Institute for Fiscal Studies

Alison Andrew
Sonya Krutikova
Gabriela Smarrelli
Hemlata Verma

Working paper

Empowering adolescent girls: does it take a village?

22/41

Economic and Social Research Council

Empowering Adolescent Girls: Does it take a Village?*†

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

May 2026

Abstract

The social environment is key to sustaining gender inequalities but many policies and programs target only women and do not involve the wider community. Can such approaches work or, by pushing women to break accepted norms, do they expose women to stress and backlash? What are the impacts of engaging the wider community? We use a 3-armed RCT covering 5000 adolescent girls across 125 communities in rural Rajasthan to explore these questions. We assess the impacts of weekly Girl Groups that worked only with adolescent girls and the impacts of additionally engaging the wider community. Both models led to a reduction in school dropout and early marriage. However, targeting adolescent girls without involving the broader community led to an increase in girls adopting a ruminative thinking style and no improvements in depression and anxiety. By contrast, when the wider community was engaged, girls' symptoms of depression and anxiety fell by 0.16 SD and 0.17 SD respectively and there were no negative impacts on rumination. We show evidence that such improvements in mental health may have resulted from the community engagement changing prevailing attitudes and internalized norms.

*We thank and acknowledge Pranita Achyut, Madhumita Das, Ravi Verma, Abhishek Gautam, Ronak Soni, and Sanjay Kumar for their fundamental role in intervention and study design. We thank all the peer mentors, field facilitators, Pankh sakhis and other members of Pradan and ICRW based in Dholpur whose effort and dedication made the Pankh program possible. We thank Abi Adams-Prassl, Oriana Bandiera, James Banks, Nathan Barker, Natalie Bau, Rossella Calvi, Rachel Cassidy, Jonathan de Quidt, Rachel Griffith, Johannes Haushofer, Seema Jayachandran, Imran Rasul, Adam Salisbury, Abhijeet Singh, Micaela Sviatschi, Michela Tincani, and Marcos Vera-Hernandez for helpful comments and feedback.

†This research was funded by the Children's Investment Fund Foundation, MacArthur Foundation and International Inspiration. Additional research support was provided from the Leverhulme Trust (RPG-2021-391), European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement no. 695300-HKADeC-ERC-2015-AdG) and the ESRC Centre for Microeconomic Analysis of Public Policy at the Institute for Fiscal Studies. Ethical clearance for the project was granted by the Sigma IRB Committee (New Delhi), ICRW IRB (Washington DC), and Oxford University IRB (Oxford, UK). We discuss ethics in more detail in the ethical appendix (Appendix A). The trial was registered with the AEA registry (AEARCTR-0002656).

‡University of Oxford & Institute for Fiscal Studies. Email: alison.andrew@economics.ox.ac.uk

§University of Manchester & Institute for Fiscal Studies. Email: sonya.krutikova@manchester.ac.uk

¶Center for Global Development. Email: gsmarrelli@cgdev.org

‖International Centre for Research on Women. Email: hverma@icrw.org

1 Introduction

The social environment – including the attitudes and beliefs of others, norms, and the enforcement of norms – plays a key role in sustaining gender inequalities (Fernández 2013; Jayachandran 2019; Field et al. 2021; Bursztyn et al. 2023). Despite this, many policies and programs do not engage directly with the wider community, instead targeting only women and often the most marginalized groups of women. Can interventions that solely work with marginalized groups achieve meaningful change? In contexts with strong community gender norms, are there hidden costs of such change? What are the impacts on women of engaging the wider community in a process of collective change?

We examine these questions in the context of empowering adolescent girls in rural Rajasthan, India. We compare and contrast the impacts of an intervention that works solely with adolescent girls with one that additionally engages with the broader community. This is a setting where there is great need for effective interventions. India in general, and Rajasthan in particular, is characterized by large gender inequalities in education, health, paid work, unpaid work, poverty, agency and political representation (Dhar et al. 2022; Field et al. 2021; Kishor and Gupta 2004; Duflo 2012). Choices made during adolescence, particularly early dropout from school and early marriage, are key to creating these inequalities (Duflo 2012; Jensen and Thornton 2003) and therefore interventions that impact these choices may have lifelong benefits. In light of this, there has been a big focus on interventions that engage with adolescent girls with the aim of encouraging them to persevere with schooling and delay marriage. Encouragingly, there is evidence that such interventions can be effective at achieving these goals (Edmonds et al. 2021; Buchmann et al. 2017).

However, much of what makes it challenging for girls to stay in school and unmarried in this context are the strict rules and restrictions that exist over many dimensions of girls' lives. For example, in our baseline data, the great majority of adolescent girls and their mothers agreed that girls should primarily stay at home, only venturing into public spaces when absolutely necessary and, ideally, only when accompanied. These rules are enforced through sanctions; our baseline data shows a widespread perception and acceptance that those who break the gendered norms will face repercussions, including violence, harassment, shaming, and disapproval. In such situations, interventions that are successful at changing adolescent girls' behaviors may risk bringing girls into conflict with widely shared norms and may unintentionally cause hardship for the girls through any costs incurred from going against these norms, including, for example, disapproval and sanctions from the broader community or psychological stress from behaviors or attitudes that diverge from internalized community norms.

This is a particular concern for programs working with adolescents. Adolescence is a highly sensitive period for the development of mental health, and mental health problems that develop at this age have remarkable persistence and adverse consequences for a broad range of outcomes (Patel et al. 2007; Currie and Stabile 2007). India is a context where the mental health of adolescents, who make up a fifth of its population, and particularly adolescent girls is a substantial concern (Patel et al. 2012). India's suicide rate among young women aged 15-24, at almost three times the global average, is the highest in the world. Whereas globally, suicide rates for young women are 40% lower than those for young men, in India, young women's rates are 50% higher (Figure B.1).[1] Adolescent girls in India also report substantially more symptoms of depression than boys (Figure B.2).[2] The risk is, therefore, that even if successful at increasing education and delaying marriage, interventions that place the burden of change on adolescent girls themselves might not allow girls to reap the mental health benefits one might expect from these gains. They might even end up harming mental health at this critical period.

In this high-stakes context, where the potential upside of interventions targeting adolescent girls is large but where adolescent girls themselves are a highly vulnerable group, we implemented a large-scale cluster randomized controlled trial in which we compare the impacts of an intervention that engages only with adolescent girls themselves to one that additionally facilitates engagement between the girls and others in their communities. The aim of both interventions was to empower girls to identify and pursue their life goals, with emphasis on the importance of remaining in education and delaying marriage. The first intervention, which we refer to as Girl Groups, consisted of group-based sessions led by young women from the same communities. In these sessions girls were encouraged to recognize constraining gendered norms and restrictions, understand their origin and arbitrariness, and consider other, more gender-equitable ideas of what constitutes appropriate behavior and success in life for women and girls. The intervention also contained sports sessions that gave girls a chance to occupy public space within the village. While members of the local community were aware of the intervention and many came to watch the sports tournaments that girls took part in, there was no structured dialogue or engagement between the girls and the community as part of the intervention.

[1] Data on crude suicide rates come from the World Health Organization's Global Health Observatory, Indicator 4664. India's crude suicide mortality rate (per 100,000) is 16.8 for women aged 15-24 and 11.07 for men. The corresponding global average rates (weighting by population size) are 5.8 for young women and 9.6 for young men. India is one of only three countries with a population greater than 5 million (the others being Haiti and the Republic of Korea) where the female youth suicide rate is greater than that for males. See Figure B.1 for more details.

[2] For instance, adolescent girls in Uttar Pradesh, which borders Rajasthan, reported experiencing each of the nine depressive symptoms asked about at a rate of between 25% and 150% higher than that of boys (Figure B.2). Analysis uses the PHQ-9 measure in the 2018-19 UDAYA Adolescent Survey in Uttar Pradesh. The raw score of the PHQ-9 measure of depression (standardized using the mean and standard deviation of the male sample) is 26% higher for girls than boys (CI: 21% - 30%).

Such engagement was the distinguishing feature of the second approach, which we refer to as the Girl Groups and Community Campaigns. This intervention *additionally* provided a platform and structure for the girls to reach out to and engage with their community. It did so by supporting the girls who were participating in the Girl Groups to organize and lead community events every two to three weeks. It was up to the girls to decide the exact content and message of these events, but they typically focused on issues affecting young women and involved girls advocating for more equitable treatment. The events also included time for discussion between community members, including community leaders who are likely to be the most influential in terms of creating and enforcing norms. Therefore, the events provided both an opportunity for girls to persuade the broader community of their perspective (Heller and Rao 2015; Finnemore and Sikkink 1998) and an opportunity for different members of the community to learn about the views of others (Bursztyn, González, and Yanagizawa-Drott 2020). We hypothesize that the events might relax some of the external constraints faced by girls, including the stringency of norms and sanctions, creating a more enabling environment for pursuing the goals and aspirations articulated in the Girl Groups.

We use a three-armed cluster randomized controlled trial covering over 5,000 adolescent girls across 125 villages, grouped into 90 clusters, in rural Rajasthan to evaluate and compare the effects of these two programs. We follow two 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 cohort and the content of the interventions varied slightly across cohorts.

We find that the stand-alone Girl Groups intervention led to a significant increase in the proportion of girls enrolled in and attending school or post-secondary education by 3.9 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 6.0 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 endline, relative to the control group (effect size 3.6 p.p.; $p = 0.035$). This effect is even larger (5.3 p.p.) once we expand the definition of "married" girls to include girls who were not yet married but for whom concrete marriage plans had been made (those who were engaged or whose marriage had been fixed by the time of the endline).

In spite of these changes, we see no improvement in the girls' psychological well-being from the stand-alone Girl Groups intervention, measured using three key dimensions of mental health: rumination, anxiety, and depression. Rumination is defined as repetitive and recurrent negative

thinking about oneself, one's situation, upsetting experiences, and concerns (Watkins 2008). It is a risk factor for the onset and persistence of depression and anxiety (Treynor et al. 2003). We see that the Girl Groups intervention led to no improvement in depression and anxiety, and a significant increase in rumination of 15% of a standard deviation of the control group (SD).

The addition of Community Campaigns in the second intervention - Girl Groups and Community Campaigns - did not change the effects of the Girl Groups on education and marriage; we see the same pattern of impacts on these outcomes in both programs. We cannot reject the hypothesis that the size of these effects on each outcome is the same in the two arms of the trial. However, there are striking differences in the effects of the two programs on psychological well-being. Whereas the Girl Groups program led to an increase in rumination, this effect was not observed when it was combined with the Community Campaigns. Further, the combined Girl Groups and Community Campaigns led to substantial and highly significant improvements in anxiety and depression, of 0.17 SD ($p = 0.002$) and 0.16 SD ($p = 0.011$) respectively. Both improvements are significantly different from the null effects in the Girl Groups arm ($p = 0.001$ in each case). Impacts are especially large for the older girls; in this group, we see an improvement of 0.23 SD in depression and 0.23 SD in anxiety relative to the control group. In light of Community Campaigns being a relatively low-intensity addition, these are substantial effects, especially for a population that was not restricted to girls who were depressed or anxious at baseline. As a comparison, the effect size we estimate among the older girls for depression is similar to what Angelucci and Bennett (2024) estimate for a combined livelihood and pharmacotherapy intervention in India for a population who were depressed at baseline.

We explore the potential mechanisms underlying these impacts. Girl Groups appear to be the critical driver of the effects on marriage and schooling, as we observe similar impacts on these outcomes regardless of whether the Girl Groups are accompanied by Community Campaigns. The effects on marriage likely operate through changes in girls' schooling decisions: while girls have little say over when they marry, they do have some influence over when they leave school, and marriage discussions typically begin only after they leave education.

The Girl Groups intervention was designed to empower girls to pursue their life goals and improve their well-being by weakening the influence of restrictive gender norms, strengthening aspirations and motivation to pursue education, and building non-cognitive skills, voice and agency. We find little evidence that changes in gender attitudes or non-cognitive skills explain the improvements in schooling. Instead, our analysis points to a potential role for motivation. At baseline, the most commonly reported reason for dropping out of school was "a lack of interest." By endline, we observe a reduction in the number of school days missed in both

treatment arms (measured six months after the program) among girls who remain enrolled. This pattern is notable in light of our heterogeneity results, which show that the schooling effects are concentrated among girls who had the highest risk of dropping out, as predicted by their baseline characteristics. As a result, the group of girls still in school at endline in the treatment arms likely includes more high-risk — and potentially less motivated — girls than the comparable group in the control.

What might explain the lack of mental health improvements in the Girl Groups only arm, despite gains in schooling and reductions in early marriage that align with girls' stated aspirations at baseline? Evidence from psychology suggests that large gaps between what can be realistically achieved and one's goals are an important trigger of rumination (Martin and Tesser 1996; Watkins and Nolen-Hoeksema 2014). The intervention may therefore have increased rumination by encouraging aspirations that remained difficult for girls to achieve, even in the presence of improvements in schooling and marriage outcomes. In addition, girls may have faced psychological stress or backlash from their communities as they began to change their behavior. Together these factors may explain why the intervention did not lead to improvements in depression and anxiety, despite its positive effects on schooling and early marriage.

Significant mental health improvements did, however, materialize following the addition of the Community Campaigns. There are two potential channels for this. First, performing and sharing ideas with their broader community might have been empowering and rewarding for the girls in and of itself. Second, community events might have led to shared gender beliefs or norms becoming more equitable. For example, the content of the campaigns could have persuaded community members to change their private views or the events could provide an opportunity for learning about others' private views.

In an ideal design, we would have measured the attitudes and beliefs of a broad range of community members, including men and community leaders, to directly assess how the community environment changed. However, we were not able to do this for reasons of cost and only have measures relating to girls and their mothers. Nevertheless, we provide two pieces of suggestive evidence that the improvements to girls' psychological well-being we see were likely driven by broader changes in the community rather than only the act of girls organizing and running community events. First, we show that the improvements in girls' well-being in communities where Community Campaigns took place are equally strong for girls who did not participate in the organization and implementation of the campaigns as for those who did. Leveraging the fact that in order to participate in running the Community Campaigns girls had to attend the Girl Group sessions, we analyze heterogeneity in the *additional* impact

of the Community Campaigns on girls' mental health by Girl Group attendance through a comparison of the two treatment arms (dropping the control). We do this under the assumption that the addition of the campaigns did not affect selection into attending the Girl Groups which is consistent with the fact we observe identical Girl Group attendance rates and identical correlations between attendance and baseline characteristics across the two arms.[3] Second, we find that the addition of Community Campaigns led to girls' reported gender attitudes, which are heavily linked to internalized norms, becoming more progressive (effect size=0.192 SD, p=0.001). These changes are evident among both girls who did and girls who did not directly participate in the Girl Groups and Community Campaigns. We take the progressive shift in attitudes, especially among girls who did not participate in running the campaigns, to be suggestive that the campaigns fostered wider norm and attitudinal change in the community.[4]

A natural question arising from our results is why the community-wide changes and improvements in girls' mental health in the Girl Groups and Community Campaigns arm did not lead to larger effects on education and marriage. Our analysis suggests that this may be partly due to differences in which groups of girls experienced these effects. While the impacts on schooling are concentrated among girls who were at high risk of school drop-out (based on their baseline characteristics), the mental health benefits of adding Community Campaigns are strongest among girls at the lowest risk of drop-out.[5] In addition, schooling decisions depend on a wide range of factors beyond psychological costs of attendance. As a result, even if Community Campaigns meaningfully reduced the psychological burden of staying in school, this may have been insufficient to shift behavior further. Taken together, these considerations help explain why the addition of Community Campaigns did not generate statistically significant additional effects on schooling and marriage outcomes. They also highlight that while Girl Groups primarily affect girls at highest risk of school dropout (and thus early marriage), Community Campaigns extend the benefits of the intervention to lower-risk girls, who were likely to have been infra-marginal on the schooling decision, through improvements in mental health.

The literature on approaches to reducing gender inequality in lower-income settings is dom-

[3]We argue that impacts on girls who did not participate in the Girl Groups are unlikely to be channeled through direct spillovers in girls' mental health between those who participated and those who did not since the impacts are *as large* for the girls who did not participate. Very little past literature finds any evidence of direct mental health spillovers. There is some evidence of spillovers between spouses (a much closer relationship than adolescent girls in a community). Even these, however, are far smaller in magnitude than would be needed to explain an equal-sized treatment effect on girls who participated in Community Campaigns and those who did not.

[4]To get a shift in girls' internalized norms among non-participant girls in the absence of any broader change at the community level, there would have to be very strong peer effects in internalized norms between participant and non-participant girls at the same time as having no spillovers to other groups. We consider this unlikely.

[5]We measure school drop-out risk by constructing a predicted dropout risk index using a lasso regression of endline school attendance on baseline covariates and their interactions. The model is estimated on the sub-sample of control group girls who were enrolled at baseline.

inated by efforts to achieve improvements by targeting women and girls themselves. Common intervention models include training schemes for women – such as life skills, technical, vocational and business skills (Bandiera et al. 2019; McKelway 2018; De Mel et al. 2014; Field et al. 2021; Edmonds et al. 2021) – support groups within the community – such as women's self-help and savings groups (Kumar et al. 2021) – and economic interventions such as asset transfer and business loan schemes for women (Armand et al. 2020; Hidrobo et al. 2016; Roy et al. 2019). There is considerable evidence that such programs can change women's attitudes, labor supply, and economic independence.

There is, however, a parallel literature showing that the safety and well-being of those who go against prevailing norms can be compromised through sanctions such as violence (Krishnan et al. 2010; Guarnieri and Rainer 2021; Bhalotra et al. 2021; Arenas-Arroyo et al. 2021) or higher risks of divorce (Bertrand et al. 2015). Beyond these tangible sanctions, a long-standing literature in psychology documents that exposure to the stigma and expectations of rejection associated with norm violation is itself a risk factor for depression and anxiety (Meyer 2003; Hatzenbuehler 2009). Recent work in economics is suggestive of such a dynamic in relation to violations of gender norms specifically (Ashraf et al. 2014; Mehmood et al. 2022). In the presence of these effects, well-intended programs that target women and girls and encourage them to act to improve their lives in spite of prevailing norms may inadvertently have negative effects on the well-being of this vulnerable group and expose them to retaliation or disapproval from their communities.

Our paper bridges these two literatures. It adds new evidence that significant behavior change can be achieved through interventions targeting information, beliefs, attitudes, preferences, and aspirations of adolescent girls themselves even in a very conservative setting. In this, our findings are in line with those in another recent study in rural Rajasthan, showing a significant reduction in school dropout and increased grade progression as the result of a life skills program for adolescent girls (Edmonds et al. 2021). Positive impacts of an empowerment program for adolescent girls in South Asia on educational outcomes have also been found in Bangladesh by Buchmann et al. (2017).[6] We additionally show important new evidence of positive impacts on age at marriage.

However, echoing the literature on sanctions, our paper also shows that without engagement with the wider context in which girls live, these improvements may come at important costs

[6]We also note a series of papers assessing the impact of combined livelihood training and provision of safe spaces for adolescent girls and young women across Sub-Saharan Africa (Adoho et al. 2014; Buehren et al. 2016; Bandiera et al. 2019; Bergstrom and Özler 2021). Both the bundled nature of these programs, including the provision of marketable skills, and a context of less stringent gender norms than ours make these papers distinct from this intervention.

to girls' psychological well-being. This aligns with emerging evidence that targeting boys and men as well as involving the wider community in addition to the girls might be more promising approaches than targeting girls alone (Boulhane et al. 2024; Amaral et al. 2024). Our findings offer a feasible, scalable way in which interventions targeting marginalized groups could be supplemented with a specific community engagement approach which shows promising impacts in our study context.[7]

Finally, this paper adds to a growing evidence base on improving mental health in LMICs. Poverty can exacerbate poor mental health (Ridley et al. 2020) and tackling poor mental health can have knock-on impacts on broader economic outcomes at the household and individual levels (see Lund et al. (2020) for a review of studies in LMICs and Patel et al. (2017) and Baranov et al. (2020) for evidence from South Asia specifically). The majority of adult mental health disorders start in adolescence and early intervention is more effective than at later stages suggesting that adolescence may be a particularly effective time for intervention (Patel et al. 2007; Currie and Stabile 2007). However, evidence on the types of interventions that can be implemented at scale in LMICs, especially for out-of-school adolescents, remains sparse (Kieling et al. 2011; Klasen and Crombag 2013; Mehra et al. 2022). Recent evaluations across several LMICs show that programs that aim to empower adolescent girls *can* be effective at improving their mental health, but this is not always the case and depends on the specific design of the program (Shah et al. 2024). This paper highlights the very substantial mental health benefits that forging a dialogue between adolescent girls and the communities they live in can have in contexts with highly restrictive gender norms. Our evidence suggests that the improvements in symptoms of depression and anxiety, particularly among the older girls, that stemmed from community engagement are comparable in magnitude to improvements resulting from far more intensive and often more targeted interventions, both in India (Angelucci and Bennett 2024) and elsewhere (Singla et al. 2017).

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, and a description of the two programs that we evaluate. We then present the study design, a description of the data, and our empirical strategy in Section 3 and our results in Section 4. Section 5 discusses potential mechanisms and Section 6 concludes.

[7]Some studies have responded to the disempowered role of adolescent girls by evaluating programs that target men alone (Barker et al. 2017), target young men and women together (Dhar et al. 2022; Edmonds et al. 2021) or by giving girls the skills to negotiate with others (Ashraf et al. 2020). However, these studies do not compare the effectiveness of different approaches.

2 Context and Interventions

The setting for this paper is 125 rural villages in Dholpur district in the state of Rajasthan. We group these villages into 90 clusters for the trial. Rajasthan is one of the poorest states in India and has some of the worst outcomes for women in the country. Dholpur district is typical of much of rural Rajasthan in this. 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.[8] 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.

This context is also characterized by highly conservative and pervasive gendered norms that reach many of the most mundane aspects of girls' daily lives. This is evident from the data we collected during the baseline for this study. Panel A of Figure 1 shows that the great majority of mothers and girls in our sample agree that girls should primarily stay at home, only venturing into public spaces when absolutely necessary and, ideally, only when accompanied.[9] 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.

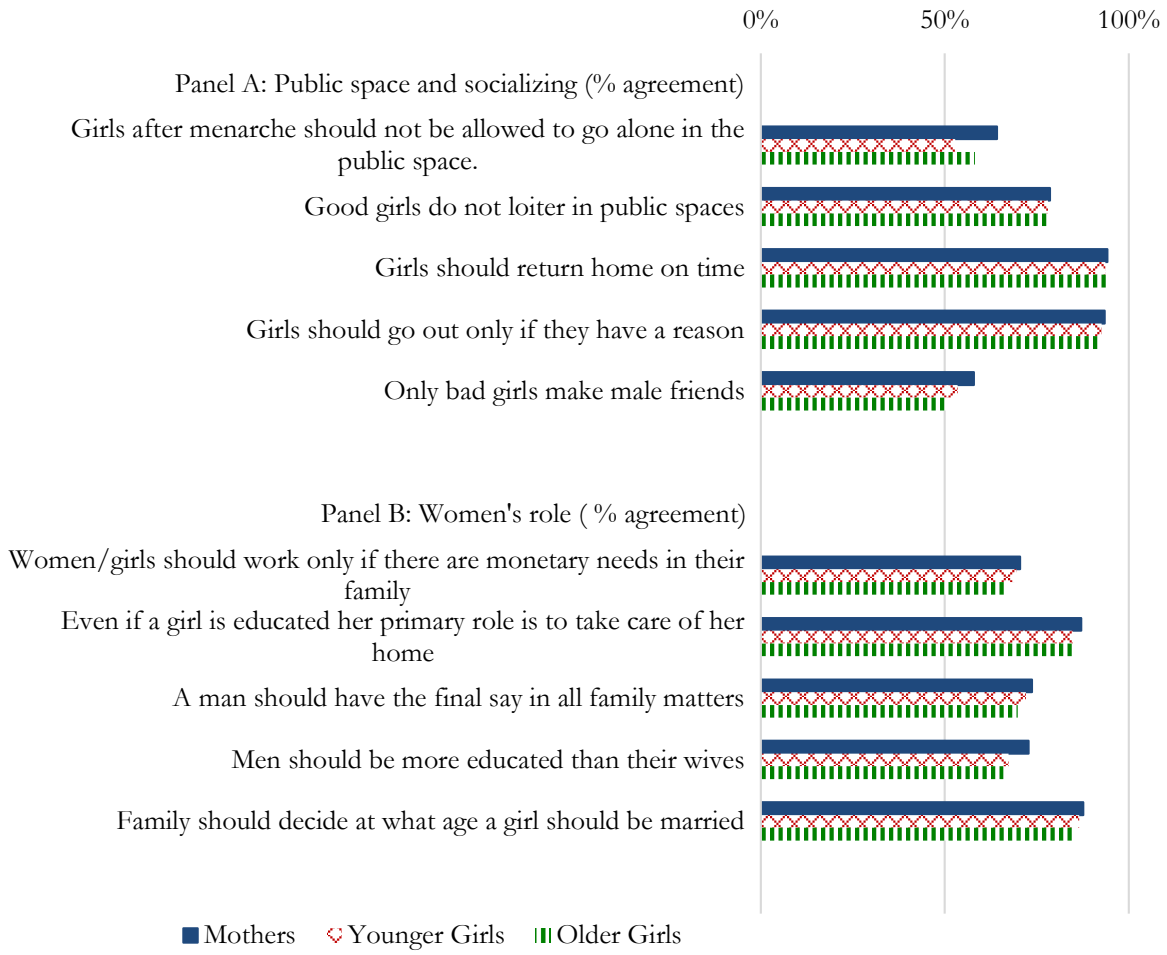
Figure 2 further suggests that women and girls living in this context also accept (Panel A) and expect (Panel B) that those who break the gendered norms of behavior are punished, including violently. For example, half of the mothers and girls in the sample believe that if a girl was to walk home from 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.

These details are 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. For example, many of these norms are at odds with the reality of the long commutes on mixed-gender public transport that are often what is required to attend education past the primary level.

[8] All figures taken from summary statistics produced using the 2015–16 National Family Health Survey. Dholpur statistics available from: `http://rchiips.org/nfhs/RJ.shtml`. National statistics available from: `http://rchiips.org/nfhs/factsheet_NFHS-4.shtml`.

