Finnemore and Sikkink (1998), in which there is a set of existing norms that govern behavior. These are so "internalized" that they are "taken for granted" by everyone involved, as are the social, economic, or physical sanctions that are commonly perceived to await those who do not conform with the norms. Norm change begins through a "norm emergence" stage in which "norm entrepreneurs" seek to persuade others that the old norm should be discarded in favor of a new and better norm, through appealing to the altruism and empathy of others and trying to convince them of the intrinsic worth of the new norm. If this process is successful, then eventually enough people will be persuaded that the new norm is correct, and that the process will reach a "tipping point" at which social pressure takes over as a key mechanism to the adoption of the new norm.[14]

The notion that new ideas, arguments, and persuasion are important in norm change is precisely how our intervention models sought to create change. Furthermore, the design embeds the idea that given how marginalized adolescent girls are, their ability to act as "norm entrepreneurs" and effectively convince others of new ideas is severely limited. This is both because adolescent girls do not have the social standing to be persuasive and because they also lack the platform, as they are so often restricted from being in public spaces or taking center stage. By backing up girls' voices with the voices of powerful members of the community and by providing a platform for girls to speak, the addition of Community Campaigns to the Girl Groups might relax these constraints and enable the girls to precipitate norm change.

3 Study Design, Data, and Empirical Strategy

3.1 Experimental Design, Timeline, and Sample

This study took place across 125 villages in three blocks – Bari, Baseri, and Dholpur – of Dholpur district, Rajasthan where Pradan, the implementing NGO, had worked for over ten years. We organized these 125 villages to form 90 clusters of roughly 1,500 households each, grouping smaller villages

---

[14]This framework is somewhat different from that adopted in some recent work in economics, which has focused on situations where conservative norms might be sustained through "pluralistic ignorance" whereby in a quickly changing legal environment most people do not privately agree with a conservative norm but believe that others do and do not voice their disagreement with the norm for fear of social sanctions (Bursztyn, González, and Yanagizawa-Drott 2018). In this type of conception of norms, norms change when people's beliefs about what others *truly* believe change.

and hamlets together. Stratifying by block, we randomized clusters in equal numbers to (i) receive the Girl Groups program as a standalone intervention, (ii) receive the Girl Groups and Community Campaigns program, and (iii) be in the control group.

The population of interest for this paper is girls who were aged 12–17 years at baseline and who were neither engaged nor married at the time of baseline. Within the study clusters, we conducted a census of all adolescent girls and used this as our sampling frame. We organized unmarried girls by age into two strata: those aged 12–14 and 15–17 at baseline.[15] In each strata, we selected 39 and 37 girls, respectively, on the basis of power calculations for main outcomes and anticipating non-response and attrition of around 20%. We obtained complete baseline data for 5,731 girls who were between the ages of 12 and 17 and neither married nor engaged at baseline. The baseline survey, which took place between January and March 2016, included measures of schooling, marriage, mental health and gender norms. In addition to interviewing girls, we interviewed their primary caregivers (typically their mothers), as well as administering a brief household survey. Details of the baseline instruments can be found in Achyut et al. (2016).

The endline survey took place between December 2017 and March 2018. As at baseline, we interviewed the girls and their caregivers and collected data about the household. Many of the girls in the sample had moved away from their parents' house for marriage or to pursue education by the time of the endline. We attempted to re-interview all girls who had moved within Dholpur district and all caregivers interviewed at baseline who had remained in the original households during the endline survey. In all, we obtained complete endline data for 5,043 of the 5,731 girls in our baseline sample, giving a follow-up rate of 88.0%. Furthermore, combining data obtained from the girls themselves and from their caregivers, we have information on marriage and education outcomes for a larger sample of 5,526 girls, whom we refer to as the extended sample; this gives us a follow-up rate of 96.4% for these outcomes. Our attrition rates of 12.0% and 3.6% for the main and extended samples, respectively, compare favorably to similar studies with adolescent girls, which report attrition rates of 42% (Buehren et al. 2017), 19.6% (Buchmann et al. 2017), 20% (Adoho et al. 2014), and 18% (Bandiera et al. 2019). Importantly, attrition from both the main and the extended samples is uncorrelated with treatment status (see Table A.1).[16]

Table 1 presents key baseline characteristics for the sample of non-attritors by treatment status while Tables A.2 and A.3 present the equivalent information for the full baseline sample (regardless of whether or not they attrited at endline) and for the extended sample, respectively. Across key socio-

---

[15]We additionally collected baseline data on 1,782 married or engaged girls aged 12–19, with the intention of including these girls in the evaluation sample. However, due to very low (<7%) rates of participation in the intervention amongst girls who were already married at the time of baseline, we dropped married girls from the quantitative endline sample. We collected additional qualitative data on married adolescent girls at endline to better understand why program participation had been so low and how future programs might be adapted to better suit the needs of this group (see Andrew et al. (2018)).

[16]Table A.1 shows how attrition varied by baseline characteristics. We find that attrition from the main sample was slightly higher among older girls and girls who were already out of school at baseline. This is likely to be driven by the fact that these girls are more likely to have moved out of the district for marriage. Attrition from the extended sample is not predicted by baseline age. This suggests that the method of using caregiver reports in cases where girls could not be interviewed directly enables us to recover a more-representative sample.

Table 1: Baseline Characteristics and Balance

	Younger Girls				Older Girls			
	(1) Control	(2) Girl Groups	(3) Girl Groups + Community	(4) (3)-(2)	(5) Control	(6) Girl Groups	(7) Girl Groups + Community	(8) (7)-(6)
Girl Age	12.93	12.89	12.92		15.82	15.87	15.88**	
	(0.858)	(0.849)	(0.847)		(0.783)	(0.809)	(0.808)	
		[p=.362]	[p=.823]	[p=.526]		[p=.152]	[p=.046]	[p=.823]
Carer's Years of Ed.	1.447	0.994*	1.240		1.196	0.781	0.907	
	(2.863)	(2.457)	(2.543)		(2.641)	(2.245)	(2.232)	
		[p=.063]	[p=.351]	[p=.278]		[p=.074]	[p=.145]	[p=.553]
SC/ST Caste	0.320	0.333	0.399		0.316	0.339	0.382	
	(0.467)	(0.472)	(0.490)		(0.465)	(0.474)	(0.486)	
		[p=.824]	[p=.194]	[p=.321]		[p=.722]	[p=.305]	[p=.515]
Asset Index	-0.0758	-0.130	-0.105		0.111	0.0187	0.00436	
	(0.844)	(0.809)	(0.811)		(0.983)	(0.897)	(0.901)	
		[p=.443]	[p=.686]	[p=.677]		[p=.205]	[p=.173]	[p=.853]
Attending School	0.927	0.903	0.858***		0.682	0.647	0.628	
	(0.260)	(0.296)	(0.349)		(0.466)	(0.478)	(0.484)	
		[p=.249]	[p<0.001]	[p=.049]		[p=.295]	[p=.129]	[p=.628]
Progressive Gender Norms	-0.0114	-0.102	0.0138		0.0763	-0.0873*	0.118	
	(0.983)	(1.006)	(0.942)		(1.008)	(1.055)	(1.000)	
		[p=.295]	[p=.759]	[p=.109]		[p=.09]	[p=.649]	[p=.020]
Carer's Progressive Gender Norms	0.0391	-0.0757	0.0661		0.0235	-0.0601	0.001	
	(1.015)	(1.020)	(0.984)		(0.981)	(1.003)	(0.995)	
		[p=.200]	[p=.77]	[p=.075]		[p=.301]	[p=.788]	[p=.459]
Mental Health	-0.0504	-0.0703	-0.0482		0.117	-0.000704	0.0705	
	(1.004)	(0.973)	(1.019)		(0.990)	(1.029)	(0.976)	
		[p=.813]	[p=.981]	[p=.817]		[p=.169]	[p=.584]	[p=.418]
Observations	931	885	909		826	744	748	

Notes: The table presents means and standard deviations (in parentheses) by treatment group and by age strata. Two-sided *p*-values test the difference between each treatment group mean and the control. The *p*-values and standard errors are constructed using a cluster bootstrap accounting for stratification by region and treatment status.

economic characteristics, the sample appears well balanced. Likewise, the sample is well balanced on baseline measures of gender norms and mental health. However, an important imbalance does emerge when we look at indicators of education: the proportion of girls attending school at baseline was lower in the Girl Groups and Community Campaigns arm than in the other two arms among the younger sample. This imbalance is present in the full baseline sample and in the extended sample (Tables A.2 and A.3) suggesting that it is a chance imbalance rather than a result of differential attrition. A few points are important to note here. First, overall, the Girl Groups and Control appear well balanced in the younger sample as do all three treatment arms in the older sample; comparisons of outcomes of these groups should thus not be affected by any concern about baseline balance. Second, we pre-specified that we would control for baseline school attendance interacted with age dummies in our published pre-analysis plan, which should help correct for these chance imbalances. Third, as a robustness exercise, we use Post Double Selection Lasso (Belloni et al. 2014) to select the control variables; reassuringly we find that our results are not sensitive to this method (Appendix B).

3.2 Implementation and Compliance

All intervention activities took place in the communities over a 14-month period between August 2016 and September 2017. Activities were run by the Dholpur branch of Pradan which had previously

worked extensively in all the study communities on programs connected to rural livelihoods.[17] All girls living in treatment communities who were aged 12–19 at the beginning of the program were invited to attend the girl groups and sports sessions, whether or not they were included in the study sample. Likewise, all community members were able to attend the "Call for Action" events.

In all, 60% of the girls in the study sample enrolled in the program and attended at least one session (61% if the Girl Groups arm and 58% in the Girl Groups + Community Campaigns arm). Enrollment was slightly higher among girls aged 12–14 (65%) than among the older girls aged 15–17 (54%). Conditional on enrolling, girls attended approximately 60% of the total number of sessions held (an average of 28 education sessions and 23 sports sessions). There was no significant difference in attendance rates between the two treatment arms. Mentors reported that the "Call for Action" community events (for which we did not collect attendance data) were well attended by girls, parents, and members of the wider community including community leaders.

3.3 Outcomes

3.3.1 Education and Marriage.

Girls' educational status and marital status at the time of the endline survey are two of our most important outcomes. At the time of the endline survey, age of marriage and final educational attainment had not yet been realized for a large part of the sample (who were aged 14–19). Therefore, we estimate treatment effects on binary indicators of whether a girl was in formal education and whether she was married at the time of the endline (6 months after the end of the interventions). We define formal education as either being in school or in formal further education (studying for either a graduate or postgraduate degree or diploma). Our indicator of marriage is defined as whether a girl has ever been married (in all but one case she was still married), including cases where the marriage ceremony had happened but the marriage had not yet been consummated (*gauna* not performed). We also show impacts on two alternative definitions: (1) excluding unconsummated marriages and (2) the combined probability of being married, engaged, or having had a marriage fixed.

We have data on marriage and education outcomes for 96.4% of girls in our baseline sample. Of these girls, this information is reported by the girls themselves in 91% of the cases while in 9% of the cases it is reported by the caregiver. We show that our main estimates are robust to using just the sample of girls who were interviewed directly

3.3.2 Mental Health.

We measured two dimensions of girls' mental health: symptoms of depression and symptoms of anxiety using the nine-item Patient Health Questionnaire (PHQ-9) and the seven-item Generalized Anxiety Disorder scale (GAD-7) to measure each respectively. We chose these scales because they have performed well with adolescent girls in Hindi-speaking areas of India in previous studies (Ganguly

---

[17]The Dholpur branch of Pradan has since become Manjari Foundation `www.manjarifoundation.in`.

et al. 2013; Leventhal et al. 2015; Leventhal et al. 2015). Following some additional adaptation to the local dialect of Hindi, piloting showed that items from both scales were well understood by respondents.

Both the PHQ-9 and GAD-7 ask about symptoms experienced during the previous two weeks and with the 4 possible responses to each symptom being "not at all", "several days", "more than half the days" and "nearly every day". We use individual item responses to construct separate factor scores for depression and anxiety, as well as combining all items to create an overall mental health factor. We provide details of the methodology we use to create these factors below in subsection 3.3.4. We also explore the robustness of our estimates to using simple raw scores and using various binary cutoffs for being at risk of mild or moderate depression and anxiety that have been used in the medical literature.

3.3.3 Internalized Gender Norms and Perceived Likelihood of Violent Sanctions.

We estimate treatment effects on gender norms and perceptions of the likelihood of girls facing violent sanctions if their behavior deviated from dominant norms, as reported by the girls themselves and their primary caregivers. We measure gender norms, we using an adapted version of the GEMS scale (Pulerwitz and Barker 2008). This scale presents respondents with 15 statements regarding gender norms, such as "Good girls do not loiter in public spaces" and asks respondents whether they "strongly agree", "agree", "disagree" or "strongly disagree" with the statement. Earlier, we presented descriptives of responses to many of these statements in the baseline data (Figures 1 and 2(A)); Appendix Table A.5 gives the full list of statements used at endline.

The scale used to measure perceptions of likelihood that girls will face violent sanctions if they do not conform to gender norms was developed and and piloted by us. We presented respondents with five statements about a fictional young women called Rama who we described as being an adolescent girl living in a village in the same district as the respondent. We described Rama facing violence and harassment after undertaking various actions that are commonly seen to be in violation of gender norms. For instance, one statement was "If Rama walked back from school alone a boy from a neighbouring village would follow her and call her names". We asked respondents whether they thought that such sanctions were "very likely to happen", "quite likely to happen", "quite unlikely to happen", or "very unlikely to happen" if Rama took this action. Figure 2(B) presented descriptives for how the control group answered these statements.

We use these item responses to create factor scores for gender norms and violent sanctions applying the procedure outlined in the following subsection and use these factor scores as our outcome measures.

3.3.4 Factor Models.

We use responses to the items in the mental health, likelihood of violence and gender norms scales to construct factor scores for each individual on each of the scales by estimating graded item response

models. This is motivated by the assumption that item responses are error-ridden measures of underlying latent constructs which a factor model can enable us to estimate (Das and Zajonc 2010).

Specifically, we index these underlying constructs by k.[18] We let the underlying level of k for individual i be denoted by θ_{ik}. We do not directly observe θ_{ik} but observe multiple noisy measures of it in the form of ordinal item responses to the scale designed to measure k. We let individual i's response to item j in the scale designed to measure construct k be y_{ijk}. For each k, we use these ordinal item responses to estimate a graded response model (which is an IRT factor model suited for ordinal items). In particular, we model the probability of individual i giving an ordinal response of more than m to item j as an ordered logit:

$$Pr(y_{ij} \geq m|\theta_i) = \frac{\exp(\alpha_{jkm} + \beta_{jk}\theta_{ik})}{1 + \exp(\alpha_{jkm} + \beta_{jk}\theta_{ik})}.$$

Here, θ_{ik} represents i's underlying level of construct k. We assume that θ_{ik} is normally distributed and impose the normalization of a zero mean and unit variance in the control group. This normalization means that effect sizes should be interpreted relative to the standard deviation of the control group. In this set-up, β_{jk} represents item j's discriminatory power and governs the rate at which the probability of a particular response changes with the underlying factor. α_{jkm} are the difficulty parameters and capture the probability of particular item responses among individuals with average (zero) levels of θ_{ik}. We estimate the measurement model by maximum likelihood using an Expectation–Maximization (EM) algorithm with Gauss-Hermite quadrature to approximate the integral over the unobserved latent factors.[19]

For each construct k, a higher value of θ_{ik} represents a higher level of the construct. Tables A.4, A.5 and A.6 present the parameter estimates for the measurement models of, respectively, mental health, gender norms, and violent sanctions. The estimates suggest that all items had a high discriminatory power; all discrimination parameters are positive and significantly different from zero with $p < 0.0001$.

For each individual i and construct k, we calculate expected a posterior (EAP) factor scores. These are the mean of the posterior distribution of θ_{ik} for each individual conditional on the observed item responses. We use EAP scores for all our main analysis. Additionally, we construct 10 "plausible values" for each individual which we use in quantile regression towards the end of the paper. It is more appropriate to use plausible values for estimating conditional quantiles, since EAP scores, while representing the best posterior guess at each individual's underlying construct, create a distribution of factor scores whose variance is understated relative to the normalized variance (one in the control group)(Das and Zajonc 2010). Plausible values are values drawn from each individual's posterior distribution of θ_i conditional on item responses. Simulated plausible values across many individuals can recover the overall population distribution. Following best practice, we draw 10 plausible values

---

[18]Specifically, k can denote: girls' depressive symptoms, girls' anxiety symptoms, girls' mental health, girls' internalized progressive gender norms, carers' internalized progressive gender norms, girls' perceived likelihood of violence or carers' percieved likelihood of violence.

