Private versus Public Support for Gender Equality: Evidence from a Social Norms Intervention for Adolescents in Somalia^{*}

Rajdev Brar^{\dagger} Niklas Buehren^{\ddagger}

Sreelakshmi Papineni[§] Munshi Sulaiman[¶]

Abstract

Gender inequality and restrictive norms are often reinforced and internalized during adolescence, influencing pivotal life choices. We present results from a randomlyassigned gender norms intervention for adolescents in Somalia that led to greater support for gender equality in privately reported attitudes among both girls and boys. Moreover, the intervention improved adolescent mental health and boys' engagement in household work. In a novel lab-in-the-field experiment designed to observe social group dynamics, we measure the likelihood of adolescents changing privately held gender attitudes when assigned to groups of peers of the same or opposite sex, and when expressing their attitudes publicly. The gender group composition matters for social influence as boys endorse more egalitarian responses when grouped with girls, while the opposite is true for girls. We find that treated adolescents are less likely to succumb to peer pressure to conform when stating their gender attitudes in public, even when faced with opposing opinions within their group. Perceptions of gender norms shift for boys, leading to greater public support for gender egalitarian ideals. These results suggest that gender norms interventions for adolescents can be effective in influencing the attitudes and public discourse around gender equality.

Impact evaluation, adolescents, gender, norms, conformity, Somalia [JEL] D63, D91, J13, J16, O12

^{*}We thank Rachael Pierotti, Estelle Koussoubé, Céline Zipfel, Verena Phipps, Shubha Chakravarty, Jeannie Annan, Dave Evans, Richmond Atta-Ankomah and several seminar participants from Yale EGC, CSAE, PacDev, Chicago AFE conferences, and the Université Paris-Dauphine for excellent comments and suggestions. We also thank Save the Children Somalia who implemented the gender norms program, *Choices*, in particular, Yahya Abdillahi, Hamse Koshin, Billow Hassan Abdi, Alice Shirley, Shelagh Possmayer, and Cianne Jones. This paper is an output of the World Bank Africa Gender Innovation Lab (GIL). We are grateful to the World Bank Group's Umbrella Facility for Gender Equality (UFGE) for financial support. The findings, interpretations, and conclusions expressed in this paper are entirely those of the authors. They do not necessarily represent the views of the World Bank and its affiliated organizations, or those of the Executive Directors of the World Bank or the governments they represent. IRB approval was obtained from HML IRB Research and Ethics, reference 574TWBG18. The pre-analysis plan is registered on the AEA RCT registry at https://doi.org/10.1257/rct.2773.

[†]University of California, Berkeley rajdevbrar@berkeley.edu

[‡]World Bank nbuehren@worldbank.org

[§]World Bank spapineni@worldbank.org (Corresponding author)

[¶]BRAC Institute of Governance and Development munshi.sulaiman@bracu.ac.bd

1 Introduction

Gender norms, a subset of social norms, are shared expectations about how women and men should behave in a social group or culture. Restrictive gender norms create social pressure that can adversely affect women and girls' decisions around education, labor supply, marriage and fertility, which are critical for women's empowerment (Jayachandran, 2021). Shifting gender attitudes and norms to encourage greater female engagement in education and income-generating activities and male involvement in household production is necessary for reducing gender inequalities and promoting economic development (Duflo, 2012).

More recent economic research in this context has put greater emphasis on understanding how gender norms are formed (Alesina et al., 2013; Jayachandran, 2021), measured (Bernhardt et al., 2018; Bursztyn et al., 2023), and influenced (Dhar et al., 2022; Bursztyn et al., 2020). Programs that directly challenge restrictive gender norms have also begun to be rigorously evaluated in various settings. For example, interventions designed to change gender attitudes and norms (e.g., evidence from India by Dhar et al., 2022) or to change perceptions of existing norms (e.g., in Saudi Arabia by Bursztyn et al., 2020) have been found to increase gender equitable attitudes and support women's work. Similarly, programs that engage men in sessions on fatherhood and caregiving have led to greater male participation in childcare and household tasks (e.g., in the Democratic Republic of Congo by Vaillant et al., 2020).

Our intention in this paper is to advance this nascent literature. To do so, we devise innovative behavioral games to examine social dynamics to supplement the survey data from a randomized controlled trial (RCT) of a gender norms intervention for adolescents in Somalia. For norms to be *social* they must be shared by other people and partly sustained by their approval and disapproval (Elster, 1989). The behavioral games were designed to measure normative social influence, drawing insights from social psychology which has long argued for the importance of social contexts and interactions for decision-making. We build on a series of conformity experiments developed by Asch (1951, 1952).¹ We modify Asch's

¹Asch provided provocative experimental findings of conformity by showing how easily individuals' judg-

experiments to measure conformity in attitudes among adolescent peers who are asked to state their gender beliefs both in *private* and in *public*.² The experiments allow us to study whether and how gender attitudes differ if expressed privately or publicly; and provide a causal link between the gender norms program and the observed behaviors among the study sample, albeit in a lab-in-the-field setting.³ We also examine the marginal effect of a complementary gender norms intervention to parents on adolescent attitudes and behaviors.