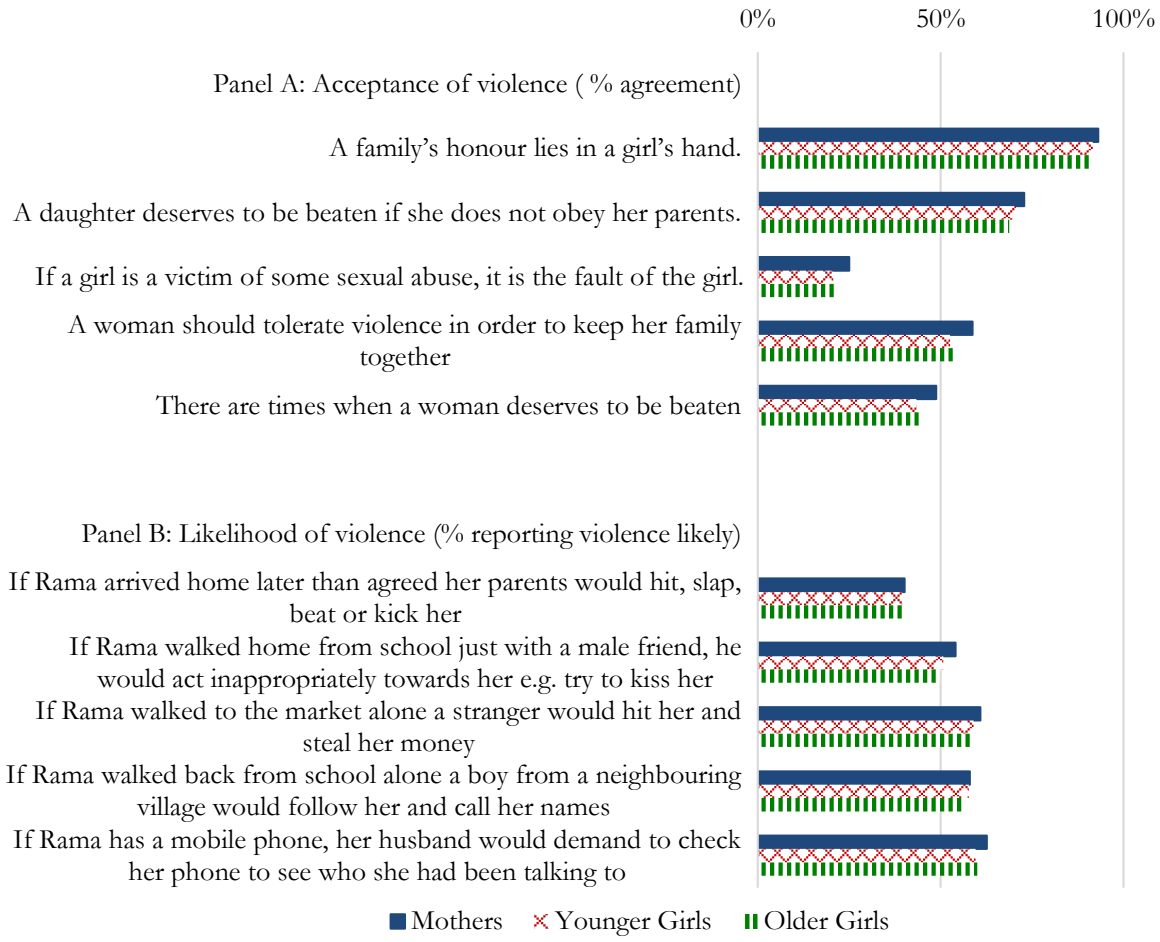
[9] Note that we split the girls' sample into younger girls (aged 12-14 at baseline) and older girls (aged 15-17). As described in Sections 2.1 and 3.1, this split reflects the sampling strata and is motivated by the fact that some of the intervention content differed across these two age-groups.

Figure 1: Attitudes towards Prevailing Gender Norms as Expressed 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.

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

2.1 Interventions

Both intervention models were designed by gender experts at the India office of the International Center for Research on Women (ICRW) - a global NGO which works on the advancement of rights and opportunities for women and girls.

2.1.1 Model 1: Girl Groups

Like many interventions aiming to improve outcomes of adolescent girls, the Girl Groups intervention worked with adolescent girls only and did not actively engage other community members. It sought to strengthen girls' motivation for pursuing schooling and delaying marriage and encouraged them to place less intrinsic value on the importance of abiding by restrictive gender norms. At the same time, it aimed to give them strategies and skills to help them pursue their goals.

One or more Girl Groups were formed in each participating community. To ensure that the content and discussions in the groups were appropriately targeted, participants were divided by age; girls aged 12–14 at the program's start were grouped together, and girls aged 15–19 were grouped together.[10] The Girl Groups were facilitated by "peer mentors" who were young women, aged 19–25, from the community. There were between one and three mentors per community, depending on the number of groups, and mentors were paid Rs. 2000 (roughly USD 25) per month which corresponded roughly to 1.5 times the minimum wage for two days of work per 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. 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.[11] In 2024 USD, the cost of the Girl Groups intervention was approximately USD 45 per eligible girl.[12]

Girl Groups met separately for their own educational and sports sessions.[13] Within the two age strata, group formation was pragmatic and girls were not randomly assigned to groups.

The discussion and education sessions followed a curriculum designed for this program. This

[10] Note that while the groups were open to girls up to 19, we only included girls up to 17 at baseline in our study.

[11] See Section 3 for information on compliance.

[12] This includes program administration, staff training, implementation, monitoring, materials, and staff wages but excludes evaluation and research activities, which would not be required for program replication. See Table B.1 for details.

[13] In the Girl Groups and Community Campaigns arm, same-age groups occasionally held joint community events.

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. For example, the girls were encouraged to consider and question the idea that it is always wrong for girls to travel to school alone or 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 different 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.[14]

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. While adolescent girls were familiar with the basic structure of kabaddi, most had not played it since childhood. The sports sessions were novel in that they gave girls the opportunity to occupy public space within the village, and to play and to be physically active in public.

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. The sakhis' role in the Girl Groups intervention was to reassure key community leaders (typically one or two per community) that there was nothing improper about the intervention and that it was appropriate for the groups to go ahead, assist the mentors in

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

making the practical arrangements for the groups (such as booking the meeting space), and reassure the parents of individual girls that the girls would benefit from attending the groups. Crucially, the role of the sakhis was limited to activities that directly enabled the groups to form and operate and did not involve broader advocacy about the situation of adolescent girls.

2.1.2 Model 2: Girl Groups and Community Campaigns

The second intervention model combined the Girl Groups program with Community Campaigns. These involved girls designing and leading "call for action" events which were intended to engage the community with the substance of the Girl Group program. As such the Community Campaigns were an *add-on* to the Girl Group program rather than a separate stand-alone intervention. The "call for action" events were organized to coincide with the completion of modules of the Girl Group curriculum. This roughly corresponded to an event after every two to three Girl Group sessions. Everybody in the community was invited. The girls invited their parents, siblings, and other people they knew; the Sakhis and Girl Group mentors reached out to community leaders and made announcements in public forums. The addition of the Community Campaigns added a further USD 6 per eligible girl, bringing the total cost of the combined intervention to USD 51 per eligible girl (in 2024 USD) (Table B.1).

The events were held in public outside spaces in a central location within the communities. During the events, the girls raised the issues that they had been considering in the Girl Group sessions. Since the Girl Groups themselves crafted the agenda, the exact content of the campaigns varied across villages but issues of mobility, education, marriage, and girls' independence were frequent topics. The girls used different communication mediums including role-play, songs, posters, and slogans. After the girls introduced the issue, the mentor and Sakhis 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 discrimination, violence and harassment of girls, reinforcing the message that these were unacceptable, regardless of the situation.

The aim of the campaigns was to give girls the rare chance to take center stage in leading a process of discussion and deliberation at the community level about the issues covered in the Girl Group program that most affected them. The design embeds the idea that, without such an intervention, the ability of adolescent girls to communicate their concerns and convince others of new ideas is severely limited because of how marginalized they are. They do not have the social standing to be persuasive or even listened to and lack a communication platform not least

because of the restrictions they face on being in public spaces or drawing attention to themselves. The addition of Community Campaigns to the Girl Groups might relax these constraints and enable the girls to shift attitudes within their community by providing a legitimate platform for girls to advocate for themselves, by backing up girls' voices with those of powerful members of the community, and by providing a space for community deliberation (Heller and Rao 2015).

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 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. Since it was integral to the design of the Community Campaigns that they built on the material covered in the Girl Groups and were led by the girls participating in the Girl Groups, it was not possible to have a treatment arm where the campaigns were implemented alone.

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.[15] We organized unmarried girls by age into two strata: those aged 12–14 and 15–17 at baseline.[16] As noted above, this age split aligns with how the Girl Groups were organized - with girls in the 12-14 age-range grouped together and those in the 15-17 age-range grouped together. In each cluster, we drew a random sample of 39 younger and 37 older girls from the census data. In communities where fewer than this number of girls were listed in the census, we included all of the girls in

[15] In rare cases where no household member was at home during the census, we collected information on whether there were adolescent girls in the household from neighbors and returned to the household at a later time to collect the details.

[16] Initially there was an ambition to involve young women who were already engaged or married in the program. To this end, an alternative education session curriculum was developed; it placed more emphasis on what to expect in married life, sexual reproductive health and family planning and less emphasis on education since in this context girls leave school when they get married. In order to evaluate the impacts of this version of the program, we additionally collected baseline data on 1,782 married or engaged girls aged 12–19. However, due to very low (<7%) rates of participation in the intervention among this group, it was not possible to evaluate the impact of the intervention on this group and they were not followed up at endline. We note that this does not bias our estimates: married and unmarried girls were always included in separate strata based on their baseline marital status and thus our estimated treatment effects for unmarried (at baseline) girls remain unbiased. We collected additional qualitative data on engaged and 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)).

the respective age groups. These target numbers were based on power calculations for main outcomes and anticipating non-response and attrition of around 20%. This sampling process gave us a list of 6,798 adolescent girls who we aimed to include in the baseline survey. During baseline, interviewers verified girls' ages which led to a small number of girls being identified as ineligible. There were also several girls (or their families) who refused to participate, either before the survey or midway through. Overall, 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. This represents 84.3% of those targeted for baseline. The process of conducting the census, sampling and baseline survey was blind to treatment allocation.

The baseline survey, which took place between January and March 2016, included measures of schooling, marriage, mental health, and gender attitudes. In addition to interviewing girls, we interviewed their primary caregivers (typically their mothers), and administered 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, beginning three months after the intervention had finished. 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 5,525 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 other studies with adolescent girls who are often a highly-mobile group, 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), but are slightly higher than the 1% attrition for a similarly-constructed "extended sample" in Edmonds et al. (2021). Importantly, attrition from both the main and the extended samples is uncorrelated with treatment status (see Table B.2).[17]

Table 1 presents key baseline characteristics for the sample of non-attritors by treatment

[17] Table B.2 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*			
	Control	Girl Groups	Girl Groups + Community	Difference	Control	Girl Groups	Girl Groups + Community	Difference
Girl Age	12.928	12.892	12.919		15.820	15.874	15.882**	
	(0.858)	(0.849)	(0.847)		(0.783)	(0.809)	(0.808)	
		[p=0.362]	[p=0.823]	[p=0.526]		[p=0.152]	[p=0.046]	[p=0.823]
Carer's Years of Ed.	1.448	0.995*	1.241		1.197	0.782*	0.908	
	(2.863)	(2.457)	(2.543)		(2.641)	(2.245)	(2.232)	
		[p=0.063]	[p=0.351]	[p=0.278]		[p=0.074]	[p=0.145]	[p=0.554]
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=0.824]	[p=0.194]	[p=0.321]		[p=0.722]	[p=0.305]	[p=0.515]
Asset Index	-0.076	-0.130	-0.105		0.111	0.019	0.004	
	(0.844)	(0.809)	(0.811)		(0.983)	(0.897)	(0.901)	
		[p=0.444]	[p=0.686]	[p=0.677]		[p=0.205]	[p=0.173]	[p=0.853]
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=0.249]	[p=0.000]	[p=0.049]		[p=0.295]	[p=0.129]	[p=0.628]
Progressive Gender Attitudes	34.722	34.010	34.658		35.447	34.429	35.464	
	(6.895)	(6.861)	(6.840)		(6.786)	(6.907)	(7.065)	
		[p=0.238]	[p=0.906]	[p=0.219]		[p=0.119]	[p=0.978]	[p=0.096]
Carer's Progressive Gender Attitudes	34.018	33.159	34.026		33.916	33.308	33.608	
	(6.632)	(6.514)	(6.580)		(6.470)	(6.399)	(6.551)	
		[p=0.139]	[p=0.989]	[p=0.087]		[p=0.255]	[p=0.599]	[p=0.569]
Mental Health	3.126	3.033	2.825		2.926	3.044	2.746	
	(2.238)	(2.102)	(2.101)		(2.221)	(2.168)	(2.123)	
		[p=0.633]	[p=0.111]	[p=0.234]		[p=0.585]	[p=0.325]	[p=0.135]
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 "Difference" column tests equality between the two intervention models. The p-values and standard errors are constructed using a cluster bootstrap accounting for stratification by region and treatment status. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

status while Tables B.3 and B.4 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-economic characteristics, the sample appears well balanced. Likewise, the sample is well balanced on baseline 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 B.3 and B.4) 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 C).

3.2 Ethics

We received ethical clearance for the project from the Sigma IRB Committee (New Delhi), ICRW IRB (Washington DC), and Oxford University IRB (Oxford, UK). In Appendix A, we follow Asiedu et al. (2021) and create a structured ethical appendix covering policy equipoise, role of the researcher, potential harms to participants and nonparticipants, conflicts of interest, intellectual freedom, feedback to participants, and foreseeable misuse of research results.

3.3 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.[18] 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.

After accounting for breaks during holidays, the median Girl Group met for 46 education

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

sessions and 40 sports sessions. Registers taken during the Girl Group sessions indicate that 49% of the girls in the study sample allocated to one of the treatment arms enrolled in the program and attended at least one session. Enrollment was slightly higher among girls aged 12–14 (54%) than among the older girls aged 15–17 (43%). Conditional on enrolling, girls attended an average of 28 group educational sessions which amounted to 61% of the sessions. Enrollment and attendance rates at the sports sessions were very similar; conditional on enrolling, the average girl attended 23 sports sessions.[19] Table B.5 shows attendance rates in more detail and documents nearly identical attendance patterns between the two treatment arms.[20]

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.

3.4 Outcomes

We pre-specified five primary outcomes: education, marriage, mental health, gender attitudes and non-cognitive skills (Andrew and Krutikova 2018). We describe each below.

3.4.1 Education and Marriage

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 a 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.

[19] Due to extreme weather during the summer months slightly fewer sports sessions were held compared to the education sessions (median of 40 vs. 46). However, Table B.5 shows that the rates of attendance are very similar.

[20] These attendance levels are comparable to those reported in other studies of adolescent girl empowerment programs. For example, Buchmann et al. (2017) report that 56.1% of adolescent girls attended at least one empowerment session in their study in Bangladesh, a similarly conservative setting. In Bandiera et al. (2019), conducted in Uganda - a much less conservative context - just 19.6% of girls in the treatment arms participated in any group activity.

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.

3.4.2 Mental Health Problems

At endline, we measured, and pre-specified, three measures relating to mental health problems. The first two – symptoms of depression and symptoms of anxiety – are measures of "disorders" and can, therefore, be considered measures of mental health problems in their own right.[21] The third, rumination, is defined as repetitive and recurrent negative thinking about oneself, one's situation, upsetting experiences, and concerns (Watkins 2008). Rumination, which is often linked to blaming oneself for one's problems, has been shown to lead to onset and worsening of symptoms of anxiety and depression and to negative emotions even in the absence of any mental disorder.[22]

To measure symptoms of depression and anxiety, respectively, we used the nine-item Patient Health Questionnaire (PHQ-9) and the seven-item Generalized Anxiety Disorder scale (GAD-7). We chose these scales because they have performed well with adolescent girls in Hindi-speaking areas of India in previous studies (Leventhal et al. 2015). We measured a ruminative response style using the 10-item Ruminative Responses Scale (RSS-10) (Treynor et al. 2003). Following adaptation to the local dialect of Hindi, piloting showed that items from all 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". Examples of symptoms of depression include experiencing "little interest or pleasure in doing things" and "feeling bad about yourself or that you are a failure and have let your family down". Symptoms of anxiety include "feeling afraid as if something awful might happen" and "not being able to stop or control worrying".

Unlike these two scales, the RSS-10 asks individuals how they respond to being "depressed or feeling sad" on a 4-point Likert scale ranging from "strongly agree" to "strongly disagree". Items ask whether respondents adopt responses including "thinking 'Why can't I handle things better?' " and "thinking 'Why do I have problems other people don't have?' ".

We use individual item responses to construct separate factor scores for depression, anxiety

[21] "Depressive disorders" and "anxiety disorders" are both major categories of mental disorders listed by the American Psychiatric Association in the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5).

[22] See summary produced by the American Psychiatric Association: `https://www.psychiatry.org/news-room/apa-blogs/rumination-a-cycle-of-negative-thinking`.

and rumination. We provide details of the methodology we use to create these factors below in subsection 3.4.6. 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.

At baseline, we did not use the PHQ-9, GAD-7 and RSS-10. Instead, we used the General Health Questionnaire (12 question version) which captures symptoms of psychological distress associated with both depression and anxiety. As seen in Table 1, this is well balanced in the baseline sample.

3.4.3 Gender Attitudes

We used an adapted version of the GEMS scale (Pulerwitz and Barker 2008) to measure attitudes of the girls and their mothers towards prevailing gender norms and pre-specified this as a primary outcome. This scale asked girls and their mothers to indicate how strongly they agreed or disagreed with "...the beliefs and assumptions held in [their] society towards boys and girls or towards men and women". Girls were presented with 14 statements such as "Women or girls should work only if there are monetary needs in their family". For each statement, possible responses were "strongly agree", "agree", "disagree" or "strongly disagree". Earlier, we presented descriptives of responses to many of these statements in the baseline data (Figures 1 and 2A); Appendix Table D.4 gives the full list of statements used at endline.

We use these item responses to create factor scores for gender attitudes applying the procedure outlined in Section 3.4.6 and use these factor scores as our outcome measures.

3.4.4 Non-Cognitive Skills

Finally, we measured and pre-specified a set of non-cognitive skills to test whether the interventions may have affected girls' ability to persevere through difficult situations and negotiate with their families (Ashraf et al. 2020) to achieve their preferred marriage and schooling outcomes. In particular, we measured the following skills using previously-validated scales: (i) self-efficacy, (ii) self-esteem, (iii) peer relations, (iv) resilience, (v) a vigilant style of decision making and (vi) a buck passing style of decision making.[23] Again, we combine items within each scale to estimate summary factors. Appendix Tables D.5 through D.10 give a full list of all statements used and the estimated factor model parameters.

[23]Self-efficacy was measured using the 10-item General Self-Efficacy Scale, self-esteem and peer relations using the relevant sub-scales of the Self Description Questionnaire, resilience using the 10-item Connor-Davidson Resilience Scale, and vigilant and buck-passing decision making using the relevant scales of the Melbourne Decision Making Questionnaire. Andrew and Krutikova (2018) provides details of how each scale has been previously validated.

3.4.5 Pre-specified secondary outcomes

In pre-specifying our five primary outcomes (education, marriage, mental health, gender attitudes and non-cognitive skills), we selected those that were most directly targeted by our interventions and which we knew could be measured well. These primary outcomes are the focus of this paper. We also pre-registered five secondary outcomes: attitudes towards school, knowledge of sexual and reproductive health, attitudes to and responses to violence, participation in and attitudes towards sports, and restrictions during menstruation. These were outcomes that were not as key to our hypothesized mechanisms and/or not as easily measured. Several of these secondary outcomes have multiple measures. For instance, we measure both girls' and their carers' attitudes and responses to violence across five different dimensions. To avoid cherry-picking results, we do not present any of the secondary outcomes (but present all of the primary outcomes) in the main text but instead present them all in Appendix B. Descriptions of the instruments used to measure each can be found in our pre-analysis plan (Andrew and Krutikova 2018) and in the table notes. Where we do bring estimated impacts on relevant secondary outcomes into our discussion of mechanisms, we make sure to clearly label the outcomes as secondary.

3.4.6 Factor Models

Constructs like mental health and gender attitudes which we attempt to capture through our scales are complex and hard to measure. The responses to each of the items in the scales are noisy measures of these complex underlying constructs. Following an increasingly common approach in economics, we use these responses to estimate the underlying constructs using factor models (Das and Zajonc 2010). These techniques combine the information contained in the available data efficiently. Furthermore, they model measurement error directly. This allows for the estimation of treatment effects scaled relative to the true variance of the underlying construct in the control group, uncontaminated by variability induced by measurement error, which would be a concern if we were to use raw aggregates.

Specifically, we index these underlying constructs by k.[24] 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

[24]Specifically, k can denote: girls' depressive symptoms, girls' anxiety symptoms, girls' rumination, girls' progressive gender attitudes, or carers' progressive gender attitudes.

item response theory (IRT) factor model suited for ordinal items). In particular, we model the probability of individual i giving an ordinal response of at least m to item j as an ordered logit:

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

We assume that θ_{ik} is normally distributed with zero mean and unit variance in the control group. We estimate the item parameters on the control group only, by maximum likelihood using an Expectation–Maximization (EM) algorithm with Gauss-Hermite quadrature to approximate the integral over the unobserved latent factors. We then use these estimated item parameters to compute factor scores for all individuals (control and treated). This procedure ensures that the effect sizes can be interpreted relative to the control-group standard deviation. 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 cutpoints, representing the value of θ_{ik} at which the probability of responding at level m or above equals 0.5.

For each construct k, a higher value of θ_{ik} represents a higher level of the construct. Tables D.1, D.2, D.3 and D.4 present the parameter estimates for the measurement models of, respectively, depression, anxiety, rumination, and gender attitudes. The estimates suggest that all items had high discriminatory power; all discrimination parameters are positive and statistically significant.

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 these factor scores as our main measures for these constructs in the remaining analysis.

3.5 Empirical Strategy

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

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

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 $\beta_{girl+comm}$, or the associated marginal effects for binary outcomes, 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, we can 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, our estimates 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 C 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 bootstrap with 1000 iterations and accounting for stratification by block and by treatment status. In line with the pre-analysis plan (Andrew and Krutikova 2018), 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. As noted above, this age split aligns with the way that the Girl Groups were organized.

4 Impacts on Education, Marriage & Mental Health

In this section, we present impacts on three of our five pre-specified primary outcomes: education, marriage and mental health. We defer discussion of gender attitudes and non-cognitive skills to Section 5.

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 at the greatest risk of getting married and dropping out of school. For these girls, we estimate that the groups increased the probability that they were still in school or post-secondary education at endline by 6.0 p.p. ($p = 0.009$) in the Girl Groups only arm and 4.7 p.p. ($p = 0.026$) 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 of younger girls from either intervention model. The point estimates, which show 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 school attendance rate in this age group (76.5% in the control group). The impacts in the Girl Groups arm for the younger sample are statistically significantly different from those for the older sample. Pooling both age groups together, the average ITT effect of Girl Groups only in the whole sample is an increase in school attendance of 3.9 p.p. ($p = 0.018$) and 3.0 p.p. ($p = 0.08$) in the Girl Groups and Community Campaigns arm. Relative to the control group, 59% of whom were attending education at the time of endline, both intervention models led to increases in education attendance in the range of 5–7%.

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 B.6 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 probability of being married by 3.6 p.p. ($p = 0.035$), 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 (2.5 p.p.) but it is not statistically significant; neither, however, is the difference between the two treatment groups (as we discuss below). Pooling the two treatment arms gives a statistically significant 3.1 p.p. reduction in the probability of older girls being married ($p = 0.049$, Appendix Table B.7). Again, we see no impacts for the younger group, among whom fewer than 6% were married at endline in the

Table 2: 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	0.060***	-0.036**	-0.053***	0.018	0.005	0.016	0.039**	-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=0.035]	[p=0.007]	[p=0.342]	[p=0.647]	[p=0.259]	[p=0.018]	[p=0.158]	[p=0.186]
Girl Groups	0.047**	-0.025	-0.057**	0.014	0.009	0.021	0.030*	-0.007	-0.015
+ Community	(0.021)	(0.019)	(0.024)	(0.021)	(0.011)	(0.016)	(0.017)	(0.012)	(0.016)
	[p=0.026]	[p=0.195]	[p=0.018]	[p=0.491]	[p=0.407]	[p=0.178]	[p=0.080]	[p=0.535]	[p=0.350]
Difference	-0.013	0.011	-0.003	-0.003	0.004	0.005	-0.009	0.008	0.003
	(0.023)	(0.019)	(0.024)	(0.021)	(0.010)	(0.016)	(0.018)	(0.012)	(0.017)
	[p=0.574]	[p=0.569]	[p=0.885]	[p=0.868]	[p=0.704]	[p=0.752]	[p=0.620]	[p=0.518]	[p=0.882]
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=0.089]	[p=0.022]	[p=0.001]			
Girl Groups + Community				[p=0.159]	[p=0.097]	[p=0.001]			
Difference				[p=0.710]	[p=0.701]	[p=0.706]			

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

control group. 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 from the definition of marriage cases where girls were formally married but had not yet moved to their husband's house (no "gauna"), as sometimes happens when young girls marry (Appendix Table B.8).

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. Alternatively, the programs may have led to a delay in initiating the whole process of marriage.

We find, in columns 3, 6 and 9 of Table 2, 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 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 an 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 among the younger girls and the pooled sample.

For both outcomes, marriage and education, the results are consistent with the two intervention models being 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. We return to this discussion below.

4.2 Mental Health Problems

Next we turn to the impacts of the interventions on girls' mental health. As detailed in Section 3.4.2, we have three measures relating to mental health problems including symptoms of

depression, anxiety, and rumination.[25]

We find that the Girl Groups only program had no significant impact on symptoms of either depression or anxiety. This holds for both younger and older girls (Table 3). However, we find evidence that the Girl Groups program, when implemented alone, led to girls adopting a more ruminative response style to negative emotions (Table 3). Estimated impacts are positive (indicating more rumination) for both younger girls (0.157 SD, $p = 0.076$) and older girls (0.135 SD, $p = 0.103$). The Control Theory in psychology posits that "goal discrepancy" - a gap between what can be realistically achieved and one's goals - is an important trigger of rumination (Martin and Tesser 1996; Watkins and Nolen-Hoeksema 2014). The intervention may thus have triggered more rumination through promoting goals that were very far from what the girls felt they could attain, even taking into account the improvements in schooling and marriage outcomes achieved. This, alongside the psychological stress and any retaliation they may have experienced from the wider community in response to changing their behavior may explain why, in spite of the improvements in schooling and early marriage, the Girl Groups intervention did not lead to improvements in depression and anxiety. We note that the null effect of the Girl Groups on depression and anxiety may mask offsetting influences: while the improvements in schooling, marriage delay, and peer contact would typically be expected to improve adolescent girls' mental health, these gains may have been countered by the psychological costs of increased rumination, goal discrepancy, and living in greater tension with prevailing community norms. We flag this possibility tentatively, since identifying each channel separately would require exogenous variation that our design does not provide.

Turning next to the Girl Groups and Community Campaigns intervention model, we find that the addition of the community engagement component resulted in significant and substantial improvements in mental health. Overall, we find that the combined intervention reduced symptoms of depression by 15.5% of an SD and reduced symptoms of anxiety by 16.5% of an SD relative to the control group (Table 3). Differences between the two intervention groups, i.e. the additional impact of the Community Campaigns, are nearly identical. Impacts on both depression and anxiety, at 23.3% SD and 23.3% SD respectively, are substantially and significantly larger for the older girls but the combined intervention also reduced anxiety by 10.4% SD for the younger girls.

Finally, we find that with the *addition* of the Community Campaigns, there is no longer an increase in rumination as there was with the Girl Groups alone. The 17.4% of an SD difference

[25]We did not measure other behavioral and cognitive symptoms commonly associated with depression and anxiety—such as anhedonia, changes in sleep and appetite, irritability, and hopelessness—which may also have been affected by the interventions. We therefore cannot assess whether the treatments altered these symptoms.