[19]We use the "uirt" Stata package to perform this estimation (Kondratek 2022).

for each individual, repeat the quantile analysis with each of the 10 samples of plausible values and then take the average as our overall estimate (Das and Zajonc 2010).[20]

3.4 Empirical Strategy

To estimate the causal impact of the girl groups, when delivered alone and in combination with community engagement, we estimate:

$$y_{ic} = \beta_0 + \beta_{girl}T_c^{girl} + \beta_{girl+comm}T_c^{girl+comm} + \gamma X_{ic} + \epsilon_{ic}, \qquad (3.1)$$

where y_{ic} is the outcome of interest for individual i in cluster c and T_c^{girl} and $T_c^{girl+comm}$ are, respectively, indicator variables describing whether cluster c was randomized into the Girl Groups standalone treatment or the Girl Groups and Community Campaigns treatment arm. For all continuous outcomes, we use Ordinary Least Squares (OLS) to estimate (3.1), while for binary outcomes we use a logit model and then calculate the corresponding average marginal effect associated with each treatment. We show robustness of our estimates to also using OLS for binary outcomes.

β_{girl} and $\beta_{girl+comm}$ are the key parameters of interest. Given that treatment was randomly assigned, so long as the joint distribution of potential outcomes and attrition are independent of treatment status, β_{girl} and $\beta_{girl+comm}$ identify the intent-to-treat (ITT) effects of, respectively, the Girl Groups alone and the Girl Groups in combination with Community Campaigns. In other words, these parameters can be interpreted as the average causal effect of being offered the interventions, regardless of whether or not girls took them up.

X_{ic} are baseline characteristics and are included to increase precision. Baseline controls were pre-specified (Andrew and Krutikova 2018); they comprise a core set of characteristics including age, caste, wealth, and maternal education, in addition to baseline variables that we anticipated would be most predictive of the outcome in question. Controls for each outcome are listed in the table notes. As a robustness check, we also estimate the main impacts using the Post Double Selection Lasso method introduced by Belloni et al. (2014) to select the control variables. We show in Appendix B that our results are not sensitive to the method of selection of controls.

We allow the random error term, ϵ_{ic}, to be arbitrarily correlated within clusters, the unit of randomization. In practice, we estimate standard errors and p-values using a cluster bootstraps with 1000 iterations and accounting for stratification by block and by treatment status. We estimate treatment effects separately for our two sampling strata – younger girls aged 12–14 at baseline and older girls aged 15–17 at baseline – before estimating a pooled treatment effect.

---

[20]We carry out this procedure using the "pv" package in Stata (Macdonald 2008).

4 Results

4.1 Education and Marriage

Table 2 presents ITT impacts of the two intervention models on girls' education and marriage. It shows the average marginal effects (calculated from a logit model) and the associated standard errors and p-values. The first two rows present estimates relative to the control group, while the final row presents estimates of the difference in average effects of the two intervention models.

The results suggest that the Girl Groups were successful at keeping adolescent girls in education, both alone and in combination with community engagement. We start with the older girls who, as outlined in Section 2, are the girls for whom gender norms are in greatest conflict with continuing their education and who are at the greatest risk of dropping out. For these girls, we estimate that the groups increased the probability that they were still in school or post-secondary education at endline by 6.1 p.p. ($p = 0.009$) in the Girl Groups only arm and 4.7 p.p. ($p = 0.025$) in the Girl Groups and Community Campaigns arm. This is a substantial effect equivalent to an increase in education attendance of around 15% compared to the control group, less than half of whom (41%) were in education at endline.

We do not find any evidence of an impact on the education for younger girls from either intervention model. The point estimates, an increase in attendance of 1.8 and 1.4 p.p. in the Girl Groups and Girl Groups and Community Campaigns arms respectively, are small and not statistically significant. This is not surprising in light of the high attendance rates in this age group (76.5% in the control group). Furthermore, the impacts in the Girl Groups arm for the younger sample are statistically significantly different from those for the older sample (p-values for difference = 0.08). Pooling both age groups together, the average ITT effect of Girl Groups only in the whole sample is an increase in school attendance of 4 p.p. ($p = 0.018$) and 3 p.p. ($p = 0.08$) in the Girls Groups and Community Campaigns arm. Relative to control group, 59% of whom were attending education at the time of endline, both intervention models led to an increase in education attendance of approximately 7%.

As noted in Section 3, our "In Education" definition includes girls who were either in school or enrolled in formal post-secondary education, such as a graduate or post-graduate degree or diploma. Appendix Table A.8 shows that the effects we find are driven by the older girls being more likely to stay in school rather than by higher take-up of post-secondary education.

Impacts on rates of marriage follow a similar pattern to the education impacts. For the older group, of whom 18.1% were married by endline in the control group, the Girl Groups intervention led to a decrease in the average probability of being married by 3.6 p.p. ($p = 0.036$), equivalent to a 20% reduction relative to the control group. The point estimate for the impact in the Girl Groups and Community Campaigns arm is similar (2.4 p.p.) but it is not statistically significant; neither, however, is the difference between the two treatment groups (as we discuss below). Again, we see no impacts for the younger group, among whom fewer than 6% were married at endline in the control

Table 2: Education and Marriage

	Older			Younger			All		
	In Education	Married	Married or Engaged or Fixed	In Education	Married	Married or Engaged or Fixed	In Education	Married	Married or Engaged or Fixed
Girl Groups	0.061***	-0.036**	-0.053***	0.018	0.010	0.017	0.040**	-0.015	-0.018
	(0.023)	(0.017)	(0.020)	(0.019)	(0.011)	(0.014)	(0.017)	(0.011)	(0.013)
	[p=0.009]	[p=0.036]	[p=0.008]	[p=0.346]	[p=0.381]	[p=0.236]	[p=0.018]	[p=0.166]	[p=0.192]
Girl Groups + Community	0.047**	-0.024	-0.056**	0.014	0.010	0.022	0.030*	-0.006	-0.014
	(0.021)	(0.020)	(0.024)	(0.021)	(0.011)	(0.016)	(0.017)	(0.012)	(0.017)
	[p=0.025]	[p=0.224]	[p=0.022]	[p=0.492]	[p=0.381]	[p=0.158]	[p=0.080]	[p=0.591]	[p=0.381]
Difference	-0.014	0.012	-0.003	-0.003	0.004	0.005	-0.009	0.009	0.003
	(0.023)	(0.019)	(0.024)	(0.021)	(0.011)	(0.016)	(0.018)	(0.013)	(0.017)
	[p=0.549]	[p=0.536]	[p=0.917]	[p=0.870]	[p=0.696]	[p=0.736]	[p=0.609]	[p=0.499]	[p=0.857]
N	2605	2605	2605	2918	2918	2918	5525	5525	5525
Control Mean	0.406	0.181	0.304	0.765	0.056	0.093	0.592	0.116	0.195
Heterogeneity by Age									
Girl Groups				[p=0.084]	[p=0.021]	[p=0.001]			
Girl Groups + Community				[p=0.156]	[p=0.103]	[p=0.001]			
Difference				[p=0.685]	[p=0.667]	[p=0.727]			

Notes: The table presents estimated average marginal effects of "Girl Groups" and "Girl Groups + Community Campaigns" interventions relative to the pure control group. "Difference" refers to differences between the two intervention models. Average marginal effects are estimated from a logit model. Standard errors associated with each marginal effect (in parentheses) and two-sided *p*-values are constructed using a cluster bootstrap accounting for stratification by block and treatment status (1000 iterations). Pre-specified controls are: a full set of interactions between dummy variables for age in years and a dummy variable for whether the girl was in school at baseline, caste, baseline wealth index, mother's years of education, district, whether the girl's family was talking about marriage at baseline, whether the girl's marriage was fixed at baseline, the girl's intended age of marriage at baseline, and the gender attitudes of the girl's caregiver. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

25

group. Future follow-ups will be required to ascertain whether the interventions were effective in delaying marriage for this cohort. Combining the two age groups, the point estimates on the average effects are negative but not statistically significantly different from zero. The same pattern of results holds if we exclude girls who were formally married but had not yet moved to their husband's house (no "gauna"), as sometimes happens especially when young girls marry (Appendix Table A.12).

There are several stages in the marriage process which take place before a girl is formally married. These include the marriage being fixed and then an official engagement taking place. A possibility is that the interventions reduced rates of formal marriage by reducing its acceptability among girls and the community, but not the rates of marriages being fixed and girls being engaged. Even if followed by a delay in formal marriage, early fixing of girls' marriage or early engagement may still limit girls' say in marriage choices and may lead to their mobility and agency being restricted within their natal home.

We find that when we expand the marital status definition to include not only girls who were married but also those who were engaged or had their marriage fixed at the time of the endline the effects of both programs become more pronounced. Now we see a significant negative impact in both arms of very similar magnitude. The reduction in the average probability of girls being married, engaged or having a fixed marriage at the time of the endline exceeds 5 p.p. and is equivalent to a 18% decrease relative to the control group. These results suggest that the interventions significantly reduced all marriage related activity for older girls not just formal marriage. Adding girls who were engaged or had a fixed marriage does not alter the findings for the younger girls - we continue to see no impacts of the programs among the younger girls and the pooled sample.

For both outcomes, marriage and education, the results strongly suggest that the two intervention models were equally effective; the third row in Table 2 shows that there are no statistically significant differences between the estimated treatment effects in the Girl Groups and Girl Groups and Community Campaigns arms. These results suggest that the improvements in the education and marriage outcomes of older girls were not driven by community engagement, but rather changes that were brought about by the girl groups (either through the education and/or sports activities). We return to this discussion below.

It is particularly striking that these activities appear to have reduced marriage, engagement and marriage fixing rates among the older cohort because girls in this context have very little say over the timing of their marriage. In our baseline data, less than 10% of girls report having a "big say" in whom and when to marry, with nearly half reporting having "no say". This is in contrast to decision-making about school, with around two-thirds of the girls reporting having a "big say" in when to leave school. Data from pre-baseline focus group discussions with mothers of adolescent girls further suggest that they consider girls who are out of school and not married at reputational risk and would start arranging a marriage once a girl has left education. This combination of factors suggests that the marriage and engagement impacts we find may be knock-on effects of changes in

26

the decisions girls are making about schooling as the result of the girl groups rather than changes in girls' say over marriage. This is also consistent with a strong negative correlation that we see between attending school and being married at endline. As we show in Table A.7 in Appendix A, a reduced-form mediation analysis, controlling for being in formal education reduces the size and significance of the marriage results.

4.2 Mental Health

Next we turn to the impacts of the interventions on girls' mental health. We find that the Girl Groups only program had no significant impact on girls' mental health - the overall measure of mental health, as well as separate measures of depression and anxiety. This holds for both younger and older girls (Table 3).

However, the addition of a community engagement component resulted in highly significant and very substantial improvements in mental health, especially for the older girls. We find a 38% of a standard deviation improvement ($p < 0.001$) in mental health of older girls and 20% of s.d. improvement for the younger girls ($p = 0.010$) (Table 3). The difference in effect sizes for older and younger girls is statistically significant. Looking separately at impacts on depression and anxiety, which make up the mental health factor, we see that for both cohorts, the symptoms of both were substantially reduced. Pooling the cohorts together, we find an overall improvement in mental health of close to 30% of s.d. ($p < 0.001$).

We test the robustness of these results to different ways of constructing the depression and anxiety measures. Tables A.13 and A.14 show estimates using both continuous raw scores (simply the sum of the likert responses given to each of the questions) for the Patient Health Questionnaire (PHQ-9) and the Generalized Anxiety Disorder (GAD-7) and binary indicators of depression and anxiety.[21] As with the factor scores, we see significant reductions in the raw depression and anxiety scores for the older girls and for the pooled sample in response to the addition of the community engagement component. Effects on younger girls also continue to be in the same direction, though only the impact of adding the community component continues to be significant ("Difference" row in the table), not the overall effect relative to the control group.

We find that the combined treatment also reduced binary indicators of both depression and anxiety. Our data suggest that 20% of the control group had symptoms consistent with (at least) mild depression while 4.2% had symptoms consistent with (at least) moderate depression; the combined treatment reduced the rate of depression defined by the milder cut-off by one quarter and that defined by the more-severe cut-off by 35%. Rates of (at least) mild anxiety fell by a quarter relative to a control group rate of 20%; anxiety defined by a more-severe cut-off was rare (3%) even in the control group and was not affected by treatment.

---

[21]The cut-off points used to define mild or moderate levels of Depression and Anxiety are based on Kroenke et al. (2001) and Kroenke et al. (2006).

Table 3: Mental Health

	Older Girls			Younger Girls			All		
	Factor	Depression	Anxiety	Factor	Depression	Anxiety	Factor	Depression	Anxiety
Girl Groups	0.015	0.008	-0.060	-0.041	-0.061	0.009	-0.014	0.035	-0.033
	(0.078)	(0.075)	(0.062)	(0.079)	(0.073)	(0.064)	(0.073)	(0.068)	(0.058)
	[p=0.847]	[p=0.918]	[p=0.336]	[p=0.601]	[p=0.402]	[p=0.884]	[p=0.849]	[p=0.602]	[p=0.563]
Girl Groups + Community	0.382***	-0.346***	-0.408***	0.196***	-0.186***	-0.266***	0.283***	-0.261***	-0.333***
	(0.077)	(0.072)	(0.065)	(0.076)	(0.070)	(0.063)	(0.070)	(0.064)	(0.057)
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.010]	[p=0.008]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
Difference	0.367***	-0.354***	-0.348***	0.237***	-0.246***	-0.257***	0.297***	-0.296***	-0.299***
	(0.067)	(0.067)	(0.056)	(0.068)	(0.062)	(0.061)	(0.061)	(0.056)	(0.051)
	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
N	2318	2318	2318	2725	2725	2725	5043	5043	5043
Control Mean	-0.063	-0.048	-0.059	0.047	0.041	0.052	-0.005	-0.001	0.000
Heterogeneity by Age									
Girl Groups				[p=0.329]	[p=0.356]	[p=0.340]			
Girl Groups + Community				[p=0.003]	[p=0.009]	[p=0.019]			
Difference				[p=0.024]	[p=0.086]	[p=0.113]			

Notes: The table presents estimated impacts of "Girl Groups" and "Girl Groups and Community Campaigns" interventions relative to the pure control group. "Difference" refers to differences between the two intervention models. Standard errors (in parentheses) and two-sided *p*-values are constructed using a cluster bootstrap accounting for stratification by block and treatment status (1000 iterations). Pre-specified controls are: a full set of interactions between dummy variables for age in years and a dummy variable for whether the girl was in school at baseline, caste, baseline wealth index, mother's years of education, district, and mental health at baseline. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

28

4.3 Robustness

We now discuss some additional sensitivity checks of our main results. First we show that the magnitude and significance of the main treatment effects on binary outcomes are very similar if we estimate an OLS model rather than our main logit specification (Table A.9).

Next, we test sensitivity of our results to selection of covariates included in the main regression models. As discussed in Section 3, the set of baseline controls that we include in our main specifications was pre-specified in a published pre-analysis plan (Andrew and Krutikova 2018). An alternative approach in the presence of a large number of baseline variables is to use machine learning to select the covariates for inclusion. We use the Post Double Selection Lasso (PDS) procedure, introduced by Belloni, Chernozhukov, and Hansen (2014), which we describe in Appendix B. Intuitively, the PDS procedure selects baseline variables that are both important predictors of treatment status *and* important predictors of the outcome of interest.

The PDS procedure selected several of the covariates that were pre-specified in the pre-analysis plan (e.g., school attendance, wealth index, whether or not elders were already talking about the girl's marriage at the time of baseline, and mothers years of education). However, it also identified a number of additional covariates. For example for the "In Education" outcome it selected hours that the girls reported studying at baseline, a measure of baseline cognitive skills as captured by a matrix reasoning test and desired age at marriage.[22] Overall, our results are robust to this alternative method of selecting covariates: we find that the estimated impacts of both intervention models remain similar in sign and magnitude whether we use the PDS-selected or a pre-specified set of covariates (Tables B.1 and B.2 in Appendix B).

Finally, we test whether our education and marriage results are robust to adjustments to who is included in the analysis sample. Our main education and marriage effects are estimated for the "extended sample", which includes girls who were directly interviewed at endline, as well as girls who were not but whose carers were re-interviewed (see Section 3). We find that very similar results hold, in magnitude and statistical significance, if we exclude the 482 girls (out of 5,525) who were not directly re-interviewed (Appendix Tables A.10 and A.11).[23]

5 Mechanisms and Discussion

Having documented the impacts of the two programs on the main outcomes of interest, we now explore the potential mechanisms underlying these findings. In doing so, we draw on the conceptual framework set out in Section 2.