In economics, conformity has been modeled in theory by Jones (1984); Bernheim (1994); Akerlof (1997), and shown to exist in educational settings (e.g., in Bursztyn and Jensen (2015); Bursztyn et al. (2017, 2019)) and in politics (e.g., Perez-Truglia (2018)).⁴ Similarly, Akerlof and Kranton (2000, 2002, 2010) translated theories of social identity into an economics framework to propose that social identity influences economic outcomes when deviating from the prescribed behavior is inherently costly. When individuals derive utility from fitting in with society or with their peers, an individual may try to avoid social penalties by conforming to perceived norms. Adolescence, in particular, is a period when the desire to conform to fit in with peers may be felt acutely.⁵ For example, Bursztyn and Jensen (2015) and Bursztyn et al. (2019) show that students in high school in the US under-invest in education to signal social compatibility to their peers. Similarly, Austen-Smith and Fryer (2005) show that minority students in the US shy away from educational investments to avoid "acting white" and thus improve their social standing with peers. Our experimental findings have potential implications for understanding different types of normative social influence more generally.

In this paper, we use random variation in individual participation in a program delivered

ment can be manipulated in a group setting using a line-match exercise (Asch, 1951, 1952).

 $^{^{2}}$ In the experiment individuals were also randomly grouped into teams with varying gender compositions and asked to select leaders for their teams.

 $^{^{3}}$ We use the experiment to observe actual behaviors and to address concerns that self-reported survey findings may reflect some social desirability bias.

⁴We examine conformity as a behavioral outcome that measures an adolescent's likelihood of changing their privately-held personal attitudes when grouped with peers of the same or opposite sex in public.

⁵The *Choices* program studied here targets young adolescents aged 10-14 and study participants may be vulnerable to peer pressure to conform. Conversely, one may expect adolescence to be a period of greater rebellion against societal expectations as individuals wrestle with gaining a sense of individuality and identity.

to young adolescents that was designed to challenge restrictive gender norms in Somalia. Early adolescence is viewed as a critical window of opportunity for formulating positive gender attitudes and behaviors as children transition into adulthood (Baird et al., 2019). The gender norms training program called *Choices* is implemented by the NGO Save the Children and targets adolescents aged 10-14 years.⁶ This is the first randomized controlled trial (RCT) experiment to examine the *Choices* program's impact.⁷ We report results from both individual-level surveys and a lab-in-the-field experiment to estimate the impact of the program on adolescents' gender attitudes and behaviors, and examine whether the effects differ for boys and girls. In the randomized experiment we also test whether the participation of their parents in a complementary gender norms program influences the effects of the *Choices* program on adolescent outcomes.⁸ Importantly, we measure persistence of program effects beyond an immediate effect using two follow-up survey rounds: 6 months and 2 years post-intervention. We test a number of key mechanisms examining whether the program affects personally held gender attitudes, perceptions of others' beliefs, the propensity to conform to gender norms, and the willingness to stand up for their own beliefs.

First, our results show that the *Choices* program in Somalia has a positive treatment impact on the personal gender equitable attitudes reported by both adolescent boys and girls. The treatment led to a 0.07 standard deviation increase in an index of gender equitable attitudes in the short-run, and a 0.05 standard deviation unit increase in the longer-run.

Next, we show that adolescents alter their stated attitudes in a social group setting. In the lab-in-the-field experiment we demonstrate significant social influence based on the gender composition of the peer group.⁹ Boys endorsed more gender-equitable responses when girls were present in their group, while the opposite was true for girls. This may mean that girls

 $^{^{6}}$ Choices is a component of the Challenging Harmful Attitudes and Norms for Gender Equality and Empowerment in Somalia (CHANGES) program under the Social Norms and Participation (SNaP) project.

⁷The *Choices* program has been adapted for use in over 10 countries. The program was evaluated in Nepal using quasi-experimental evaluation methods by Lundgren et al. (2013, 2020).

⁸Parents are invited to attend separate training sessions in an Engaging Men and Boys (EMB) component of the CHANGES program attended by adults in the community.

⁹In the experiment we randomly grouped participants into teams of five with different gender compositions (all boys, all girls, and mixed gender groups).

and boys experience different types of social influence and that social desirability bias might go in opposite directions. We find that treated adolescents were also more likely to stand by their gender egalitarian attitudes when responding in public.¹⁰ Treated adolescents were less likely to succumb to group peer pressure to conform, even when faced with opposing gender views.¹¹ We find changes in perceived gender norms for boys but not for girls, which suggests normative pressures from peers may be different for boys and girls. However, we find no impacts on the desire to conform in general. This result supports the notion that what we observe are social dynamics specific to gender attitudes and norms taught by the *Choices* program and that these dynamics are not reflective of conformity in other domains or of outright rebellion.

Moreover, we examine additional self-reported behavioral outcomes and find that the *Choices* program increased boys' reported participation in household chores, and both boys and girls report enhanced caring behaviors towards siblings of the opposite sex. The program also leads to improved mental health outcomes whereby both adolescent girls and boys report a reduction in the emotional and behavioral problems that they face. This finding supports the results described in Baird et al. (2019) who show a correlation between restrictive gender attitudes and poor mental health outcomes among adolescents in Bangladesh and Ethiopia, and in Andrew et al. (2022) who find positive impacts of a norms program on mental health outcomes for girls in India after engaging the wider community.