Table 3: Mental Health Problems

	Older Girls			*Younger Girls*			*All*		
	Depression	Anxiety	Rumination	Depression	Anxiety	Rumination	Depression	Anxiety	Rumination
Girl Groups	0.001	-0.038	0.135	0.049	0.008	0.157*	0.025	-0.014	0.148*
	(0.074)	(0.061)	(0.083)	(0.074)	(0.064)	(0.089)	(0.068)	(0.057)	(0.082)
	[p=0.991]	[p=0.535]	[p=0.103]	[p=0.504]	[p=0.895]	[p=0.076]	[p=0.716]	[p=0.800]	[p=0.073]
Girl Groups	-0.233***	-0.233***	-0.050	-0.085	-0.104*	-0.007	-0.155**	-0.165***	-0.026
+ Community	(0.067)	(0.059)	(0.103)	(0.068)	(0.060)	(0.092)	(0.061)	(0.052)	(0.094)
	[p=0.001]	[p=0.000]	[p=0.630]	[p=0.213]	[p=0.081]	[p=0.941]	[p=0.011]	[p=0.002]	[p=0.780]
Difference	-0.234***	-0.195***	-0.185*	-0.134**	-0.113**	-0.164	-0.180***	-0.151***	-0.174*
	(0.063)	(0.050)	(0.102)	(0.061)	(0.057)	(0.104)	(0.054)	(0.047)	(0.098)
	[p=0.000]	[p=0.000]	[p=0.069]	[p=0.027]	[p=0.048]	[p=0.114]	[p=0.001]	[p=0.001]	[p=0.077]
N	2,318	2,318	2,318	2,725	2,725	2,724	5,043	5,043	5,042
Control Mean	0.001	0.001	0.000	-0.000	0.000	0.000	0.000	0.001	0.000
Heterogeneity by Age									
Girl Groups				[p=0.393]	[p=0.378]	[p=0.663]			
Girl Groups + Community				[p=0.013]	[p=0.024]	[p=0.410]			
Difference				[p=0.091]	[p=0.117]	[p=0.723]			

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 (as measured by the GHQ-12). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

in rumination between the two treatment arms is statistically significant while the combined treatment mean is indistinguishable from the control group.

We test the robustness of these results to different ways of constructing the measures. Tables B.9 and B.10 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. 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. Estimated effect sizes on these raw scores are, in terms of control group standard deviations, slightly larger than for the factor score estimates.

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 note that for rumination, both the increase in rumination raw scores in the Girl Groups arm and the difference between the two arms are at the margins of statistical significance at conventional levels ($p = 0.124$ and $p = 0.122$; Table B.11).

We find that the combined treatment also reduced binary indicators of both depression and anxiety (Tables B.9 and B.10).[26] 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%.[27] 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.

4.3 Additional Robustness

We now discuss some additional robustness and sensitivity checks of our main results.

First, we investigate the role of social desirability bias. A potential concern with our findings on marriage and education is that respondents may give socially desirable answers that align with the interventions' messaging. We investigate this by exploiting the fact that girls and their carers were interviewed separately and privately, with the girl interviewed first whenever possible, making it unlikely that both would coordinate on the same socially desirable report.

[26] We could not find any cutoff for the RSS-10 that has been proposed as a binary indicator of rumination.

[27] 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 B.12 cross-tabulates girl and carer reports on marriage and education attendance, showing high agreement rates across both outcomes. We then test directly whether treatment increases the rate of disagreement between girl and carer reports. If social desirability bias were driving our results, we would expect treatment to increase divergence between reports. Table B.13 shows no evidence that treatment affects reporting discrepancies for either outcome. Finally, we construct "worst case" outcome measures that take the most socially undesirable report from either respondent: for marriage, a girl is coded as married if *either* she or her carer reports so; for education, a girl is coded as attending only if *both* reporters agree. Treatment effects using these conservative measures, reported in Table B.14, are very similar to our main estimates in Table 2, indicating that social desirability bias is unlikely to be a major driver of our findings.

We next assess sensitivity of our estimates to the empirical choices we made. 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 a logit (Table B.15).

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 et al. (2014), which we describe in Appendix C. 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 mother's 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.[28] 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 C.1 and C.2 in Appendix C).

Finally, we test whether our education and marriage results are robust to adjustments in 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

[28] Full sets of covariates selected for each outcome are listed in Table Notes under Tables C.1 and C.2

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 B.16 and B.17).[29]

5 Mechanisms

We have found so far that both intervention models – working with girls alone and combining this with structured engagement between girls and their broader communities – led to meaningful improvements in girls' schooling and marriage outcomes. However, the two approaches had strikingly different effects on mental health. Engaging with girls alone increased the extent to which girls ruminated and blamed themselves for their distress. In contrast, adding community engagement led to substantial improvements in mental health. Notably, however, these mental health gains did not feed back into stronger impacts on schooling and marriage outcomes.

What mechanisms might explain these patterns? As discussed in Section 2.1, the program was designed to empower girls to pursue their life goals by shaping their behavior and well-being through several channels. These include weakening the influence of restrictive gender norms, strengthening motivation to pursue education, and building non-cognitive skills, voice and agency. In the Girl Groups arm the aim was to activate these mechanisms by working directly with the girls. The addition of Community Campaigns was intended to amplify these effects by creating a more enabling environment and empowering the girls through giving them a platform to engage with the wider community.

We now explore these potential mechanisms for impacts of each of these interventions in turn.

5.1 Girl Groups

The Girl Groups were the critical ingredient for the impacts that we find on schooling and marriage since the addition of the Community Campaigns had no additional impact on these outcomes. Reduction in marriage through the Girl Groups is especially striking because girls in this context have very little say over the timing of their marriage; at baseline, less than 10% of girls reported 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, where around two-thirds of the girls reported 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

[29] 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 B.17).

of school and unmarried as being at reputational risk and would start arranging a marriage once a girl has left education. A negative marriage market return to age once girls are out of school and the fact that parents' choices for their daughters' education are shaped by their daughters' attitude towards school was also found by Andrew and Adams (2025) using vignette data from this same sample. This combination of factors suggests that the marriage and engagement impacts may reflect knock-on effects of changes in the decisions girls made about schooling as the result of the Girl Groups. This is consistent with a strong negative correlation that we see between attending school and being married at endline. As we show in Appendix Table B.18, controlling for being in formal education reduces the size and significance of the marriage results.

Table 4: Progressive Gender Attitudes

	Older Girls	*Younger Girls*	*All Girls*	*Mothers*
Girl Groups	-0.060	-0.130**	-0.097*	-0.070
	(0.063)	(0.058)	(0.053)	(0.059)
	[p=0.338]	[p=0.025]	[p=0.067]	[p=0.236]
Girl Groups + Community	0.167**	0.031	0.095*	0.054
	(0.067)	(0.063)	(0.057)	(0.063)
	[p=0.013]	[p=0.627]	[p=0.098]	[p=0.386]
Difference	0.227***	0.161**	0.192***	0.125**
	(0.064)	(0.065)	(0.057)	(0.063)
	[p=0.000]	[p=0.014]	[p=0.001]	[p=0.046]
N	2,318	2,725	5,043	4,805
Control Mean	-0.000	-0.000	-0.000	-0.003
Heterogeneity by Age				
Girl Groups			[p=0.235]	
Girl Groups + Community			[p=0.026]	
Difference			[p=0.276]	

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

What then might be driving the increase in schooling? As set out above, one potential mechanism is relaxing the influence of gender norms on the girls. We thus examine impacts on the progressivity of girls' attitudes towards prevailing gender norms (Table 4). The pattern of impacts is not consistent with change in gender attitudes being a mechanism. We do not find any evidence of an increase in progressivity of attitudes in the Girl Groups only arm.

If anything, we find evidence of a *reduction* in progressivity of girls' attitudes when the Girl Groups program is implemented alone, which is driven by a significant negative effect on the younger girls. One explanation might be that the discussions in the Girl Groups increased girls' awareness of just how conservative norms were in their communities, shifting their attitudes to be more in line with these norms [30]

Table 5: Estimated Impacts on the Number of School Days Missed in the Last Month and on Non-Cognitive Skills

	All						
	Schooldays missed	Self Efficacy	Self Esteem	Peer Relations	Resilience	Vigilant Dec. Mak.	Buck-passing Dec. Mak.
Girl Groups	-0.779*	-0.001	0.025	0.014	-0.117	-0.071	0.032
	(0.453)	(0.088)	(0.084)	(0.072)	(0.083)	(0.061)	(0.051)
	[p=0.086]	[p=0.991]	[p=0.768]	[p=0.842]	[p=0.159]	[p=0.245]	[p=0.527]
Girl Groups	-0.712*	0.030	0.070	0.007	-0.007	0.037	0.035
+ Community	(0.400)	(0.089)	(0.078)	(0.062)	(0.084)	(0.072)	(0.052)
	[p=0.075]	[p=0.734]	[p=0.372]	[p=0.916]	[p=0.935]	[p=0.605]	[p=0.502]
Difference	0.067	0.031	0.045	-0.008	0.110	0.108*	0.003
	(0.254)	(0.081)	(0.072)	(0.061)	(0.072)	(0.062)	(0.046)
	[p=0.792]	[p=0.701]	[p=0.535]	[p=0.897]	[p=0.127]	[p=0.081]	[p=0.948]
N	3,084	5,043	5,043	5,043	5,043	5,042	5,041
Control Mean	3.538	-0.002	-0.001	-0.000	-0.001	-0.001	0.000

Notes: The table presents estimated 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). For non-cognitive skill outcomes, 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 baseline levels of non-cognitive skills. For the number of schooldays missed, 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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Another potential mechanism is strengthening the girls' motivation to remain in education. Interestingly, when asked in the abstract, girls' stated aspirations for education in our sample were very high. For instance, at baseline 83% of girls who were in school reported wanting to finish upper secondary school (12th standard); this was also the case for 92% of girls in the control group who were still in school at endline. However, the most common explanation for dropping out of school (at both baseline and endline) was "a lack of interest in school" suggesting that girls' own day-to-day motivation may be an important driver of education decisions.[31] This

[30] Another possibility is that the increase in rumination from the Girl Groups (see Section 4.2) led girls to adopt more conservative gender attitudes to lessen the incongruence between their own attitudes and those of their community as a coping mechanism. In line with this we see a strong correlation between girls reporting more symptoms of rumination and holding more conservative gender attitudes (see Figure B.4).

[31] 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 "lack of interest" as their reason for being out of school, making this the most commonly chosen category. The second and third most-prominent reasons included "domestic

was also stressed as a key factor in focus group discussions with mothers and daughters.[32] In the first column of Table 5, we see suggestive evidence that both treatment arms reduce the number of days girls reported missing school in the last month (six months after the end of the program) by around 20% (0.7–0.8 days) which may be indicative of an increase in the girls' day-to-day educational motivation as the result of the interventions. However, notably we do not see any change in the "positive attitudes towards school" scale which we constructed and was pre-registered as a secondary outcome (Appendix Table B.19). We note that since the data on absences and school attitudes is only available for girls enrolled in school at the time of the endline survey, we may be underestimating the true impacts of the interventions on attendance and attitudes if the marginal girls induced to remain in school due to the interventions have lower attendance and enthusiasm.

Finally, the program also targeted girls' non-cognitive skills. Two recent studies (Edmonds et al. 2021; Ashraf et al. 2020) find evidence that, in India and Zambia respectively, developing non-cognitive skills of adolescent girls may contribute to improvements in schooling. We find little evidence that in our context the Girl Groups intervention impacted any of the non-cognitive skills that we measured, including self-efficacy, self-esteem, peer relations, resilience, and decision making skills (Table 5).[33] Importantly, "non-cognitive skills" are a multidimensional construct and are hard to measure. Hence, we cannot rule out that the intervention impacted a dimension that our measures did not capture. Of particular interest are negotiation skills (Ashraf et al. 2020) and skills around planning for the future (Edmonds et al. 2021), especially since one of the aims of the Girl Groups curriculum was to help girls with skills they would need to stay in school in spite of various external constraints.

5.2 Girl Groups + Community Campaigns

As noted above, the addition of community engagement was intended to amplify the effects of the Girl Groups. It had no *additional* impacts on the schooling and marriage outcomes. However, it was critical to the achievement of mental health improvements which were not observed in the Girl Groups only arm.

There are two classes of potential mechanisms for this. First, there could be direct impacts stemming from the *process* of girls organizing and running the community campaigns. This process involved the girls creating, rehearsing and performing role-plays and songs in front of

responsibilities" and school being "too expensive".

[32]For example, one mother mentioned that "if the girl is hardworking and good, the parents will not have to ask her to study or to go to school, they themselves do" and another said "the ones who are good in studies, they themselves get up each day, get ready and look forward to going to school".

[33]Note that here we present combined impacts on older and younger girls together for reasons of space. We present results separately in Appendix Table B.20 but do not find any notable heterogeneity by age.

their communities. It is possible that this was enjoyable and empowering and thus, in itself, beneficial for the girls' mental health. Activities such as story-telling and role play were also included in the Girl Groups' sessions in both arms, therefore, any direct additional impact of running the campaigns on girls' mental health would have to be associated with the impact of performing in front of and engaging with the audience of community members. Second, the aim of the campaigns was to foster discussion of issues that affect adolescent girls between girls and their community and between community members. Thus they may have led to broader changes within the community in attitudes and/or behavior, as well as in the relationship between the girls and the wider community (Baird et al. 2019).

We use heterogeneity analysis to explore these mechanisms. In particular, we assess heterogeneity in the *additional* impacts of the Community Campaigns by whether or not girls attended the Girl Groups. Since Girl Groups attendance was a pre-requisite to participation in the Community Campaigns, girls who did not participate in the Girl Groups also did not get involved in running the Community Campaigns in the combined intervention arm. The motivation here is that *if* our intent-to-treat impacts on mental health are driven by the process of running the campaigns, they should be concentrated among the girls who attended the Girl Groups and hence would have participated in the Community Campaigns. If broader changes in the community play a role, we would expect effects to be present even for girls who did not participate in the Girl Groups.

In this analysis, we leverage the fact that within our sample, roughly half the girls in the treatment arms participated in the Girl Groups and half did not. The issue is that selection into participating in the Girl Groups is not random so we cannot use girls in the control group as a comparison to estimate heterogeneity in the *overall* effects of the two treatment arms by attendance. Instead, we drop the control group and estimate heterogeneity in the *additional* impact of the Community Campaigns by whether or not girls attended the Girl Groups.[34] As we show formally in Appendix E using a Potential Outcomes Framework, we can interpret these estimates causally under the assumption that selection into participation in Girl Groups is invariant with which of the two treatment groups the girl was assigned to.

We offer three pieces of evidence in support of this assumption. First, when girls and their

[34]Specifically, using girls assigned to either treatment group, we estimate the following regression:

$$y_{ic} = \alpha + \beta T_c^{girl+comm} + \kappa T_c^{girl+comm} Attended_{ic} + \theta Attended_{ic} + \gamma X_{ic} + \epsilon_{ic}, \tag{5.1}$$

where $Attended_{ic}$ is an indicator equal to one if girl i attended the Girl Groups and all other variables defined as in equation (3.1). β represents the additional effect of the Community Campaigns (over and above the Girl Groups intervention) for girls who did not themselves attend the Girl Groups. κ is the difference in the additional impact of the campaigns between girls who did attend the Girl Groups and those who did not and $\beta + \kappa$ is the total additional impact of the campaigns on those who attended the groups.

Table 6: Heterogeneity in Mental Health Impacts of Community Campaigns by Attendance

	Panel A: Depression					
	Older Girls			*Younger Girls*		
Attendance threshold:	any	50%	75%	any	50%	75%
Community Campaigns	-0.261***	-0.247***	-0.216***	-0.121*	-0.120*	-0.131**
	(0.070)	(0.064)	(0.063)	(0.069)	(0.066)	(0.060)
Attended	0.057	-0.010	0.130	0.027	-0.019	0.002
	(0.068)	(0.098)	(0.127)	(0.077)	(0.070)	(0.071)
Community Campaigns × Attended	0.096	0.100	-0.036	-0.007	-0.013	0.027
	(0.091)	(0.123)	(0.140)	(0.096)	(0.089)	(0.104)
N	1,492	1,492	1,492	1,794	1,794	1,794
	Panel B: Anxiety					
	Older Girls			*Younger Girls*		
Attendance threshold:	any	50%	75%	any	50%	75%
Community Campaigns	-0.220***	-0.202***	-0.175***	-0.158**	-0.117*	-0.115**
	(0.054)	(0.053)	(0.049)	(0.069)	(0.063)	(0.056)
Attended	0.042	-0.039	0.068	-0.053	-0.043	-0.005
	(0.058)	(0.078)	(0.091)	(0.059)	(0.062)	(0.075)
Community Campaigns × Attended	0.103	0.104	-0.004	0.092	0.029	0.043
	(0.082)	(0.105)	(0.114)	(0.084)	(0.081)	(0.103)
N	1,492	1,492	1,492	1,794	1,794	1,794
	Panel C: Rumination					
	Older Girls			*Younger Girls*		
Attendance threshold:	any	50%	75%	any	50%	75%
Community Campaigns	-0.210**	-0.193*	-0.179*	-0.200*	-0.166	-0.175*
	(0.106)	(0.112)	(0.105)	(0.109)	(0.105)	(0.105)
Attended	-0.072	-0.068	-0.017	-0.064	0.031	0.090
	(0.072)	(0.077)	(0.071)	(0.076)	(0.073)	(0.095)
Community Campaigns × Attended	0.077	0.067	0.017	0.041	-0.032	-0.021
	(0.122)	(0.123)	(0.125)	(0.097)	(0.108)	(0.156)
N	1,492	1,492	1,492	1,793	1,793	1,793

Notes: The table presents estimated heterogeneous impacts of the addition of "Community Campaigns" over and above the Girl Groups intervention by whether or not the girl attended the Girl Groups. Analysis contains only the two treatment arms and excludes the control group. Attendance was measured using implementation data and re-coded as a binary variable based on three thresholds: (i) any attendance (columns 1 and 4); (ii) 50% attendance (columns 2 and 5); and (iii) 75% attendance (columns 3 and 6). Standard errors (in parentheses) are constructed using a cluster bootstrap accounting for stratification by block and treatment status (1000 iterations). Pre-specified controls are those listed in the notes to Table 3. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

families were first told about the program and girls were invited to join a Girl Group, there was no mention of the Community Campaigns in the combined arm and, indeed, mentors in this arm had not yet undergone any of the training on this component. Hence, there is a strong reason to expect that on the extensive margin, selection should be independent of the inclusion of Community Campaigns.[35] Second, in Section 3.3 and Table B.5 we documented identical

[35]When it comes to the intensive margin of participation, fueled by dropouts and attendance rates, it is possible that the addition of the Community Campaigns could have affected attendance in either direction.

attendance rates, both on the intensive and extensive margin, across the two treatment arms. Third, in Table B.21, we additionally show that the patterns of selection based on baseline characteristics (including age, school attendance, wealth, caste, mental health) are the same across the two treatment groups.[36] The additional assumption that we need to interpret our estimates of heterogeneity by participation as causal is that the selection based on the unobserved drivers of potential outcomes is the same across the two treatment groups. Overall, the fact that girls were not informed about the Community Campaigns at the time they were making enrollment decisions makes this highly plausible for heterogeneity analysis by the extensive margin of participation, which we consider as the most theoretically-sound split.

Table 6 shows that the positive impacts of adding the Community Campaigns on girls' mental health (as measured by depression symptoms in Panel A and anxiety symptoms in Panel B and rumination in Panel C) were experienced equally by girls who attended the groups and those who did not. This holds regardless of the attendance threshold: the coefficients on "Community Campaigns" in Columns 1 and 4 show the impact of adding the Community Campaigns on the older and younger girls, respectively, who did not attend any Girl Groups. In Columns 2 and 5 the coefficient shows the effect for girls who attended less than half of the sessions; in Columns 3 and 6 it shows the effect for those who attended less than 75% of the sessions.[37]

This pattern of heterogeneity is most consistent with Community Campaigns altering the day-to-day experiences of *all* adolescent girls in the community in ways that improved their mental health. In particular, the campaigns may have led to broader shifts in community attitudes and behaviors and/or changes in the relationship between girls and their communities.[38] Several pieces of evidence support this interpretation.

First, we find that the combined intervention increased the progressivity of older girls' gender attitudes relative to the control (effect size=0.167 SD, p=0.013). The estimated difference between the two treatment arms is statistically significant for both the younger and older girls although, at around 23% of an SD, the point estimate is slightly larger for the older girls.

[36] As expected, some of these baseline covariates (e.g. school attendance) predict attendance but they do so in the same way between treatment groups.

[37] In Appendix Tables B.22 and B.23 we use the same approach to assess heterogeneity in the additional effects of the Community Campaigns on education and marriage. We find no evidence of such heterogeneity. We note that since we did not see any overall additional effects of the Community Campaigns on education and marriage, we would not expect to see heterogeneous effects here. Any heterogeneous effects would have to be caused by the unlikely scenario that the overall zero effect masks offsetting effects.

[38] Note that one alternative reason why we might see impacts on non-participating girls is if there were direct peer effects in mental health from participating girls. However, for this to explain our findings would require that the net peer effect of participation in the Community Campaigns was *as large* as the direct effect. This is unlikely especially in light of other studies that have examined causal peer effects in adolescent mental health and have found either no (Eisenberg et al. 2013; Zhang 2019) or small (Giulietti et al. 2022) peer effects.

Among the younger girls, the Community Campaigns more than offset the negative impacts of the Girl Groups alone (discussed above). While there is no significant shift in mothers' gender attitudes relative to the control in either arm, we do see some evidence that the addition of the Community Campaigns led to attitudes becoming more progressive.

Second, just as we did for mental health above, and under the same caveats and assumptions, we analyze heterogeneity in the additional impact of the Community Campaigns on gender attitudes by girls' attendance at the Girl Groups (and therefore by whether or not they would have been involved in the campaigns). Table 7 shows that for older girls, younger girls, and mothers the addition of Community Campaigns led to gender attitudes becoming more progressive not only for those who attended the Girl Groups, but also those who did not.

As with the mental health results, we consider peer effects an unlikely explanation for the attitudinal shifts among girls who did not participate in the Community Campaigns since impacts on these non-participant girls are just as large as on participant girls. To the extent that girls' stated attitudes can be interpreted as a reflection of internalized community norms (e.g. McKelway (2022)), this result could be suggestive of wider changes in the attitudes and norms of the wider community in response to the Community Campaigns. The campaigns could lead to norm change through causing community members to change their private beliefs (Finnemore and Sikkink 1998) or serving as an opportunity to learn about the beliefs of others in the community. In particular, recent work, including in India, has highlighted how both men and women, on average, overestimate how conservative others in their community are (Bursztyn et al. 2023) and that learning the true private beliefs of their peers can lead to a change in the community-level norm (Bursztyn et al. 2020). In our context these channels may be interrelated; if individuals change their private beliefs in response to the campaigns, doing so in a public setting where people get signals that others are also changing their private beliefs might be important in allowing these changes to translate into changed norms.

Finally, we note that estimated impacts on the attitudes toward violence of the girls and their mothers also appear consistent with this mechanism (with the caveat that we only pre-registered these attitudes as secondary rather than primary outcomes). In Appendix Table B.24 we see that, if anything, the Girl Groups alone led to (older) girls becoming more reluctant to retaliate against or report incidents of violence. However, the combined intervention of the Girl Groups and Community Campaigns reduced the degree to which the girls reported that the female victims were to blame for the violence they faced. We also see a reduction in victim blaming from caregivers in response to the addition of Community Campaigns (Appendix Table B.25).

Table 7: Heterogeneity in Gender Attitudes by Attendance

	Older Girls			*Younger Girls*			*Mothers*		
Attendance threshold:	any	50%	75%	any	50%	75%	any	50%	75%
Community Campaigns	0.206***	0.207***	0.221***	0.146	0.131*	0.165**	0.128	0.116	0.157**
	(0.071)	(0.067)	(0.067)	(0.098)	(0.079)	(0.074)	(0.078)	(0.071)	(0.071)
Attended	-0.048	-0.054	-0.012	0.007	0.018	0.080	0.048	0.032	0.142**
	(0.065)	(0.080)	(0.100)	(0.090)	(0.075)	(0.069)	(0.054)	(0.059)	(0.069)
Community Campaigns x Attended	0.041	0.066	0.024	0.021	0.068	-0.047	0.004	0.036	-0.166
	(0.096)	(0.108)	(0.169)	(0.122)	(0.112)	(0.128)	(0.080)	(0.085)	(0.104)
N	1,492	1,492	1,492	1,794	1,794	1,794	3,114	3,114	3,114

Notes: The table presents estimated heterogeneous impacts of the addition of "Community Campaigns" over and above the Girl Groups intervention by whether or not the girl attended the Girl Groups. Analysis contains only the two treatment arms and excludes the control group. Attendance was measured using implementation data and re-coded as a binary variable based on three thresholds: (i) any attendance (columns 1, 4, and 7); (ii) 50% attendance (columns 2, 5, and 8); and (iii) 75% attendance (columns 3, 6, and 9). Standard errors (in parentheses) are constructed using a cluster bootstrap accounting for stratification by block and treatment status (1000 iterations). Pre-specified controls are those listed in the notes to Table 4. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

In sum, changes in the progressivity of gender attitudes among girls irrespective of whether they participated in the Community Campaigns provide suggestive evidence that Community Campaigns led to community-wide changes. Such changes can explain the large mental health improvements that we see and why we see them even in girls who were not involved in leading the Community Campaigns.

These patterns are less consistent with the mental health effects being driven by girls finding the act of organizing the campaigns itself empowering. First, if this were the case we would expect to see larger effects on mental health of girls who participated in the campaigns. Second, we find no improvements in girls' self-efficacy, self-esteem or peer relations - outcomes we would expect to respond to the Community Campaigns if the mental health effects were driven by an empowerment boost from running the events (Table 5).

5.3 Why Community Campaigns did not increase education and marriage effects

A natural question arising from our results is why the community wide changes and improvements in girls' mental health in the Girl Groups and Community Campaigns arm did not lead to larger effects on education and marriage. If the Community Campaigns substantially reduced the psychological burden of staying in school and delaying marriage, we might expect more girls to remain in school and unmarried in the Girl Groups and Community Campaigns arm than in the Girl Groups only arm. However, this is not what we see in Table 2. We explore potential explanations here.

The first thing to note is that our design is not powered to detect *small* differences in effect sizes between the two treatment arms. The standard errors on our estimated difference between the treatment arms imply that the minimum detectable effects (size 5%, power 80%) for the difference in schooling and marriage outcomes between arms are 0.05 and 0.03, respectively.[39] In other words, the addition of Community Campaigns would need to almost double the effect size, relative to the Girl Groups alone, in order for us to detect a statistically significant difference.

There are several reasons why such large additional effects may not arise. One possibility is that the mental health benefits from Community Campaigns accrued primarily to girls who were already very likely to remain in school and delay marriage (i.e. girls who were infra-marginal on the schooling and marriage margins). In this case, improvements in mental health would have limited scope to further change schooling and marriage outcomes. To examine this possibility, we analyze treatment effect heterogeneity by predicted school dropout risk. We construct a predicted dropout risk index using a lasso regression of school attendance at endline on baseline covariates and their interactions (Table B.26). The model is estimated on the sub-sample of control group girls who were enrolled at baseline. We then split the sample at the median of predicted dropout risk and estimate treatment effects separately for each group.

Table 8 reports the results. The effects on schooling are concentrated among girls with above-median predicted dropout risk. For these girls, the estimated treatment effect of the Girl Groups is a 16.7 percentage point increase in likelihood of being at school (17.4 percentage points for the combined intervention). For girls with below-median predicted risk, the effects are statistically indistinguishable from zero. Marriage effects are not statistically distinguishable from zero in either risk group, which could reflect a reduction in power from splitting the sample.