---

[22]Full sets of covariates selected for each outcome are listed in Table Notes under Tables B.1, B.2 and B.3

[23]Additionally, the marriage results are robust to inclusion of girls whose marital status was captured during the pre-endline respondent tracking exercise but who are not included in the "extended sample" as neither they nor their caregivers were directly interviewed at endline (Table A.11).

5.1 Individual Preferences and Motivation

First we focus on the mechanisms behind education and marriage effects for the older girls. The lack of difference in impacts on these outcomes between the two interventions suggests that effects are driven by the Girl Groups program. As illustrated through our conceptual framework, if effects are achieved through intervening with girls alone, then girls' own preferences (captured through $u_S S + u_y y + vS$) are likely to be the mechanism. Such preference shifts may have arisen due to changes in girls immediate enjoyment of education and/or actions y (actions girls have to take to go to school which are at odds with what is considered "appropriate behavior"), as well as from changes in a wider set of consideration such as perceived returns to education.

While it is difficult to measure girls' valuation of education, we do measure the number of days of education per month that girls reported missing in the last month among those still at school at endline (six months after the end of the program). We think of this as a good proxy for girls' enjoyment and motivation to attend school despite the various obstacles they faced to doing so, especially since girls' "lack of interest" in school was the most prominent reason given for dropping out of school in our baseline data.[24] The first column of Table 4 shows that indeed both of the programs appear to have reduced school absences among the older girls.[25] There is also some indication of a smaller effect in the same direction among the younger girls (Column 4).

To assess whether the interventions may have affected girls' preferences in relation to actions y we assess impacts on the gender norms scale (described in Section 3.3.3). On this scale higher values indicate more progressive gender norms. While we do see evidence that the addition of community campaigns changed girls' perceived norms (and we return to this in the next subsection), we see no evidence that the Girl Group program changed older girls' stated norms.[26]

Taking stock, the analysis so far suggests that the Girl Groups (encompassing both the education and the sports sessions) were effective at keeping older girls in education *despite* strong external gender norms because they increased girls' day-to-day enthusiasm for school and/or their perception of the later rewards for pursuing education. Encouragingly, we see evidence consistent with Girl Groups achieving similar impacts on younger girls' attitudes to school. However, it is only among the older girls, who are at much higher risk of school drop out, that this attitudinal change translates into an increase in likelihood of staying at school.

---

[24]In a multiple-select question inquiring about the reasons they had dropped out, 30.1% of girls who were already out of school at baseline answered a "lack of interest" in school making this the most commonly chosen category. The second and third most-prominent reasons included "domestic responsibilities" and school being "too expensive".

[25]While the coefficient for the Girl Groups and Community Campaigns program is not statistically significant, it is also not significantly different from the significant coefficient for the Girl Groups only program.

[26]However, we also find an interesting *reduction* in progressivity of younger girls' norms in response to the Girl Groups only program, which is not there for the older girls. One explanation for this might be that before the start of the program the younger girls were not as aware of prevailing gender norms as the older girls. While the Girl Groups Program simply increased their awareness reflected in them reporting more patriarchal norms, in the Girl Groups and Community Arm this effect was offset by the more critical approach encouraged through discussions of these norms with the community.

5.2 Shared Norms

Next we turn to exploring the mechanisms behind the mental health effects. Specifically, we consider ways in which this outcome may have been impacted through the introduction Community Campaigns as we found that improvements in mental health were only present once these were added to the Girl Groups. Since the addition of community engagement activities did not change girls' behavior (education or marriage), our conceptual framework suggests two potential mechanisms through which community engagement may have led to improvements in mental health.[27] First, the community engagement activities may have changed community norms or what is considered "appropriate behaviour" (i.e. changing y^* in our model). In our model, keeping girls' behavior and educational choices fixed, if community norms become more progressive, the gap between girls' behavior y and norms y^* will be reduced which will lead to improvements in mental health. Second, the campaigns may have reduced the sanctions girls faced for deviation from norms (P) which would again lead to an improvement in mental health.[28]

We find direct evidence that both of these mechanisms were at play. Column 2 of Table 4 shows that the addition of the Community Campaigns led to a significant shift in girls' stated gender norms, rendering them, on average, almost 30 % of a standard deviation more progressive than those in the Girl Groups only arm. Furthermore, in Column 3 of Table 4, we show that it also resulted in a significant reduction in girls' perceptions of the likelihood that they would face violence and harassment if they broke with traditionally gendered ways of behaving. We show in Table 5 that the community engagement activities had a similar impact on the norms and perceived threat of violent sanctions of the girls' mothers, suggesting that the changes in girls' norms and perceptions are indicative of wider-ranging shifts in their communities.

So far, this narrative does not resolve *why* girls' behavior was left unchanged by the community campaigns, i.e. why did we not see a further increase in educational attendance in the combined treatment arm given that the campaigns relaxed the gender norms and the severity of punishments associated with breaking these. Indeed, our framework suggests that girls who would otherwise not attend education because of prohibitively restrictive norms may be induced to attend education if community campaigns caused norms to become more progressive. Our framework further highlights that for this group of girls the impact on mental health is *ambiguous*.[29]

One reason why we did not see a further shift in education (or marriage) from the addition of community campaigns may be that they resulted in norm shifts primarily for those girls who were *already* facing relatively more progressive norms. If this is the case then we would not necessarily expect a further increase in girls' education since norms would only be changing for girls who were

---

[27]Implicitly, this argument assumes that the *same* girls would have been moved into education by both treatments.

[28]Note that in Section 2, we presented evidence that norms relating to adolescent girls' behavior are enforced, at least in part, by threats of violence or harassment against girls who do not follow the rules.

[29]The marginal girl who moves into school does so at a cost to her mental health; the direction of the impact on mental health from the shift in norms arising from the community campaigns would, therefore, depend on whether the positive direct impacts are counterbalanced by the negative effects of more girls entering school.

already in education at very high rates. Table 6 shows the results of quantile regressions for girls' gender norms and perceived likelihood of sanctions and directly supports this explanation.[30] The addition of community engagement primarily shifted higher quantiles (i.e. more progressive) of the distribution of gender norms. This can explain why the campaigns had large impacts of the mental health: they caused a reduction in the sanctions faced by girls who were already facing (relatively) more-progressive norms and, therefore, were already in school at high rates, but did not shift norms sufficiently for girls facing less-progressive norms for any changes in educational choices to take place. Unfortunately, we do not have data on who attended the community engagement sessions so we are not able to test whether impacts are concentrated at the more progressive end of the distribution because events were better attended by communities where norms were already more progressive, or that these groups found the events more persuasive since the material did not conflict as much with their prior beliefs. We also see the biggest mental health impacts in higher quantiles of the mental health distribution (Panel C, Table 6). This is consistent with the community campaigns primarily shifting norms for girls who were already facing more progressive norms.

---

[30]As discussed in subsection 3.3.4, we use "plausible values" to estimate these conditional quantiles.

Table 4: Girl-Reported Missed School, Progressive Norms and Likelihood of Violent Sanctions

	Older Girls			Younger Girls			All		
	Schooldays missed	Progressive Norms	Violent Sanctions	Schooldays missed	Progressive Norms	Violent Sanctions	Schooldays missed	Progressive Norms	Violent Sanctions
Girl Groups	-1.506**	-0.089	-0.036	-0.476	-0.144**	0.000	-0.779*	-0.118**	-0.015
	(0.672)	(0.064)	(0.072)	(0.422)	(0.057)	(0.068)	(0.456)	(0.053)	(0.064)
	[p=0.025]	[p=0.166]	[p=0.616]	[p=0.259]	[p=0.012]	[p=0.996]	[p=0.087]	[p=0.025]	[p=0.814]
Girl Groups + Community	-0.867	0.197***	-0.127**	-0.694*	0.050	-0.121*	-0.712*	0.119**	-0.123**
	(0.700)	(0.071)	(0.064)	(0.382)	(0.063)	(0.071)	(0.399)	(0.060)	(0.062)
	[p=0.216]	[p=0.006]	[p=0.046]	[p=0.070]	[p=0.431]	[p=0.089]	[p=0.075]	[p=0.046]	[p=0.047]
Difference	0.639	0.285***	-0.091	-0.218	0.194***	-0.120*	0.067	0.237***	-0.108
	(0.609)	(0.068)	(0.077)	(0.207)	(0.065)	(0.072)	(0.260)	(0.059)	(0.068)
	[p=0.294]	[p=0.000]	[p=0.240]	[p=0.293]	[p=0.003]	[p=0.093]	[p=0.796]	[p=0.000]	[p=0.113]
N	996	2318	2318	2088	2725	2724	3084	5043	5042
Control Mean	4.938	-0.017	-0.018	2.856	0.018	0.018	3.538	0.001	0.002
Heterogeneity by Age									
Girl Groups				[p=0.043]	[p=0.343]	[p=0.516]			
Girl Groups + Community				[p=0.787]	[p=0.017]	[p=0.907]			
Difference				[p=0.155]	[p=0.138]	[p=0.611]			

Notes: The table presents estimated impacts of "Girl Groups" and "Girl Groups and Community Campaigns" interventions relative to the pure control group. "Difference" refers to differences between the two intervention models. Standard errors (in parentheses) and two-sided *p*-values are constructed using a cluster bootstrap accounting for stratification by block and treatment status (1000 iterations). Pre-specified controls are: a full set of interactions between dummy variables for age in years and a dummy variable for whether the girl was in school at baseline, caste, baseline wealth index, mother's years of education, district, whether the girl's family was talking about marriage at baseline, whether the girl's marriage was fixed at baseline, the girl's intended age of marriage at baseline, and the girl's mother's gender attitudes. * *p* < 0.1, ** *p* < 0.05, *** *p* < 0.01.

Table 5: Mother-Reported Progressive Norms and Likelihood of Violent Sanctions

	Older		Younger		All	
	Progressive Norms	Violent Sanctions	Progressive Norms	Violent Sanctions	Progressive Norms	Violent Sanctions
Girl Groups	-0.076	0.035	-0.095	-0.084	-0.086	-0.027
	(0.072)	(0.086)	(0.065)	(0.069)	(0.060)	(0.069)
	[p=0.292]	[p=0.687]	[p=0.143]	[p=0.221]	[p=0.155]	[p=0.699]
Girl Groups	0.093	-0.112	0.028	-0.096	0.060	-0.101
+ Community	(0.081)	(0.070)	(0.067)	(0.073)	(0.064)	(0.064)
	[p=0.251]	[p=0.112]	[p=0.675]	[p=0.186]	[p=0.349]	[p=0.116]
Difference	0.169**	-0.146	0.123*	-0.012	0.146**	-0.074
	(0.075)	(0.094)	(0.069)	(0.079)	(0.064)	(0.078)
	[p=0.024]	[p=0.119]	[p=0.076]	[p=0.876]	[p=0.023]	[p=0.343]
N	2608	2606	2197	2197	4805	4803
Control Mean	-0.016	-0.020	0.017	0.019	0.002	0.001
Heterogeneity by Age						
Girl Groups			[p=0.773]	[p=0.081]		
Girl Groups + Community			[p=0.361]	[p=0.814]		
Difference			[p=0.464]	[p=0.061]		

Notes: The table presents estimated impacts of "Girl Groups" and "Girl Groups and Community Campaigns" interventions relative to the pure control group. "Difference" refers to differences between the two intervention models. Standard errors (in parentheses) and two-sided p-values are constructed using a cluster bootstrap accounting for stratification by block and treatment status (1000 iterations). Pre-specified controls are: a full set of interactions between dummy variables for age in years and a dummy variable for whether the girl was in school at baseline, caste, baseline wealth index, mother's years of education, district, whether the girl's family was talking about marriage at baseline, whether the girl's marriage was fixed at baseline, the girl's intended age of marriage at baseline, and the girl's mother's gender attitudes. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table 6: Quantile Treatment Effects on Girls' Internalized Gender Norms, Percieved Likelihood of Violence and Mental Health

	10th	20th	30th	40th	50th	60th	70th	80th	90th
	Panel A: Gender Norms								
Girl Groups	-0.041	-0.064	-0.062	-0.065	-0.063	-0.085	-0.106	-0.112	-0.109
	(0.101)	(0.074)	(0.062)	(0.060)	(0.062)	(0.067)	(0.080)	(0.096)	(0.126)
	[p=0.688]	[p=0.391]	[p=0.322]	[p=0.281]	[p=0.308]	[p=0.203]	[p=0.185]	[p=0.245]	[p=0.390]
Girl Groups	0.030	0.055	0.110	0.134**	0.152**	0.175**	0.225***	0.290***	0.413***
+ Community	(0.098)	(0.079)	(0.067)	(0.066)	(0.070)	(0.079)	(0.086)	(0.099)	(0.132)
	[p=0.759]	[p=0.487]	[p=0.100]	[p=0.041]	[p=0.031]	[p=0.027]	[p=0.009]	[p=0.003]	[p=0.002]
Difference	0.070	0.119	0.172***	0.199***	0.214***	0.260***	0.331***	0.401***	0.521***
	(0.093)	(0.077)	(0.065)	(0.062)	(0.063)	(0.072)	(0.082)	(0.098)	(0.133)
	[p=0.447]	[p=0.120]	[p=0.008]	[p=0.001]	[p=0.001]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
N	2318	2318	2318	2318	2318	2318	2318	2318	2318
Control Quantile	-1.222	-0.810	-0.533	-0.298	-0.078	0.157	0.422	0.758	1.281
	Panel B: Perceived Likelihood of Violence								
Girl Groups	-0.146	-0.035	-0.016	-0.002	-0.013	-0.011	-0.023	-0.042	0.007
	(0.133)	(0.109)	(0.093)	(0.082)	(0.079)	(0.078)	(0.080)	(0.104)	(0.141)
	[p=0.273]	[p=0.746]	[p=0.861]	[p=0.981]	[p=0.871]	[p=0.886]	[p=0.775]	[p=0.686]	[p=0.962]
Girl Groups	-0.214	-0.133	-0.100	-0.111	-0.121	-0.106	-0.100	-0.102	-0.094
+ Community	(0.133)	(0.101)	(0.088)	(0.077)	(0.074)	(0.073)	(0.078)	(0.090)	(0.118)
	[p=0.107]	[p=0.188]	[p=0.253]	[p=0.149]	[p=0.105]	[p=0.146]	[p=0.197]	[p=0.254]	[p=0.423]
Difference	-0.068	-0.098	-0.084	-0.109	-0.108	-0.095	-0.078	-0.060	-0.101
	(0.160)	(0.120)	(0.094)	(0.086)	(0.081)	(0.082)	(0.084)	(0.109)	(0.133)
	[p=0.670]	[p=0.414]	[p=0.372]	[p=0.206]	[p=0.185]	[p=0.249]	[p=0.357]	[p=0.580]	[p=0.447]
N	2318	2318	2318	2318	2318	2318	2318	2318	2318
Control Quantile	-1.337	-0.820	-0.500	-0.240	-0.003	0.234	0.485	0.791	1.266
	Panel C: Mental Health								
Girl Groups	0.027	0.003	0.014	0.017	0.039	0.042	0.041	0.023	0.004
	(0.092)	(0.084)	(0.084)	(0.088)	(0.094)	(0.100)	(0.103)	(0.101)	(0.097)
	[p=0.769]	[p=0.976]	[p=0.864]	[p=0.847]	[p=0.677]	[p=0.672]	[p=0.694]	[p=0.823]	[p=0.970]
Girl Groups	0.229**	0.247***	0.311***	0.370***	0.434***	0.465***	0.472***	0.480***	0.500***
+ Community	(0.095)	(0.086)	(0.086)	(0.096)	(0.100)	(0.100)	(0.099)	(0.098)	(0.092)
	[p=0.016]	[p=0.004]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
Difference	0.201**	0.244***	0.296***	0.353***	0.395***	0.423***	0.431***	0.458***	0.496***
	(0.092)	(0.081)	(0.081)	(0.090)	(0.091)	(0.088)	(0.081)	(0.086)	(0.089)
	[p=0.028]	[p=0.002]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]	[p=0.000]
N	2318	2318	2318	2318	2318	2318	2318	2318	2318
Control Quantile	-1.183	-0.822	-0.544	-0.277	-0.002	0.302	0.630	0.987	1.459

Notes: The table presents estimated impacts of "Girl Groups" and "Girl Groups and Community Campaigns" interventions relative to the pure control group on quantiles of the distribution of girls' internalized gender norms, girls' perceived likelihood of violence, and mental health. "Difference" refers to differences between the two intervention models. Impacts for older girls only (15–17 at baseline). Standard errors (in parentheses) and two-sided p-values are constructed using a cluster bootstrap accounting for stratification by block and treatment status (1000 iterations). Pre-specified controls are: a full set of interactions between dummy variables for age in years and a dummy variable for whether the girl was in school at baseline, caste, baseline wealth index, mother's years of education, district, whether the girl's family was talking about marriage at baseline, whether the girl's marriage was fixed at baseline, the girl's intended age of marriage at baseline, and the gender attitudes of the girl's primary caregiver. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

6   Conclusions

If economic development alone is not enough to eliminate the striking gender gaps that exist in resource-poor settings characterized by conservative gender norms, a key question becomes whether there exist approaches that can help girls and women thrive in spite of these norms or whether a broader challenging of these norms is necessary to create meaningful change. In this paper, we use a randomized field experiment in order to compare two approaches to designing interventions that aim to improve the well-being of girls living in such circumstances. We focus on adolescent girls living in rural Rajasthan – a part of India with especially conservative gender norms. Like many programs targeting adolescent girls in contexts with restrictive gender norms, the first approach encouraged girls to question prevailing gender norms and act to improve their life outcomes in spite of these. The second additionally targeted the prevailing norm structures by helping girls engage with the wider community and, in particular, with community leaders who have the greatest influence in setting and enforcing gender norms. In order to capture well-being effects, we measured not only specific key outcomes targeted by the programs, such as education and early marriage, but also broader measures of mental health. Therefore, we do not assume that improvements in education and marriage age necessarily enhance all aspects of well-being.