Our paper makes three important contributions. The first contribution is the development of experimental measures that capture social influences and norms. This aspect of our work is novel because previous work have mostly investigated the effects of gender norms programs using self-reported survey data. A key concern in evaluating the effects of norms

¹⁰In the experiment each adolescent was privately interviewed and asked if they agreed or disagreed with three gender attitudes statements and play two line-match games where participants were asked to match lines drawn on paper according to their length in private. This private interview was done with an enumerator asking questions that are typical of self-reported survey questions. We subsequently repeated the two exercises in the team setting. Participants were asked to publicly express their gender attitudes and the line choice in their teams, one after the other, where they could deviate from their private answers.

¹¹In our analysis, we consider both the influence of responses of team members in the positions before (position effects) and potential peer effects from teammates who attended the program (*Choices* peer effects).

programs on self-reported attitudes is that they may reflect some form of social desirability bias among the treatment group rather than actual changes in attitudes. Using a lab-inthe-field experiment, we try to link self-reports to observed behaviors to address this issue.¹² While the reference group in our experiment is a randomly-assigned group of adolescent peers, future work could explore social dynamics with alternative reference groups.¹³

Our second main contribution to the literature is that this study is one of the first RCTs in Somalia; a conflict-affected country with restrictive female labor norms and where publicly available data is scarce. Somalia is also an example of a country where the prevailing norms support more restrictive roles for women and girls. It ranks as the fourth lowest in the world in terms of gender equality, and female labor force participation remains low and stagnant at just 22% (UNDP, 2012; ILO, 2019).¹⁴ Patriarchal practices persist, including female genital cutting (FGC), son-preference, and child and forced marriage.¹⁵ Our study collects unique data on attitudes, norms and behaviors from both adolescents and adults to estimate the causal effect of programs to promote gender equality in an understudied setting.

The third contribution is that we provide direct evidence of the effects of a gender norms program for adolescents in- and out-of-school and test the additional effect of a parental treatment arm.¹⁶ Recent research has demonstrated that engaging the broader community in normative change can be effective (Andrew et al., 2022; Chetty and Hendren, 2018; Dahl et al., 2020). However, the causal impact of gender norms programs for parents on adolescent outcomes is largely unexplored.¹⁷ Programs that encourage behaviors that go against societal expectations may cause backlash effects (Aizer, 2010). Engaging participants' parents was

 $^{^{12}}$ We also collect a short-form version of the Marlowe-Crowne social desirability scale to check that our main results are robust to a social desirability bias correction as in Dhar et al. (2022).

¹³For example, alternative reference groups may include other influential actors in the daily lives of adolescents such as community elders/leaders, partners, teachers and/or parents.

¹⁴Somalia scored 0.776 (1 is complete inequality) in the Global Gender Inequality Index (UNDP, 2012).

¹⁵UNFPA estimates that 97% of Somali girls (15-19 years) are circumcised, and 36% of women in Somalia are married before they turn 18 (UNFPA, 2020).

¹⁶Previous research have shown effects of gender norms programs delivered in school (Dhar et al., 2022).

¹⁷Dhar et al. (2019) provide descriptive evidence that mothers' influence on children's gender attitudes is greater than that of fathers in India.

meant to elicit greater support from the family for adolescents and limit possible backlash.¹⁸ We show that training parents leads to an increase in reported gender egalitarian attitudes among mothers and fathers in the short-run, and attitudes around equal gender rights to education are maintained in the longer-run. However, we find limited additional impact of the parent treatment on adolescent-level attitudes and behaviors. Importantly, we find no immediate backlash effects from the *Choices* program and the relationship quality with both parents is similar between treatment and control groups.

The remainder of this paper is organized as follows. We begin, in Section 2, by defining the interrelationship between attitudes, norms and conformity that are the focus of this paper. Section 3 describes the gender norms interventions to adolescents and parents, while section 4 outlines details of the evaluation design. Section 5 describes the data and our sample, including the lab-in-the-field experiment. Section 6 outlines the estimation strategy and section 7 presents the results for adolescents; the effects on parents' attitudes and perceived norms are reported in the Appendix. We provide concluding thoughts in Section 8.

¹⁸We examine impacts on parental attitudes and norms towards gender equality, FGC, intimate partner violence, and early marriage; and ask whether this influences *Choices*' effect on adolescent outcomes.

2 Attitudes, Norms and Conformity

Before turning to the empirical analysis, we first define key terms and concepts as they are used in this paper, and discuss their interrelationship when considering the effects of a program that tackles restrictive gender norms.

Social norms are implicit, specific rules shared by a group of individuals on how they should behave. Gender norms, more specifically, refer to the expected behaviors of women and men in a social group or culture.¹⁹ We refer to personal attitudes as to what individuals think themselves, and perceived norms are what individuals think others think.²⁰ Recent studies have emphasized the importance of perceived norms for decision-making (see for example, Bicchieri, 2016; Tankard and Paluck, 2016; Field et al., 2021; Bursztyn et al., 2023). When social norms change, individuals' perceptions about what others do and approve of could change collectively; although not necessarily uniformly within a population.

Conformity refers to how individuals are influenced by what their respective social group thinks. Social norms are often sustained by conformist behavior through multiple mechanisms, including a desire to coordinate, fear of being sanctioned, signaling membership in a group, or simply following the lead of others (Young, 2015).²¹ In the conformity experiments studied by Asch (1952), participants were asked to match lines of the same length. Actors or "confederates" deliberately chose the incorrect line. This led participants, influenced by conformity, to also select the incorrect line, even when aware of its inaccuracy.²² In this context, conformity is defined as a type of social influence involving the tendency to alter

¹⁹Norms around the acceptability of women's work out of the home, roles and responsibilities in the home, women's mobility, capacity to interact with men and/or the male breadwinner status could all be at play when it comes to a woman's decision to enter the labor force (Jayachandran, 2021).