In contrast, the mental health improvements from the combined intervention relative to the control are largest among girls with below-median predicted school dropout risk. Depression falls by 0.353 SD for the low-risk group, compared with 0.104 SD (which is not statistically different from zero) for the high-risk group (p-value for the difference between risk groups = 0.030). The same qualitative pattern holds for anxiety, with reductions of 0.301 SD for low-risk girls and 0.163 SD for high-risk girls, though the difference between the two groups is not statistically significant (p-value = 0.187). We see a consistent pattern when we compare estimated effects of adding the Community Campaigns on girls who did in fact remain enrolled in school until endline with those who had left school. This comparison assumes that selection into remaining

[39] We calculate ex-post minimum detectable effects as $2.8 \times$ SE, where the standard error is taken from the estimated difference between the two treatment arms. The factor 2.8 derives from the sum of the critical values for a two-sided test at the 5% significance level and 80% power: $z_{0.975} + z_{0.80} = 1.96 + 0.84 = 2.8$ (Duflo, Glennerster, and Kremer 2007).

in school is the same across the two treatment arms (but may differ from control).[40] Among girls still enrolled, depression falls by 0.318 SD with the addition of the campaigns, compared with 0.161 SD among those who have dropped out (p-value on the difference = 0.060) (Table 8). This pattern is consistent with Community Campaigns reducing the psychological costs of behaviors promoted in the Girl Groups, particularly remaining in school, leading to the largest mental health gains among girls who continued their education. Taken together, the heterogeneity pattern suggests that the addition of the Community Campaigns extends the reach of the program to girls who would not benefit from the Girl Groups alone, at modest additional cost of around USD 6 per eligible girl, or approximately 14% of the cost of the Girl Groups intervention.

These findings suggest that the addition of the Community Campaigns improved mental health most among girls whose education and marriage decisions were least responsive to the Girl Groups program because they were already likely to remain in school. However, it should be noted that Table 8 suggests that high-risk girls did experience some, albeit much smaller, improvements to their mental health from the Community Campaigns. These improvements did not translate into additional changes in schooling decisions large enough to be detected at this study's power. One explanation is that the mental health gains are not large enough to shift behavior at the margin, especially if schooling decisions depend on many factors beyond psychological costs. Schooling decisions are known to depend on perceived economic returns in labor and marriage markets, expectations about future well-being, parental preferences, bargaining power within the household, school quality, peer effects, direct costs of attendance, and academic progression requirements. Even if Community Campaigns meaningfully reduced the psychological cost of schooling, this may only have proved marginal for a small subset of high-dropout-risk girls.[41]

[40] We show that treatment effects on schooling are similar across arms (Table 2), including across baseline-predicted dropout-risk strata (Table 8). However, equality of selection on unobservables remains an assumption.

[41] It is also worth noting that educational choices are often made at discrete points, particularly at the transition between school years, while mental health effects may accumulate more gradually over the course of the intervention. Our study timeline implies that many of the educational choices may have been "locked in" midway through the intervention while the mental health effects may have accumulated throughout. Our baseline was completed in March 2016, with the intervention running until September 2017 and the new academic year beginning in June 2017—at which point many schooling decisions would have been effectively locked in. If mental health benefits from community engagement crystallized primarily toward the end of the intervention, when Community Campaigns had covered more nuanced material and accumulated sufficient community exposure, they may not have been timed to influence critical educational decision points.

Table 8: Treatment Effects by Predicted and Realized Dropout

	School	Married/Engaged	Depression		Anxiety		Rumination	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Girl Groups								
× High dropout risk	0.167***	-0.029	0.032		-0.024		0.115	
	(0.043)	(0.035)	(0.073)		(0.071)		(0.127)	
× Low dropout risk	0.015	-0.004	0.035		0.000		0.100	
	(0.041)	(0.031)	(0.124)		(0.112)		(0.098)	
Girl Groups + Community								
× High dropout risk	0.174***	-0.046	-0.104		-0.163**		-0.185	
	(0.040)	(0.036)	(0.087)		(0.078)		(0.137)	
× Low dropout risk	0.007	-0.028	-0.353***		-0.301***		0.023	
	(0.039)	(0.031)	(0.100)		(0.093)		(0.115)	
Difference (GG+C − GG)								
× High dropout risk	0.008	-0.017	-0.137*		-0.139*		-0.300**	
	(0.043)	(0.037)	(0.080)		(0.073)		(0.126)	
× Low dropout risk	-0.008	-0.023	-0.388***		-0.301***		-0.077	
	(0.044)	(0.030)	(0.106)		(0.099)		(0.118)	
× Dropped out				-0.161**		-0.127**		-0.199*
				(0.068)		(0.058)		(0.100)
× Stayed in school				-0.318***		-0.255***		-0.139
				(0.088)		(0.078)		(0.145)
P-value: GG High = Low	0.003	0.565	0.980		0.848		0.905	
P-value: GG+C High = Low	0.003	0.681	0.030		0.187		0.134	
P-value: Diff High = Low	0.770	0.868	0.035	.	0.191	.	0.057	.
P-value: Dropout = Stayer				0.060		0.151		0.607
Control mean (high risk)	0.231	0.294	-0.070		-0.070		-0.000	
Control mean (low risk)	0.672	0.151	0.023		0.012		-0.024	
GG mean (dropouts)				-0.000		-0.046		0.153
GG mean (stayers)				0.026		-0.021		0.126
Control group included in sample	Y	Y	Y	N	Y	N	Y	N
Observations	1676	1676	1514	1492	1514	1492	1514	1492

Notes: The table presents estimated heterogeneous impacts of "Girl Groups" and "Girl Groups + Community Campaigns" interventions for older girls. Columns 1, 2, 3, 5, and 7 report heterogeneity by predicted dropout risk among girls who were enrolled in school at baseline, estimated relative to the pure control group; columns 4, 6, and 8 report heterogeneity by realised endline school status for older girls in the two treatment arms. Predicted dropout risk is constructed from a lasso regression of endline school attendance on baseline covariates, estimated on the control group sample, with girls then split at the median into above-median ("High dropout risk") and below-median ("Low dropout risk") groups. Columns 4, 6, and 8 compare the two treatment arms among girls who remained in school ("Stayed in school") and those who dropped out ("Dropped out"). "Difference (GG+C − GG)" refers to the additional effect of the Community Campaigns over and above the Girl Groups intervention. "P-value" rows report tests of equality of coefficients across groups. Standard errors (in parentheses) are clustered at the cluster level. Pre-specified controls are as described in the notes to Tables 2 and 3. GG = Girl Groups; GG+C = Girl Groups + Community. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

6 Conclusions

In this paper, we used a randomized field experiment in order to compare two approaches to designing interventions that aim to improve the well-being of adolescent girls living in rural Rajasthan, a context with especially conservative gender norms. Like many programs, the first approach engaged only with adolescent girls themselves. It encouraged them to question prevailing gender norms and ideas and, in spite of these, to stay in school and delay marriage. The second additionally provided a platform and structure for girls to engage with their broader community, including the community leaders who have the greatest influence in setting and enforcing gender norms, and to involve them in the process of change.

We showed that while both approaches resulted in a significant and substantial reduction in school dropout and early marriage among the older girls, these were only accompanied by an improvement in girls' mental health in the program that also engaged the community in addition to the girls themselves. This approach led to a significant reduction in depression (0.16 SD) and anxiety (0.17 SD), especially among older girls who were at an age when norms become more constraining. In contrast, engaging only with adolescent girls themselves led to an *increase* in a ruminative thinking style and had no impact on symptoms of depression and anxiety.

Furthermore, we find that the mental health improvements that stemmed from the addition of the community engagement were accompanied by an increase in the progressivity of girls' attitudes towards prevailing norms. These impacts on girls' attitudes and mental health were present both for girls who attended the Girl Groups and those who did not. This provides suggestive evidence that the community engagement impacted mental health through changing the social environment at the community level. Heterogeneity analysis further shows that while Girl Groups primarily affected girls at highest risk of school dropout (and thus early marriage), Community Campaigns extend the benefits of the intervention to lower-risk girls, who were likely to have been infra-marginal on the schooling decision, through improvements in mental health.

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 mental health associated with increased schooling or delayed marriage may be offset by the stress or backlash arising from breaking the community-level norms required to achieve these or by an increased discrepancy between aspirations and what can realistically be achieved.

The latter point may be particularly important for explaining the increase in rumination when girls were targeted alone. Our evidence suggests these costs can be reduced through directly engaging the broader community. Involving the broader community in the process of change may be an effective add-on for many types of empowerment interventions targeting different marginalized groups. The addition of the community engagement events organized by members of the marginalized group was relatively low cost and straightforward to organize yet it yielded improvements in mental health that are comparable to more intensive and targeted interventions.

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*.

Amaral, S., A. Garcia-Ramos, S. Gulesci, S. Oré, A. Ramos, and M. M. Sviatschi (2024). Gender-based violence in schools and girls' education: Experimental evidence from mozambique.

Andrew, A. and A. Adams (2025). Revealed beliefs and the marriage market return to education. *The Quarterly Journal of Economics 140*(3), 2107–2162.

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*.

Angelucci, M. and D. Bennett (2024). The economic impact of depression treatment in India: Evidence from community-based provision of pharmacotherapy. *American Economic*

Review 114(1), 169–198.

Arenas-Arroyo, E., D. Fernandez-Kranz, and N. Nollenberger (2021). Intimate partner violence under forced cohabitation and economic stress: Evidence from the COVID-19 pandemic. *Journal of Public Economics 194*, 104350.

Armand, A., O. Attanasio, P. Carneiro, and V. Lechene (2020). The effect of gender-targeted conditional cash transfers on household expenditures: Evidence from a randomized experiment. *The Economic Journal 130*(631), 1875–1897.

Ashraf, N., N. Bau, C. Low, and K. McGinn (2020, 5). Negotiating a better future: How interpersonal skills facilitate intergenerational investment*. *The Quarterly Journal of Economics 135*, 1095–1151.

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.

Asiedu, E., D. Karlan, M. Lambon-Quayefio, and C. Udry (2021). A call for structured ethics appendices in social science papers. *Proceedings of the National Academy of Sciences 118*(29), e2024570118.

Baird, S., Z. A. Bhutta, B. A. Hamad, J. H. Hicks, N. Jones, and J. Muz (2019). Do restrictive gender attitudes and norms influence physical and mental health during very young adolescence? evidence from Bangladesh and Ethiopia. *SSM-population health 9*, 100480.

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*.

Baranov, V., S. Bhalotra, P. Biroli, and J. Maselko (2020). Maternal depression, women's empowerment, and parental investment: Evidence from a randomized controlled trial. *American Economic Review 110*(3), 824–859.

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

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*.

Bhalotra, S., U. Kambhampati, S. Rawlings, and Z. Siddique (2021). Intimate partner violence: The influence of job opportunities for men and women. *The World Bank Economic Review 35*(2), 461–479.

Boulhane, O., C. Boxho, D. Kanga, E. Koussoube, and L. Rouanet (2024). Empowering adolescent girls through safe spaces and accompanying measures in Cote d'Ivoire. *World Bank Working Paper 10721*.

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. Cappelen, B. Tungodden, A. Voena, and D. Yanagizawa-Drott (2023, 3). How are gender norms perceived?

Bursztyn, L., A. L. González, and D. Yanagizawa-Drott (2020). Misperceived social norms: Women working outside the home in Saudi Arabia. *American Economic Review 110*(10), 2997–3029.

Currie, J. and M. Stabile (2007). Mental health in childhood and human capital. In *The problems of disadvantaged youth: An economic perspective*, pp. 115–148. University of Chicago Press.

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.

De Mel, S., D. McKenzie, and C. Woodruff (2014). Business training and female enterprise start-up, growth, and dynamics: Experimental evidence from Sri Lanka. *Journal of Development Economics 106*, 199–210.

Dhar, D., T. Jain, and S. Jayachandran (2022). Reshaping adolescents' gender attitudes: Evidence from a school-based experiment in India. *American Economic Review 112*(3),

899–927.

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

Duflo, E., R. Glennerster, and M. Kremer (2007). Using randomization in development economics research: A toolkit. *Handbook of development economics 4*, 3895–3962.

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

Eisenberg, D., E. Golberstein, J. L. Whitlock, and M. F. Downs (2013). Social contagion of mental health: evidence from college roommates. *Health economics 22*(8), 965–986.

Fernández, R. (2013, February). Cultural change as learning: The evolution of female labor force participation over a century. *American Economic Review 103*(1), 472–500.

Field, E., R. Pande, N. Rigol, S. Schaner, and C. Troyer Moore (2021, July). On her own account: How strengthening women's financial control impacts labor supply and gender norms. *American Economic Review 111*(7), 2342–75.

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.

Giulietti, C., M. Vlassopoulos, and Y. Zenou (2022). Peers, gender, and long-term depression. *European Economic Review 144*, 104084.

Guarnieri, E. and H. Rainer (2021). Colonialism and female empowerment: A two-sided legacy. *Journal of Development Economics 151*, 102666.

Hatzenbuehler, M. L. (2009). How does sexual minority stigma "get under the skin"? a psychological mediation framework. *Psychological Bulletin 135*(5), 707–730.

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

Hidrobo, M., A. Peterman, and L. Heise (2016). The effect of cash, vouchers, and food transfers on intimate partner violence: evidence from a randomized experiment in Northern Ecuador. *American Economic Journal: Applied Economics 8*(3), 284–303.

Jayachandran, S. (2019). Social Norms as a Barrier to Women's Employment in Developing Countries. *Working Paper*.

Jensen, R. and R. Thornton (2003). Early female marriage in the developing world. *Gender & Development 11*(2), 9–19.

Kieling, C., H. Baker-Henningham, M. Belfer, G. Conti, I. Ertem, O. Omigbodun, L. A. Rohde, S. Srinath, N. Ulkuer, and A. Rahman (2011). Child and adolescent mental health worldwide: evidence for action. *The Lancet 378*(9801), 1515–1525.

Kishor, S. and K. Gupta (2004). Women's empowerment in India and its states: evidence from the NFHS. *Economic and Political weekly*, 694–712.

Klasen, H. and A.-C. Crombag (2013). What works where? a systematic review of child and adolescent mental health interventions for low and middle income countries. *Social psychiatry and psychiatric epidemiology 48*, 595–611.

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. Löwe (2006). A brief measure for assessing generalized anxiety disorder: the GAD-7. *Archives of Internal Medicine 166*(10), 1092–1097.

Kumar, N., K. Raghunathan, A. Arrieta, A. Jilani, and S. Pandey (2021). The power of the collective empowers women: Evidence from self-help groups in India. *World Development 146*, 105579.

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.

Lund, C., K. Orkin, M. Witte, T. Davies, J. Haushofer, J. Bass, and V. Patel (2020). Economic impacts of mental health interventions in low and middle-income countries: a systematic review and meta-analysis.

Martin, L. L. and A. Tesser (1996). Some ruminative thoughts. In *Ruminative thoughts*, pp. 1–47. Psychology Press.

McKelway, M. (2018). Women's self-efficacy and women's employment: Experimental evidence from India. *Working Paper*.

McKelway, M. (2022). Women's employment and empowerment: Descriptive evidence. In

AEA Papers and Proceedings, Volume 112, pp. 541–545. American Economic Association 2014 Broadway, Suite 305, Nashville, TN 37203.

Mehmood, B. S., S. Naseer, and D. L. Chen (2022). Why are rights revolutions rare?

Mehra, D., T. Lakiang, N. Kathuria, M. Kumar, S. Mehra, and S. Sharma (2022). Mental health interventions among adolescents in India: a scoping review. In *Healthcare*, Volume 10, pp. 337. MDPI.

Meyer, I. H. (2003). Prejudice, social stress, and mental health in lesbian, gay, and bisexual populations: conceptual issues and research evidence. *Psychological Bulletin 129*(5), 674–697.

Patel, V., A. J. Flisher, S. Hetrick, and P. McGorry (2007). Mental health of young people: a global public-health challenge. *The lancet 369*(9569), 1302–1313.

Patel, V., C. Ramasundarahettige, L. Vijayakumar, J. Thakur, V. Gajalakshmi, G. Gururaj, W. Suraweera, and P. Jha (2012). Suicide mortality in India: a nationally representative survey. *The lancet 379*(9834), 2343–2351.

Patel, V., B. Weobong, H. A. Weiss, A. Anand, B. Bhat, B. Katti, S. Dimidjian, R. Araya, S. D. Hollon, M. King, et al. (2017). The healthy activity program (hap), a lay counsellor-delivered brief psychological treatment for severe depression, in primary care in India: a randomised controlled trial. *The Lancet 389*(10065), 176–185.

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.

Ridley, M., G. Rao, F. Schilbach, and V. Patel (2020). Poverty, depression, and anxiety: Causal evidence and mechanisms. *Science 370*(6522), eaay0214.

Roy, S., M. Hidrobo, J. Hoddinott, and A. Ahmed (2019). Transfers, behavior change communication, and intimate partner violence: Postprogram evidence from rural Bangladesh. *Review of Economics and Statistics 101*(5), 865–877.

Shah, M., S. Baird, J. Seager, B. Avuwadah, J. Hamory, S. Sabarwal, and A. Vyas (2024). Improving mental health of adolescent girls in low-and middle-income countries. *The Journal of Human Resources 59*(S), S317–S364.

Singla, D. R., B. A. Kohrt, L. K. Murray, A. Anand, B. F. Chorpita, and V. Patel (2017). Psychological treatments for the world: lessons from low-and middle-income countries. *Annual review of clinical psychology 13*, 149–181.

Treynor, W., R. Gonzalez, and S. Nolen-Hoeksema (2003). Rumination reconsidered: A psychometric analysis. *Cognitive Therapy and Research 27*(3), 247–259.

Watkins, E. R. (2008). Constructive and unconstructive repetitive thought. *Psychological bulletin 134*(2), 163.

Watkins, E. R. and S. Nolen-Hoeksema (2014). A habit-goal framework of depressive rumination. *Journal of abnormal psychology 123*(1), 24.

Zhang, A. (2019). Peer effects on mental health: Evidence from random assignment into classrooms. *SSRN Working Paper*.

A Ethics Appendix

We executed a three-armed cluster randomized controlled trial covering adolescent girls aged 12 to 17 years old in rural Rajasthan. Details of the two interventions are provided in Section 2.1. In this appendix, we discuss key ethical considerations of this research, grouping them into 7 categories proposed by Asiedu et al. (2021). For each category, we first set out the questions posed by Asiedu et al. (2021) before discussing our answers.

A.1 Policy equipoise and scarcity.

"Is There Policy Equipoise? That Is, Is There Uncertainty Regarding Participants' Net Benefits from Each Arm of the Study Relative to the Other Arms and to the Best Possible Policy to Which Participants Could Have Access? If Not, Ethical Randomization Requires Two Conditions Related to Scarcity: 1) Was There Scarcity, i.e., Did the Inclusion of Multiple Arms Change the Expected Aggregate Value of the Programs Delivered, and 2) Do All Ex-ante Identifiable Participants Have Equal Moral or Legal Claims to the Scarce Programs?"

Asiedu et al. (2021) make the case that ethical randomization of participants to different treatment conditions requires either: (1) policy equipoise (i.e. that there is "uncertainty regarding participants' net benefits from each arm of the study relative to the other arms and [relative] to the best possible policy which participants could have access") or (2) scarcity (i.e. that "no participant can be predicted to be worse off in any arm of the study than under counterfactual policy", that there is "scarcity of the resources required for the arms in which participants are better off" and that "all ex-ante identifiable participants have equal moral or legal claims to the scarce programs"). We consider that in our case, both conditions (1) and (2) are met for the reasons outlined below.

1. **Policy equipoise.** After an examination of the prior evidence, we considered that there was no expert consensus that one treatment (or control) arm dominated the others when considering participants' well-being overall. As discussed in the introduction of the paper, there is *some* evidence that interventions similar to our "girl group only" intervention that work with adolescent girls in group settings *can* have positive impacts on education (Edmonds et al. 2021; Buchmann et al. 2017; Adoho et al. 2014; Buehren et al. 2016; Bandiera et al. 2019). So while there was encouraging evidence that such approaches may be effective, there are two reasons why there was no consensus in the literature. First, many of the previous studies have evaluated girl group interventions "bundled" with other components (such as providing girls with marketable skill, financial resources etc) (Adoho

et al. 2014; Buehren et al. 2016; Bandiera et al. 2019). This means there was considerable uncertainty over whether simply running educational groups alone could be useful. No study that we know of considered the additional impact of involving the community. The uncertainty in this literature is summarized by the fact that a comprehensive literature review conducted by the World Bank concluded that the state of the evidence on the impact of "life skills training, mentoring and empowerment programs via Girls Clubs" is "mixed" on education, marriage, and fertility and does not consider impacts on mental health or broader aspects of well-being.

Second, as discussed at length in our paper, the impact of intervening with a marginalized group, such as adolescent girls, on potentially contentious issues on girls' broader well-being is far from obvious. On the one hand, past evidence is suggestive that such interventions may increase girls' schooling which may be beneficial both for their long-term material wellbeing and empowerment and their short-term mental wellbeing. On the other hand, we were very aware that interventions that introduced girls to ideas that were not broadly accepted in their communities or which pushed them to take actions (e.g. attending school) that violated accepted norms may put them in a stressful situation or expose them to sanctions from their broader community. Likewise, we didn't know ex-ante whether involving the broader community in the second treatment arm would be beneficial for girls' wellbeing (e.g. if the community became more supportive) or harmful (e.g. by drawing attention to adolescent girls' activities). No related papers that we know of have evaluated impacts of similar programs on mental health so we had little evidence to guide us here ex-ante and hence contributing to this evidence on the impacts of such programs on broader measures of wellbeing was a top priority for us in designing our study.

2. **Scarcity.** As discussed above, we consider that we conducted this study against a backdrop of policy equipoise. Even if there were pre-existing evidence that the Girl Groups were beneficial or that the addition of the community component was beneficial, both of these elements cost additional resources. We consider it implausible that without further rigorous evidence on the effectiveness of these programs either would be considered as a route for policy.

   In this context of scarcity, we also note that communities were randomly allocated to treatment and control and that, although we randomly selected some adolescent girls to be part of the surveyed sample, *all* adolescent girls in the community were invited to participate in the intervention. Therefore, all participants had ex-ante equal claims to the

programs.

A.2 Role of the researcher

Are Researchers "Active" Researchers, i.e., Did the Researchers Have Direct Decision-Making Power over Whether and How to Implement the Program? If Yes, What Was the Disclosure to Participants and Informed Consent Process for Participation in the Program? Providing IRB Approval Details May Be Sufficient but Further Clarification of Any Important Issues Should Be Discussed Here. If No, i.e., Implementation Was Separate, Explain the Separation.

The researchers in the team were active researchers, meaning that the researchers secured funding, hired staff and made decisions regarding the design of the intervention. Pradan, an NGO that had worked in the area for over a decade, was in charge of the implementation of the intervention, and the researchers had decision-making power regarding changes in the implementation of the intervention.

Intervention design was led by researchers based at ICRW-India and supported by researchers based at IFS (London). In designing the intervention, researchers evaluated potential impacts or harms, as well as its ethical validity, and took actions to minimize risks of harm (see next point).

The involvement of all three organizations in the program was disclosed to participants. We took informed consent/assent (assent from girls below 18 and consent from their parent) from each participant prior to their involvement in any element of our study – formative research (inc. focus groups), baseline survey, participation in intervention, endline survey. At each stage, we explained to participants that the study was a collaboration between ICRW, IFS and Pradan and that "we" (ICRW, IFS, Pradan) were the ones running the program and the surveys. All consent/assent forms received ethical approval from the International Center for Research on Women Institutional Review Board, Sigma Institutional Review Board in New Delhi and Oxford University. Moreover, the project was reviewed by the Federal Ministry of Health and Family Affairs in India and the board of the Children's Investment Fund Foundation.

A.3 Potential harms to participants and non participants

Does the Intervention, Policy, or Product Being Studied Pose Potential Harm to Participants or Nonparticipants? Related to This, Are Participants or Likely Affected Nonparticipants Particularly Vulnerable? Also Related to This, Are Participants' Access to Future Services or Policies Changed Because of Participation in the Study? If the Answer Is Yes to Any of the above, What Is Being Done to Mitigate Such Risks?

As discussed in our discussion of policy equipoise above, we undertook this study against a backdrop of uncertainty about the benefits and potential unintended consequences of our two different treatment arms relative to one another and relative to the control group. In particular, promising existing evidence had suggested that participating in Girl Groups similar to the ones in these interventions could lead to girls staying on in school longer, an outcome that brings with it a multitude of longer-term benefits for wages, decision making and health to name a few. We hypothesized that introducing the community component in the second treatment arm could further strengthen such positive impacts on education leading to further benefits. At the same time, we thought carefully about the risks the two interventions might pose to participants and non-participants. Below, beginning with the Girl Groups intervention and then moving onto the addition of the community engagement and then finally participation in data collection, we list these risks and the steps we took to mitigate them.

Potential harms and mitigation from Girl Groups intervention

- Discomfort, stress or sanctions: We considered that it was possible that girls may have found participating in the Girl Groups uncomfortable, stressful or faced sanctions for their participation. These harms may have arisen directly from (and even during) the group sessions or indirectly from their community's reaction.
- The sessions often discussed sensitive topics and so a potential harm was that private information that girls had shared in the group would be shared more broadly
- It is possible that the ideas that girls were exposed to in the group sessions could lead them to take actions or express views in their families and communities that might have put them in opposition to others. For instance, girls may have faced backlash if they expressed some of the ways of thinking about gender that they discussed with peers in the sessions.
- Physical injuries during sports activities

Mitigation:

- We designed the curriculum and training for the program staff in a way that was mindful of these potential risks and that tried to minimize them. For instance, the interventions in no way encouraged girls to enter into conflict with family or community members. Many of the activities involved thinking about negotiation and communication as approaches girls could take to further their interests. For instance, often activities would involve role-playing discussions with parents or family members about decisions such as schooling.
- The mentors who ran the Girl Groups sessions were given a total of 28 full days of training

over the intervention. They were also supported through frequent mentoring visits from program staff and from designated older women – Pankh sakhis – from their own village. In addition to training on the specific program content, much of the training focused on creating productive and supportive group dynamics. The training was adapted to the needs of the mentors and when, for example, some mentors reported feeling uncomfortable discussing sensitive topics in their groups, a further two-day capacity building workshop was organized.

- We organized sessions so as to minimize creating friction or conflict for girls. For example, initially girls and caregivers often reported that sessions held on weekday evenings were not convenient as girls had to help in the home after school. In response to this, many groups switched to holding sessions on Sundays. It was made clear to girls and parents that girls could miss sessions or stop attending if they wished to. No pressure was put on girls to attend. Overall, as we show in Table B.5, attendance was relatively high (49% attended at least one session, and subsequent dropout was low).
- Pradan, alongside the mentor(s) and Pankh sakhi ran community information sessions about the Girl Groups to explain the aims and to seek consent from the community at large. In addition to promoting support for the program, these sessions were intended to encourage community members to address any concerns about the program directly to Pradan rather than to the participants themselves.
- We took informed consent from all parents and assent from all girls for participation in the Girl Groups.
- The sessions were mainly held in the community centres or in closed spaces to ensure a private and safe environment. Most of the first session was dedicated to groups establishing a set of "ground rules" that defined how they would work together. Core principles, including maintaining confidentiality and being sensitive of people with different viewpoints, were repeatedly stressed by the mentors. These principles were reinforced throughout the entire intervention and were designed to ensure the groups felt like a safe space for girls.
- There were clear protocols for how mentors and other program staff should feed back any worries, concerns or incidents to the central team. All mentors were regularly in contact with the intervention managers over WhatsApp and when any issue arose the intervention managers endeavored to visit the community in question rapidly.
- The mentors received first aid training and a first aid box was provided for emergencies. The research established protocols to refer the participants to the Primary Health Center if needed.