This is the first paper to provide experimental evidence on the importance of challenging restrictive gender norms in interventions aiming to improve women's well-being. We show that while both approaches resulted in a significant and substantial reduction in school dropout and early marriage, these were only accompanied by an improvement in girls' mental health in the program that targeted the prevailing gender norms in the community in addition to the behavior of the girls themselves. This approach resulted in a significant reduction in mental health problems, especially among older girls who were at an age when norms become more constraining. In contrast, there was a striking lack of mental health improvement among the girls who participated in the program that targeted only them, in spite of a 15% increase in school attendance and a 20% reduction in marriage rates in this group.

Furthermore, we see that the mental health improvement is accompanied by changes in attitudes and perceptions that are consistent with a broader change in (perceived) prevailing gender norms achieved through community engagement activities. We see that the stated gender norms and attitudes became more progressive among the girls and their mothers and that the perceived likelihood of experiencing violence and harassment if girls deviated from traditionally gendered ways of behaving falls in both of these groups.

This is one of a handful of studies to provide new encouraging evidence that interventions which focus on the motivation, attitudes, and preferences of adolescent girls in conservative settings can have significant impacts on girls' trajectories without providing any additional resources or marketable skills. However, critically, our results show that any enhancements in well-being associated with increased schooling or delayed marriage may be offset by the costs of backlash incurred through

deviation from the community-level norms required to achieve these. These costs can be reduced through directly targeting the prevailing norms and sanctions. This suggests that, to improve the welfare of the adolescent girls programs that encourage them to challenge restrictive gender norms should also involve the wider community. Different approaches to engaging the different actors within the broader community is a key topic for future research.

References

Achyut, P., A. Andrew, M. Das, A. Gautam, M. Huepe, S. Krutikova, S. Kumar, S. Sharma, R. Soni, H. Verma, and R. Verma (2016). Promoting Adolescent Engagement, Knowledge and Health Evaluation of PAnKH : an adolescent girl intervention in Rajasthan, India. Baseline Report. Technical report.

Adoho, F., S. Chakravarty, D. T. Korkoyah, M. Lundberg, and A. Tasneem (2014). The Impact of an Adolescent Girls Employment Program: The EPAG Project in Liberia.

Ahrens, A., C. B. Hansen, and M. E. Schaffer (2018). PDSLASSO: Stata module for post-selection and post-regularization OLS or IV estimation and inference. *Statistical Software Components, Boston College Department of Economics*.

Ahrens, A. and M. E. Schaer (2018). Prediction, model selection, and causal inference with regularized regression.

Aizer, A. (2010). The gender wage gap and domestic violence. *American Economic Review 100* (4), 1847–1859.

Andrew, A. and A. Adams-Prassl (2021). Revealed Beliefs and the Marriage Market Return to Education. *Working Paper*.

Andrew, A. and S. Krutikova (2018, jan). Evaluation of an adolescent girl intervention in Rajasthan, India - Pre-analysis plan. Technical report, The Institute for Fiscal Studies.

Andrew, A., S. Krutikova, G. Smarrelli, H. Verma, A. Gautam, M. Das, P. Achyut, R. Soni, and S. Sharma (2018). Programme Implementation and Cost-Effectiveness Analysis Report: Promoting Adolescent Engagement, Knowledge and Health (PAnKH) in Rajasthan, India. *IFS Report*.

Ashraf, N., E. Field, and J. Lee (2014, jul). Household Bargaining and Excess Fertility: An Experimental Study in Zambia . *American Economic Review 104* (7), 2210–2237.

Bandiera, O., N. Buehren, R. Burgess, M. Goldstein, S. Gulesci, I. Rasul, and M. Sulaiman (2019). Women's Empowerment in Action: Evidence from a Randomized Control Trial in Africa. *American Economic Journal: Applied Economics*.

Barker, G., C. Ricardo, and M. Nascimento (2017). Engaging men and boys in changing gender-based inequity in health. *World Health Organization*.

Beaman, L., E. Duflo, R. Pande, and P. Topalova (2012, feb). Female Leadership Raises Aspirations and Educational Attainment for Girls: A Policy Experiment in India. *Science 335*(6068), 582–586.

Belloni, A., V. Chernozhukov, and C. Hansen (2014, May). High-Dimensional Methods and Inference on Structural and Treatment Effects. *Journal of Economic Perspectives 28*(2), 29–50.

Bergstrom, K. and B. Özler (2021). Improving the Well-Being of Adolescent Girls in Developing Countries. *World Bank Working Paper* (November).

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

Buchmann, N., E. Field, R. Glennerster, S. Nazneen, S. Pimkina, and I. Sen (2017). Power vs Money: Alternative Approaches to Reducing Child Marriage in Bangladesh, a Randomized Control Trial.

Buehren, N., M. Goldstein, S. Gulesci, M. Sulaiman, and V. Yam (2017). Evaluation of an Adolescent Development Program for Girls in Tanzania.

Buehren, N., K. Leonard, J. Montalvao, K. Vasilaky, and M. P. Goldstein (2016). Women's empowerment, sibling rivalry, and competitiveness: evidence from a lab experiment and a randomized control trial in Uganda.

Bursztyn, L., A. L. González, and D. Yanagizawa-Drott (2018). Misperceived Social Norms: Female Labor Force Participation in Saudi Arabia.

Das, J. and T. Zajonc (2010). India shining and Bharat drowning: Comparing two Indian states to the worldwide distribution in mathematics achievement. *Journal of Development Economics 92*(2), 175–187.

Dhar, D., T. Jain, and S. Jayachandran (2018). Reshaping Adolescents' Gender Attitudes: Evidence from a School-Based Experiment in India.

Duflo, E. (2012, dec). Women Empowerment and Economic Development. *Journal of Economic Literature 50*(4), 1051–1079.

Edmonds, E., B. Feigenberg, and J. Leight (2021, 07). Advancing the Agency of Adolescent Girls. *Review of Economics and Statistics*, 1–46.

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 100*(2), 125–129.

Field, E., S. Jayachandran, R. Pande, and N. Rigol (2016). Friendship at work: Can peer effects catalyze female entrepreneurship? *American Economic Journal: Economic Policy 8*(2), 125–153.

Field, E., R. Pande, N. Rigol, S. Schaner, and C. Moore (2019). On Her Own Account: How Strengthening Women's Financial Control Impacts Labor Supply and Gender Norms. *SSRN Electronic Journal* (2201).

Finnemore, M. and K. Sikkink (1998). International Norm Dynamics and Political Change. *International Organization 52*, 887–917.

Ganguly, S., M. Samanta, P. Roy, S. Chatterjee, D. W. Kaplan, and B. Basu (2013). Patient health questionnaire-9 as an effective tool for screening of depression among indian adolescents. *Journal of Adolescent Health 52*(5), 546–551.

Guarnieri, E. and H. Rainer (2018). Female Empowerment and Male Backlash. *Working Paper*, 1–55.

Heller, P. and V. Rao (2015, jul). *Deliberation and Development: Rethinking the Role of Voice and Collective Action in Unequal Societies*. The World Bank.

Jayachandran, S. (2015). The roots of gender inequality in developing countries. *Annual Review of Economics 7*(1), 63–88.

Jayachandran, S. (2021). Social Norms as a Barrier to Women's Employment in Developing Countries. *IMF Economic Review 69*, 576–595.

Jensen, R. (2012). Do labor market opportunities affect young women's work and family decisions? Experimental evidence from India. *Quarterly Journal of Economics 127*(2), 753–792.

Kandpal, E. and K. Baylis (2019, sep). The social lives of married women: Peer effects in female autonomy and investments in children. *Journal of Development Economics 140*(April), 26–43.

Kondratek, B. (2022). UIRT: Stata module to fit unidimensional Item Response Theory models. *Statistical Software Components S458247*.

Krishnan, S., C. H. Rocca, A. E. Hubbard, K. Subbiah, J. Edmeades, and N. S. Padian (2010, jan). Do changes in spousal employment status lead to domestic violence? Insights from a prospective study in Bangalore, India. *Social Science & Medicine 70*(1), 136–143.

Kroenke, K., R. L. Spitzer, and J. B. W. Williams (2001). The PHQ-9. *Journal of General Internal Medicine 16*(9), 606–613.

Kroenke, K., R. L. Spitzer, J. B. W. Williams, and B. Lwe (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine 166*(10), 1092–1097.

Leventhal, K. S., L. M. DeMaria, J. Gillham, G. Andrew, J. W. Peabody, and S. Leventhal (2015). Fostering emotional, social, physical and educational wellbeing in rural India: the methods of a multi-arm randomized controlled trial of Girls First. *Trials 16*(1), 481.

Leventhal, K. S., J. Gillham, L. DeMaria, G. Andrew, J. Peabody, and S. Leventhal (2015). Building psychosocial assets and wellbeing among adolescent girls: A randomized controlled trial. *Journal of Adolescence 45*, 284–295.

Macdonald, K. (2008). PV: Stata module to perform estimation with plausible values. *Statistical Software Components S456951*.

Muralidharan, K. and N. Prakash (2017, July). Cycling to school: Increasing secondary school enrollment for girls in india. *American Economic Journal: Applied Economics 9*(3), 321–50.

Pulerwitz, J. and G. Barker (2008). Measuring attitudes toward gender norms among young men in brazil: development and psychometric evaluation of the gem scale. *Men and masculinities 10*(3), 322–338.

Riley, E. (2022, jan). Role Models in Movies: The Impact of Queen of Katwe on Students' Educational Attainment. *The Review of Economics and Statistics*, 1–48.

Sanyal, P. (2009). From credit to collective action: The role of microfinance in promoting women's social capital and normative influence. *American Sociological Review 74*(4), 529–550.

A Additional Tables and Figures

A.1 Additional Figures

Figure A.1: Girls' Stated Aspirations versus Reality in Education and Marriage
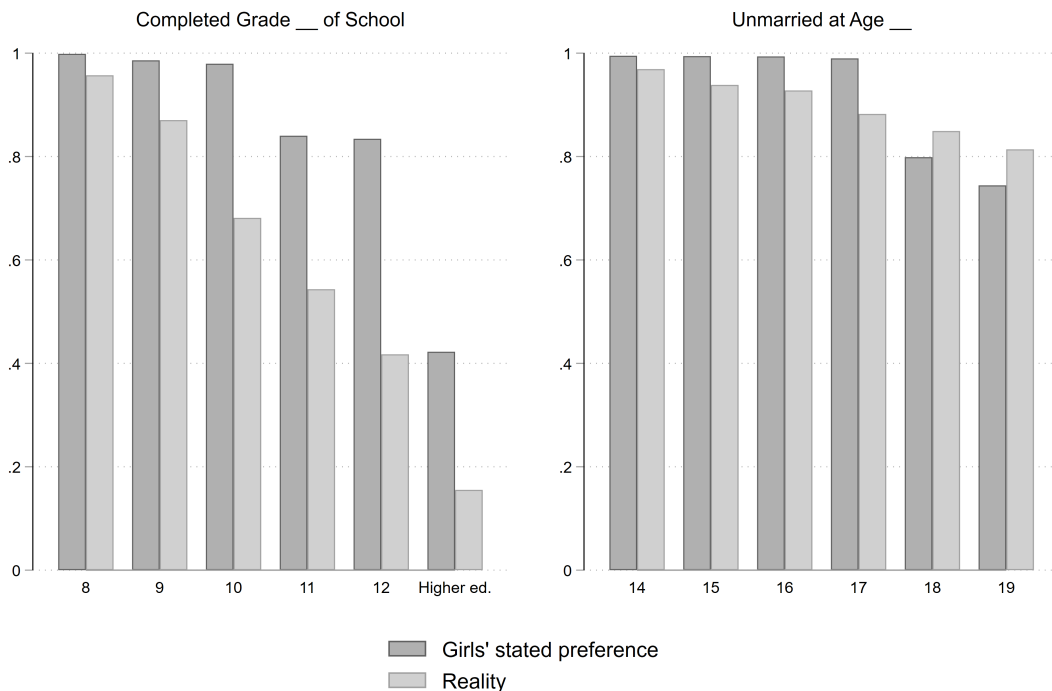
Notes: Figure shows that gap between aspirations measured at baseline and reality at endline amongst sample girls for education and marriage outcomes. In dark gray, we plot the proportion of girls who at baseline said that they wanted to study until at least that grade (left hand figure) and the proportion who said they still wanted to be unmarried at that age (right hand figure). In pale gray, we plot the proportion who did indeed remain in school/education until at least that grade and the proportion who did indeed remain unmarried until at least that age.

Figure A.2: Descriptive Relationship between Gender Norms, Violence and Education and Mental Health in the Control Group (Older Sample)

((a)) Norms & Education

((b)) Norms & Mental Health

((c)) Violence & Education

((d)) Violence & Mental Health
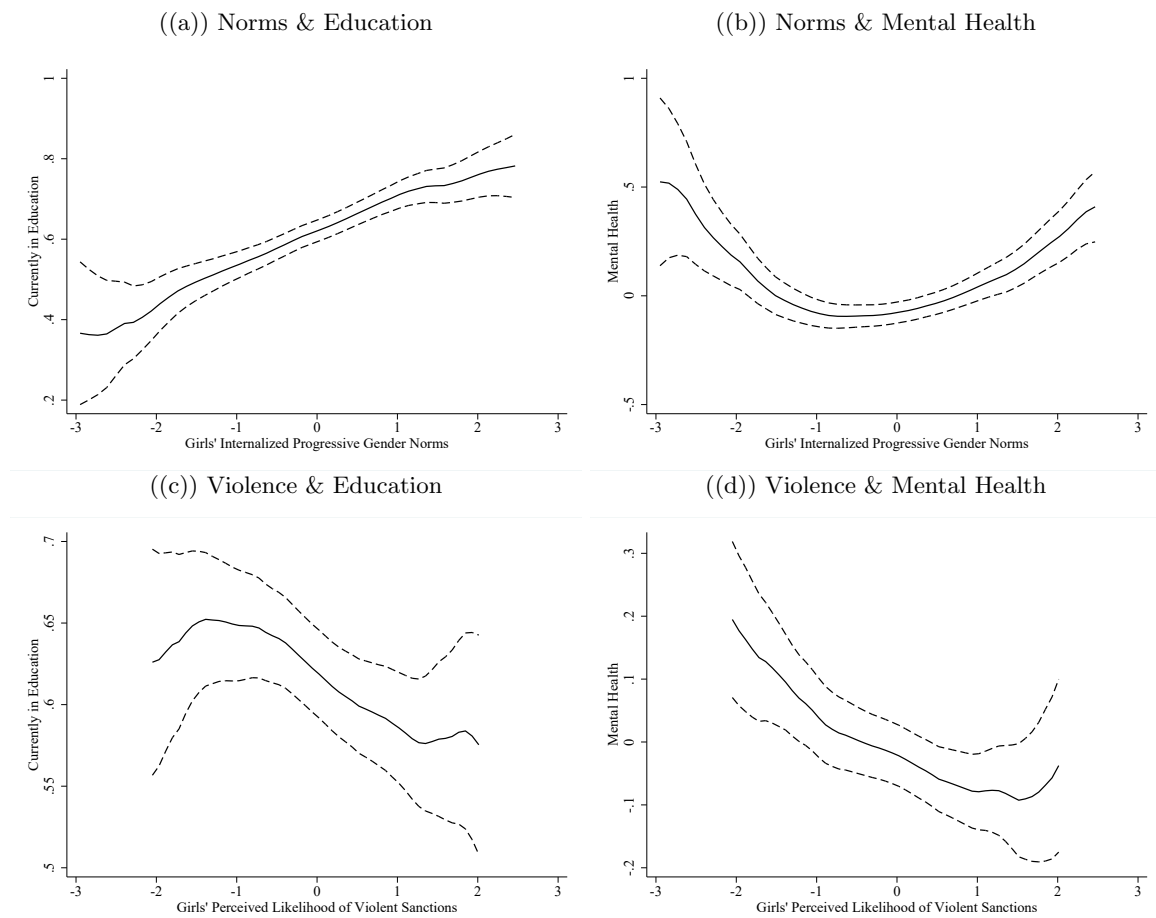
Notes: Older sample only (15-17 at baseline). Solid lines plot relationships between girls' perceived gender norms, girls' perceived likelihood of facing sanctions for breaking norms, girls' school enrollment and girls' mental health in the control group using kernel-weighted local polynomial smoothing with an Epanechnikov and a bandwidth of 0.5. Dashed lines plot 95% confidence intervals.