 $^{^{20}}$ Beliefs about what others do are referred to as *descriptive* norms and beliefs about what others approve of are referred to as *injunctive* norms in social norms theory (Cialdini et al., 1991; Heinicke et al., 2022). In our measure of perceived norms we are capturing injunctive norms rather than the descriptive norms.

²¹Social sanctions may include the prospect of social disapproval, ostracism, loss of status, and other forms of social punishment (Young, 2015).

 $^{^{22}}$ Asch (1952) finds that in a situation where an individual is asked to decide on the correct line alone in private, less than 1% make a mistake and pick the incorrect line. However, in group situations with confederates, individuals are wrong about one-third of the time.

one's personal perceptions, opinions, or behaviors to align with group norms.²³

Norm shifting can eventually occur when a sufficient number of individuals do not conform to an existing norm, making it evident that a previously held perceived norm no longer prevails. For example, Bertrand (2020) proposes that direct exposure to a proscribed counter-stereotypical behavior, such as women's work outside the home in some settings, may eventually reduce the costs of going against the norm and help erode the restrictive norm. In this way, we may consider individual non-conformist behavior as a first-order condition to allow for norms change at the societal level. The *Choices* program aims to shift personal gender attitudes, by making the returns to gender equality more salient, and can be seen as an attempt to kickstart such change.²⁴

We measure personal gender attitudes, perceived community norms and conformity through self-reported survey questions and complement these constructs with behavioral measures of perceived norms and social conformity from a lab-in-the-field experiment. Conformity is considered a behavior that is situationally determined, rather than an individual trait or characteristic. In our experiment, we define conformity as adolescents taking actions or stating beliefs in a public group setting that deviate from what they privately considered to be the optimal action or belief (what they would say or do if peers would not observe them). Specifically, we first measure conformity with respect to gender beliefs (the focus of the *Choices* training content), whereby we assess whether adolescents switch their private response when they are in public as part of a group of randomly-selected peers. That is, we explore individuals' likelihood of conforming in public where they may be confronted with an opposing view to their own. Next, we measure conformity in an abstract setting by asking respondents to match the length of lines as in the Asch conformity experiments. These line match games are used to bolster the argument that the observed social dynamics

 $^{^{23}}$ Social conformity is often explained by either peer pressure or information i.e. normative social influence (altering your behavior to fit in with the group), informational social influence (assuming others are better informed) or perceptual error (believing the group gave the correct answer). See Sunstein (2019) for a summary discussion of these concepts.

²⁴Studies in low- and middle-income countries have typically focused on programs that try to change social norms that are restrictive or sustain harmful or discriminatory practices.

are specific to the gender norms addressed by the program, rather than reflecting a general propensity for conformity. Finally, we contrast these experimental measures with self-reports of conformist behavior from the surveys.

Evidence shows differences in the way men and women conform to social influence. See, for example, Eagly and Carli (2007) who conduct a meta-analysis and find that women are more persuadable and conforming than men in group pressure situations that involve surveillance. This implies that the gender composition of the group may play a role in conformity. Indeed, Reitan and Shaw (1964) show that men and women conformed more when there were participants of both sexes involved versus when all participants were of the same sex. In our conformity experiments we examine different types of social influence for both boys and girls and explore the influence of: (a) the effect of *Choices* peers in an individual's team, (b) mixed-gender vs. single-gender teams, and (c) what responses were given by other peers directly before an individual responds in public.

3 Gender Norms Program

The *Choices* program aims to encourage the development of positive, mutually-respectful gender attitudes and behaviors among adolescent boys and girls and promotes greater gender equality in the home.²⁵ The program is targeted to adolescents aged 10 to 14 years old and the program's curriculum is designed to address, discuss and reflect on relevant gender norms, in particular those norms related to gender roles pertaining to both work outside of the home and domestic responsibilities within the household.

The *Choices* program involves adolescents' participation in ten 2-hour sessions over three months in an after-school program. In these sessions adolescents engage in interactive, age-

 $^{^{25}}$ The *Choices* training curriculum was developed by Save the Children International, and has been adapted for Somalia and for over 10 other countries including Egypt, Bangladesh, and Bolivia. In Somalia, the core *Choices* model has been complemented by activities that expand its influence into families and wider communities (the programs are called *Voices* and *Promises*) that are offered together as part of the *CHANGES* project. *CHANGES* is a community-level program implemented by a consortium led by Save the Children in partnership with CARE International and International Rescue Committee. *Choices* is the only component that targets adolescents.

appropriate, exploratory activities which are meant to initiate a dialogue about their notions of respect, communication, fairness and their dreams for the future. Each session typically includes two trainers (one male and one female) who lead mixed-gender activities such as drama, poetry and debate competitions. While most *Choices* sessions are mixed-gender, some chapters of the training are conducted for boys and girls separately.²⁶ Similarly, the goal of the *EMB* training is to promote gender equality by engaging adult men and women in community discussions on a range of topics such as gender norms, fatherhood, violence, and child marriage.²⁷ Both mothers and fathers are invited to attend these discussions and instructed to identify, analyze and address the role they play in sustaining gender inequalities. Participants are also encouraged to self-reflect and consider making changes. The interventions often use existing community infrastructure such as youth clubs or community centers to deliver the training.