Potential harms and mitigation from the addition of the Community Engagement intervention

- Explicitly involving the community could make any controversial subjects raised by the intervention more salient. This could lead to families withdrawing girls from the Girl Groups (we didn't see this in practice as attendance was the same across treatment arms). It could have plausibly led to increased conflict within the community.
- Girls may have found performing in front of their community difficult and stressful. Note that in practice we saw the opposite as mental health improved.

Mitigation:

- The community engagement sessions were designed by the Girl Groups in collaboration with the mentors. Girls were never pressurized to include content that they were not comfortable sharing with the broader community. Indeed, while girls frequently discussed sensitive personal issues during the Girl Group sessions, the material they presented during the community engagement was typically fictionalized and more abstract.
- The Pankh sakhis took charge of facilitating the community discussion during the community sessions. This is because these older women had a higher status within the community than either the girls or the mentors and generally felt much more comfortable taking on this role.

Potential harms and mitigation from data collection:

- A violation of participants' privacy and confidentiality via a data breach
- Participants' responses to sensitive survey questions being overheard by family or community members

Mitigation:

- We followed a strict protocol to ensure data protection. During baseline and endline, the interviewers used a password-protected tablet. The data collected every day was downloaded at ICRW Delhi on a password-protected computer. Once the survey was completed, identifiers and records of consent/assent were separated from the main dataset after giving a unique ID to each respondent. Considering that we followed the same respondents for endline, we kept the identifiers linked to a unique ID on a password-protected computer and on a password-protected hard drive. For the analysis, we only used the data set without identifiers. The focus group sessions were organized in places that provided enough privacy for participants to feel comfortable engaging in the discussions. The moderator emphasized the need to keep the conversation limited to the group and that their names and personal information were not being recorded. The focus group sessions had a note taker but did not record any identifier. The notes were kept separate from consent and

assent forms in a locked cabinet.

- Interviewers always tried to conduct the interviews with only the respondent present and received training on how to facilitate this. Where there were other people present, e.g. due to household members wandering in and out of a one-roomed dwelling, the interviewer always explicitly paused and asked the respondent if they were happy to continue. We never asked sensitive questions (e.g. about contraceptives, attitudes to violence, savings) in front of other individuals. Instead, interviewers were trained how to leave sections blank if another person was present and return to them once they were alone with the respondent.

A.4 Conflicts of interest and intellectual freedom

The researchers did not have financial conflicts of interest with regard to the results of the research, nor did they have potential reputational conflicts of interest. Moreover, the researchers had intellectual freedom over the research. In particular, we had permission to name the implementation partner (Pradan). We discussed findings at length with the team at Pradan but there was no commitment to incorporating their views into the research nor was there any agreement whereby they (or the funders) could see the results before others.

A.5 Feedback to participants

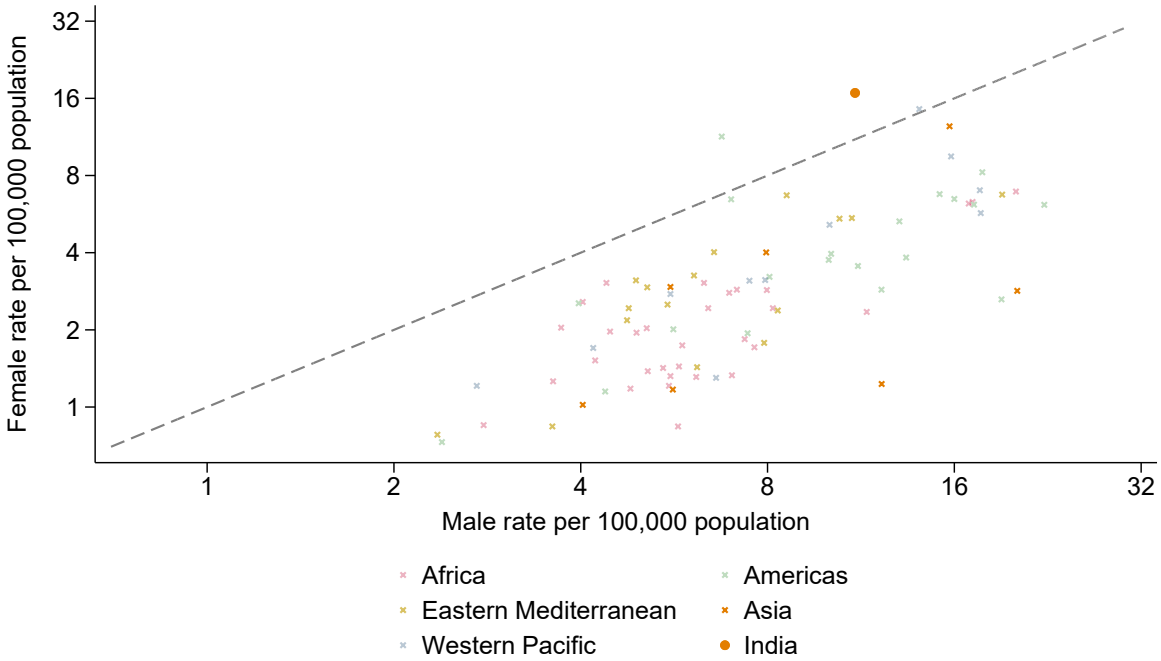
We did not feed back findings directly to individual participants about the research results. This choice was motivated by the wish to continue following the sample over the longer run. Nevertheless, we did complete feedback activities in the district of Dholpur in two ways. First, we organized dissemination events at the district level, including elected representatives, members of community-based organizations, frontline workers, and religious and other community leaders. We also organized smaller-group meetings with district officials as well as running events at the state and national level. The objective of these events was to share the findings of the research. To specifically involve adolescent girls and young women from the district in the discussion of the findings, in 2022 we formed a Youth Advisory Group comprising young women of the same age as the study participants from the study area. This group met three times with members of the research team to discuss findings as well as to brainstorm future research.

A.6 Foreseeable misuse of research results

We do not see foreseeable or plausible risks that the results will be misused. The main concern could be related to generalizing the results of this research to other contexts. However, we have clearly explained the limitations regarding external validity.

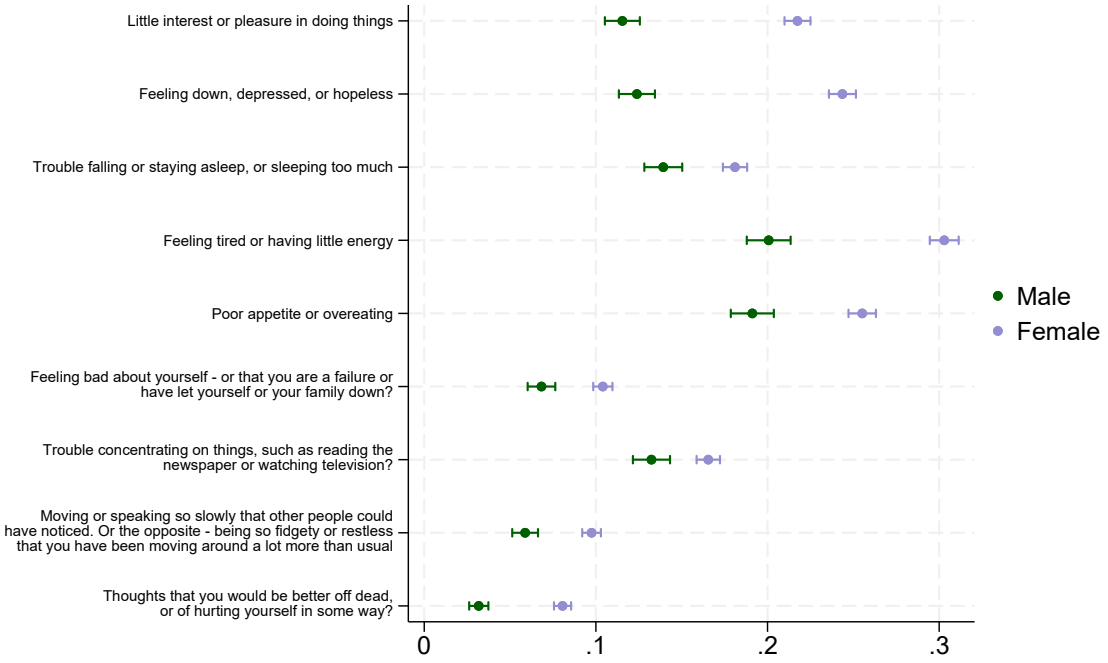
B Additional Figures and Tables

Figure B.1: Male vs. Female 15-24 Year Old Suicide Rates by Country



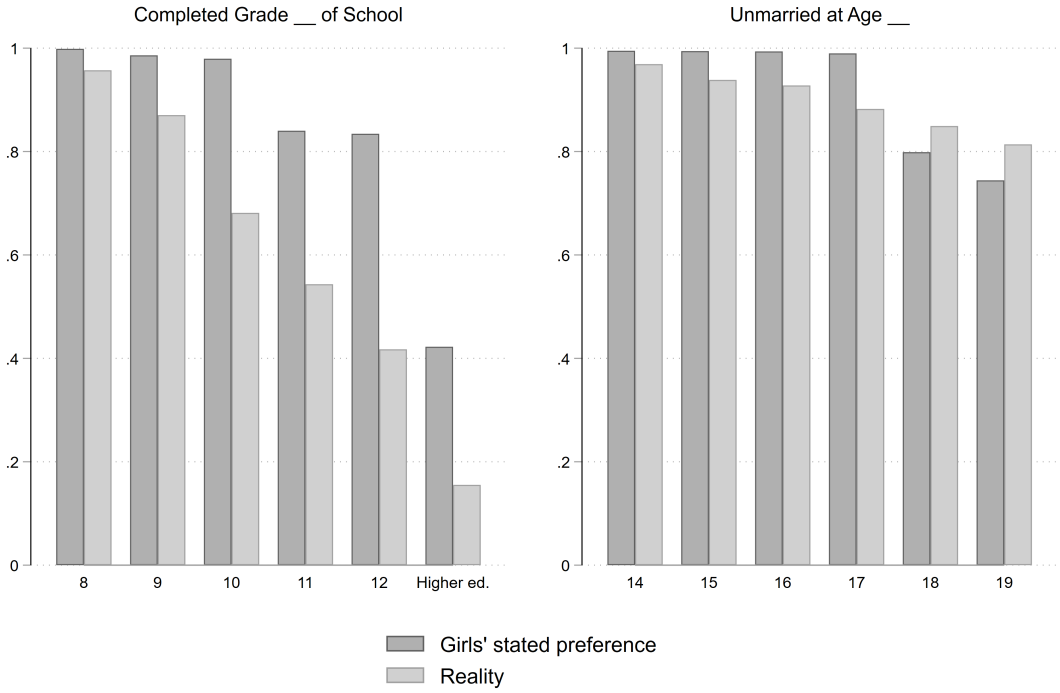
Notes: Figure plots the male vs. female suicide rates for young people aged 15 to 24 by country. Only countries with a population of over 5 million people are included in the figure.

Figure B.2: Proportion of Adolescent Girls vs. Boys Reporting Symptoms of Depression



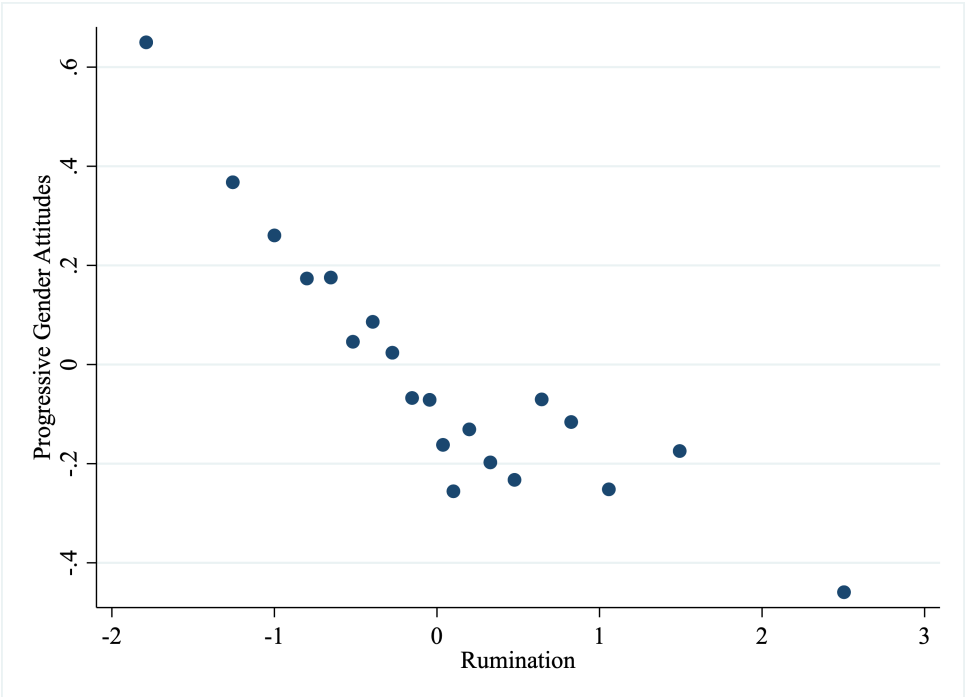
Notes: Figure plots the proportion of adolescent girls vs boys in the Uttar Pradesh Youth Survey reporting that they had experienced each of the 9 indicators of depression measured by the PHQ-9 during the last two weeks. Data source: Santhya, K.G. 2020. "UDAYA, Adolescent Survey, Bihar and Uttar Pradesh, 2018–19." Harvard Dataverse, V2.

Figure B.3: Girls' Stated Aspirations versus Reality in Education and Marriage



Notes: Figure shows the gap between aspirations measured at baseline and reality at endline among 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 B.4: Binscatter between Progressive Gender Attitudes and Rumination



B.1 Additional Tables

Table B.1: Total Costs and Costs per Eligible Girl, in 2024 US$

	Total Costs			Costs per Eligible Girl	
	Girl Groups	**Girl Groups + Community Campaigns**	**Total**	**Girl Groups**	**Girl Groups + Community Campaigns**
Administration	7,136.42	7,136.42	14,272.85	2.33	2.22
Staff training and recruitment	40,361.64	43,288.32	83,649.96	13.21	13.50
Implementation	80,506.00	103,941.12	184,447.13	26.34	32.41
Monitoring	9,487.72	9,487.72	18,975.43	3.11	2.96
Total	**137,491.79**	**163,853.58**	**301,345.37**	**44.99**	**51.09**

Notes: The table presents the total costs "Girl Groups" and "Girl Groups + Community Campaigns" interventions in Columns 1 and 2. Columns 4 and 5 report the costs per eligible girl, using as a reference the universe of unmarried girls aged 12 to 17 years old identified in the baseline census: 3,056 in the Girl Groups and 3,207 "Girl Groups + Community Campaigns". Program administration includes the salaries of key personnel supervising the program. Staff training and recruitment cover the costs associated with training delivery. Implementation costs include all expenses related to running the intervention, including staff salaries, transportation, materials, venue hire, and food. Monitoring costs include expenses incurred in visiting villages to monitor the program and collect data. The Table excludes costs related to evaluation and research activities (e.g., listing, baseline and endline data collection, and social mapping), as these are not required for program replication. All the costs were converted into US$ and are presented in US$ 2024: they account for time preferences (using a base year of 2016 and a 10% discount rate) and inflation (using average Gross Domestic Product (GDP) deflators).

Table B.2: Predictors of Attrition

	(1) Main Sample	(2) Main Sample	(3) Extended Sample	(4) Extended Sample
Girl Groups	-0.001	-0.000	-0.008	-0.008
	(0.012)	(0.012)	(0.007)	(0.006)
Girl Groups + Community	-0.005	-0.003	-0.003	-0.001
	(0.014)	(0.014)	(0.008)	(0.008)
Age=13		0.016		0.015**
		(0.013)		(0.007)
Age=14		-0.016		0.008
		(0.015)		(0.009)
Age=15		-0.014		0.015*
		(0.015)		(0.008)
Age=16		-0.032**		0.013
		(0.014)		(0.008)
Age=17		-0.060***		0.007
		(0.015)		(0.009)
Carer's Years of Education		-0.005**		-0.001
		(0.002)		(0.001)
SC/ST Caste		-0.021*		-0.016***
		(0.011)		(0.006)
Asset Index		0.016***		0.003
		(0.006)		(0.004)
Attending School at Baseline		0.057***		0.023***
		(0.012)		(0.008)
Progressive Gender Attitudes		0.001		0.000
		(0.001)		(0.000)
Carer's Progressive Gender Attitudes		-0.001		-0.001
		(0.001)		(0.000)
Mental Health		-0.004**		-0.001
		(0.002)		(0.001)
Constant	0.882***	0.886***	0.968***	0.967***
	(0.006)	(0.034)	(0.004)	(0.017)
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 B.3: Baseline Characteristics and Balance (Full Baseline Sample)

	Younger Girls				*Older Girls*			
	Control	Girl Groups	Girl Groups + Community	Difference	Control	Girl Groups	Girl Groups + Community	Difference
Girl Age	12.929	12.893	12.934		15.857	15.887	15.889	
	(0.859)	(0.852)	(0.854)		(0.793)	(0.810)	(0.808)	
		[p=0.361]	[p=0.902]	[p=0.301]		[p=0.393]	[p=0.342]	[p=0.950]
Carer's Years of Ed.	1.438	1.056	1.228		1.231	0.797*	0.894*	
	(2.867)	(2.548)	(2.570)		(2.735)	(2.218)	(2.218)	
		[p=0.131]	[p=0.342]	[p=0.458]		[p=0.060]	[p=0.085]	[p=0.638]
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=0.842]	[p=0.169]	[p=0.284]		[p=0.471]	[p=0.154]	[p=0.479]
Asset Index	-0.093	-0.151	-0.124		0.100	0.002	0.003	
	(0.849)	(0.798)	(0.806)		(0.976)	(0.883)	(0.905)	
		[p=0.413]	[p=0.659]	[p=0.645]		[p=0.150]	[p=0.173]	[p=0.998]
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=0.526]	[p=0.001]	[p=0.049]		[p=0.335]	[p=0.335]	[p=0.965]
Progressive Gender Attitudes	34.624	34.018	34.593		35.471	34.324*	35.445	
	(6.945)	(6.810)	(6.860)		(6.764)	(6.859)	(7.029)	
		[p=0.328]	[p=0.956]	[p=0.289]		[p=0.074]	[p=0.966]	[p=0.050]
Carer's Progressive Gender Attitudes	34.030	33.229	33.929		34.014	33.253	33.734	
	(6.621)	(6.472)	(6.598)		(6.600)	(6.399)	(6.553)	
		[p=0.157]	[p=0.859]	[p=0.164]		[p=0.146]	[p=0.618]	[p=0.319]
Mental Health	3.140	3.064	2.833*		2.958	3.065	2.796	
	(2.245)	(2.084)	(2.090)		(2.215)	(2.177)	(2.150)	
		[p=0.684]	[p=0.093]	[p=0.170]		[p=0.617]	[p=0.355]	[p=0.149]
Observations	1,030	985	1,016		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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.4: Baseline Characteristics and Balance (Extended Sample)

	Younger Girls				*Older Girls*			
	Control	Girl Groups	Girl Groups + Community	Difference	Control	Girl Groups	Girl Groups + Community	Difference
Girl Age	12.935	12.896	12.935		15.853	15.888	15.881	
	(0.858)	(0.850)	(0.854)		(0.791)	(0.813)	(0.805)	
		[p=0.317]	[p=0.999]	[p=0.323]		[p=0.320]	[p=0.390]	[p=0.862]
Carer's Years of Ed.	1.433	1.046	1.227		1.203	0.804*	0.916	
	(2.842)	(2.537)	(2.562)		(2.667)	(2.242)	(2.252)	
		[p=0.111]	[p=0.338]	[p=0.438]		[p=0.089]	[p=0.151]	[p=0.599]
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=0.850]	[p=0.180]	[p=0.290]		[p=0.466]	[p=0.207]	[p=0.592]
Asset Index	-0.089	-0.144	-0.120		0.093	0.003	0.016	
	(0.840)	(0.799)	(0.805)		(0.961)	(0.885)	(0.908)	
		[p=0.429]	[p=0.660]	[p=0.681]		[p=0.191]	[p=0.287]	[p=0.862]
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=0.556]	[p=0.001]	[p=0.042]		[p=0.325]	[p=0.351]	[p=0.985]
Progressive Gender Attitudes	34.620	33.957	34.626		35.435	34.344*	35.505	
	(6.914)	(6.770)	(6.818)		(6.790)	(6.888)	(7.068)	
		[p=0.272]	[p=0.991]	[p=0.206]		[p=0.096]	[p=0.911]	[p=0.049]
Carer's Progressive Gender Attitudes	33.965	33.173	33.972		33.911	33.270	33.769	
	(6.607)	(6.492)	(6.581)		(6.587)	(6.400)	(6.564)	
		[p=0.152]	[p=0.991]	[p=0.109]		[p=0.231]	[p=0.804]	[p=0.318]
Mental Health	3.127	3.049	2.830		2.958	3.077	2.773	
	(2.235)	(2.097)	(2.100)		(2.222)	(2.183)	(2.135)	
		[p=0.676]	[p=0.103]	[p=0.197]		[p=0.572]	[p=0.287]	[p=0.097]
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. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.5: Differential Girl Group Attendance between Treatment Arms

Panel A: Older Girls – Educational Sessions

	Attended any session	Attended at least…			Number of sessions	# conditional on enrollment
		25%	50%	75%		
Community Campaigns	0.000768	0.00956	0.0165	0.0132	0.486	1.073
	(0.0612)	(0.0574)	(0.0496)	(0.0407)	(2.126)	(2.214)
N	1738	1738	1738	1738	1738	753
Mean in GG only arm	0.433	0.365	0.256	0.142	11.8	27.3

Panel B: Younger Girls – Educational Sessions

	Attended any session	Attended at least…			Number of sessions	# conditional on enrollment
		25%	50%	75%		
Community Campaigns	-0.0318	-0.000872	0.0119	0.0122	0.233	2.113
	(0.0554)	(0.0531)	(0.0502)	(0.0413)	(2.013)	(2.102)
N	2001	2001	2001	2001	2001	1087
Mean in GG only arm	0.559	0.480	0.351	0.180	15.50	27.7

Panel C: Older Girls – Sports Sessions

	Attended any session	Attended at least…			Number of sessions	# conditional on enrollment
		25%	50%	75%		
Community Campaigns	-0.00265	-0.00307	0.00635	-0.00864	0.172	0.538
	(0.0609)	(0.0564)	(0.0524)	(0.0397)	(1.722)	(1.978)
N	1738	1738	1738	1738	1738	748
Mean in GG only arm	0.432	0.368	0.245	0.155	9.542	22.1

Panel D: Younger Girls – Sports Sessions

	Attended any session	Attended at least…			Number of sessions	# conditional on enrollment
		25%	50%	75%		
Community Campaigns	-0.0278	-0.0204	-0.00466	0.00149	0.0355	1.281
	(0.0551)	(0.0564)	(0.0554)	(0.0463)	(1.796)	(1.921)
N	2001	2001	2001	2001	2001	1083
Mean in GG only arm	0.555	0.475	0.346	0.209	12.80	23.05

Notes: The table reports OLS regressions of measures of Girl Group session attendance on an indicator for whether the cluster was assigned to the combined Girl Groups + Community Campaigns arm (vs. Girl Groups only). The sample excludes the control group. In each panel, the "Mean in GG only arm" row gives the mean of the outcome among girls in clusters assigned to Girl Groups only (the omitted category); the "Community Campaigns" row reports the coefficient on the indicator for being in the combined arm, i.e. the difference in attendance between the two treatment arms. Panels A and B report results for Educational Sessions (Panel A: older girls aged 15–17 at baseline; Panel B: younger girls aged 12–14). Panels C and D report the same for Sports Sessions. The six columns use different attendance markers: (1) attended any session; (2) attended at least 12 sessions (25% of the total); (3) attended at least 23 sessions (50%); (4) attended at least 35 sessions (75%); (5) number of sessions attended; (6) number attended conditional on attending any. Analytical standard errors (in parentheses) are clustered at the cluster level. Attendance data come from registers recorded and filed by the peer mentors. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.6: Treatment effects on school enrollment and post-secondary education

	Older Girls		*Younger Girls*		*All*	
	School Enrollment	Post-secondary Education	School Enrollment	Post-secondary Education	School Enrollment	Post-secondary Education
Girl Groups	0.064***	0.001	0.017	0.003	0.041***	0.001
	(0.022)	(0.014)	(0.018)	(0.004)	(0.015)	(0.008)
	[p=0.004]	[p=0.927]	[p=0.356]	[p=0.428]	[p=0.009]	[p=0.847]
Girl Groups	0.070***	-0.013	0.012	0.006	0.040**	-0.004
+ Community	(0.021)	(0.013)	(0.020)	(0.005)	(0.017)	(0.008)
	[p=0.001]	[p=0.323]	[p=0.564]	[p=0.203]	[p=0.018]	[p=0.604]
Difference	0.006	-0.015	-0.005	0.003	-0.001	-0.005
	(0.024)	(0.014)	(0.020)	(0.005)	(0.018)	(0.008)
	[p=0.805]	[p=0.297]	[p=0.808]	[p=0.542]	[p=0.974]	[p=0.485]
N	2,605	2,605	2,920	2,920	5,525	5,525
Control Mean	0.316	0.094	0.761	0.004	0.547	0.047
Heterogeneity by Age						
Girl Groups			[p=0.067]	[p=0.892]		
Girl Groups + Community			[p=0.017]	[p=0.145]		
Difference			[p=0.694]	[p=0.198]		

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 B.7: Pooled Treatment Effects on Education and Marriage

	Older Girls			*Younger Girls*		
	In Education	Married	Married or Engaged or Fixed	In Education	Married	Married or Engaged or Fixed
Any Treatment	0.053***	-0.031**	-0.055***	0.016	0.007	0.019
	(0.019)	(0.016)	(0.018)	(0.017)	(0.010)	(0.013)
	[p=0.005]	[p=0.049]	[p=0.003]	[p=0.336]	[p=0.462]	[p=0.144]
N	2,605	2,605	2,605	2,920	2,920	2,920
Control Mean	0.406	0.181	0.304	0.765	0.056	0.093

Notes: The table reports estimated effects of being assigned to either of the two treatment arms (Girl Groups or Girl Groups + Community Campaigns) relative to the pure control group, pooling across the two treatment arms. Estimates are average marginal effects from logit regressions. 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 as in Table 2. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.8: Married - excluding married but gauna not performed (GNP)

	Older Girls	*Younger Girls*	*All*
	Married excl. GNP	Married excl. GNP	Married excl. GNP
Girl Groups	-0.033**	-0.002	-0.017*
	(0.016)	(0.009)	(0.009)
	[p=0.042]	[p=0.785]	[p=0.053]
Girl Groups	-0.024	0.010	-0.006
+ Community	(0.018)	(0.010)	(0.011)
	[p=0.191]	[p=0.314]	[p=0.603]
Difference	0.009	0.012	0.012
	(0.018)	(0.009)	(0.011)
	[p=0.615]	[p=0.169]	[p=0.300]
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=0.075]	
Difference		[p=0.852]	