Figure A.3: Descriptive Relationship between Gender Norms, Violence and Education and Mental Health in the Control Group (Younger Sample)

((a)) Norms & Education

((b)) Norms & Mental Health

((c)) Violence & Education

((d)) Violence & Mental Health
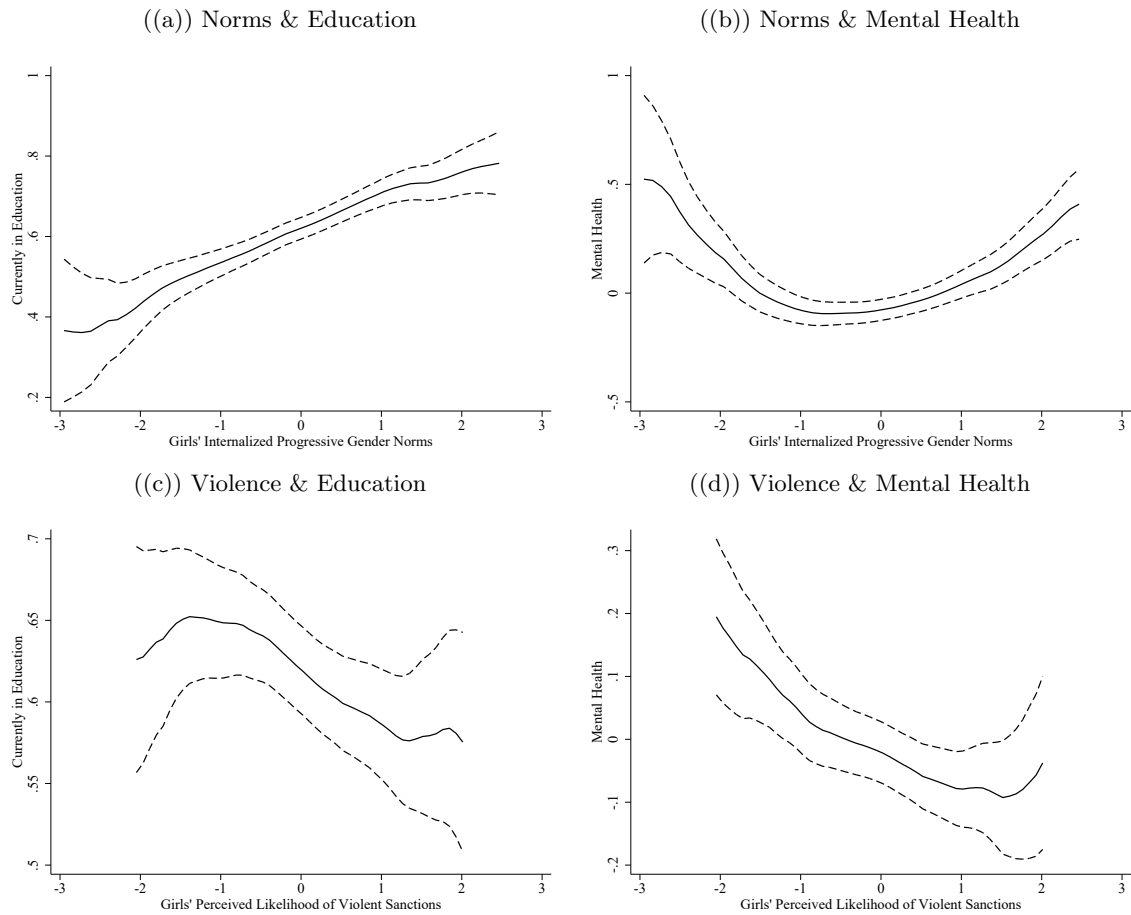
Notes: Younger sample only (12-14 at baseline). Solid lines plot relationships between girls' perceived gender norms, girls' perceived likelihood of facing sanctions for breaking norms, girls' school enrollment and girls' mental health in the control group using kernel-weighted local polynomial smoothing with an Epanechnikov and a bandwidth of 0.5. Dashed lines plot 95% confidence intervals.

Figure A.4: Descriptive Relationship between *Mothers'* Gender Norms, and *Daughter's* Education and Mental Health in the Control Group

((a)) Norms & Education

((b)) Norms & Mental Health

((c)) Violence & Education

((d)) Violence & Mental Health
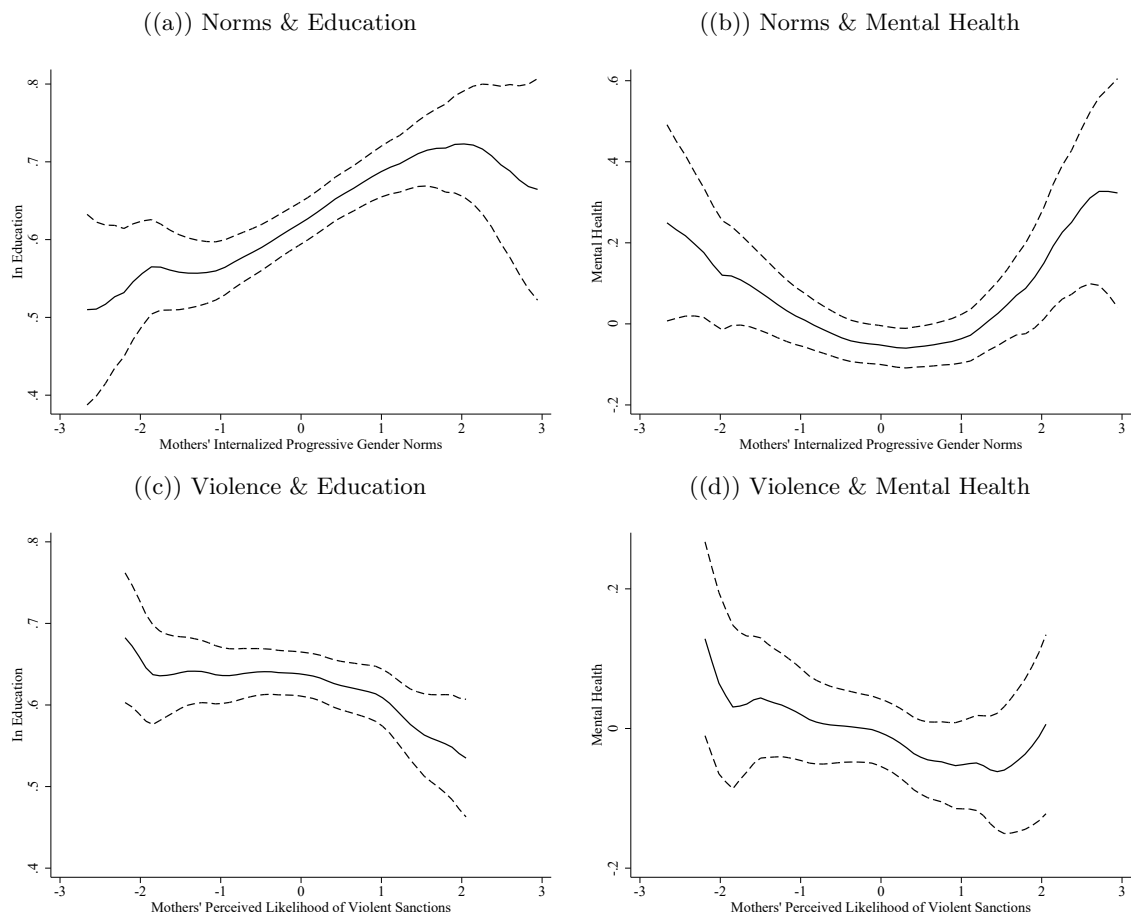
Notes: Solid lines plot relationships between mothers' perceived gender norms, mothers' perceived likelihood of facing sanctions for breaking norms, girls' school enrollment and girls' mental health in the control group using kernel-weighted local polynomial smoothing with an Epanechnikov and a bandwidth of 0.5. Dashed lines plot 95% confidence intervals.

A.2 Additional Tables

Table A.1: Predictors of Attrition

	(1) Main Sample	(2) Main Sample	(3) Extended Sample	(4) Extended Sample
Girl Groups	-0.00101 (0.0120)	0.000280 (0.0115)	-0.00789 (0.00674)	-0.00717 (0.00647)
Girl Groups + Community	-0.00531 (0.0141)	-0.00133 (0.0139)	-0.00332 (0.00786)	-0.000815 (0.00751)
Baseline Age=13		0.0157 (0.0128)		0.0150** (0.00705)
Baseline Age=14		-0.0161 (0.0149)		0.00836 (0.00918)
Baseline Age=15		-0.0139 (0.0147)		0.0154* (0.00791)
Baseline Age=16		-0.0318** (0.0142)		0.0132 (0.00803)
Baseline Age=17		-0.0598*** (0.0151)		0.00702 (0.00952)
Carer's Years of Education		-0.00348* (0.00183)		-0.000873 (0.00117)
SC/ST Caste		-0.0208* (0.0110)		-0.0162*** (0.00614)
Asset Index		0.0148** (0.00572)		0.00240 (0.00365)
Attending School at Baseline		0.0570*** (0.0122)		0.0232*** (0.00825)
Progressive Gender Norms		0.00277 (0.00494)		-0.000851 (0.00313)
Carer's Progressive Gender Norms		-0.00643 (0.00539)		-0.00279 (0.00303)
Mental Health		0.00929** (0.00379)		0.00332 (0.00210)
Constant	0.882*** (0.00645)	0.864*** (0.0164)	0.968*** (0.00426)	0.946*** (0.0104)
Observations	5731	5731	5731	5731

Notes: The table presents OLS coefficients and standard errors (in parentheses, corrected for clustering) for regressions of a girl being in the main sample (columns 1 and 2) and the extended sample (columns 3 and 4) on treatment status. In columns 2 and 4 we additionally control for key baseline characteristics. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2: Baseline Characteristics and Balance (Full Baseline Sample)

	Younger Girls				Older Girls			
	(1) Control	(2) Girl Groups	(3) Girl Groups + Community	(4) (3)-(2)	(5) Control	(6) Girl Groups	(7) Girl Groups + Community	(8) (7)-(6)
Girl Age	12.93	12.89	12.93		15.86	15.89	15.89	
	(0.859)	(0.852)	(0.854)		(0.793)	(0.810)	(0.808)	
		[p=.361]	[p=.902]	[p=.301]		[p=.393]	[p=.342]	[p=.95]
Carer's Years of Ed.	1.433	1.045	1.222		1.222	0.783*	0.886*	
	(2.869)	(2.550)	(2.572)		(2.737)	(2.219)	(2.219)	
		[p=.125]	[p=.339]	[p=.447]		[p=.058]	[p=.086]	[p=.62]
SC/ST Caste	0.327	0.339	0.409		0.305	0.348	0.394	
	(0.469)	(0.474)	(0.492)		(0.460)	(0.477)	(0.489)	
		[p=.842]	[p=.169]	[p=.284]		[p=.471]	[p=.154]	[p=.479]
Asset Index	-0.0927	-0.151	-0.124		0.100	0.00241	0.00260	
	(0.849)	(0.798)	(0.806)		(0.976)	(0.883)	(0.905)	
		[p=.412]	[p=.659]	[p=.645]		[p=.15]	[p=.173]	[p=.998]
Attending School	0.913	0.898	0.852***		0.657	0.627	0.626	
	(0.283)	(0.302)	(0.355)		(0.475)	(0.484)	(0.484)	
		[p=.526]	[p=.001]	[p=.049]		[p=.335]	[p=.335]	[p=.965]
Progressive	-0.0203	-0.0961	0.0107		0.0793	-0.0932*	0.112	
Gender Norms	(0.981)	(0.991)	(0.934)		(0.989)	(1.028)	(0.989)	
		[p=.39]	[p=.713]	[p=.141]		[p=.065]	[p=.71]	[p=.012]
Carer's Progressive	0.0400	-0.0612	0.0549		0.0345	-0.0642	0.0151	
Gender Norms	(1.002)	(1.001)	(0.972)		(0.978)	(0.987)	(0.981)	
		[p=.239]	[p=.867]	[p=.134]		[p=.200]	[p=.819]	[p=.277]
Mental Health	-0.0579	-0.0857	-0.0584		0.0981	-0.00833	0.0594	
	(0.988)	(0.957)	(0.999)		(0.967)	(1.005)	(0.965)	
		[p=.73]	[p=.996]	[p=.765]		[p=.201]	[p=.637]	[p=.423]
Observations	1030	985	1016		962	864	874	

Notes: The table presents means and standard deviations (in parentheses) by treatment group and by age strata for the entire baseline sample, regardless of whether or not we re-interviewed them at endline. Two-sided *p*-values test the difference between each treatment group mean and the control. The *p*-values and standard errors are constructed using a cluster bootstrap accounting for stratification by region and treatment status.

Table A.3: Baseline Characteristics and Balance (Extended Sample)

	Younger Girls				Older Girls			
	(1) Control	(2) Girl Groups	(3) Girl Groups + Community	(4) (3)-(2)	(5) Control	(6) Girl Groups	(7) Girl Groups + Community	(8) (7)-(6)
Girl Age	12.93	12.90	12.93		15.85	15.89	15.88	
	(0.858)	(0.850)	(0.854)		(0.791)	(0.813)	(0.805)	
		[p=.317]	[p=.999]	[p=.323]		[p=.32]	[p=.39]	[p=.862]
Carer's Years of Ed.	1.431	1.039	1.225		1.197	0.794*	0.911	
	(2.842)	(2.539)	(2.562)		(2.668)	(2.242)	(2.252)	
		[p=.106]	[p=.338]	[p=.425]		[p=.086]	[p=.152]	[p=.582]
SC/ST Caste	0.323	0.334	0.404		0.305	0.349	0.385	
	(0.468)	(0.472)	(0.491)		(0.460)	(0.477)	(0.487)	
		[p=.85]	[p=.18]	[p=.29]		[p=.466]	[p=.207]	[p=.592]
Asset Index	-0.0890	-0.144	-0.120		0.0933	0.00302	0.0156	
	(0.840)	(0.799)	(0.805)		(0.961)	(0.885)	(0.908)	
		[p=.429]	[p=.66]	[p=.681]		[p=.191]	[p=.287]	[p=.862]
Attending School	0.914	0.901	0.855***		0.663	0.633	0.632	
	(0.281)	(0.299)	(0.353)		(0.473)	(0.482)	(0.483)	
		[p=.556]	[p=.001]	[p=.042]		[p=.325]	[p=.351]	[p=.985]
Progressive	-0.0222	-0.108	0.0139		0.0741	-0.0913*	0.120	
Gender Norms	(0.981)	(0.991)	(0.935)		(0.998)	(1.039)	(0.997)	
		[p=.318]	[p=.668]	[p=.087]		[p=.085]	[p=.606]	[p=.012]
Carer's Progressive	0.0316	-0.0695	0.0602		0.0212	-0.0624	0.0204	
Gender Norms	(1.004)	(1.008)	(0.973)		(0.981)	(0.991)	(0.986)	
		[p=.233]	[p=.748]	[p=.092]		[p=.29]	[p=.993]	[p=.272]
Mental Health	-0.0536	-0.0788	-0.0534		0.103	-0.0143	0.0688	
	(0.996)	(0.968)	(1.009)		(0.978)	(1.019)	(0.965)	
		[p=.759]	[p=.998]	[p=.786]		[p=.161]	[p=.684]	[p=.322]
Observations	999	939	983		929	836	840	

Notes: The table presents means and standard deviations (in parentheses) by treatment group and by age strata for the extended sample. This includes all girls for whom we either directly interviewed at endline or we directly interviewed their mother/caregiver. Two-sided *p*-values test the difference between each treatment group mean and the control. The *p*-values and standard errors are constructed using a cluster bootstrap accounting for stratification by region and treatment status.