This impact evaluation assesses the effectiveness of the *Choices* program as well as the marginal impact of layering the *EMB* program on top. An important aspect of this paper is evaluating whether the effects persist beyond an immediate effect right after the training, with follow-ups after six months and after two years.

All adolescents aged between 10 and 14 years living in targeted program communities were eligible for the *Choices* intervention irrespective of whether they were in- or out-ofschool. There are no corresponding eligibility criteria for taking part in the *EMB* community discussions that were strictly enforced. Typically, these discussions were attended by around 30 participants. For the purpose of this study, however, we specifically invited parents or guardians of a subset of the adolescents to attend the EMB sessions in accordance pf the RCT study design described in the next section.²⁸

 $^{^{26}}$ Out of the 10 chapters, there were 2-3 chapters where boys and girls were separated.

 $^{^{27}}$ For a full description of modules for the *Choices* and *EMB* interventions, refer to the Appendix.

 $^{^{28}{\}rm The}$ implementing partner was instructed to adhere to treatment lists to avoid potential contamination concerns.

4 Study Design

This study uses an RCT evaluation design to measure impacts of the *Choices* and *EMB* programs. The evaluation strategy builds on randomizing households into one of the following three groups: (1) *Choices Only*: adolescents invited to participate in the *Choices* training; (2) *Choices+EMB*: adolescents invited to participate in the *Choices* training and adult members of their household invited to participate in *EMB* sessions; and (3) Control group: no invitation to participate in any of the two interventions.

For those households that were assigned to the Choices+EMB treatment arm, the biological parents of the adolescents were invited to attend EMB sessions if they were living in the household. If there were no biological parents present in the household, any male and/or female primary caregiver was invited to attend EMB sessions.

The randomization was stratified by community, gender of the adolescent, school enrollment, and gender of the household head. For households with siblings of similar age, there could be more than one eligible adolescent per household.²⁹ In those cases, household level randomization ensured that adolescents within a household were not assigned to different treatments which minimizes possible spillover or contamination effects. Our study design is not set up to capture spillover effects which can arise if treated adolescents or parents share information and material from the trainings within their networks. If knowledge spillovers occur between treated and control households then our results are likely biased downward which means that our analysis will underestimate the program's effect. Table A8 in the Appendix shows that gender attitudes in the control group are fairly stable over time from baseline to midline and endline which suggests that spillover effects are not a major factor.³⁰

²⁹Strata for mixed-gender or all enrolled in school were considered for the multiple adolescent households.

 $^{^{30}}$ The few attitudes that change over time may be driven by cohort effects as adolescents are 3 years older at the time of the endline survey.

5 Data

5.1 Census

In order to establish a sampling frame, the implementing partner conducted a community mapping exercise and identified project sites covering urban, peri-urban, rural and internally displaced persons (IDP) communities. For the evaluation, 33 community clusters were selected across the six project districts based on accessibility and population size.³¹ The research team completed a census that listed all children aged 10-14 years in at least 400 households located within each of the targeted communities. The census also collected information on the adolescents' willingness to participate in the *Choices* training to predict program take-up. The listing also asked for parental consent to participate in the program to establish eligibility. Overall, the census listing contains data from 8,572 adolescents living in 4,341 households.

5.2 Adolescent and Parent Surveys

The main source of data we use for the evaluation consists of three rounds of individual-level surveys administered to both adolescents and their parents: one baseline, and two follow-up surveys (midline and endline). The baseline survey was conducted from February to July 2018 just before program implementation started. The midline survey was administered between February 2019 and January 2020 after approximately half of those assigned to the *Choices* training had completed the training. The endline survey was completed between September 2020 and October 2021, after all *Choices* training and *EMB* sessions were com-

³¹The study took place in six districts across Somalia: Hargeisa, Burao, Erigavo and Badhan in Somaliland; and Galkayo and Galdogob in Puntland. In these districts, the program team had originally identified 37 communities. However, following the household listing exercise, 5 communities which had fewer than 80 young adolescents listed were excluded. In addition, 5 of the larger communities were divided into 2 separate clusters each: Daami B, October IDP, Bacadweyn, Waaberi and Statehouse. Approximately 100 households were dropped from the census listing for having incorrect GPS coordinates and another 50 households did not provide study consent.

pleted and after a government-mandated lockdown in response to the coronavirus had been removed in Somalia.³² The follow-up surveys were timed to take place approximately 6 months and 2 years post-intervention. To participate in the surveys, written or verbal consent, depending on literacy level, was obtained from parents or guardians, as well as from the adolescents themselves. Surveys were translated into Somali language and data collection took place through face-to-face interviews with enumerators. Wherever possible, the enumerator was from the same region as the adolescent. Questions and scales were piloted extensively and enumerators were trained carefully in the wording of the questions and how to appropriately interact with young adolescents.