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 B.9: 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	-0.013	0.001	-0.010	0.170	0.006	0.007	0.079	0.003	-0.001
	(0.272)	(0.028)	(0.014)	(0.238)	(0.025)	(0.010)	(0.229)	(0.024)	(0.010)
	[p=0.962]	[p=0.966]	[p=0.484]	[p=0.475]	[p=0.819]	[p=0.494]	[p=0.730]	[p=0.888]	[p=0.925]
Girl Groups	-0.755***	-0.071***	-0.025**	-0.262	-0.035	-0.005	-0.498***	-0.052***	-0.015*
+ Community	(0.230)	(0.023)	(0.012)	(0.216)	(0.023)	(0.009)	(0.193)	(0.020)	(0.008)
	[p=0.001]	[p=0.002]	[p=0.033]	[p=0.224]	[p=0.123]	[p=0.583]	[p=0.010]	[p=0.007]	[p=0.085]
Difference	-0.742***	-0.072***	-0.015	-0.432**	-0.041*	-0.012	-0.577***	-0.056***	-0.014*
	(0.226)	(0.024)	(0.011)	(0.196)	(0.023)	(0.009)	(0.173)	(0.020)	(0.008)
	[p=0.001]	[p=0.002]	[p=0.198]	[p=0.027]	[p=0.069]	[p=0.184]	[p=0.001]	[p=0.005]	[p=0.075]
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=0.417]	[p=0.857]	[p=0.196]			
Girl Groups + Community				[p=0.027]	[p=0.145]	[p=0.121]			
Difference				[p=0.195]	[p=0.206]	[p=0.860]			

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 $\geq$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 $\geq$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 B.10: 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	-0.149	-0.030	0.001	0.056	0.005	0.006	-0.042	-0.011	0.004
	(0.214)	(0.026)	(0.012)	(0.195)	(0.024)	(0.007)	(0.180)	(0.021)	(0.007)
	[p=0.484]	[p=0.243]	[p=0.911]	[p=0.774]	[p=0.844]	[p=0.386]	[p=0.814]	[p=0.599]	[p=0.545]
Girl Groups	-0.775***	-0.078***	-0.010	-0.247	-0.028	0.010	-0.496***	-0.051***	0.001
+ Community	(0.195)	(0.025)	(0.009)	(0.182)	(0.021)	(0.007)	(0.160)	(0.019)	(0.005)
	[p=0.000]	[p=0.002]	[p=0.283]	[p=0.175]	[p=0.173]	[p=0.155]	[p=0.002]	[p=0.006]	[p=0.882]
Difference	-0.626***	-0.047**	-0.011	-0.303*	-0.033	0.004	-0.454***	-0.040**	-0.004
	(0.176)	(0.021)	(0.011)	(0.182)	(0.020)	(0.008)	(0.146)	(0.016)	(0.007)
	[p=0.000]	[p=0.027]	[p=0.325]	[p=0.096]	[p=0.103]	[p=0.664]	[p=0.002]	[p=0.015]	[p=0.601]
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=0.291]	[p=0.166]	[p=0.710]			
Girl Groups + Community				[p=0.008]	[p=0.069]	[p=0.103]			
Difference				[p=0.122]	[p=0.574]	[p=0.300]			

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

Table B.11: Estimated Impacts on Rumination Raw Scores

	Older Girls	*Younger Girls*	*All*
	Rumination	Rumination	Rumination
Girl Groups	0.761	1.040	0.914
	(0.598)	(0.639)	(0.594)
	[p=0.203]	[p=0.103]	[p=0.124]
Girl Groups	-0.350	-0.086	-0.206
+ Community	(0.735)	(0.665)	(0.675)
	[p=0.635]	[p=0.897]	[p=0.760]
Difference	-1.110	-1.126	-1.121
	(0.749)	(0.755)	(0.725)
	[p=0.138]	[p=0.136]	[p=0.122]
N	2,318	2,724	5,042
Control Mean	21.530	21.818	21.683
Heterogeneity by Age			
Girl Groups		[p=0.443]	
Girl Groups + Community		[p=0.461]	
Difference		[p=0.968]	

Notes: The table presents estimated 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, and baseline measures of mental health. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.12: Cross-Tabulation of Girl and Carer Reports on Marriage and Education Attendance

Panel A: Marriage

	Carer's report	
Girl's report	Not married	Married
Not married	4388	5
Married	11	335
Agreement rate	99.7%	

Panel B: Education attendance

	Carer's report	
Girl's report	Not in education	In education
Not in education	1723	68
In education	66	2882
Agreement rate	97.2%	

Notes: The table cross-tabulates reports from girls and their carers on marriage status (Panel A) and education attendance (Panel B). Girls and carers were interviewed separately and privately, with the girl interviewed first whenever possible. Agreement rate is the percentage of girl–carer pairs for whom both reports coincide.

Table B.13: Treatment Effects on Girl–Carer Reporting Discrepancies

	(1) Marriage	(2) Education attendance
Girl Groups	-0.003	0.000
	(0.002)	(0.006)
Girl Groups + Community	-0.003	0.004
	(0.002)	(0.006)
Control mean	0.005	0.028
Observations	4739	4739

Notes: The table presents estimated effects of "Girl Groups" and "Girl Groups + Community" interventions on the probability that the girl's and carer's reports disagree. The dependent variable is an indicator equal to one when the girl and carer give different reports on marriage status or education attendance. Standard errors (in parentheses) are clustered at the cluster level. For marriage, 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. For education attendance, 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 survey month fixed effects. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.14: Treatment Effects on "Worst-Case" Education and Marriage Reports

	Older Girls			*Younger Girls*		
	In Education	Married	Married or Engaged or Fixed	In Education	Married	Married or Engaged or Fixed
Girl Groups	0.054**	-0.037**	-0.053***	0.012	0.004	0.016
	(0.025)	(0.017)	(0.020)	(0.019)	(0.011)	(0.014)
	[p=0.031]	[p=0.034]	[p=0.008]	[p=0.522]	[p=0.735]	[p=0.259]
Girl Groups + Community	0.049**	-0.027	-0.057**	0.010	0.009	0.021
	(0.021)	(0.020)	(0.024)	(0.020)	(0.011)	(0.016)
	[p=0.021]	[p=0.167]	[p=0.017]	[p=0.622]	[p=0.425]	[p=0.178]
Difference	-0.004	0.010	-0.004	-0.002	0.005	0.005
	(0.025)	(0.019)	(0.024)	(0.021)	(0.010)	(0.016)
	[p=0.867]	[p=0.609]	[p=0.855]	[p=0.922]	[p=0.631]	[p=0.752]
N	2,605	2,605	2,605	2,920	2,920	2,920
Control Mean	0.386	0.183	0.305	0.763	0.057	0.093
Heterogeneity by Age						
Girl Groups				[p=0.120]	[p=0.027]	[p=0.001]
Girl Groups + Community				[p=0.101]	[p=0.087]	[p=0.001]
Difference				[p=0.940]	[p=0.793]	[p=0.675]

Notes: The table presents estimated treatment effects using "worst case" outcome measures constructed from girl and carer reports. For marriage, a girl is coded as married if *either* she or her carer reports that she is married. For education attendance, a girl is coded as attending only if *both* she and her carer report that she is in education. 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). Controls are as described in Table 2. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.15: 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	0.060***	-0.036**	-0.053***	0.015	0.007	0.017	0.037**	-0.013	-0.016
	(0.023)	(0.018)	(0.020)	(0.019)	(0.011)	(0.014)	(0.017)	(0.011)	(0.014)
	[p=0.009]	[p=0.038]	[p=0.008]	[p=0.414]	[p=0.536]	[p=0.222]	[p=0.029]	[p=0.232]	[p=0.250]
Girl Groups	0.048**	-0.025	-0.057**	0.013	0.010	0.023	0.030*	-0.006	-0.014
+ Community	(0.021)	(0.019)	(0.024)	(0.021)	(0.011)	(0.016)	(0.018)	(0.012)	(0.017)
	[p=0.024]	[p=0.199]	[p=0.019]	[p=0.545]	[p=0.336]	[p=0.153]	[p=0.089]	[p=0.621]	[p=0.405]
Difference	-0.012	0.012	-0.004	-0.003	0.004	0.006	-0.007	0.008	0.002
	(0.023)	(0.020)	(0.025)	(0.021)	(0.011)	(0.016)	(0.018)	(0.013)	(0.017)
	[p=0.611]	[p=0.553]	[p=0.878]	[p=0.898]	[p=0.735]	[p=0.715]	[p=0.684]	[p=0.565]	[p=0.920]
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=0.063]	[p=0.016]	[p=0.001]			
Girl Groups + Community				[p=0.132]	[p=0.083]	[p=0.001]			
Difference				[p=0.712]	[p=0.662]	[p=0.663]			

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 B.16: Education - Main Sample

	Older Girls	*Younger Girls*	*All*
	In Education	In Education	In Education
Girl Groups	0.060**	0.022	0.042**
	(0.024)	(0.019)	(0.018)
	[p=0.011]	[p=0.245]	[p=0.018]
Girl Groups	0.048**	0.014	0.030
+ Community	(0.022)	(0.022)	(0.019)
	[p=0.030]	[p=0.531]	[p=0.103]
Difference	-0.012	-0.009	-0.012
	(0.024)	(0.021)	(0.019)
	[p=0.608]	[p=0.676]	[p=0.537]
N	2,318	2,725	5,043
Control Mean	0.431	0.782	0.617
Heterogeneity by Age			
Girl Groups		[p=0.123]	
Girl Groups + Community		[p=0.142]	
Difference		[p=0.892]	

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 education 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 B.17: Marriage - Main Sample and Extended Sample + Tracking Data

	Older Girls		*Younger Girls*		*All*	
	Main Sample	Extended + Tracking	Main Sample	Extended + Tracking	Main Sample	Extended + Tracking
Girl Groups	-0.032**	-0.037**	0.015	0.004	-0.008	-0.016
	(0.016)	(0.017)	(0.010)	(0.011)	(0.010)	(0.011)
	[p=0.046]	[p=0.030]	[p=0.136]	[p=0.699]	[p=0.447]	[p=0.125]
Girl Groups	-0.015	-0.027	0.015*	0.009	0.000	-0.008
+ Community	(0.019)	(0.019)	(0.009)	(0.011)	(0.012)	(0.012)
	[p=0.420]	[p=0.167]	[p=0.094]	[p=0.413]	[p=0.985]	[p=0.476]
Difference	0.017	0.010	0.000	0.005	0.008	0.008
	(0.018)	(0.018)	(0.009)	(0.010)	(0.011)	(0.012)
	[p=0.361]	[p=0.579]	[p=0.975]	[p=0.646]	[p=0.487]	[p=0.507]
N	2,318	2,656	2,725	2,990	5,043	5,646
Control Mean	0.114	0.178	0.030	0.055	0.069	0.114
Heterogeneity by Age						
Girl Groups			[p=0.005]	[p=0.021]		
Girl Groups + Community			[p=0.096]	[p=0.084]		
Difference			[p=0.347]	[p=0.744]		

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 B.18: Does the Reduction in Education Dropouts Mediate Delayed Marriage?

	(1) Married/Engaged/Fixed	(2) Married/Engaged/Fixed
Girl Groups	-0.053***	-0.039**
	(0.020)	(0.019)
	[p=0.008]	[p=0.039]
Girl Groups + Community	-0.057**	-0.046*
	(0.024)	(0.023)
	[p=0.019]	[p=0.050]
In formal education		-0.224***
		(0.021)
		[p<0.001]
N	2,605	2,605
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). The dependent variable is a binary indicator for whether the girl was married, engaged, or had her marriage fixed at endline. Column (1) reports OLS estimates of the impact of the two treatment arms relative to the pure control group; these match the estimates in Table B.15. Column (2) re-runs the same regression but additionally includes a binary indicator for endline formal-education enrollment as a control variable. The two p-values at the bottom of column (2) test the equality of the "Girl Groups" and "Girl Groups + Community" coefficients between columns (1) and (2), respectively. 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 B.19: Estimated Impacts on Secondary Outcomes

Panel A: Older Girls						
	Positive attitude to school	Contraception and sexual health knowledge	Puberty and menstruation knowledge	Minutes of sports played per day	Positive attitudes to sport	Restrictions during menstruation
Girl Groups	0.052	0.032	0.003	10.151***	0.022	0.046
	(0.083)	(0.067)	(0.053)	(2.872)	(0.069)	(0.068)
	[p=0.533]	[p=0.632]	[p=0.962]	[p=0.000]	[p=0.747]	[p=0.494]
Girl Groups + Community	0.054	0.095	-0.046	6.854**	0.095	0.000
	(0.075)	(0.064)	(0.053)	(2.819)	(0.076)	(0.065)
	[p=0.467]	[p=0.138]	[p=0.382]	[p=0.015]	[p=0.213]	[p=0.994]
Difference	0.002	0.063	-0.049	-3.297	0.073	-0.046
	(0.078)	(0.068)	(0.055)	(3.535)	(0.070)	(0.079)
	[p=0.976]	[p=0.352]	[p=0.375]	[p=0.351]	[p=0.300]	[p=0.561]
N	997	2,301	2,307	2,315	2,315	2,287
Control Mean	0.000	0.000	-0.000	10.927	-0.001	-0.000
Panel B: Younger Girls						
	Positive attitude to school	Contraception and sexual health knowledge	Puberty and menstruation knowledge	Minutes of sports played per day	Positive attitudes to sport	Restrictions during menstruation
Girl Groups	0.014	0.059	-0.063	13.351***	0.053	0.070
	(0.065)	(0.073)	(0.052)	(3.236)	(0.071)	(0.068)
	[p=0.830]	[p=0.419]	[p=0.225]	[p=0.000]	[p=0.454]	[p=0.301]
Girl Groups + Community	0.010	0.004	-0.006	10.245***	0.053	0.028
	(0.066)	(0.072)	(0.054)	(3.048)	(0.073)	(0.072)
	[p=0.881]	[p=0.956]	[p=0.910]	[p=0.001]	[p=0.470]	[p=0.699]
Difference	-0.004	-0.055	0.057	-3.106	-0.000	-0.042
	(0.063)	(0.073)	(0.052)	(3.588)	(0.067)	(0.080)
	[p=0.948]	[p=0.449]	[p=0.267]	[p=0.387]	[p=0.998]	[p=0.596]
N	2,088	1,750	2,570	2,724	2,724	2,309
Control Mean	-0.000	-0.000	-0.000	19.281	-0.001	0.000

Notes: The table presents estimated 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 and district. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Measures constructed as follows (as described in pre-analysis plan):

- **Positive attitude towards school**: Factor score from the following author developed scale: 1)I enjoy school; 2) I am motivated to work hard at school; 3) I am bored in school; 4) What I learn at school will be useful for my future; 5) I want to quit school; 6) I learn lots of new things at school; 7) I look forward to going to school; 8) My teachers at school want me to do well; 9) I feel my studies have no meaning. As discussed in Section 5, this scale was only collected for girls in school/education at the time of endline.
- **Puberty and menstruation knowledge:** Score predicted using IRT on puberty and menstruation knowledge scale (19 items). This is a series of 19 indicators of whether girls answered each question on puberty and menstruation correctly. This outcome is missing girls who reported that they had not heard of menstruation. There is no effect of either treatment on whether or not girls had heard of menstruation.
- **Contraception and sexual health knowledge:** Score predicted using IRT on contraception and sexual health knowledge scale (19 items). These questions were only collected for girls who were 15 or older at the time of endline.
- **Minutes of sport played per day:** Typical time spent playing sports per day (NB. measure taken several months after intervention end date).
- **Positive attitude to sports:** Factor score from author-created sports attitudes scale: 1) I can play sports with the other girls in my village. 2) I can make friends by playing sports 3) I can be healthier by playing sports. 4) I can encourage other girls to take up sports in my community. 5) I can ask for play time after I complete household chores. 6) I can be comfortable with my body while playing sport 7) I can compete in sports just as well as boys. 8) I enjoy playing sports 9) I feel safe when I play sports 10) When I play sports I feel I have time for myself.
- **Restrictions during menstruation:** Score predicted using IRT on 7 indicators of whether girls face the following restrictions during menstruation: 1) Attend religious function 2) Cook 3) Touch stored food 4) Sleep in your usual bed/ place 5) Touch family members 6) Play outside/ see friends outside house 7) Visit relatives

Table B.20: Estimated Impacts on Non-Cognitive Skills and School Attendance by Age

	Schooldays missed	Self Efficacy	Self Esteem	Peer Relations	Resilience	Vigilant Decision Making	Buck-passing Decision Making
Panel A: Older Girls							
Girl Groups	-1.445**	0.008	-0.001	-0.056	-0.142	-0.116	0.048
	(0.673)	(0.092)	(0.088)	(0.080)	(0.086)	(0.082)	(0.056)
	[p=0.032]	[p=0.932]	[p=0.994]	[p=0.483]	[p=0.100]	[p=0.157]	[p=0.395]
Girl Groups + Community	-0.876	0.076	0.086	-0.029	0.010	0.065	0.038
	(0.698)	(0.094)	(0.083)	(0.072)	(0.090)	(0.083)	(0.058)
	[p=0.210]	[p=0.415]	[p=0.301]	[p=0.690]	[p=0.907]	[p=0.439]	[p=0.515]
Difference	0.568	0.068	0.086	0.027	0.152**	0.180**	-0.010
	(0.602)	(0.084)	(0.079)	(0.067)	(0.071)	(0.069)	(0.050)
	[p=0.345]	[p=0.418]	[p=0.278]	[p=0.685]	[p=0.032]	[p=0.009]	[p=0.844]
N	996	2,318	2,318	2,318	2,318	2,317	2,317
Panel B: Younger Girls							
Girl Groups	-0.479	-0.008	0.047	0.076	-0.097	-0.036	0.016
	(0.421)	(0.093)	(0.087)	(0.072)	(0.088)	(0.060)	(0.061)
	[p=0.256]	[p=0.929]	[p=0.593]	[p=0.288]	[p=0.273]	[p=0.543]	[p=0.788]
Girl Groups + Community	-0.680*	-0.012	0.056	0.038	-0.022	0.015	0.030
	(0.384)	(0.094)	(0.081)	(0.062)	(0.090)	(0.077)	(0.060)
	[p=0.076]	[p=0.901]	[p=0.495]	[p=0.536]	[p=0.805]	[p=0.842]	[p=0.616]
Difference	-0.201	-0.003	0.009	-0.038	0.075	0.052	0.014
	(0.204)	(0.089)	(0.074)	(0.064)	(0.082)	(0.070)	(0.055)
	[p=0.325]	[p=0.970]	[p=0.904]	[p=0.533]	[p=0.362]	[p=0.462]	[p=0.804]
N	2,088	2,725	2,725	2,725	2,725	2,725	2,724

Notes: The table presents estimated 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 as described in Table 5. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.21: Differential Selection into Attendance by Treatment Arm

	Attended at least 50% of education sessions		Attended at least 50% of sport sessions	
	(1) Without Interactions	(2) With Interactions	(3) Without Interactions	(4) With Interactions
Age	-0.019***	-0.019**	-0.023***	-0.023***
	(0.006)	(0.008)	(0.005)	(0.008)
Attending School at Baseline	0.071***	0.045	0.061**	0.057
	(0.022)	(0.036)	(0.024)	(0.040)
Carer's Years of Education	-0.002	-0.006	-0.001	-0.004
	(0.003)	(0.005)	(0.004)	(0.005)
SC/ST Caste	0.023	0.046	0.046	0.051
	(0.041)	(0.065)	(0.039)	(0.064)
Caste OBC EBC BL	0.060	0.115*	0.059	0.106
	(0.044)	(0.067)	(0.046)	(0.073)
Asset Index	0.007	0.017	0.013	0.022
	(0.014)	(0.022)	(0.014)	(0.021)
Mental Health	-0.000	0.003	-0.006	-0.005
	(0.004)	(0.006)	(0.004)	(0.006)
Progressive Gender Attitudes	-0.003**	-0.005**	-0.003**	-0.004**
	(0.002)	(0.002)	(0.002)	(0.002)
Girl Groups + Community Campaigns (CC)		-0.003		0.000
		(0.209)		(0.205)
CC x Age		-0.001		-0.000
		(0.011)		(0.010)
CC x Attending School at Baseline		0.048		0.006
		(0.044)		(0.048)
CC x Carer's Years of Education		0.008		0.006
		(0.007)		(0.007)
CC x SC/ST Caste		-0.034		-0.004
		(0.083)		(0.081)
CC x Caste OBC EBC BL		-0.095		-0.088
		(0.086)		(0.092)
CC x Asset Index		-0.020		-0.018
		(0.028)		(0.028)
CC x Mental Health		-0.005		-0.000
		(0.008)		(0.009)
CC x Progressive Gender Attitudes		0.002		0.001
		(0.003)		(0.003)
Constant	0.624***	0.625***	0.662***	0.660***
	(0.105)	(0.129)	(0.102)	(0.128)
Observations	3739	3739	3739	3739

Notes: The table presents coefficients and associated standard errors from regressing indicators for attending at least 50% of education sessions (columns 1–2) and at least 50% of sport sessions (columns 3–4) on baseline characteristics and interactions of baseline characteristics with an indicator of whether the cluster was assigned to the Community Campaigns (CC). We use only data from the two treatment arms (i.e. we drop the control group). 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). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.22: Heterogeneity in Education and Marriage by Attendance (Younger Girls)

	Married		Married, Engaged or Fixed		In Education	
	Any	50%	Any	50%	Any	50%
Community Campaigns	0.019	0.012	0.000	-0.002	-0.029	-0.003
	(0.0174)	(0.0145)	(0.0249)	(0.0199)	(0.0246)	(0.0235)
Attended	-0.026	-0.027	-0.040**	-0.044**	-0.008	0.028
	(0.0172)	(0.0172)	(0.0193)	(0.0186)	(0.0223)	(0.0269)
Attended × Community Campaigns	-0.035	-0.026	0.007	0.021	0.042	-0.010
	(0.0223)	(0.0221)	(0.0285)	(0.0262)	(0.0333)	(0.0364)
N	1921	1921	1921	1921	1921	1921

Notes: The table presents estimated heterogeneous impacts of the addition of "Community Campaigns" over and above the Girl Groups intervention by whether or not the girl attended the Girl Groups. Analysis contains only the two treatment arms and excludes the control group. Attendance was measured using implementation data and re-coded as a binary variable based on two thresholds: (i) any attendance; (ii) 50% attendance. Pre-specified controls are as in Table 2. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.23: Heterogeneity in Education and Marriage by Attendance (Older Girls)

	Married		Married, Engaged or Fixed		In Education	
	Any	50%	Any	50%	Any	50%
Community Campaigns	0.001	0.012	0.012	0.004	-0.014	-0.020
	(0.0298)	(0.0253)	(0.0353)	(0.0301)	(0.0274)	(0.0261)
Attended	-0.113***	-0.102***	-0.042	-0.082**	0.008	-0.012
	(0.0256)	(0.0269)	(0.0329)	(0.0367)	(0.0272)	(0.0304)
Attended × Community Campaigns	0.020	-0.003	-0.035	-0.021	0.005	0.032
	(0.0352)	(0.0369)	(0.0420)	(0.0486)	(0.0366)	(0.0447)
N	1676	1676	1676	1676	1676	1676

Notes: The table presents estimated heterogeneous impacts of the addition of "Community Campaigns" over and above the Girl Groups intervention by whether or not the girl attended the Girl Groups. Analysis contains only the two treatment arms and excludes the control group. Attendance was measured using implementation data and re-coded as a binary variable based on two thresholds: (i) any attendance; (ii) 50% attendance. Pre-specified controls are as in Table 2. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table B.24: Estimated Impacts on Secondary Outcomes (Girls' Attitudes towards Violence)

Panel A: Older Girls					
	Victim Blaming	Perpetrator Blaming	Avoidance	Retaliation	Reporting
Girl Groups	-0.004	-0.026	0.024	-0.117**	-0.065***
	(0.041)	(0.047)	(0.043)	(0.050)	(0.024)
	[p=0.926]	[p=0.576]	[p=0.580]	[p=0.019]	[p=0.006]
Girl Groups + Community	-0.073*	0.016	-0.013	-0.005	-0.005
	(0.043)	(0.049)	(0.045)	(0.050)	(0.029)
	[p=0.092]	[p=0.736]	[p=0.771]	[p=0.920]	[p=0.871]
Difference	-0.069*	0.043	-0.037	0.112**	0.060*
	(0.041)	(0.048)	(0.047)	(0.052)	(0.032)
	[p=0.091]	[p=0.379]	[p=0.434]	[p=0.031]	[p=0.058]
N	2,318	2,318	2,318	2,318	2,318
Control Mean	0.000	-0.004	-0.008	0.002	0.000
Panel B: Younger Girls					
	Victim Blaming	Perpetrator Blaming	Avoidance	Retaliation	Reporting
Girl Groups	-0.027	0.013	0.053	-0.010	-0.047*
	(0.039)	(0.042)	(0.042)	(0.045)	(0.029)
	[p=0.483]	[p=0.760]	[p=0.210]	[p=0.818]	[p=0.096]
Girl Groups + Community	-0.069**	0.027	0.004	-0.004	-0.007
	(0.033)	(0.037)	(0.043)	(0.040)	(0.025)
	[p=0.038]	[p=0.466]	[p=0.931]	[p=0.926]	[p=0.779]
Difference	-0.041	0.014	-0.049	0.007	0.040
	(0.038)	(0.047)	(0.048)	(0.051)	(0.032)
	[p=0.278]	[p=0.771]	[p=0.308]	[p=0.896]	[p=0.208]
N	2,725	2,724	2,724	2,724	2,724
Control Mean	-0.000	-0.000	-0.005	0.000	0.001

Notes: The table presents estimated 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 and district. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.
As described in pre-analysis plan, each measure is constructed using girls' responses to a series of 6 vignettes describing situations of violence against women and girls. Girls were then asked a series of binary questions about their attitudes and response to this violence. We combined these binary indicators using IRT to create each measure. "Victim blaming" measures the extent to which girls reported that the female victim of violence was to blame. "Perpetrator blaming" measures the extent to which girls reported that the perpetrator was to blame. "Avoidance" measures the extent to which girls indicated that the female victim should avoid that situation in future. "Retaliation" measures the extent to which girls reported that the female victim should seek retaliation. "Reporting" measures the extent to which girls reported that the female victim should report what had happened to others.

Table B.25: Estimated Impacts on Secondary Outcomes (Mothers' Attitudes towards Violence)

	All				
	Victim Blaming	Perpetrator Blaming	Avoidance	Retaliation	Reporting
Girl Groups	0.046	-0.018	-0.018	-0.083**	-0.058*
	(0.038)	(0.025)	(0.039)	(0.041)	(0.033)
	[p=0.231]	[p=0.469]	[p=0.652]	[p=0.043]	[p=0.077]
Girl Groups	-0.047	0.011	-0.033	-0.008	-0.032
+ Community	(0.036)	(0.024)	(0.040)	(0.034)	(0.032)
	[p=0.196]	[p=0.647]	[p=0.406]	[p=0.820]	[p=0.314]
Difference	-0.093**	0.029	-0.016	0.075*	0.026
	(0.038)	(0.028)	(0.041)	(0.039)	(0.033)
	[p=0.014]	[p=0.304]	[p=0.707]	[p=0.051]	[p=0.439]
N	4,804	4,804	4,804	4,804	4,804
Control Mean	-0.023	0.006	-0.004	0.011	0.008

Notes: The table presents estimated 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 mother controls are: age, age squared, years of schooling, caste, baseline wealth index and district. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. As described in pre-analysis plan, each measure is constructed using carers' responses to a series of 3 vignettes describing situations of violence against women and girls. Carers were then asked a series of binary questions about their attitudes and response to this violence. We combined these binary indicators using IRT to create each measure (given we just have three items per scale we use a one-parameter IRT model and constrain the discrimination parameter to be constant across items). "Victim blaming" measures the extent to which carers reported that the female victim of violence was to blame. "Perpetrator blaming" measures the extent to which carers reported that the perpetrator was to blame. "Avoidance" measures the extent to which carers indicated that the female victim should avoid that situation in future. "Retaliation" measures the extent to which carers reported that the female victim should seek retaliation. "Reporting" measures the extent to which carers reported that the female victim should report what had happened to others.