Table A.4: Parameters of IRT measurement model for mental health scales

	(1) Combined		(2) Depression (PHQ-9)		(3) Anxiety (GAD-7)
Over the last 2 weeks, how often have you had little interest or pleasure in doing things?					
$\hat{\beta}_j$	2.187***	(0.0864)	1.987***	(0.0856)	
$\hat{\alpha}_{j1}$	-2.799***	(0.101)	-2.978***	(0.115)	
$\hat{\alpha}_{j2}$	-1.809***	(0.0578)	-1.916***	(0.0655)	
$\hat{\alpha}_{j3}$	-0.739***	(0.0337)	-0.765***	(0.0360)	
Over the last 2 weeks, how often have you been feeling down, depressed, or hopeless?					
$\hat{\beta}_j$	2.206***	(0.0859)	1.933***	(0.0827)	
$\hat{\alpha}_{j1}$	-2.832***	(0.102)	-3.071***	(0.120)	
$\hat{\alpha}_{j2}$	-1.976***	(0.0626)	-2.129***	(0.0732)	
$\hat{\alpha}_{j3}$	-0.612***	(0.0325)	-0.636***	(0.0349)	
Over the last 2 weeks, how often have you had trouble falling or staying asleep, or sleeping too much?					
$\hat{\beta}_j$	2.158***	(0.0924)	2.267***	(0.104)	
$\hat{\alpha}_{j1}$	-2.921***	(0.111)	-2.888***	(0.112)	
$\hat{\alpha}_{j2}$	-2.096***	(0.0696)	-2.091***	(0.0711)	
$\hat{\alpha}_{j3}$	-1.067***	(0.0390)	-1.057***	(0.0397)	
Over the last 2 weeks, how often have you been feeling tired or had little energy?					
$\hat{\beta}_j$	1.756***	(0.0676)	1.768***	(0.0747)	
$\hat{\alpha}_{j1}$	-3.144***	(0.119)	-3.164***	(0.124)	
$\hat{\alpha}_{j2}$	-1.897***	(0.0633)	-1.916***	(0.0667)	
$\hat{\alpha}_{j3}$	-0.476***	(0.0335)	-0.468***	(0.0344)	
Over the last 2 weeks, how often have you had poor appetite or been overeating?					
$\hat{\beta}_j$	1.897***	(0.0773)	2.157***	(0.0942)	
$\hat{\alpha}_{j1}$	-3.074***	(0.118)	-2.904***	(0.110)	
$\hat{\alpha}_{j2}$	-2.052***	(0.0690)	-1.968***	(0.0661)	
$\hat{\alpha}_{j3}$	-0.826***	(0.0363)	-0.788***	(0.0355)	
Over the last 2 weeks, how often have you been feeling bad about yourself or that you are a failure or have let yourself or your family down?					
$\hat{\beta}_j$	2.365***	(0.117)	2.641***	(0.141)	
$\hat{\alpha}_{j1}$	-3.637***	(0.185)	-3.516***	(0.178)	
$\hat{\alpha}_{j2}$	-2.485***	(0.0892)	-2.425***	(0.0869)	
$\hat{\alpha}_{j3}$	-1.493***	(0.0494)	-1.460***	(0.0486)	
Over the last 2 weeks, how often have you had trouble concentrating on things, such as reading the newspaper or watching television?					
$\hat{\beta}_j$	2.153***	(0.102)	2.295***	(0.116)	
$\hat{\alpha}_{j1}$	-3.423***	(0.155)	-3.361***	(0.153)	
$\hat{\alpha}_{j2}$	-2.432***	(0.0872)	-2.409***	(0.0874)	
$\hat{\alpha}_{j3}$	-1.415***	(0.0482)	-1.397***	(0.0483)	
Over the last 2 weeks, how often have you been moving or speaking so slowly that other people could have noticed? Or the opposite being so fidgety or restless that you have been moving around a lot more than usual?					
$\hat{\beta}_j$	2.245***	(0.114)	2.536***	(0.139)	
$\hat{\alpha}_{j1}$	-3.665***	(0.185)	-3.516***	(0.176)	
$\hat{\alpha}_{j2}$	-2.621***	(0.0983)	-2.535***	(0.0942)	
$\hat{\alpha}_{j3}$	-1.602***	(0.0540)	-1.554***	(0.0523)	
Over the last 2 weeks, how often have you had thoughts that you would be better off dead or of hurting yourself in some way?					
$\hat{\beta}_j$	2.018***	(0.104)	2.249***	(0.122)	
$\hat{\alpha}_{j1}$	-3.796***	(0.194)	-3.630***	(0.183)	
$\hat{\alpha}_{j2}$	-2.685***	(0.105)	-2.591***	(0.0995)	

	Col 1		Col 2		Col 3	
$\hat{\alpha}_{j3}$	-1.676***	(0.0588)	-1.622***	(0.0565)		

Over the last 2 weeks, how often have you been feeling nervous, anxious or on edge?

	Col 1		Col 2		Col 3	
$\hat{\beta}_j$	1.645***	(0.0672)			1.912***	(0.0842)
$\hat{\alpha}_{j1}$	-2.757***	(0.102)			-2.562***	(0.0939)
$\hat{\alpha}_{j2}$	-2.135***	(0.0749)			-1.989***	(0.0698)
$\hat{\alpha}_{j3}$	-0.739***	(0.0369)			-0.660***	(0.0356)

Over the last 2 weeks, how often have you not been able to stop or control worrying?

	Col 1		Col 2		Col 3	
$\hat{\beta}_j$	2.001***	(0.0847)			2.491***	(0.117)
$\hat{\alpha}_{j1}$	-2.823***	(0.105)			-2.594***	(0.0937)
$\hat{\alpha}_{j2}$	-2.103***	(0.0710)			-1.945***	(0.0650)
$\hat{\alpha}_{j3}$	-1.015***	(0.0390)			-0.921***	(0.0370)

Over the last 2 weeks, how often have you been worrying too much about different things?

	Col 1		Col 2		Col 3	
$\hat{\beta}_j$	1.859***	(0.0742)			2.199***	(0.0962)
$\hat{\alpha}_{j1}$	-2.651***	(0.0942)			-2.472***	(0.0872)
$\hat{\alpha}_{j2}$	-1.944***	(0.0652)			-1.818***	(0.0615)
$\hat{\alpha}_{j3}$	-0.722***	(0.0351)			-0.644***	(0.0341)

Over the last 2 weeks, how often have you been trouble relaxing?

	Col 1		Col 2		Col 3	
$\hat{\beta}_j$	2.255***	(0.0964)			2.315***	(0.109)
$\hat{\alpha}_{j1}$	-2.690***	(0.0973)			-2.702***	(0.101)
$\hat{\alpha}_{j2}$	-2.036***	(0.0667)			-2.034***	(0.0694)
$\hat{\alpha}_{j3}$	-1.043***	(0.0381)			-1.013***	(0.0392)

Over the last 2 weeks, how often have you been being so restless that it is hard to sit still?

	Col 1		Col 2		Col 3	
$\hat{\beta}_j$	2.411***	(0.104)			2.283***	(0.109)
$\hat{\alpha}_{j1}$	-2.799***	(0.104)			-2.927***	(0.114)
$\hat{\alpha}_{j2}$	-2.113***	(0.0684)			-2.188***	(0.0756)
$\hat{\alpha}_{j3}$	-1.037***	(0.0373)			-1.035***	(0.0398)

Over the last 2 weeks, how often have you been becoming easily annoyed or irritable?

	Col 1		Col 2		Col 3	
$\hat{\beta}_j$	1.908***	(0.0735)			1.699***	(0.0729)
$\hat{\alpha}_{j1}$	-2.983***	(0.110)			-3.217***	(0.129)
$\hat{\alpha}_{j2}$	-1.861***	(0.0609)			-1.976***	(0.0706)
$\hat{\alpha}_{j3}$	-0.557***	(0.0333)			-0.544***	(0.0357)

Over the last 2 weeks, how often have you been feeling afraid as if something awful might happen?

	Col 1		Col 2		Col 3	
$\hat{\beta}_j$	2.165***	(0.0934)			1.706***	(0.0806)
$\hat{\alpha}_{j1}$	-3.090***	(0.123)			-3.562***	(0.158)
$\hat{\alpha}_{j2}$	-2.215***	(0.0744)			-2.513***	(0.0956)
$\hat{\alpha}_{j3}$	-1.103***	(0.0398)			-1.200***	(0.0476)

| N | 5043 | | 5043 | | 5043 | |

Notes: Table presents estimated parameters and standard errors (in parentheses) for IRT measurement model of girls' mental health. Measurement model and estimation procedure described in Section 3.3.4. Column 1 presents the measurement model for overall mental health which combines items from both the depression and anxiety scales. Column 2 presents the measurement model for depression while column 3 does the same for anxiety. Items in column 1 are reverse coded so that the combined mental health factor is increasing in mental health while the depression and anxiety factors are increasing in symptoms of depression and anxiety.

Table A.5: Parameters of IRT measurement model for gender norms scales

	(1) Girls		(2) Carers	
Boys should be given more privilege as compared to girls.				
$\hat{\beta}_j$	1.070***	(0.0407)	0.961***	(0.0389)
$\hat{\alpha}_{j1}$	-2.295***	(0.0868)	-1.712***	(0.0744)
$\hat{\alpha}_{j2}$	-1.098***	(0.0514)	-0.485***	(0.0442)
$\hat{\alpha}_{j3}$	0.423***	(0.0418)	1.267***	(0.0633)
Women/girls should work only if there are monetary needs in their family				
$\hat{\beta}_j$	1.093***	(0.0402)	1.076***	(0.0419)
$\hat{\alpha}_{j1}$	-1.137***	(0.0530)	-0.809***	(0.0483)
$\hat{\alpha}_{j2}$	0.158***	(0.0381)	0.651***	(0.0446)
$\hat{\alpha}_{j3}$	1.471***	(0.0617)	2.053***	(0.0823)
Only bad girls make male friends				
$\hat{\beta}_j$	1.057***	(0.0398)	1.105***	(0.0428)
$\hat{\alpha}_{j1}$	-1.669***	(0.0675)	-0.794***	(0.0472)
$\hat{\alpha}_{j2}$	-0.673***	(0.0436)	0.238***	(0.0393)
$\hat{\alpha}_{j3}$	0.936***	(0.0509)	1.778***	(0.0732)
A man should have the final say in all family matters				
$\hat{\beta}_j$	1.306***	(0.0459)	1.371***	(0.0508)
$\hat{\alpha}_{j1}$	-0.667***	(0.0403)	-0.173***	(0.0357)
$\hat{\alpha}_{j2}$	0.499***	(0.0380)	0.918***	(0.0445)
$\hat{\alpha}_{j3}$	1.646***	(0.0616)	2.111***	(0.0764)
Men should be more educated than their wives				
$\hat{\beta}_j$	1.458***	(0.0497)	1.407***	(0.0515)
$\hat{\alpha}_{j1}$	-0.741***	(0.0397)	-0.305***	(0.0364)
$\hat{\alpha}_{j2}$	0.384***	(0.0353)	1.098***	(0.0476)
$\hat{\alpha}_{j3}$	1.486***	(0.0552)	2.233***	(0.0795)
Boys are naturally better than girls in studies				
$\hat{\beta}_j$	1.507***	(0.0506)	1.291***	(0.0463)
$\hat{\alpha}_{j1}$	-1.578***	(0.0556)	-1.400***	(0.0562)
$\hat{\alpha}_{j2}$	-0.519***	(0.0355)	-0.105**	(0.0359)
$\hat{\alpha}_{j3}$	0.759***	(0.0408)	1.345***	(0.0564)
A daughter deserves to be beaten if she does not obey her parents.				
$\hat{\beta}_j$	1.072***	(0.0397)	1.267***	(0.0469)
$\hat{\alpha}_{j1}$	-0.930***	(0.0488)	-0.411***	(0.0389)
$\hat{\alpha}_{j2}$	0.412***	(0.0405)	0.885***	(0.0451)
$\hat{\alpha}_{j3}$	1.736***	(0.0696)	2.143***	(0.0789)
Girls should be married early to protect them from sexual harassment.				
$\hat{\beta}_j$	2.207***	(0.0729)	2.046***	(0.0694)
$\hat{\alpha}_{j1}$	-1.237***	(0.0434)	-0.861***	(0.0390)
$\hat{\alpha}_{j2}$	-0.386***	(0.0311)	0.0718*	(0.0311)
$\hat{\alpha}_{j3}$	0.588***	(0.0343)	1.163***	(0.0443)
Girls should be married early to ease familys financial burden.				
$\hat{\beta}_j$	2.098***	(0.0700)	1.773***	(0.0610)
$\hat{\alpha}_{j1}$	-1.537***	(0.0501)	-1.201***	(0.0466)
$\hat{\alpha}_{j2}$	-0.642***	(0.0336)	-0.249***	(0.0328)
$\hat{\alpha}_{j3}$	0.413***	(0.0330)	0.979***	(0.0428)
Instead of spending money on a girls education, it should be saved for her dowry.				
$\hat{\beta}_j$	1.804***	(0.0613)	1.445***	(0.0516)
$\hat{\alpha}_{j1}$	-2.020***	(0.0642)	-1.855***	(0.0655)
$\hat{\alpha}_{j2}$	-1.107***	(0.0417)	-0.735***	(0.0399)

$\hat{\alpha}_{j3}$	0.161***	(0.0324)	0.812***	(0.0430)

If a girl is a victim of some sexual abuse, it is the fault of the girl.

$\hat{\beta}_j$	1.716***	(0.0589)	1.530***	(0.0538)
$\hat{\alpha}_{j1}$	-1.925***	(0.0627)	-1.473***	(0.0546)
$\hat{\alpha}_{j2}$	-1.062***	(0.0416)	-0.457***	(0.0357)
$\hat{\alpha}_{j3}$	0.0611*	(0.0324)	0.742***	(0.0411)

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

$\hat{\beta}_j$	1.335***	(0.0457)	1.410***	(0.0499)
$\hat{\alpha}_{j1}$	-1.318***	(0.0521)	-0.722***	(0.0414)
$\hat{\alpha}_{j2}$	0.117***	(0.0349)	0.875***	(0.0428)
$\hat{\alpha}_{j3}$	1.239***	(0.0513)	2.087***	(0.0733)

There are times when a woman deserves to be beaten

$\hat{\beta}_j$	1.412***	(0.0479)	1.400***	(0.0491)
$\hat{\alpha}_{j1}$	-2.076***	(0.0702)	-1.638***	(0.0602)
$\hat{\alpha}_{j2}$	-0.613***	(0.0374)	-0.0479	(0.0348)
$\hat{\alpha}_{j3}$	0.743***	(0.0414)	1.200***	(0.0510)

Girls who are highly educated indulge in improper behaviour

$\hat{\beta}_j$	1.332***	(0.0463)	1.249***	(0.0456)
$\hat{\alpha}_{j1}$	-2.149***	(0.0740)	-1.929***	(0.0714)
$\hat{\alpha}_{j2}$	-0.844***	(0.0415)	-0.472***	(0.0387)
$\hat{\alpha}_{j3}$	0.516***	(0.0391)	1.017***	(0.0498)
N	5043		4805	

Notes: Table presents estimated parameters and standard errors (in parentheses) for IRT measurement model of girls' and carers' internalized progressive gender norms. Measurement model and estimation procedure described in Section 3.3.4. Column 1 presents the measurement model for girls' norms while 2 does the same for carers. Items are coded such that a higher value of the factor represents more progressive gender norms.