Each respondent was interviewed in private and adolescents and parents were asked two different questionnaires. Typically the adolescent's mother and/or father were interviewed for the adult survey.³³ At baseline, adolescents were asked to provide information on basic demographic characteristics (such as age, education and health), attitudes towards gender equality and perceived community norms (particularly on the role of women in- and outside of the household), opinions on marriage, hope and aspirations for the future, as well as their answers to a strengths and difficulties questionnaire (SDQ) as a measure of mental health.³⁴ Adolescents were also asked to provide information on the quality of the relationship with their parents, time use, and their involvement in household chores. Parents were asked about their own marriage history, decision-making power and attitude towards intimate partner violence (IPV), gender attitudes and perceived norms, behavioral characteristics (conformity, risk preferences), aspirations for their children, and opinions on FGC and early

³²The COVID-19 pandemic reached Somalia in March 2020 and the country went into lockdown for several months. In July 2020 most of the government restrictions in Somalia had been lifted and schools reopened in August 2020. The lockdown also explains the relatively long survey period as survey teams tried to match the timing of surveys to the timing of the rollout of the intervention in each of the six districts.

 $^{^{33}}$ Enumerators were instructed to survey the mother and whenever possible, the father too. The baseline sample consists of 57% of households with one female respondent, 9% with one male respondent and 34% with both a female and a male respondent. Therefore, data from fathers should be considered a highly selective sample as it was difficult to track and survey fathers for all rounds.

³⁴Scales to measure attitudes and mental health outcomes were changed from a 3-point Likert scale at baseline to agree/disagree options in the follow-up surveys to make it easier for adolescents to respond accurately. In the Appendix, we detail the individual statements in the attitudes index and present the Cronbach's alpha to test for scale reliability (Cronbach's alpha of the gender attitudes index is 0.798).

marriage.³⁵ The midline and endline questionnaires were very similar to the baseline but collected additional information on certain topics.³⁶

The survey was conducted with all adolescents living in sampled households who fell into the targeted age range. In total, information from 3,237 adolescents across 2,393 randomly sampled households was collected at baseline.

5.3 Lab-in-the-Field Experiments

Lab-in-the-field experiments supplement the survey data at endline. The experiments allow us to study gender attitudes expressed privately and publicly. To do this, we first randomly grouped participants into teams of five with different gender compositions (same and mixedgender) and asked individual participants to choose a leader and deputy for the team. Next, each adolescent in the experiment sample was privately interviewed and asked to choose if they agree or disagree with three gender attitudes statements and play two line-match games inspired by Asch (1951, 1952) in private. During the line-match game, participants were shown a set of lines and they were asked to indicate which of these lines had the same length as the reference line which was also shown to them. Then, we repeated the two exercises in a team setting. More precisely, all participants were asked to express their gender attitudes and the line choice in their teams one after the other in public where they could deviate from their private answers. We use the private and public statements to measure social conformity. Further details of the lab-in-the-field experiments are included in the Appendix.

Each community cluster in the study sample hosted at least two experimental sessions.

³⁵Parents were also asked to provide basic information on the household including demographic information, food security, household assets and consumption modules.

³⁶The midline survey also included questions on participation in the *Choices* training, additional beliefs and community norms, time use, household chores, and a module to capture social desirability bias. We capture social desirability bias through the 13-item social desirability scale, which Reynolds (1982) demonstrates is a viable substitute for the 33-item Marlowe-Crowne scale. The endline survey included additional questions on marital status, fertility, and girls' menstruation status in the adolescent questionnaire as these questions may have become applicable to some of the older adolescents in the study. For both adolescents and adults, we included a measure of generalized self-efficacy (Schwarzer and Jerusalem, 1995). Parents' expectations on fertility, marriage, and FGC for adolescents was also collected.

Each of these sessions consisted of 20 adolescents (10 girls and 10 boys) divided into four teams, each with one of four gender compositions (all boys, all girls, mixed majority boys, mixed majority girls). Each team included both control group and treatment group participants. Games were completed for a subset of the study sample randomly selected.³⁷ The lab-in-the-field experimental games were conducted between November 2020 and June 2021 to coincide with the endline.

5.4 Balance Check and Sample Description

In Table 1 we show that the randomization produced balanced groups across a set of observable characteristics at baseline. The sample consists of different types of communities with approximately 40% IDP settlements, 40% urban, and 20% rural. The average number of household members is 6.5 and the majority of adolescents are enrolled in school: 81% at baseline (79% of girls and 84% of boys). While the program was often delivered using school buildings, adolescents out-of-school were also eligible (19% of the sample were not enrolled in school at the time of the survey). The average age of children in the sample is just under 12 years old (targeted age range is 10-14 years). The average age of FGC for girls in the community is 8.5 years. In terms of child marriage, adult respondents estimate 43% of girls in the community are married before turning 18 years old (descriptive norm) and about 36% of mothers were under 18 years when first married. The majority of parents are married in monogamous relationships, with 12% in polygamous marriages and 12% are divorced or widowed.

 $^{^{37}}$ The sampling for the lab-in-the-field experiment was stratified by community cluster, gender of the adolescent and treatment status. Participant replacement lists were used to ensure that each session had exactly 20 adolescents with an equal representation of male and female adolescents as well as treatment and control adolescents.