Table B.26: Post-Lasso OLS: Predicting Endline School Attendance

	(1)	
What is the highest standard of education that you have completed?	-0.093***	(0.016)
Girls can hold leadership position in school.	-0.002	(0.034)
Only bad girls make male friends	0.045**	(0.022)
Family should decide till when the daughters should be educated.	-0.054***	(0.018)
Instead of spending money on a girl's education, it should be saved for her dowry	0.054***	(0.019)
If the decision were yours alone, at what age would you like to marry? (years)	-0.002**	(0.001)
Carer age BL	-0.005**	(0.002)
Men should be more educated than their wives	-0.046	(0.029)
Girls should be married early to protect them from sexual harassment.	0.045**	(0.019)
If a girl is a victim of some sexual abuse, it is the fault of the girl.	-0.045*	(0.023)
It is okay for boys to tease girls in public spaces	-0.051**	(0.022)
Girls should return home on time	0.054*	(0.030)
Caste OBC EBC BL	-0.048	(0.033)
Asset Index	0.027*	(0.016)
Probable age of marriage [unmarried] at BL	0.018	(0.013)
Elders talking about marriage [unmarried]	-0.273***	(0.086)
Would prefer an arranged marriage if decision was hers alone [unmarried]	-0.313**	(0.128)
Maximum say in arranging marriage: MOTHER [unmarried]	-0.098	(0.085)
Attend private school [currently in school]	0.061	(0.048)
Literate	0.285***	(0.058)
Expected highest level of education: 8th GRADE OR LESS [currently in school] (g1	-0.178	(0.122)
Expected highest level of education: 9th-10th GRADE [currently in school] (g126)	-0.106**	(0.045)
Expected highest level of education: GRADUATE [currently in school] (g126)	0.030	(0.041)
Time spent studying outside of school [currently in school] (g148_7)	0.031*	(0.017)
Baseline Raven's Score	0.047**	(0.023)
Is it safe for [NAME] to attend a school outside village?	0.096*	(0.047)
Baseri block	-0.053	(0.032)
Dholpur block	0.096**	(0.044)
Missing indicator for: Sports attitudes raw score [baseline]	-0.113***	(0.039)
Missing indicator for: Menstruation restrictions raw score [baseline]	0.095	(0.090)
Missing indicator for: Menstruation knowledge raw score [baseline]	0.177*	(0.098)
Missing indicator for: If a girl is a victim of some sexual abuse, it is the fault of the girl	-0.588***	(0.209)
Missing indicator for: A woman should tolerate violence in order to keep her family together	0.527***	(0.163)
Missing indicator for: Girls who are highly educated indulge in improper behaviour	0.372*	(0.187)
Missing indicator for: Girl education aspiration (caregiver) at BL	-0.058	(0.057)
Constant	1.122***	(0.287)
Dep. var. mean	0.458	
R-squared	0.264	
N	616	

Notes: The table presents post-Lasso OLS estimates from a regression of endline school attendance on baseline covariates selected by a cross-validated Lasso. The sample consists of control-group older girls (aged 15–17 at baseline) who were enrolled in school at baseline. Missing baseline covariates are imputed at their median and accompanied by missing indicators. The fitted values from this regression are used to construct the predicted school dropout risk index used in Table 8. Standard errors clustered at the cluster level are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C Robustness Check: Post Double Selection Lasso

C.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 et al. 2018)[42].

Step 1: LASSO on the treatment variable[43]

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

Step 2: LASSO on the dependent variable

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

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

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

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

Before estimating equations (C.1), (C.2), and (C.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:

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

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

- 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;
- 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.

C.2 Results

We include all potential control variables and estimate Steps 1–3. Tables C.1 and C.2 show the results of the PDS Lasso estimation. The lasso selects some control variables that we prespecified in the pre-analysis plan, including school attendance, wealth index, elders talking about marriage, and mother's 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 3.5).

Table C.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*	-0.031*	-0.049**	0.010	0.007	0.015	0.029*	-0.014	-0.017
	(0.027)	(0.018)	(0.022)	(0.016)	(0.011)	(0.014)	(0.016)	(0.012)	(0.014)
	[p=0.076]	[p=0.095]	[p=0.024]	[p=0.560]	[p=0.492]	[p=0.286]	[p=0.067]	[p=0.241]	[p=0.216]
Girl Groups	0.039*	-0.018	-0.057**	0.010	0.009	0.022	0.023	-0.007	-0.015
+ Community	(0.021)	(0.024)	(0.027)	(0.021)	(0.012)	(0.015)	(0.017)	(0.014)	(0.016)
	[p=0.068]	[p=0.460]	[p=0.033]	[p=0.635]	[p=0.446]	[p=0.126]	[p=0.170]	[p=0.616]	[p=0.352]
Difference	-0.009	0.013	-0.008	0.001	0.002	0.007	-0.005	0.007	0.002
	(0.023)	(0.024)	(0.028)	(0.023)	(0.011)	(0.013)	(0.018)	(0.013)	(0.016)
	[p=0.715]	[p=0.593]	[p=0.779]	[p=0.980]	[p=0.858]	[p=0.576]	[p=0.767]	[p=0.594]	[p=0.882]
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=0.201]	[p=0.041]	[p=0.005]			
Girl Groups + Community				[p=0.286]	[p=0.257]	[p=0.004]			
Difference				[p=0.763]	[p=0.658]	[p=0.593]			

Notes: We used the Stata command pdslasso by (Ahrens et al. 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: wealth index, caste, and a dummy variable for whether the girl's family was talking about marriage at baseline. 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, and wealth index. In addition to those, PDS Lasso selected other variables including RAVENs test score, hours studying, and dummy variables for age interacted with other variables. Standard errors (in parentheses) are based on 50 bootstrap replications clustered at the cluster level.

Table C.2: Post-double LASSO estimates for Impacts on Depression, Anxiety and Rumination

	Older Girls			*Younger Girls*			*All*		
	Depression	Anxiety	Rumination	Depression	Anxiety	Rumination	Depression	Anxiety	Rumination
Girl Groups	0.009	-0.038	0.136**	0.045	0.011	0.170**	0.028	-0.013	0.144**
	(0.065)	(0.058)	(0.068)	(0.076)	(0.063)	(0.083)	(0.066)	(0.055)	(0.073)
	[p=0.893]	[p=0.515]	[p=0.046]	[p=0.553]	[p=0.867]	[p=0.041]	[p=0.665]	[p=0.811]	[p=0.048]
Girl Groups	-0.224***	-0.222***	-0.026	-0.089	-0.108**	-0.022	-0.150***	-0.160***	-0.010
+ Community	(0.059)	(0.053)	(0.130)	(0.059)	(0.052)	(0.103)	(0.053)	(0.047)	(0.114)
	[p=0.000]	[p=0.000]	[p=0.843]	[p=0.129]	[p=0.037]	[p=0.830]	[p=0.005]	[p=0.001]	[p=0.927]
Difference	-0.233***	-0.184***	-0.162	-0.134**	-0.118**	-0.192*	-0.179***	-0.147***	-0.155
	(0.057)	(0.045)	(0.120)	(0.054)	(0.057)	(0.100)	(0.048)	(0.046)	(0.105)
	[p=0.000]	[p=0.000]	[p=0.177]	[p=0.013]	[p=0.038]	[p=0.054]	[p=0.000]	[p=0.001]	[p=0.141]
N	2,318	2,318	2,318	2,725	2,725	2,724	5,043	5,043	5,042
Control Mean	0.001	0.001	0.000	-0.000	0.000	0.000	0.000	0.001	0.000
Heterogeneity by Age									
Girl Groups				[p=0.503]	[p=0.331]	[p=0.555]			
Girl Groups + Community				[p=0.008]	[p=0.013]	[p=0.954]			
Difference				[p=0.091]	[p=0.183]	[p=0.667]			

Notes: We used the Stata command pdslasso by (Ahrens et al. 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. PDS Lasso did not select any control variables for depression. It selected one control for anxiety, and three controls for rumination. Standard errors (in parentheses) are based on 50 bootstrap replications clustered at the cluster level.

D Measurement Appendix

Table D.1: Depression: Parameters of IRT measurement model

	Young Girls	Old Girls
Little interest or pleasure in doing things		
$\hat{\beta}_j$	1.956***	1.905***
	(0.162)	(0.169)
$\hat{\alpha}_{j1}$	0.698	0.822
	(0.060)	(0.069)
$\hat{\alpha}_{j2}$	1.895	1.938
	(0.115)	(0.127)
$\hat{\alpha}_{j3}$	2.970	3.421
	(0.220)	(0.291)
Feeling down, depressed, or hopeless		
$\hat{\beta}_j$	1.970***	1.894***
	(0.167)	(0.164)
$\hat{\alpha}_{j1}$	0.687	0.526
	(0.060)	(0.061)
$\hat{\alpha}_{j2}$	2.184	2.130
	(0.139)	(0.141)
$\hat{\alpha}_{j3}$	3.049	3.260
	(0.232)	(0.264)
Trouble falling or staying asleep, or sleeping too much		
$\hat{\beta}_j$	2.215***	1.954***
	(0.204)	(0.179)
$\hat{\alpha}_{j1}$	1.062	0.969
	(0.068)	(0.073)
$\hat{\alpha}_{j2}$	2.144	2.146
	(0.134)	(0.144)
$\hat{\alpha}_{j3}$	2.792	3.201
	(0.203)	(0.265)
Feeling tired or having little energy		
$\hat{\beta}_j$	1.655***	1.868***
	(0.141)	(0.157)
$\hat{\alpha}_{j1}$	0.652	0.417
	(0.065)	(0.060)
$\hat{\alpha}_{j2}$	2.230	1.779
	(0.151)	(0.116)
$\hat{\alpha}_{j3}$	3.481	3.079
	(0.299)	(0.240)
Poor appetite or overeating		
$\hat{\beta}_j$	2.061***	2.160***
	(0.179)	(0.191)
$\hat{\alpha}_{j1}$	0.893	0.755
	(0.064)	(0.063)
$\hat{\alpha}_{j2}$	2.066	1.889
	(0.127)	(0.117)
$\hat{\alpha}_{j3}$	2.934	2.934
	(0.218)	(0.221)
Feeling bad about yourself or that you are a failure or have let yourself or your family down		
$\hat{\beta}_j$	2.729***	2.812***
	(0.316)	(0.312)
$\hat{\alpha}_{j1}$	1.468	1.321
	(0.084)	(0.078)
$\hat{\alpha}_{j2}$	2.493	2.264
	(0.172)	(0.148)
$\hat{\alpha}_{j3}$	3.632	3.600
	(0.441)	(0.377)
Trouble concentrating on things reading newspaper or watching television		
$\hat{\beta}_j$	2.644***	2.586***
	(0.299)	(0.272)
$\hat{\alpha}_{j1}$	1.474	1.294
	(0.085)	(0.079)
$\hat{\alpha}_{j2}$	2.417	2.069
	(0.163)	(0.131)
$\hat{\alpha}_{j3}$	3.401	3.291

Continued on next page

Table D.1: Depression: Parameters of IRT measurement model

	Young Girls	Old Girls
	(0.349)	(0.294)
Moving or speaking so slowly that other people could have noticed, or the opposite,		
$\hat{\beta}_j$	3.313***	2.469***
	(0.426)	(0.279)
$\hat{\alpha}_{j1}$	1.418	1.479
	(0.076)	(0.093)
$\hat{\alpha}_{j2}$	2.390	2.422
	(0.158)	(0.173)
$\hat{\alpha}_{j3}$		3.516
		(0.346)
Thoughts that you would be better off dead or of hurting yourself in some way		
$\hat{\beta}_j$	2.462***	2.335***
	(0.304)	(0.260)
$\hat{\alpha}_{j1}$	1.629	1.493
	(0.102)	(0.096)
$\hat{\alpha}_{j2}$	2.753	2.449
	(0.225)	(0.178)
$\hat{\alpha}_{j3}$		3.573
		(0.361)

Notes: Table presents estimated parameters and standard errors (in parentheses) for IRT measurement model of girls' depression (PHQ-9). Measurement model and estimation procedure described in Section 3.4.6. Column 1 presents the measurement model for depression for younger girls, and Column 2 for older girls. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D.2: Anxiety: Parameters of IRT measurement model

	Young Girls	Old Girls
Feeling nervous, anxious or on edge		
$\hat{\beta}_j$	2.110***	2.025***
	(0.193)	(0.173)
$\hat{\alpha}_{j1}$	0.780	0.497
	(0.061)	(0.059)
$\hat{\alpha}_{j2}$	2.064	1.916
	(0.129)	(0.122)
$\hat{\alpha}_{j3}$	2.590	2.467
	(0.179)	(0.171)
Not being able to stop or control worrying		
$\hat{\beta}_j$	2.696***	2.596***
	(0.271)	(0.235)
$\hat{\alpha}_{j1}$	0.948	0.770
	(0.060)	(0.058)
$\hat{\alpha}_{j2}$	1.908	1.797
	(0.109)	(0.104)
$\hat{\alpha}_{j3}$	2.418	2.501
	(0.155)	(0.167)
Worrying too much about different things		
$\hat{\beta}_j$	2.126***	2.387***
	(0.187)	(0.203)
$\hat{\alpha}_{j1}$	0.706	0.561
	(0.059)	(0.056)
$\hat{\alpha}_{j2}$	1.806	1.690
	(0.108)	(0.099)
$\hat{\alpha}_{j3}$	2.530	2.238
	(0.170)	(0.142)
Trouble relaxing		
$\hat{\beta}_j$	2.001***	2.793***
	(0.204)	(0.269)
$\hat{\alpha}_{j1}$	1.216	0.855
	(0.081)	(0.059)
$\hat{\alpha}_{j2}$	2.195	1.815
	(0.149)	(0.105)
$\hat{\alpha}_{j3}$	2.760	2.423
	(0.209)	(0.159)
Being so restless that it is hard to sit still		

Continued on next page

Table D.2: Anxiety: Parameters of IRT measurement model

	Young Girls	Old Girls
$\hat{\beta}_j$	1.965***	2.784***
	(0.203)	(0.275)
$\hat{\alpha}_{j1}$	1.242	0.894
	(0.084)	(0.060)
$\hat{\alpha}_{j2}$	2.438	1.995
	(0.175)	(0.120)
$\hat{\alpha}_{j3}$	3.019	2.690
	(0.245)	(0.192)
Becoming easily annoyed or irritable		
$\hat{\beta}_j$	1.432***	1.687***
	(0.133)	(0.146)
$\hat{\alpha}_{j1}$	0.665	0.460
	(0.072)	(0.063)
$\hat{\alpha}_{j2}$	2.345	1.897
	(0.179)	(0.131)
$\hat{\alpha}_{j3}$	3.717	3.153
	(0.337)	(0.256)
Feeling afraid as if something awful might happen		
$\hat{\beta}_j$	1.348***	1.875***
	(0.148)	(0.187)
$\hat{\alpha}_{j1}$	1.417	1.154
	(0.120)	(0.084)
$\hat{\alpha}_{j2}$	2.920	2.447
	(0.265)	(0.183)
$\hat{\alpha}_{j3}$	4.858	3.436
	(0.608)	(0.321)

Notes: Table presents estimated parameters and standard errors (in parentheses) for IRT measurement model of girls' anxiety (GAD-7). Measurement model and estimation procedure described in Section 3.4.6. Column 1 presents the measurement model for anxiety for younger girls, and Column 2 for older girls. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D.3: Rumination: Parameters of IRT measurement model

	Young Girls	Old Girls
Think What am I doing to deserve this		
$\hat{\beta}_j$	2.011***	1.960***
	(0.128)	(0.135)
$\hat{\alpha}_{j1}$	-0.448	-0.348
	(0.056)	(0.058)
$\hat{\alpha}_{j2}$	0.821	0.938
	(0.065)	(0.074)
$\hat{\alpha}_{j3}$	1.583	1.583
	(0.094)	(0.102)
Analyze recent events to try to understand why you are depressed		
$\hat{\beta}_j$	2.198***	2.223***
	(0.138)	(0.152)
$\hat{\alpha}_{j1}$	-0.621	-0.490
	(0.056)	(0.057)
$\hat{\alpha}_{j2}$	0.748	0.842
	(0.061)	(0.068)
$\hat{\alpha}_{j3}$	1.508	1.561
	(0.088)	(0.096)
Go away by yourself and think about why you feel this way		
$\hat{\beta}_j$	2.061***	2.012***
	(0.128)	(0.136)
$\hat{\alpha}_{j1}$	-0.533	-0.627
	(0.057)	(0.063)
$\hat{\alpha}_{j2}$	0.745	0.747
	(0.062)	(0.067)
$\hat{\alpha}_{j3}$	1.632	1.603
	(0.094)	(0.102)
Write down what you are thinking about and analyze it		
$\hat{\beta}_j$	0.599***	0.651***

Continued on next page

Table D.3: Rumination: Parameters of IRT measurement model

	Young Girls	Old Girls
	(0.075)	(0.081)
$\hat{\alpha}_{j1}$	-3.335	-3.323
	(0.427)	(0.418)
$\hat{\alpha}_{j2}$	-1.500	-1.359
	(0.221)	(0.203)
$\hat{\alpha}_{j3}$	0.504	0.517
	(0.130)	(0.130)
Think about a recent situation, wishing it had gone better		
$\hat{\beta}_j$	2.133***	2.537***
	(0.133)	(0.171)
$\hat{\alpha}_{j1}$	-0.574	-0.507
	(0.056)	(0.055)
$\hat{\alpha}_{j2}$	0.771	0.812
	(0.062)	(0.063)
$\hat{\alpha}_{j3}$	1.702	1.601
	(0.096)	(0.093)
Think Why do I have problems other people do not have		
$\hat{\beta}_j$	1.949***	2.019***
	(0.123)	(0.136)
$\hat{\alpha}_{j1}$	-0.591	-0.531
	(0.059)	(0.061)
$\hat{\alpha}_{j2}$	0.671	0.824
	(0.062)	(0.069)
$\hat{\alpha}_{j3}$	1.472	1.578
	(0.090)	(0.100)
Think Why can not I handle things better		
$\hat{\beta}_j$	2.215***	2.205***
	(0.139)	(0.148)
$\hat{\alpha}_{j1}$	-0.479	-0.550
	(0.054)	(0.059)
$\hat{\alpha}_{j2}$	0.780	0.811
	(0.061)	(0.066)
$\hat{\alpha}_{j3}$	1.553	1.512
	(0.089)	(0.094)
Think about how sad you feel		
$\hat{\beta}_j$	2.425***	2.347***
	(0.154)	(0.157)
$\hat{\alpha}_{j1}$	-0.617	-0.611
	(0.054)	(0.058)
$\hat{\alpha}_{j2}$	0.762	0.869
	(0.060)	(0.067)
$\hat{\alpha}_{j3}$	1.628	1.675
	(0.089)	(0.098)
Analyze your personality to try to understand why you are depressed		
$\hat{\beta}_j$	1.791***	1.975***
	(0.114)	(0.132)
$\hat{\alpha}_{j1}$	-0.984	-0.951
	(0.072)	(0.073)
$\hat{\alpha}_{j2}$	0.283	0.277
	(0.057)	(0.057)
$\hat{\alpha}_{j3}$	1.302	1.219
	(0.086)	(0.085)
Go someplace alone to think about your feelings		
$\hat{\beta}_j$	1.676***	1.524***
	(0.108)	(0.109)
$\hat{\alpha}_{j1}$	-0.887	-0.744
	(0.072)	(0.076)
$\hat{\alpha}_{j2}$	0.567	0.748
	(0.064)	(0.076)
$\hat{\alpha}_{j3}$	1.551	1.620
	(0.099)	(0.116)

Notes: Table presents estimated parameters and standard errors (in parentheses) for IRT measurement model of girls' rumination. Measurement model and estimation procedure described in Section 3.4.6. Column 1 presents the measurement model for rumination for younger girls, and Column 2 for older girls. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D.4: Gender attitudes: Parameters of IRT measurement model

	Young Girls	Old Girls	Carers
Boys should be given more privilege as compared to the girls [recoded]			
$\hat{\beta}_j$	1.008***	1.102***	0.903***
	(0.088)	(0.095)	(0.065)
$\hat{\alpha}_{j1}$	-2.616	-2.162	-2.011
	(0.224)	(0.185)	(0.148)
$\hat{\alpha}_{j2}$	-1.357	-0.996	-0.699
	(0.127)	(0.107)	(0.081)
$\hat{\alpha}_{j3}$	0.375	0.431	1.226
	(0.085)	(0.085)	(0.104)
Women or girls should work only if there are monetary needs in their family [recoded]			
$\hat{\beta}_j$	1.034***	1.168***	1.089***
	(0.085)	(0.097)	(0.071)
$\hat{\alpha}_{j1}$	-1.106	-1.007	-0.724
	(0.108)	(0.103)	(0.071)
$\hat{\alpha}_{j2}$	0.292	0.193	0.693
	(0.080)	(0.076)	(0.070)
$\hat{\alpha}_{j3}$	1.553	1.388	2.046
	(0.139)	(0.125)	(0.132)
Only bad girls make male friends [recoded]			
$\hat{\beta}_j$	0.937***	1.147***	1.069***
	(0.083)	(0.097)	(0.071)
$\hat{\alpha}_{j1}$	-2.011	-1.499	-0.828
	(0.181)	(0.133)	(0.077)
$\hat{\alpha}_{j2}$	-0.866	-0.719	0.132
	(0.104)	(0.091)	(0.059)
$\hat{\alpha}_{j3}$	0.996	0.830	1.823
	(0.116)	(0.097)	(0.120)
A man should have the final say in all family matters [recoded]			
$\hat{\beta}_j$	1.395***	1.472***	1.421***
	(0.100)	(0.113)	(0.086)
$\hat{\alpha}_{j1}$	-0.683	-0.579	-0.147
	(0.073)	(0.074)	(0.050)
$\hat{\alpha}_{j2}$	0.471	0.399	0.893
	(0.070)	(0.069)	(0.065)
$\hat{\alpha}_{j3}$	1.584	1.392	2.064
	(0.115)	(0.109)	(0.116)
Men should be more educated than their wives [recoded]			
$\hat{\beta}_j$	1.661***	1.652***	1.573***
	(0.113)	(0.120)	(0.091)
$\hat{\alpha}_{j1}$	-0.710	-0.668	-0.297
	(0.067)	(0.071)	(0.049)
$\hat{\alpha}_{j2}$	0.337	0.363	0.941
	(0.061)	(0.064)	(0.062)
$\hat{\alpha}_{j3}$	1.303	1.386	2.024
	(0.091)	(0.100)	(0.106)
Boys are naturally better than girls in studies [recoded]			
$\hat{\beta}_j$	1.404***	1.455***	1.308***
	(0.101)	(0.109)	(0.076)
$\hat{\alpha}_{j1}$	-1.653	-1.497	-1.408
	(0.117)	(0.114)	(0.087)
$\hat{\alpha}_{j2}$	-0.606	-0.487	-0.200
	(0.070)	(0.071)	(0.053)
$\hat{\alpha}_{j3}$	0.767	0.763	1.272
	(0.079)	(0.081)	(0.080)
A daughter deserves to be beaten if she does not obey her parents [recoded]			
$\hat{\beta}_j$	1.035***	1.192***	1.113***
	(0.085)	(0.098)	(0.074)
$\hat{\alpha}_{j1}$	-0.803	-0.872	-0.316
	(0.094)	(0.096)	(0.061)
$\hat{\alpha}_{j2}$	0.591	0.362	1.067
	(0.088)	(0.077)	(0.084)
$\hat{\alpha}_{j3}$	1.945	1.617	2.515
	(0.162)	(0.137)	(0.161)
Girls should be married early to protect them from sexual harassment [recoded]			
$\hat{\beta}_j$	2.112***	2.189***	2.056***

Continued on next page

Table D.4: Gender attitudes: Parameters of IRT measurement model

	Young Girls	Old Girls	Carers
	(0.141)	(0.154)	(0.110)
$\hat{\alpha}_{j1}$	-1.260	-1.125	-0.815
	(0.077)	(0.077)	(0.052)
$\hat{\alpha}_{j2}$	-0.319	-0.341	0.140
	(0.052)	(0.055)	(0.041)
$\hat{\alpha}_{j3}$	0.663	0.630	1.178
	(0.062)	(0.063)	(0.061)
Girls should be married early to ease family financial burden [recoded]			
$\hat{\beta}_j$	1.960***	2.107***	1.658***
	(0.137)	(0.151)	(0.092)
$\hat{\alpha}_{j1}$	-1.700	-1.473	-1.252
	(0.102)	(0.094)	(0.070)
$\hat{\alpha}_{j2}$	-0.719	-0.543	-0.277
	(0.062)	(0.060)	(0.047)
$\hat{\alpha}_{j3}$	0.425	0.476	1.005
	(0.059)	(0.060)	(0.062)
Money should be saved for a girls dowry and not her education [recoded]			
$\hat{\beta}_j$	1.842***	1.765***	1.463***
	(0.134)	(0.134)	(0.085)
$\hat{\alpha}_{j1}$	-2.089	-2.003	-1.838
	(0.132)	(0.135)	(0.100)
$\hat{\alpha}_{j2}$	-1.248	-1.120	-0.799
	(0.084)	(0.087)	(0.061)
$\hat{\alpha}_{j3}$	0.117	0.179	0.752
	(0.056)	(0.060)	(0.060)
If a girl is a victim of some sexual abuse, it is the fault of the girl [recoded]			
$\hat{\beta}_j$	1.802***	1.553***	1.693***
	(0.132)	(0.123)	(0.093)
$\hat{\alpha}_{j1}$	-1.920	-2.067	-1.357
	(0.121)	(0.148)	(0.073)
$\hat{\alpha}_{j2}$	-1.084	-1.221	-0.444
	(0.077)	(0.098)	(0.049)
$\hat{\alpha}_{j3}$	0.046	0.016	0.732
	(0.056)	(0.063)	(0.054)
A woman should tolerate violence in order to keep her family together [recoded]			
$\hat{\beta}_j$	1.456***	1.366***	1.362***
	(0.101)	(0.104)	(0.079)
$\hat{\alpha}_{j1}$	-1.424	-1.297	-0.794
	(0.100)	(0.106)	(0.063)
$\hat{\alpha}_{j2}$	-0.023	0.181	0.901
	(0.061)	(0.068)	(0.066)
$\hat{\alpha}_{j3}$	1.069	1.280	2.171
	(0.088)	(0.106)	(0.120)
There are times when a woman deserves to be beaten [recoded]			
$\hat{\beta}_j$	1.420***	1.386***	1.512***
	(0.101)	(0.105)	(0.083)
$\hat{\alpha}_{j1}$	-2.074	-2.058	-1.570
	(0.140)	(0.151)	(0.085)
$\hat{\alpha}_{j2}$	-0.578	-0.577	-0.068
	(0.069)	(0.076)	(0.048)
$\hat{\alpha}_{j3}$	0.798	0.798	1.119
	(0.080)	(0.084)	(0.069)
Girls who are highly educated indulge in improper behaviour [recoded]			
$\hat{\beta}_j$	1.267***	1.188***	1.265***
	(0.097)	(0.098)	(0.075)
$\hat{\alpha}_{j1}$	-2.270	-2.261	-1.876
	(0.165)	(0.182)	(0.110)
$\hat{\alpha}_{j2}$	-0.923	-0.906	-0.513
	(0.086)	(0.097)	(0.059)
$\hat{\alpha}_{j3}$	0.566	0.561	0.989
	(0.078)	(0.084)	(0.072)