Table A.6: Paramters of IRT measurement model for Violent Sanctions scales

	(1) Girls		(2) Carers	
If Rama arrived home later than agreed her parents would hit, slapped, beat or kicked her				
$\hat{\beta}_j$	1.127***	(0.0441)	1.070***	(0.0432)
$\hat{\alpha}_{j1}$	-1.281***	(0.0575)	-1.597***	(0.0684)
$\hat{\alpha}_{j2}$	0.429***	(0.0406)	0.403***	(0.0424)
$\hat{\alpha}_{j3}$	2.080***	(0.0821)	2.256***	(0.0918)
If Rama walked home from school alone with a male friend, he would act inappropriately towards her e.g. try to kiss her				
$\hat{\beta}_j$	1.675***	(0.0601)	1.677***	(0.0607)
$\hat{\alpha}_{j1}$	-1.101***	(0.0458)	-1.357***	(0.0519)
$\hat{\alpha}_{j2}$	-0.0421	(0.0329)	-0.167***	(0.0338)
$\hat{\alpha}_{j3}$	1.383***	(0.0521)	1.508***	(0.0560)
If Rama arrives home to her husband after meeting her sister, her husband would demand to check her phone to see who she had been talking to.				
$\hat{\beta}_j$	1.335***	(0.0492)	1.209***	(0.0462)
$\hat{\alpha}_{j1}$	-1.729***	(0.0641)	-2.164***	(0.0806)
$\hat{\alpha}_{j2}$	-0.402***	(0.0372)	-0.526***	(0.0407)
$\hat{\alpha}_{j3}$	1.285***	(0.0543)	1.356***	(0.0596)
If Rama walked to the market alone a stranger would hit her and steal her money.				
$\hat{\beta}_j$	2.258***	(0.0846)	2.326***	(0.0882)
$\hat{\alpha}_{j1}$	-1.224***	(0.0451)	-1.399***	(0.0490)
$\hat{\alpha}_{j2}$	-0.290***	(0.0316)	-0.385***	(0.0325)
$\hat{\alpha}_{j3}$	0.884***	(0.0392)	0.954***	(0.0413)
If Rama walked back from school alone a boy from a neighbouring village would follow her and call her names.				
$\hat{\beta}_j$	2.469***	(0.0955)	2.648***	(0.109)
$\hat{\alpha}_{j1}$	-1.144***	(0.0429)	-1.259***	(0.0451)
$\hat{\alpha}_{j2}$	-0.227***	(0.0307)	-0.276***	(0.0311)
$\hat{\alpha}_{j3}$	0.897***	(0.0389)	0.927***	(0.0402)
N	5043		4805	

Notes: Table presents estimated parameters and standard errors (in parentheses) for IRT measurement model of girls' and carers' perceived likelihood of violent sanctions when girls and young women break gender norms. Measurement model and estimation procedure described in Section 3.3.4. Column 1 presents the measurement model for girls' perceptions while 2 does the same for carers. Items are coded such that a higher value of the factor represents a higher percieved likelihood of violence.

Table A.7: Does the Reduction in Education Dropouts Mediate Delayed Marriage?

	(1)	(2)
	Married/Engaged/Fixed	Married/Engaged/Fixed
Girl Groups	-0.053***	-0.040**
	(0.020)	(0.019)
	[p=0.008]	[p=0.040]
Girl Groups + Community	-0.056**	-0.045*
	(0.025)	(0.024)
	[p=0.023]	[p=0.058]
In formal education		-0.224***
		(0.021)
		[p<0.001]
N	2605	2605
P-value: diff.in Girl Groups		[p=0.010]
P-value: diff. in Girl Groups+Community		[p=0.028]

Notes: Mediation analysis for older girls (age 15–17 at baseline). Column (1) is the OLS estimate of the treatment of the two treatment arms on a girl being either married, or engaged or having her marriage fixed in the older group. It is identical to the estimate shown in Table A.9. Column (2) displays the same regression but adds our current educational status as a "control" variable. In particular

Table A.9: Education and Marriage (OLS)

	Older Girls			Younger Girls			All		
	In Education	Married	Married or Engaged or Fixed	In Education	Married	Married or Engaged or Fixed	In Education	Married	Married or Engaged or Fixed
Girl Groups	0.061***	-0.037**	-0.053***	0.015	0.007	0.017	0.037**	-0.013	-0.015
	(0.023)	(0.018)	(0.020)	(0.019)	(0.011)	(0.014)	(0.017)	(0.011)	(0.014)
	[p=0.009]	[p=0.037]	[p=0.008]	[p=0.415]	[p=0.515]	[p=0.207]	[p=0.030]	[p=0.240]	[p=0.256]
Girl Groups + Community	0.048**	-0.024	-0.056**	0.013	0.011	0.024	0.030*	-0.005	-0.013
	(0.021)	(0.020)	(0.025)	(0.021)	(0.011)	(0.016)	(0.018)	(0.012)	(0.017)
	[p=0.025]	[p=0.221]	[p=0.023]	[p=0.546]	[p=0.326]	[p=0.141]	[p=0.091]	[p=0.670]	[p=0.436]
Difference	-0.012	0.013	-0.003	-0.003	0.004	0.006	-0.008	0.008	0.002
	(0.023)	(0.020)	(0.025)	(0.021)	(0.011)	(0.016)	(0.018)	(0.013)	(0.018)
	[p=0.589]	[p=0.520]	[p=0.917]	[p=0.897]	[p=0.732]	[p=0.700]	[p=0.673]	[p=0.547]	[p=0.894]
N	2,605	2,605	2,605	2,920	2,920	2,920	5,525	5,525	5,525
Control Mean	0.406	0.181	0.304	0.765	0.056	0.093	0.592	0.116	0.195
Heterogeneity by Age									
Girl Groups	[p=0.062]	[p=0.015]	[p=0.001]						
Girl Groups + Community	[p=0.132]	[p=0.089]	[p=0.001]						
Difference	[p=0.691]	[p=0.624]	[p=0.693]						

Notes: The table presents estimated impacts of "Girl Groups" and "Girl Groups and Community Campaigns" interventions relative to the pure control group. "Difference" refers to differences between the two intervention models. Standard errors (in parentheses) and two-sided *p*-values constructed using a cluster *t*-bootstrap accounting for stratification by region and treatment status (1,000 iterations). Pre-specified controls are: a full set of interactions between dummy variables for age in years and a dummy variable for whether the girl was in school at baseline, caste, baseline wealth index, mother's years of education, district, whether the girl's family was talking about marriage at baseline, whether the girl's marriage was fixed at baseline, the girl's intended age of marriage at baseline, and the girl's mother's gender attitudes. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.8: Treatment effects on school enrollment and post-secondary education

	Older Girls		Younger Girls		All	
	In School	Post-Secondary Education	In School	Post-Secondary Education	In School	Post-Secondary Education
Girl Groups	0.065*** (0.023) [p=0.004]	-0.009 (0.016) [p=0.585]	0.017 (0.018) [p=0.361]	0.007 (0.031) [p=0.815]	0.041*** (0.016) [p=0.009]	-0.005 (0.013) [p=0.675]
Girl Groups + Community	0.070*** (0.021) [p=0.001]	-0.025* (0.013) [p=0.064]	0.012 (0.02) [p=0.565]	0.011 (0.027) [p=0.677]	0.040** (0.017) [p=0.017]	-0.016 (0.012) [p=0.158]
Difference	0.005 (0.024) [p=0.834]	-0.016 (0.014) [p=0.246]	-0.005 (0.02) [p=0.810]	0.004 (0.033) [p=0.904]	-0.001 (0.018) [p=0.961]	-0.011 (0.012) [p=0.378]
N	2605	2605	2920	2920	5525	5525
Control Mean	0.761	0.089	0.316	0.004	0.547	0.045
Heterogeneity by Age						
Girl Groups	[p=0.064]	[p=0.620]				
Girl Groups + Community	[p=0.017]	[p=0.192]				
Difference	[p=0.720]	[p=0.538]				

Notes: The table presents estimated average marginal effects of "Girl Groups" and "Girl Groups + Community" interventions relative to the pure control group. "Difference" refers to differences between the two intervention models. Average marginal effects are estimated from a logit model. Standard errors (in parentheses) and two-sided *p*-values are constructed using a cluster *t*-bootstrap accounting for stratification by region and treatment status (1,000 iterations). Pre-specified controls are: a full set of interactions between dummy variables for age in years and a dummy variable for whether the girl was in school at baseline, caste, baseline wealth index, mother's years of education, district, whether the girl's family was talking about marriage at baseline, whether the girl's marriage was fixed at baseline, the girl's intended age of marriage at baseline, and the girl's mother's gender attitudes. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.10: Education - Main Sample

	Older Girls Main Sample (Only Girls)	*Younger Girls* Main Sample (Only Girls)	*All* Main Sample (Only Girls)
Girl Groups	0.062*** (0.024) [p=0.010]	0.022 (0.019) [p=0.253]	0.042** (0.018) [p=0.018]
Girl Groups + Community	0.048** (0.022) [p=0.028]	0.014 (0.022) [p=0.526]	0.030 (0.019) [p=0.102]
Difference	-0.013 (0.024) [p=0.581]	-0.008 (0.021) [p=0.692]	-0.012 (0.019) [p=0.534]
N	2,318	2,725	5,043
Control Mean	0.431	0.782	0.617
Heterogeneity by Age			
Girl Groups	[p=0.112]		
Girl Groups + Community	[p=0.140]		
Difference	[p=0.851]		

Notes: The table presents estimated average marginal effects of "Girl Groups" and "Girl Groups + Community" interventions relative to the pure control group. "Difference" refers to differences between the two intervention models. Standard errors (in parentheses) and two-sided p-values constructed using a cluster t-bootstrap accounting for stratification by region and treatment status (1,000 iterations). Main Sample refers to the sample that includes only the marriage data reported from the girls themselves. Pre-specified controls are: a full set of interactions between dummy variables for age in years and a dummy variable for whether the girl was in school at baseline, caste, baseline wealth index, mother's years of education, district, whether the girl's family was talking about marriage at baseline, whether the girl's marriage was fixed at baseline, the girl's intended age of marriage at baseline, and the girl's mother's gender attitudes. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.11: Marriage - Main Sample and Extended Sample + Tracking Data

	Older Girls		Younger Girls		All	
	Main Sample (Only Girls)	Extended Sample +Tracking data	Main Sample (Only Girls)	Extended Sample +Tracking data	Main Sample (Only Girls)	Extended Sample +Tracking data
Girl Groups	-0.032**	-0.037**	0.015	0.004	-0.008	-0.017
	(0.016)	(0.017)	(0.010)	(0.011)	(0.010)	(0.011)
	[p=0.045]	[p=0.026]	[p=0.133]	[p=0.732]	[p=0.439]	[p=0.117]
Girl Groups + Community	-0.016	-0.025	0.016*	0.009	0.000	-0.008
	(0.019)	(0.019)	(0.009)	(0.011)	(0.012)	(0.012)
	[p=0.413]	[p=0.195]	[p=0.088]	[p=0.424]	[p=0.984]	[p=0.507]
Difference	0.017	0.012	0.001	0.005	0.008	0.009
	(0.018)	(0.019)	(0.009)	(0.010)	(0.011)	(0.012)
	[p=0.364]	[p=0.523]	[p=0.923]	[p=0.623]	[p=0.482]	[p=0.478]
N	2,318	2,679	2,725	3,010	5,043	5,689
Control Mean	0.114	0.182	0.030	0.057	0.069	0.117
Heterogeneity by Age						
Girl Groups	[p=0.005]	[p=0.020]				
Girl Groups + Community	[p=0.086]	[p=0.099]				
Difference	[p=0.360]	[p=0.684]				

Notes: The table presents estimated average marginal effects of "Girl Groups" and "Girl Groups + Community" interventions relative to the pure control group. "Difference" refers to differences between the two intervention models. Standard errors (in parentheses) and two-sided *p*-values constructed using a cluster *t*-bootstrap accounting for stratification by region and treatment status (1,000 iterations). Main Sample refers to the sample that includes only the marriage data reported from the girls themselves. Extended Sample includes the marriage data collected from the girls and the caregiver, as well as the data collected during the tracking phase (just before endline data collection). Pre-specified controls are: a full set of interactions between dummy variables for age in years and a dummy variable for whether the girl was in school at baseline, caste, baseline wealth index, mother's years of education, district, whether the girl's family was talking about marriage at baseline, whether the girl's marriage was fixed at baseline, the girl's intended age of marriage at baseline, and the girl's mother's gender attitudes. * *p* < 0.1, ** *p* < 0.05, *** *p* < 0.01.

Table A.12: Married - excluding married but gauna not perfomed (GNP)

	Older Girls	Younger Girls	All
	Married (excludes GNP)	Married (excludes GNP)	Married (excludes GNP)
Girl Groups	-0.033** (0.016) [p=0.044]	-0.002 (0.009) [p=0.840]	-0.017* (0.009) [p=0.057]
Girl Groups + Community	-0.023 (0.019) [p=0.221]	0.011 (0.010) [p=0.288]	-0.005 (0.011) [p=0.661]
Difference	0.01 (0.018) [p=0.584]	0.013 (0.009) [p=0.171]	0.012 (0.011) [p=0.291]
N	2,605	2,920	5,525
Control Mean	0.171	0.045	0.106
Heterogeneity by Age			
Girl Groups	[p=0.084]		
Girl Groups + Community	[p=0.082]		
Difference	[p=0.887]		

Notes: The table presents estimated average marginal effects of "Girl Groups" and "Girl Groups + Community" interventions relative to the pure control group. "Difference" refers to differences between the two intervention models. Standard errors (in parentheses) and two-sided p-values constructed using a cluster t-bootstrap accounting for stratification by region and treatment status (1,000 iterations). Pre-specified controls are: a full set of interactions between dummy variables for age in years and a dummy variable for whether the girl was in school at baseline, caste, baseline wealth index, mother's years of education, district, whether the girl's family was talking about marriage at baseline, whether the girl's marriage was fixed at baseline, the girl's intended age of marriage at baseline, and the girl's mother's gender attitudes. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.13: Depression: Raw Scores and Indicators for Mild and Moderate Depression

	Older Girls			Younger Girls			All		
	Depression Raw Score (PHQ-9)	Mild Depression (PHQ-9)	Moderate Depression (PHQ-9)	Depression Raw Score (PHQ-9)	Mild Depression (PHQ-9)	Moderate Depression (PHQ-9)	Depression Raw Score (PHQ-9)	Mild Depression (PHQ-9)	Moderate Depression (PHQ-9)
Girl Groups	-0.023 (0.270) [p=0.933]	0.000 (0.028) [p=0.997]	-0.010 (0.014) [p=0.476]	0.162 (0.241) [p=0.502]	0.005 (0.026) [p=0.850]	0.007 (0.011) [p=0.507]	0.075 (0.229) [p=0.744]	0.003 (0.024) [p=0.899]	-0.001 (0.010) [p=0.924]
Girl Groups + Community	-0.764*** (0.232) [p=0.001]	-0.071*** (0.024) [p=0.002]	-0.025** (0.012) [p=0.034]	-0.282 (0.218) [p=0.195]	-0.038 (0.023) [p=0.104]	-0.006 (0.010) [p=0.553]	-0.508*** (0.195) [p=0.009]	-0.053*** (0.020) [p=0.008]	-0.015* (0.009) [p=0.084]
Difference	-0.741*** (0.225) [p=0.001]	-0.071*** (0.024) [p=0.003]	-0.015 (0.011) [p=0.194]	-0.444** (0.198) [p=0.025]	-0.043* (0.023) [p=0.060]	-0.013 (0.009) [p=0.170]	-0.583*** (0.173) [p=0.001]	-0.056*** (0.020) [p=0.004]	-0.014* (0.008) [p=0.071]
N	2,318	2,318	2,318	2,725	2,725	2,725	5,043	5,043	5,043
Control Mean	2.565	0.207	0.051	2.247	0.191	0.034	2.397	0.199	0.042
Heterogeneity by Age									
Girl Groups	[p=0.416]	[p=0.852]	[p=0.198]						
Girl Groups + Community	[p=0.030]	[p=0.165]	[p=0.128]						
Difference	[p=0.222]	[p=0.251]	[p=0.876]						

Notes: The table presents the estimated impacts of "Girl Groups" and "Girl Groups + Community" interventions relative to the pure control group. "Difference" refers to differences between the two intervention models. Standard errors (in parentheses) and two-sided *p*-values constructed using a cluster *t*-bootstrap accounting for stratification by region and treatment status (1,000 iterations). The table shows the estimated impacts on three alternative measures for the outcome of Depression. First, the raw scores obtained by aggregating the responses to the items of the PHQ-9 scale. Second, a dummy variable that takes the value of 1 if the girl has mild depression (i.e., if the raw scores are ≥5) (Kroenke et al. 2001; Ganguly et al. 2013). Third, a dummy variable that takes the value of 1 if the girl has moderate depression (i.e., if the raw scores are above ≥10) (Kroenke et al. 2001; Ganguly et al. 2013). Pre-specified controls are: a full set of interactions between dummy variables for age in years and a dummy variable for whether the girl was in school at baseline, caste, baseline wealth index, mother's years of education, district, and mental health at baseline. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A.14: Anxiety: Raw Scores and Indicators for Mild and Moderate Anxiety