(1) (2) (3) (4) (5) (7) (6) (7) (6) (7) Control ChOICES CHOICES CHOICES CHOICES Heat diff CHOICES Heat diff Mean Mean (1) (2) (3) Mean (1) (2) (3) (4) Mean (1) (2) (3) (4) (1) Mean (1) (2) (3) (2) (3) (2) (3) (4) (3) (2) (3) (4) (3) (1) (2) (3) (4) (3)<					urvev sample				mple	
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $		(1)	(2)	(3)	(1)	(5)	(6)	(7)	(8)	(0)
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $		(1)			(4)	(J) t toot diff	(0)	(7)	(0)	(⁹)
$\begin{array}{ c c c c c c c c c c c c c c c c c c c$		Control	CHUICES			(Normalized diff)		Control	CHOICES	Normalized diff
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $		Moon	Moon	EIVID	(1) (2)	(1) (2)	(2) (2)	Moon	Moon	(7) (9)
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	Panel A: Adolescent characteristics	Wedn	Wedn	Iviedi	(1)-(2)	(1)-(3)	(2)-(3)	Iviedii	Iviedi	(7)-(0)
Age (real) 11.823 11.803 <td>Age (Veere)</td> <td>11 001</td> <td>11 025</td> <td>11 066</td> <td>0.024</td> <td>0.064</td> <td>0.041</td> <td>11 750</td> <td>11 702</td> <td>0.049</td>	Age (Veere)	11 001	11 025	11 066	0.024	0.064	0.041	11 750	11 702	0.049
$ \begin{array}{c} Has sibling(s) (Yes=1) \\ \mbox{Has sibling(s) (Y$	Age (Teals)	11.001	11.025	11.000	-0.024	-0.004	(0.041	11.750	11.702	(0.040
nas staining(y) (18=1) 0.933 0.974 0.933 0.934 0.935 0.934 0.935 0.934 0.935 0.934 0.035 0.934 0.035 0.934 0.035 0.934 0.035 0.934 0.035 0.934 0.035 0.934 0.035 0.034 0.006 0.014 0.832 0.844 0.028 (0.033) Gender Equitable Index Score (0-1) 0.525 0.529 0.531 0.004 -0.006 -0.002 (0.033) (0.035) (0.005) (0.034) Number of chores in past week (of 8) 3.241 3.149 3.154 0.092 0.086 -0.005 3.373 3.223 0.150 Mental Health Index (SDQ Total 0-20) 6.880 6.841 6.641 0.039 0.233 0.199 7.142 6.746 0.396 Panal B: Household (Yes=1) 0.534 0.534 0.534 0.556 -0.001 -0.023 -0.022 0.540 -0.007 (0.035) (0.015) (0.017) (0.025) (0.017) (0.015) (0.017) (0.025) (0.017) (0.026) (0.007) (0.051) (0.00	l_{100} sibling(s) (l_{100} - 1)	0.002	0.074	0.000	1-0.0103	1-0.0437	{-0.027} 0.015**	0.001	0.004	10.0323
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	Has sibility(s) (res-1)	0.965	0.974	0.969	0.009	-0.000	-0.015^{m}	0.901	0.964	-0.003
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	Envelled in each cal ()(as -1)	0.000	0.010	0 700	{0.002}	{-0.049}	{-0.109}	0.000	0.004	{-0.024}
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	Enrolled in school (Yes=1)	0.800	0.813	0.799	-0.014	0.000	0.014	0.832	0.804	0.028
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	Orandez Franklada (0.1)	0.505	0 500	0 501	{-0.034}	{0.000}	{0.035}	0.51.4	0 501	{0.071}
Number of chores in past week (of 8) 3.241 3.149 3.154 0.092 0.086 -0.005 3.373 3.223 0.150 Mental Health Index (SDQ Total 0-20) 6.880 6.841 6.641 0.039 0.239 0.199 7.142 6.746 0.396 Panel B: Household characteristics (0.010) (0.059) (0.049) (0.037) (0.037) Multiple VYA in household (Yes=1) 0.534 0.556 0.001 -0.023 -0.022 0.522 0.540 -0.017 Type of residence: IDP camp (Yes=1) 0.427 0.413 0.406 0.014 0.021 0.007 0.466 0.473 -0.003 Type of residence: Urban (Yes=1) 0.398 0.400 -0.002 -0.006 -0.004 -0.002 -0.008 (0.007) Type of residence: Rural (Yes=1) 0.175 0.186 0.190 -0.017 -0.004 -0.003 0.221 0.211 0.107 Household bietary Diversity Score (0-12) 6.471 6.400 6.514 0.071 -0.042 -0.114	Gender Equitable Index Score (0-1)	0.525	0.529	0.531	-0.004	-0.006	-0.002	0.514	0.521	-0.007
Number of chores in past week (of b) 3.241 3.194 0.092 0.0086 -0.005 3.373 3.223 0.150 Mental Health Index (SDQ Total 0-20) 6.880 6.841 6.641 0.039 0.239 0.199 7.142 6.746 0.396 Panel B: Household characteristics Multiple VYA in household (Yes=1) 0.534 0.556 -0.001 -0.022 0.046 -0.044 (0.035) Type of residence: IDP camp (Yes=1) 0.427 0.413 0.406 0.014 0.021 0.007 0.466 0.473 -0.007 Type of residence: Urba (Yes=1) 0.398 0.400 0.404 -0.002 -0.006 -0.004 .0303 0.306 -0.003 Type of residence: Rural (Yes=1) 0.175 0.186 0.190 -0.012 -0.003 0.231 0.221 0.010 Household bitary Diversity Score (0-12) 6.471 6.400 6.514 0.071 -0.042 -0.014 6.303 0.304 -0.068 Household bitary Diversity Score (0-12) 6.471 6.408					{-0.020}	{-0.028}	{-0.008}	0.070		{-0.034}