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 3.4.6. Column 1 and 2 present the measurement model for girls' norms while 3 does the same for carers. Items are coded such that a higher value of the factor represents more progressive gender norms. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D.5: Self-esteem: Parameters of IRT measurement model

	Young Girls	Old Girls
I do lots of important things		
$\hat{\beta}_j$	1.433***	1.776***
	(0.119)	(0.147)
$\hat{\alpha}_{j1}$	-3.963	-3.533
	(0.356)	(0.312)
$\hat{\alpha}_{j2}$	-2.653	-2.225
	(0.191)	(0.150)
$\hat{\alpha}_{j3}$	-0.099	-0.157
	(0.063)	(0.060)
In general, I like being the way I am		
$\hat{\beta}_j$	1.011***	1.268***
	(0.101)	(0.120)
$\hat{\alpha}_{j1}$	-4.674	-3.907
	(0.485)	(0.372)
$\hat{\alpha}_{j2}$	-3.375	-2.798
	(0.312)	(0.233)
$\hat{\alpha}_{j3}$	-0.661	-0.694
	(0.090)	(0.082)
Overall, I have a lot to be proud of		
$\hat{\beta}_j$	1.082***	1.012***
	(0.092)	(0.093)
$\hat{\alpha}_{j1}$	-3.325	-3.275
	(0.277)	(0.291)
$\hat{\alpha}_{j2}$	-1.808	-1.694
	(0.145)	(0.149)
$\hat{\alpha}_{j3}$	0.430	0.328
	(0.083)	(0.090)
I can do things as well as most people		
$\hat{\beta}_j$	2.237***	2.432***
	(0.188)	(0.204)
$\hat{\alpha}_{j1}$	-3.031	-2.984
	(0.227)	(0.219)
$\hat{\alpha}_{j2}$	-2.212	-2.070
	(0.132)	(0.121)
$\hat{\alpha}_{j3}$	-0.193	-0.209
	(0.052)	(0.053)
Other people think I am a good person		
$\hat{\beta}_j$	1.805***	1.979***
	(0.145)	(0.159)
$\hat{\alpha}_{j1}$	-3.471	-3.132
	(0.289)	(0.240)
$\hat{\alpha}_{j2}$	-2.515	-2.182
	(0.165)	(0.137)
$\hat{\alpha}_{j3}$	-0.036	-0.024
	(0.056)	(0.057)
A lot of things about me are good		
$\hat{\beta}_j$	2.073***	2.294***
	(0.172)	(0.193)
$\hat{\alpha}_{j1}$	-3.508	-3.656
	(0.313)	(0.370)
$\hat{\alpha}_{j2}$	-2.529	-2.153
	(0.163)	(0.131)
$\hat{\alpha}_{j3}$	-0.115	-0.130
	(0.053)	(0.054)
I'm as good as most other people		
$\hat{\beta}_j$	2.192***	2.073***
	(0.184)	(0.173)
$\hat{\alpha}_{j1}$	-3.402	-3.468
	(0.297)	(0.311)
$\hat{\alpha}_{j2}$	-2.478	-2.427
	(0.157)	(0.158)
$\hat{\alpha}_{j3}$	-0.096	-0.117
	(0.052)	(0.056)
When I do something, I do it well		
$\hat{\beta}_j$	2.017***	2.343***

Continued on next page

Table D.5: Self-esteem: Parameters of IRT measurement model

	Young Girls	Old Girls
	(0.167)	(0.202)
$\hat{\alpha}_{j1}$	-3.725	-3.625
	(0.358)	(0.366)
$\hat{\alpha}_{j2}$	-2.325	-2.376
	(0.146)	(0.150)
$\hat{\alpha}_{j3}$	-0.207	-0.265
	(0.054)	(0.054)

Notes: Table presents estimated parameters and standard errors (in parentheses) for IRT measurement model of girls' self-esteem. Measurement model and estimation procedure described in Section 3.4.6. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D.6: Self-efficacy: Parameters of IRT measurement model

	Young Girls	Old Girls
I can usually handle whatever comes my way.		
$\hat{\beta}_j$	1.056***	1.325***
	(0.098)	(0.121)
$\hat{\alpha}_{j1}$	-4.970	-4.580
	(0.519)	(0.511)
$\hat{\alpha}_{j2}$	-3.521	-3.059
	(0.317)	(0.262)
$\hat{\alpha}_{j3}$	-0.445	-0.506
	(0.081)	(0.075)
I can always manage to solve difficult problems if I try hard enough.		
$\hat{\beta}_j$	1.593***	1.941***
	(0.118)	(0.145)
$\hat{\alpha}_{j1}$	-3.589	-3.129
	(0.278)	(0.231)
$\hat{\alpha}_{j2}$	-2.209	-1.973
	(0.141)	(0.121)
$\hat{\alpha}_{j3}$	-0.213	-0.123
	(0.059)	(0.057)
If someone opposes me, I can find the means and ways to get what I want.		
$\hat{\beta}_j$	1.843***	1.930***
	(0.123)	(0.134)
$\hat{\alpha}_{j1}$	-2.583	-2.382
	(0.154)	(0.147)
$\hat{\alpha}_{j2}$	-1.520	-1.307
	(0.090)	(0.084)
$\hat{\alpha}_{j3}$	0.094	0.096
	(0.056)	(0.058)
It is easy for me to stick to my aims and accomplish my goals.		
$\hat{\beta}_j$	1.932***	2.115***
	(0.132)	(0.150)
$\hat{\alpha}_{j1}$	-3.121	-2.928
	(0.209)	(0.198)
$\hat{\alpha}_{j2}$	-1.715	-1.465
	(0.098)	(0.089)
$\hat{\alpha}_{j3}$	0.032	0.056
	(0.054)	(0.055)
I am confident that I could deal efficiently with unexpected events.		
$\hat{\beta}_j$	2.502***	2.273***
	(0.168)	(0.159)
$\hat{\alpha}_{j1}$	-2.633	-2.760
	(0.147)	(0.179)
$\hat{\alpha}_{j2}$	-1.469	-1.480
	(0.078)	(0.087)
$\hat{\alpha}_{j3}$	0.140	0.089
	(0.050)	(0.054)
Thanks to my resourcefulness, I know how to handle unforeseen situations.		
$\hat{\beta}_j$	2.547***	2.290***
	(0.173)	(0.163)
$\hat{\alpha}_{j1}$	-2.750	-2.751
	(0.159)	(0.176)

Continued on next page

Table D.6: Self-efficacy: Parameters of IRT measurement model

	Young Girls	Old Girls
$\hat{\alpha}_{j2}$	-1.572	-1.644
	(0.082)	(0.094)
$\hat{\alpha}_{j3}$	0.152	0.104
	(0.049)	(0.054)
I can solve most problems if I invest the necessary effort.		
$\hat{\beta}_j$	2.442***	3.196***
	(0.170)	(0.251)
$\hat{\alpha}_{j1}$	-3.067	-2.452
	(0.205)	(0.141)
$\hat{\alpha}_{j2}$	-1.809	-1.644
	(0.095)	(0.085)
$\hat{\alpha}_{j3}$	-0.059	-0.067
	(0.049)	(0.048)
relying on my coping abilities to remain calm		
$\hat{\beta}_j$	2.327***	2.619***
	(0.156)	(0.191)
$\hat{\alpha}_{j1}$	-2.758	-2.674
	(0.163)	(0.167)
$\hat{\alpha}_{j2}$	-1.605	-1.520
	(0.086)	(0.085)
$\hat{\alpha}_{j3}$	0.015	-0.027
	(0.050)	(0.051)
When I am confronted with a problem, I can usually find several solutions.		
$\hat{\beta}_j$	2.293***	2.247***
	(0.155)	(0.160)
$\hat{\alpha}_{j1}$	-2.900	-2.962
	(0.178)	(0.208)
$\hat{\alpha}_{j2}$	-1.636	-1.624
	(0.088)	(0.094)
$\hat{\alpha}_{j3}$	0.123	0.063
	(0.051)	(0.054)
If I am in trouble, I can usually think of a solution.		
$\hat{\beta}_j$	2.202***	1.873***
	(0.154)	(0.143)
$\hat{\alpha}_{j1}$	-2.971	-3.476
	(0.191)	(0.288)
$\hat{\alpha}_{j2}$	-1.941	-1.983
	(0.106)	(0.125)
$\hat{\alpha}_{j3}$	-0.012	-0.163
	(0.051)	(0.058)

Notes: Table presents estimated parameters and standard errors (in parentheses) for IRT measurement model of girls' self-efficacy. Measurement model and estimation procedure described in Section 3.4.6. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D.7: Peer relations: Parameters of IRT measurement model

	Young Girls	Old Girls
I have lots of friends		
$\hat{\beta}_j$	1.215***	1.212***
	(0.131)	(0.123)
$\hat{\alpha}_{j1}$	-3.642	-3.349
	(0.354)	(0.310)
$\hat{\alpha}_{j2}$	-2.354	-1.958
	(0.209)	(0.170)
$\hat{\alpha}_{j3}$	-0.600	-0.284
	(0.081)	(0.075)
Other people of my age want me to be their friend		
$\hat{\beta}_j$	1.762***	1.902***
	(0.162)	(0.170)
$\hat{\alpha}_{j1}$	-2.703	-2.753
	(0.194)	(0.199)
$\hat{\alpha}_{j2}$	-1.876	-1.781
	(0.125)	(0.116)
$\hat{\alpha}_{j3}$	0.089	0.061

Continued on next page

Table D.7: Peer relations: Parameters of IRT measurement model

	Young Girls	Old Girls
	(0.058)	(0.059)
I have more friends than most other people my age		
$\hat{\beta}_j$	1.705***	1.820***
	(0.160)	(0.159)
$\hat{\alpha}_{j1}$	-2.618	-2.634
	(0.190)	(0.187)
$\hat{\alpha}_{j2}$	-1.451	-1.252
	(0.102)	(0.089)
$\hat{\alpha}_{j3}$	0.330	0.223
	(0.062)	(0.062)
I am popular with people of my own age		
$\hat{\beta}_j$	1.509***	1.931***
	(0.151)	(0.191)
$\hat{\alpha}_{j1}$	-4.003	-3.932
	(0.400)	(0.453)
$\hat{\alpha}_{j2}$	-2.425	-2.148
	(0.189)	(0.151)
$\hat{\alpha}_{j3}$	-0.004	-0.042
	(0.062)	(0.058)
Most other people my age like me		
$\hat{\beta}_j$	1.394***	1.775***
	(0.136)	(0.163)
$\hat{\alpha}_{j1}$	-3.467	-3.080
	(0.304)	(0.246)
$\hat{\alpha}_{j2}$	-2.290	-1.879
	(0.178)	(0.129)
$\hat{\alpha}_{j3}$	0.119	0.003
	(0.065)	(0.060)

Notes: Table presents estimated parameters and standard errors (in parentheses) for IRT measurement model of girls' peer relations. Measurement model and estimation procedure described in Section 3.4.6. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D.8: Resilience: Parameters of IRT measurement model

	Young Girls	Old Girls
I am able to adapt when changes occur		
$\hat{\beta}_j$	1.932***	1.865***
	(0.132)	(0.138)
$\hat{\alpha}_{j1}$	-2.361	-2.619
	(0.140)	(0.176)
$\hat{\alpha}_{j2}$	-1.652	-1.742
	(0.096)	(0.110)
$\hat{\alpha}_{j3}$	0.033	-0.095
	(0.054)	(0.058)
I can deal with whatever comes my way		
$\hat{\beta}_j$	1.921***	2.151***
	(0.131)	(0.154)
$\hat{\alpha}_{j1}$	-2.468	-2.546
	(0.144)	(0.160)
$\hat{\alpha}_{j2}$	-1.659	-1.573
	(0.095)	(0.094)
$\hat{\alpha}_{j3}$	0.130	0.095
	(0.055)	(0.055)
I try to see the humorous side of things when I am faced with problems		
$\hat{\beta}_j$	0.576***	0.697***
	(0.073)	(0.078)
$\hat{\alpha}_{j1}$	-3.835	-3.194
	(0.475)	(0.351)
$\hat{\alpha}_{j2}$	-1.477	-1.187
	(0.202)	(0.155)
$\hat{\alpha}_{j3}$	2.062	1.504
	(0.282)	(0.199)
Having to cope with stress can make me stronger		

Continued on next page

Table D.8: Resilience: Parameters of IRT measurement model

	Young Girls	Old Girls
$\hat{\beta}_j$	1.827***	1.808***
	(0.122)	(0.127)
$\hat{\alpha}_{j1}$	-2.367	-2.615
	(0.138)	(0.168)
$\hat{\alpha}_{j2}$	-1.454	-1.431
	(0.087)	(0.092)
$\hat{\alpha}_{j3}$	0.306	0.250
	(0.058)	(0.061)
I tend to bounce back after illness, injury, or other hardships		
$\hat{\beta}_j$	1.689***	1.701***
	(0.112)	(0.120)
$\hat{\alpha}_{j1}$	-2.404	-2.569
	(0.144)	(0.166)
$\hat{\alpha}_{j2}$	-1.314	-1.391
	(0.084)	(0.092)
$\hat{\alpha}_{j3}$	0.330	0.273
	(0.061)	(0.063)
Even when things look hopeless, I do not give up		
$\hat{\beta}_j$	2.381***	2.721***
	(0.161)	(0.195)
$\hat{\alpha}_{j1}$	-2.474	-2.510
	(0.138)	(0.147)
$\hat{\alpha}_{j2}$	-1.541	-1.388
	(0.082)	(0.078)
$\hat{\alpha}_{j3}$	0.104	0.070
	(0.050)	(0.051)
Under pressure, I stay focused and think clearly		
$\hat{\beta}_j$	2.455***	2.143***
	(0.164)	(0.152)
$\hat{\alpha}_{j1}$	-2.478	-2.599
	(0.137)	(0.163)
$\hat{\alpha}_{j2}$	-1.425	-1.507
	(0.076)	(0.090)
$\hat{\alpha}_{j3}$	0.217	0.082
	(0.051)	(0.055)
I am not easily discouraged by failure		
$\hat{\beta}_j$	2.021***	2.212***
	(0.133)	(0.151)
$\hat{\alpha}_{j1}$	-2.683	-2.416
	(0.161)	(0.142)
$\hat{\alpha}_{j2}$	-1.455	-1.374
	(0.084)	(0.081)
$\hat{\alpha}_{j3}$	0.273	0.189
	(0.055)	(0.056)
I think of myself as a strong person when dealing with challenges or difficulties		
$\hat{\beta}_j$	2.195***	2.474***
	(0.147)	(0.176)
$\hat{\alpha}_{j1}$	-2.669	-2.530
	(0.157)	(0.151)
$\hat{\alpha}_{j2}$	-1.487	-1.462
	(0.083)	(0.083)
$\hat{\alpha}_{j3}$	0.222	0.129
	(0.053)	(0.053)
I am able to handle unpleasant or painful feelings like sadness or fear or anger		
$\hat{\beta}_j$	1.857***	2.192***
	(0.125)	(0.155)
$\hat{\alpha}_{j1}$	-2.718	-2.799
	(0.167)	(0.181)
$\hat{\alpha}_{j2}$	-1.522	-1.551
	(0.090)	(0.091)
$\hat{\alpha}_{j3}$	0.247	0.183
	(0.057)	(0.056)

Notes: Table presents estimated parameters and standard errors (in parentheses) for IRT measurement model of girls' resilience. Measurement model and estimation procedure described in Section 3.4.6. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D.9: Vigilance: Parameters of IRT measurement model

	Young Girls	Old Girls
Do you like to consider all of the alternatives when you make decisions		
$\hat{\beta}_j$	1.936***	1.824***
	(0.134)	(0.138)
$\hat{\alpha}_{j1}$	-2.100	-2.489
	(0.123)	(0.166)
$\hat{\alpha}_{j2}$	-0.718	-0.951
	(0.063)	(0.075)
$\hat{\alpha}_{j3}$	0.188	0.102
	(0.055)	(0.060)
Do you try to find out the disadvantages of all alternatives		
$\hat{\beta}_j$	1.964***	2.106***
	(0.131)	(0.152)
$\hat{\alpha}_{j1}$	-1.956	-2.308
	(0.114)	(0.143)
$\hat{\alpha}_{j2}$	-0.460	-0.614
	(0.057)	(0.061)
$\hat{\alpha}_{j3}$	0.648	0.449
	(0.062)	(0.061)
Do you consider how best to carry out a decision		
$\hat{\beta}_j$	2.498***	2.240***
	(0.171)	(0.165)
$\hat{\alpha}_{j1}$	-2.101	-2.213
	(0.113)	(0.135)
$\hat{\alpha}_{j2}$	-0.531	-0.759
	(0.053)	(0.063)
$\hat{\alpha}_{j3}$	0.531	0.364
	(0.054)	(0.058)
Do you like to collect a lot of information when you make decisions		
$\hat{\beta}_j$	2.374***	2.598***
	(0.162)	(0.200)
$\hat{\alpha}_{j1}$	-1.976	-2.196
	(0.108)	(0.130)
$\hat{\alpha}_{j2}$	-0.600	-0.769
	(0.055)	(0.060)
$\hat{\alpha}_{j3}$	0.331	0.238
	(0.052)	(0.053)
Do you like to be clear about your objectives before choosing		
$\hat{\beta}_j$	2.244***	2.073***
	(0.151)	(0.152)
$\hat{\alpha}_{j1}$	-2.114	-2.322
	(0.116)	(0.147)
$\hat{\alpha}_{j2}$	-0.501	-0.782
	(0.055)	(0.066)
$\hat{\alpha}_{j3}$	0.591	0.423
	(0.058)	(0.061)
Do you take a lot of care before choosing		
$\hat{\beta}_j$	1.882***	1.453***
	(0.127)	(0.114)
$\hat{\alpha}_{j1}$	-2.286	-2.842
	(0.136)	(0.209)
$\hat{\alpha}_{j2}$	-0.618	-0.900
	(0.062)	(0.084)
$\hat{\alpha}_{j3}$	0.520	0.337
	(0.061)	(0.070)

Notes: Table presents estimated parameters and standard errors (in parentheses) for IRT measurement model of girls' vigilance in decision making. Measurement model and estimation procedure described in Section 3.4.6. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D.10: Buck passing: Parameters of IRT measurement model

	Young Girls	Old Girls
Do you avoid making decisions		
$\hat{\beta}_j$	1.102***	1.229***

Continued on next page

Table D.10: Buck passing: Parameters of IRT measurement model

	Young Girls	Old Girls
	(0.107)	(0.127)
$\hat{\alpha}_{j1}$	-0.184	-0.041
	(0.075)	(0.073)
$\hat{\alpha}_{j2}$	1.594	1.568
	(0.147)	(0.145)
$\hat{\alpha}_{j3}$	2.650	2.417
	(0.230)	(0.212)
Do you not make decisions unless you really have to		
$\hat{\beta}_j$	0.730***	0.748***
	(0.087)	(0.094)
$\hat{\alpha}_{j1}$	-2.997	-2.932
	(0.346)	(0.356)
$\hat{\alpha}_{j2}$	0.476	0.338
	(0.110)	(0.109)
$\hat{\alpha}_{j3}$	2.483	2.175
	(0.285)	(0.265)
Do you prefer to leave decision to others		
$\hat{\beta}_j$	2.582***	2.501***
	(0.285)	(0.317)
$\hat{\alpha}_{j1}$	-0.218	-0.178
	(0.050)	(0.053)
$\hat{\alpha}_{j2}$	0.717	0.797
	(0.060)	(0.068)
$\hat{\alpha}_{j3}$	1.454	1.550
	(0.090)	(0.106)
Do you not like to take responsibility for making decisions		
$\hat{\beta}_j$	1.077***	1.132***
	(0.103)	(0.115)
$\hat{\alpha}_{j1}$	-0.585	-0.458
	(0.087)	(0.085)
$\hat{\alpha}_{j2}$	1.102	1.180
	(0.113)	(0.121)
$\hat{\alpha}_{j3}$	2.304	2.345
	(0.199)	(0.211)
If a decision can be made by you or another person do you let the other person make		
$\hat{\beta}_j$	1.503***	1.070***
	(0.129)	(0.111)
$\hat{\alpha}_{j1}$	-1.207	-1.584
	(0.095)	(0.155)
$\hat{\alpha}_{j2}$	0.355	0.432
	(0.064)	(0.087)
$\hat{\alpha}_{j3}$	1.544	1.956
	(0.115)	(0.185)
Do you prefer that people who are better informed decide for you		
$\hat{\beta}_j$	0.804***	0.626***
	(0.090)	(0.093)
$\hat{\alpha}_{j1}$	-3.161	-4.105
	(0.340)	(0.591)
$\hat{\alpha}_{j2}$	-0.364	-0.211
	(0.100)	(0.125)
$\hat{\alpha}_{j3}$	1.485	2.166
	(0.172)	(0.316)

Notes: Table presents estimated parameters and standard errors (in parentheses) for IRT measurement model of girls' buck passing in decision making. Measurement model and estimation procedure described in Section 3.4.6. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

E Heterogeneity by Attendance in a Potential Outcomes/Potential Attendance Framework

This appendix provides the formal justification for the heterogeneity analysis reported in Tables 6 and 7. The analysis is restricted to the two treatment arms; the control group is excluded throughout, as in the empirical specifications of those tables.

Setup. Clusters c are randomized to one of two arms: $Z_c \in \{G, GC\}$, where G denotes "Girl Groups only" and GC denotes "Girl Groups + Community Campaigns." For individual i in cluster c, let $Y_i(z)$ be the outcome of interest if the cluster receives arm $z \in \{G, GC\}$. Fix an attendance threshold h (e.g., any/50%/75% of sessions), and define the potential attendance indicator

$$A_{i,h}(z) \in \{0, 1\}, \qquad z \in \{G, GC\},$$

which equals 1 if i would meet attendance threshold h under arm z. The four possible "potential attendance" combinations at threshold h are

$$S_{i,h} \equiv \big(A_{i,h}(G),\, A_{i,h}(GC)\big) \in \{(0,0), (1,1), (0,1), (1,0)\}.$$

That is, individual i might not meet the threshold in either treatment regime, might meet it in both, might meet it only in G, or might meet it only in GC. Cluster randomization implies

$$Z_c \perp \big(Y_i(G),\, Y_i(GC),\, A_{i,h}(G),\, A_{i,h}(GC)\big).$$

Assumption of attendance invariance. Our analysis rests on the following assumption.

Assumption 1 (Attendance Invariance)**.** *For all i,* $A_{i,h}(G) = A_{i,h}(GC)$.

In words, adding Community Campaigns does not change whether a girl meets attendance threshold h. This rules out girls who would meet the threshold under G but not GC, or vice versa, so that $S_{i,h} \in \{(0,0), (1,1)\}$. Under Assumption 1 we may write $A_{i,h} \equiv A_{i,h}(G) = A_{i,h}(GC)$ for the common potential attendance indicator. The assumption is supported empirically: Table B.5 documents nearly identical attendance rates across the two treatment arms at every threshold, and Table B.21 shows that the pattern of selection into attendance on observed baseline characteristics is the same across arms.

Average treatment effects. For $s \in \{(0,0),(1,1)\}$, define

$$\tau_{s,h} \;=\; \mathbb{E}[Y_i(GC) - Y_i(G) \mid S_{i,h} = s]\,.$$

$\tau_{00,h}$ is the average effect of Community Campaigns for girls who would not have attended the Girl Groups under either arm. $\tau_{11,h}$ is the corresponding effect for girls who would have attended under both arms.

Identification. Given Assumption 1, cluster randomization, and SUTVA (the Stable Unit Treatment Value Assumption, which requires no interference in outcomes between clusters), we have:

$$\begin{aligned}
\mathbb{E}[Y_i(GC) \mid S_{i,h} = (0,0)] &= \mathbb{E}[Y_i \mid Z_c = GC,\ A_{i,h} = 0]\,,\\
\mathbb{E}[Y_i(G) \mid S_{i,h} = (0,0)] &= \mathbb{E}[Y_i \mid Z_c = G,\ A_{i,h} = 0]\,,\\
\mathbb{E}[Y_i(GC) \mid S_{i,h} = (1,1)] &= \mathbb{E}[Y_i \mid Z_c = GC,\ A_{i,h} = 1]\,,\\
\mathbb{E}[Y_i(G) \mid S_{i,h} = (1,1)] &= \mathbb{E}[Y_i \mid Z_c = G,\ A_{i,h} = 1]\,.
\end{aligned}$$

Thus both treatment effects of interest are identified by differences in means within observed attendance cells:

$$\begin{aligned}
\tau_{00,h} \;&=\; \mathbb{E}[Y_i \mid Z_c = GC,\ A_{i,h} = 0] \;-\; \mathbb{E}[Y_i \mid Z_c = G,\ A_{i,h} = 0]\,,\\
\tau_{11,h} \;&=\; \mathbb{E}[Y_i \mid Z_c = GC,\ A_{i,h} = 1] \;-\; \mathbb{E}[Y_i \mid Z_c = G,\ A_{i,h} = 1]\,.
\end{aligned}$$

These parameters are identified without reliance on control variables. The interaction regression

$$Y_{ic} \;=\; \alpha + \beta\,\mathbf{1}\{Z_c = GC\} + \kappa\,\mathbf{1}\{Z_c = GC\}\cdot A_{i,h} + \theta\,A_{i,h} + X_{ic}'\gamma + u_{ic}, \tag{E.1}$$

which additionally conditions on a vector of baseline covariates X_{ic}, recovers the same treatment effects under standard regression-adjustment arguments; the inclusion of X_{ic} does not alter identification but improves precision. The coefficients satisfy

$$\beta = \tau_{00,h}, \qquad \beta + \kappa = \tau_{11,h}, \qquad \kappa = \tau_{11,h} - \tau_{00,h}.$$

Interpretation and connection to the paper's findings. The coefficient κ measures the differential additional effect of Community Campaigns between attenders and non-attenders. A finding of $\kappa \approx 0$ implies that the additional effect of the Community Campaigns is the same

for girls who attended the Girl Groups and for those who did not. As discussed in Section 5, this is precisely what Tables 6 and 7 show: the improvements in mental health and the shift towards more progressive gender attitudes driven by the addition of Community Campaigns are equally large for non-participating girls as for participants. We interpret this as evidence that the campaigns operated through community-wide changes in norms and perceived sanctions rather than through the direct experience of planning and running the events.

Consequences of a failure of Assumption 1. If the invariance assumption does not hold, observed attendance $A_{i,h}$ under each arm would reflect a mixture of the four potential-attendance types, and the regression in equation (E.1) would no longer recover $\tau_{00,h}$ and $\tau_{11,h}$ cleanly. In particular, if Community Campaigns induced some girls to attend the Girl Groups who would not otherwise have done so, i.e. if there exist girls with $S_{i,h} = (0, 1)$, then the non-attender cell in the GC arm would be selected differently from the non-attender cell in the G arm, biasing the estimate of $\tau_{00,h}$. The direction of this bias would depend on whether induced attenders have above- or below-average potential outcomes under the Girl Groups arm. The near-identical attendance rates and selection patterns documented in Tables B.5 and B.21 make this failure unlikely in our case.