	Older Girls			Younger Girls			All		
	Anxiety Raw Score (GAD-7)	Mild Anxiety (GAD-7)	Moderate Anxiety (GAD-7)	Anxiety Raw Score (GAD-7)	Mild Anxiety (GAD-7)	Moderate Anxiety (GAD-7)	Anxiety Raw Score (GAD-7)	Mild Anxiety (GAD-7)	Moderate Anxiety (GAD-7)
Girl Groups	-0.159 (0.212) [p=0.453]	-0.031 (0.026) [p=0.231]	0.001 (0.012) [p=0.932]	0.047 (0.196) [p=0.812]	0.004 (0.024) [p=0.851]	0.006 (0.007) [p=0.408]	-0.049 (0.179) [p=0.784]	-0.012 (0.021) [p=0.590]	0.004 (0.007) [p=0.569]
Girl Groups + Community	-0.790*** (0.197) [p=0.000]	-0.078*** (0.026) [p=0.002]	-0.010 (0.009) [p=0.266]	-0.261 (0.183) [p=0.153]	-0.028 (0.021) [p=0.175]	0.009 (0.007) [p=0.206]	-0.509*** (0.162) [p=0.002]	-0.051*** (0.019) [p=0.006]	0.000 (0.005) [p=0.995]
Difference	-0.631*** (0.176) [p=0.000]	-0.047** (0.021) [p=0.027]	-0.011 (0.011) [p=0.315]	-0.308* (0.182) [p=0.092]	-0.032 (0.020) [p=0.107]	0.003 (0.008) [p=0.734]	-0.460*** (0.145) [p=0.002]	-0.040** (0.016) [p=0.014]	-0.004 (0.007) [p=0.544]
N	2,318	2,318	2,318	2,725	2,725	2,725	5,043	5,043	5,043
Control Mean	2.599	0.232	0.039	2.176	0.171	0.026	2.375	0.200	0.032
Heterogeneity by Age									
Girl Groups	[p=0.296]	[p=0.160]	[p=0.706]						
Girl Groups + Community	[p=0.009]	[p=0.068]	[p=0.114]						
Difference	[p=0.129]	[p=0.574]	[p=0.319]						

Online Appendix - p20

Notes: The table presents the estimated impacts of "Girl Groups" and "Girl Groups + Community" interventions relative to the pure control group. "Difference" refers to differences between the two intervention models. Standard errors (in parentheses) and two-sided p-values constructed using a cluster t-bootstrap accounting for stratification by region and treatment status (1,000 iterations). The table shows the estimated impacts on three alternative measures for the outcome of Anxiety. First, the raw scores obtained by aggregating the responses to the items of the GAD-7 scale. Second, a dummy variable that takes the value of 1 if the girl has mild anxiety (i.e., if the raw scores are ≥ 5) (Kroenke et al. 2006). Third, a dummy variable that takes the value of 1 if the girl has moderate anxiety (i.e., if the raw scores are ≥ 10) (Kroenke et al. 2006). Pre-specified controls are: a full set of interactions between dummy variables for age in years and a dummy variable for whether the girl was in school at baseline, caste, baseline wealth index, mother's years of education, district, and mental health at baseline. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

B Robustness Check: Post Double Selection Lasso

B.1 Methodology

As a robustness check, we use Post Double Selection (PDS) Lasso to estimate the impact of the intervention. This method was introduced by Belloni et al. (2014) and it allows us to systematically choose covariates in the presence of a large number of baseline variables. The method selects the control variables that minimize the sum of squared errors, setting the coefficients on some variables to be exactly zero and allowing to perform variable selection. PDS Lasso involves three steps. The first step consists of estimating a lasso regression with our treatment variable as a dependent variable and all the available baseline variables as regressors. In the second step, the same procedure is followed but on the dependent variable. In the third step, we run the treatment effects specification including the choice of controls selected in the first two steps (Belloni et al. 2014; Ahrens and Schaer 2018)[31].

Step 1: LASSO on the treatment variable[32]

$$T_j = \beta_0 + \beta_1 X_{ij,1} + \beta_2 X_{ij,2} + \ldots + \beta_p X_{ij,p} + \epsilon_j. \tag{B.1}$$

Step 2: LASSO on the dependent variable

$$y_{ij} = \beta_0 + \beta_1 X_{ij,1} + \beta_2 X_{ij,2} + \ldots + \beta_p X_{ij,p} + \epsilon_{ij}. \tag{B.2}$$

Step 3: OLS using selected controls from Step 1 and 2

$$y_{ij} = \alpha_0 + \alpha_{girl} T_j^{girl} + \alpha_{girl+comm} T_j^{girl+comm} + \gamma W_{ij} + \epsilon_{ij}, \tag{B.3}$$

where W_{ij} is the union of the selected controls from step 1 and 2.

Before estimating equations (B.1), (B.2), and (B.3), we gather all potential control variables from the baseline data, removing those that were only applicable for married girls as we did not collect data of these girls for the endline, as well as variables that were applicable only to a small subset of the sample (e.g. reasons for saving money for girls who said they had a bank account). Then:

- we generated a set of indicators (e.g. dummy variables or numeric variables) for all categorical variables;

- we added the squared of each numeric variable and two-way interactions between the age variable (dummies for age) and the other available variables;

---

[31]We use the Stata command *pdslasso* to implement this procedure. To select the optimal penalty level (λ) the command uses the "rigorous' theory driven penalization approach that is estimated with the companion Stata package *rlasso* (Ahrens et al. 2018; Ahrens and Schaer 2018).

[32]Considering that we have two treatments, the PDS Lasso estimates equation (B.1) for each treatment arm.

- we imputed the missing covariate value with the average (mean for continuous controls and the median for discrete controls) of the non-missing observations and created dummy variables equal to one for imputed observations;

- we dropped one variable from any pair of perfectly collinear variables;

- we standardized all the variables.

B.2 Results

We include all potential control variables and estimate Steps 1–3. Tables B.1, B.2 and B.3 show the results of the PDS Lasso estimation. The lasso selects some control variables that we pre-specified in the pre-analysis plan, including school attendance, wealth index, elders talking about marriage, intended age of marriage (reported by the carer), and mothers years of education. In addition to this, lasso selects other variables that are good predictors of the dependent variables. Overall, the treatment effects from PDS Lasso are similar in sign and magnitude to the treatment effects obtained in our main specification (see Section 3.4).

Table B.1: Education and Marriage

	Older Girls			Younger Girls			All		
	In Education	Married	Married or Engaged or Fixed	In Education	Married	Married or Engaged or Fixed	In Education	Married	Married or Engaged or Fixed
Girl Groups	0.048* (0.025) [p=0.058]	-0.032* (0.018) [p=0.081]	-0.046** (0.020) [p=0.024]	0.009 (0.016) [p=0.582]	0.008 (0.010) [p=0.434]	0.020 (0.015) [p=0.192]	0.030* (0.016) [p=0.057]	-0.0013 (0.012) [p=0.254]	-0.010 (0.015) [p=0.496]
Girl Groups + Community	0.038* (0.023) [p=0.093]	-0.021 (0.020) [p=0.291]	-0.056** (0.025) [p=0.023]	0.010 (0.020) [p=0.605]	.009 (0.011) [p=0.417]	0.026 (0.016) [p=0.110]	0.024 (0.017) [p=0.155]	-0.006 (0.012) [p=0.591]	-0.012 (0.017) [p=0.454]
Difference	-0.01 [p=0.678]	0.011 [p=0.596]	-0.01 [p=0.687]	0.001 [p=0.955]	0.001 [p=0.929]	0.006 [p=0.724]	-0.006 [p=0.76]	0.007 [p=0.599]	-0.003 [p=0.886]
N	2,605	2,605	2,605	2,920	2,920	2,920	5,525	5,525	5,525
Control Mean	0.406	0.181	0.304	0.765	0.056	0.093	0.592	0.116	0.195
Heterogeneity by Age									
Girl Groups	[p=0.197]	[p=0.057]	[p=0.010]						
Girl Groups + Community	[p=0.354]	[p=0.187]	[p=0.006]						

Notes: We used the Stata command pdslasso by (Ahrens and Schaer 2018). We enter the command pdslasso "outcome variable" "treatment variable" (control variables) lopt(prestd) cluster(cluster) noisily, where cluster indicates that the parameters of the lasso have to be computed using clustered standard errors, and lopt is used to standardize the covariates. For the marriage outcome, PDS Lasso selected some control variables that we pre-specified in the pre-analysis plan: school attendance, wealth index, intended age of marriage (reported by the carer), dummy variable for whether the girls family was talking about marriage at baseline, and mother's years of education. In addition to those, PDS Lasso selected other variables including desired age for marriage and dummy variables for age interacted with other variables. For the education outcome, PDS Lasso selected some control variables that we pre-specified in the pre-analysis plan: school attendance, carer's years of education, wealth index, dummy variable for whether the girls family was talking about marriage at baseline,and intended age of marriage (reported by the carer). In addition to those, PDS Lasso selected other variables including RAVENs test score, hours studying, desired age for marriage, and and dummy variables for age interacted with other variables.

Table B.2: Mental Health

Post Double Lasso

	Older Girls	Younger Girls	All
	Factor Mental Health	Factor Mental Health	Factor Mental Health
Girl Groups	0.007	-0.040	-0.017
	(0.079)	(0.079)	(0.073)
	[p=0.926]	[p=0.610]	[p=0.816]
Girl Groups + Community	0.365***	0.201***	0.277***
	(0.079)	(0.076)	(0.070)
	[p=0.000]	[p=0.008]	[p=0.000]
Difference	0.358***	0.241***	0.294***
	[p=0.000]	[p=0.000]	[p=0.000]
N	2,318	2,725	5,043
Control Mean	-0.063	0.047	-0.005
Heterogeneity by Age			
Girl Groups	[p=0.670]		
Girl Groups + Community	[p=0.133]		

Notes: We used the Stata command pdslasso by (Ahrens and Schaer 2018). We enter the command pdslasso "outcome variable" "treatment variable" (control variables) lopt(prestd) cluster(cluster) noisily, where cluster indicates that the parameters of the lasso have to be computed using clustered standard errors, and lopt is used to standardize the covariates. For the mental health outcome, PDS Lasso did not select any control variable.

Table B.3: Depression and Anxiety

Post Double Lasso

	Older Girls		Younger Girls		All	
	Depression (PHQ-9)	Anxiety (GAD-7)	Depression (PHQ-9)	Anxiety (GAD-7)	Depression (PHQ-9)	Anxiety (GAD-7)
Girl Groups	0.019 (0.075) [p=0.799]	-0.056 (0.064) [p=0.378]	0.059 (0.073) [p=0.425]	-0.007 (0.062) [p=0.912]	0.039 (0.068) [p=0.564]	-0.031 (0.057) [p=0.587]
Girl Groups + Community	-0.334*** (0.073) [p=0.000]	-0.391*** (0.066) [p=0.000]	-0.186*** (0.070) [p=0.008]	-0.268*** (0.062) [p=0.000]	-0.254*** (0.064) [p=0.000]	-0.326*** (0.057) [p=0.000]
Difference	-0.353*** [p=0.000]	-0.335*** [p=0.000]	-0.244*** [p=0.000]	-0.261*** [p=0.000]	-0.293*** [p=0.000]	-0.295*** [p=0.000]
N	2,318	2,318	2,725	2,725	5,043	5,043
Control Mean	0.048	0.059	-0.041	-0.052	0.001	0.000
Heterogeneity by Age						
Girl Groups	[p=0.706]	[p=0.579]				
Girl Groups + Community	[p=0.143]	[p=0.174]				

Notes: We used the Stata command pdslasso by (Ahrens and Schaer 2018). We enter the command pdslasso "outcome variable" "treatment variable" (control variables) lopt(prestd) cluster(cluster) noisily, where cluster indicates that the parameters of the lasso have to be computed using clustered standard errors, and lopt is used to standardize the covariates. For the mental health outcomes, PDS Lasso did not select any control variable.

C   Appendix to Model

In this Appendix we provide more detail on the implications of our model. Recall that we have that decisions regarding schooling S and behavior y are made according to:

$$\max_{S,y} \quad \underbrace{u_S S + u_y y}_{\substack{\text{Girls' immediate} \\ \text{enjoyment/dislike} \\ \text{of } S \text{ and } y}} + \underbrace{vS}_{\substack{\text{Future benefits} \\ \text{from schooling}}} - \underbrace{\max(0, y - y^*)P}_{\substack{\text{Punishment for Deviation} \\ \text{from norm}}} \quad s.t. \quad \underbrace{S \leq \mathbf{1}(y > \bar{y})}_{\substack{\text{School only possible} \\ \text{if } y > \bar{y}}} \quad \text{(C.1)}$$

while mental health takes the form:

$$M(S, y, \bar{y}, y^*, P) = u_S S + u_y y - \max(0, y - y^*)P \quad \text{(C.2)}$$

We make the following assumptions:

- $v > 0$. This is motivated by the fact that past evidence, including in this sample, suggests that girls and parents perceive strong future returns (financial and in the marriage market) to education (Andrew and Adams-Prassl 2021; Jensen 2012). This suggests that their motivation for pursuing school is greater than simply the mental wellbeing they derive from it in the here and now. This assumption implies that the marginal girl who moves into school due to a change in community-level norms y^* will do so at an immediate cost to her own mental health.

- $u_y \leq 0$. This assumption states girls have internalized gendered norms of behavior sufficiently such that they dislike y other than, for instance, the instrumental value in terms of allowing them to pursue schooling. This assumption implies that if girls do not attend school, they will always choose $y = 0$ while if they do attend school they will always choose $y = \bar{y}$.

Given these assumptions, choice of optimal y which we denote y^{**}, and the optimal schooling choice, S^{**} will follow the decision rule:

$$y^{**}(y^*) = \begin{cases} \bar{y} & \text{if } u_S + u_y \bar{y} + v - \max(0, \bar{y} - y^*)P \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{(C.3)}$$

$$S^{**}(y^*) = \begin{cases} 1 & \text{if } u_S + u_y \bar{y} + v - \max(0, \bar{y} - y^*)P \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{(C.4)}$$

This implies that mental health follows:

$$M^{**}(y^*) = \begin{cases} u_S + u_y - \max(0, \bar{y} - y^*)P & \text{if } u_S + u_y \bar{y} + v - \max(0, \bar{y} - y^*)P \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad \text{(C.5)}$$

In words, mental wellbeing is constant below the norms threshold at which the girl attends school. At the threshold value of y^* wellbeing will fall by v. This is because at this threshold value $\tilde{y}^*$ we have that $u_S + u_y \bar{y} + v - \max(0, \bar{y} - \tilde{y}^*)P = 0$ which gives that $M^{**}(\tilde{y}^*) = u_S + u_y \bar{y} - \max(0, \bar{y} - \tilde{y}^*)P = -v$. Intuitively, girls are induced to enter school due to a marginal change in y^* at the overall net present value of schooling makes them indifferent between pursuing schooling or not despite the fact that their immediate wellbeing is made worse off from pursuing schooling.

As norms get more progressive beyond $\tilde{y}^*$, immediate wellbeing improves at a rate of P up until the point $(y^* = \bar{y})$ that norms no longer conflict with the behaviors that are practically required to attend school. After this point, wellbeing remains constant in community norms. These are the patterns that are plotted in Figure 4(a).

The above analysis considers how the choices and mental wellbeing of one particular girl (facing a particular constraint $\bar{y}$) might be affected by changes in community-level gendered norms. In reality, different girls are likely to face different constraints driven by, for example, the fact that different girls face very different commutes to school. Averaging over girls who face different levels of $\bar{y}$ gives a monotonically positive relationship between y^* and average school enrollment. It gives a convex relationship, with the possibility of a U-shaped relationship, between community norms y^* and average mental wellbeing. This is shown in Figure 4(b).

Our model has very similar implications for how changes in expected punishments P will affect girls' schooling and wellbeing. Averaging across girls with different external constraints suggests that an increase in the severity of such punishments will decrease school attendance. It will form a convex relationship with mental wellbeing. When punishments are less severe and most girls are in school then increases in the severity of punishments will worsen average wellbeing. However, they will do so at a decreasing rate as eventually the punishments get so severe that they lead some girls to drop out from school which, our model suggests, comes with a short-term improvements wellbeing even if it has large long-term costs. In Figure 5, we see that the patterns we observe in our control group fit these patterns well.