(0.037) (0.035) (-0.002) (-0.006) Mental Health Index (SDQ Total 0-20) 6.881 6.641 0.039 0.239 0.199 7.142 6.746 0.396 Panel B: Household characteristics (0.010) (0.059) (0.049) (0.093) (0.037) (0.059) (0.049) Valitple VYA in household (Yes=1) 0.534 0.534 0.556 -0.001 -0.022 -0.022 0.522 0.540 -0.017 Type of residence: IDP camp (Yes=1) 0.427 0.413 0.406 -0.014 -0.021 -0.007 0.466 0.473 -0.007 Type of residence: Urban (Yes=1) 0.398 0.400 -0.002 -0.006 -0.004 -0.017 -0.008 (-0.007) Type of residence: Rural (Yes=1) 0.175 0.186 0.190 -0.017 -0.004 -0.009 (-0.007) (-0.007) Household bizer (Number) 6.471 6.400 6.514 0.071 -0.042 -0.114 6.432 6.500 -0.028 Household bizeray Diversity Score (0-12)<	Number of chores in past week (of 8)	3.241	3.149	3.154	0.092	0.086	-0.005	3.373	3.223	0.150
Mental Health Index (SQC Total 0-20) 6.880 6.841 6.641 0.039 0.239 0.199 7.142 6.746 0.396 Panel B: Household characteristics (0.010) (0.059) (0.049) (0.093) (0.093) Type of residence: IDP camp (Yes=1) 0.427 0.413 0.406 0.014 0.021 0.007 0.466 0.473 -0.007 Type of residence: Urban (Yes=1) 0.398 0.400 0.404 -0.002 -0.006 -0.004 0.303 0.306 -0.003 Type of residence: Urban (Yes=1) 0.398 0.400 0.404 -0.002 -0.006 -0.004 -0.003 -0.0					{0.037}	{0.035}	{-0.002}			{0.061}
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	Mental Health Index (SDQ Total 0-20)	6.880	6.841	6.641	0.039	0.239	0.199	7.142	6.746	0.396
Panel B: Household (Yes=1) 0.534 0.534 0.534 0.556 -0.001 -0.023 -0.024 0.522 0.522 0.502 -0.017 Type of residence: IDP camp (Yes=1) 0.427 0.413 0.406 0.014 0.021 0.007 0.466 0.473 -0.007 Type of residence: Urban (Yes=1) 0.398 0.400 0.404 -0.002 -0.006 -0.004 0.303 0.306 -0.003 Type of residence: Urban (Yes=1) 0.175 0.186 0.190 -0.012 -0.003 0.231 0.221 0.010 Type of residence: Rural (Yes=1) 0.175 0.186 0.190 -0.012 -0.013 0.040 -0.003 0.231 0.221 0.010 Household size (Number) 6.471 6.400 6.514 0.029 -0.017 -0.045 -0.028 -0.028 Household Dietary Diversity Score (0-12) 6.471 6.490 6.514 0.019 -0.017 0.012 6.106 6.345 -0.239 Household Extery Diversity Score (0-12)					{0.010}	{0.059}	{0.049}			{0.093}
Multiple VYA in household (Yes=1) 0.534 0.534 0.536 -0.001 -0.023 -0.022 0.522 0.540 -0.017 Type of residence: IDP camp (Yes=1) 0.427 0.413 0.406 0.014 0.021 0.007 0.466 0.473 -0.007 Type of residence: Irban (Yes=1) 0.398 0.400 0.404 -0.002 -0.006 -0.004 0.303 0.306 -0.003 Type of residence: Rural (Yes=1) 0.175 0.186 0.190 -0.012 -0.015 -0.003 0.231 0.221 0.010 Household Size (Number) 6.471 6.400 6.514 0.071 -0.042 -0.114 6.432 6.500 -0.068 Household Dietary Diversity Score (0-12) 6.471 6.490 6.478 -0.019 -0.012 6.106 6.345 -0.239 You an umber of types of assets (0-24) 4.313 4.468 4.165 -0.155 0.148 0.303** 3.947 4.434 -0.487** You anumber of types of assets (0-24) 4.313 4.468 4.165 -0.155 0.148 0.303** 3.947 4.434 <td>Panel B: Household characteristics</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td>	Panel B: Household characteristics									
	Multiple VYA in household (Yes=1)	0.534	0.534	0.556	-0.001	-0.023	-0.022	0.522	0.540	-0.017
Type of residence: IDP camp (Yes=1) 0.427 0.413 0.406 0.014 0.021 0.007 0.466 0.473 -0.007 Type of residence: Urban (Yes=1) 0.398 0.400 0.404 -0.002 -0.006 -0.004 0.303 0.306 -0.003 Type of residence: Rural (Yes=1) 0.175 0.186 0.190 -0.012 -0.003 0.231 0.221 0.010 Household size (Number) 6.471 6.400 6.514 0.071 -0.042 -0.114 6.432 6.500 -0.068 Household Dietary Diversity Score (0-12) 6.471 6.490 6.478 -0.019 -0.007 0.012 6.106 6.345 -0.239 Total number of types of assets (0-24) 4.313 4.468 4.165 -0.155 0.148 0.303** 3.947 4.434 -0.487** {0.023} 100.615 (0.04) (0.039) (-0.173) (-0.178) (-0.078) (-0.078) Total number of types of assets (0-24) 4.313 4.468 4.165 -0.155 (0.148 0.303** 3.947 4.434 -0.487** <t< td=""><td></td><td></td><td></td><td></td><td>-0.002</td><td>-0.046</td><td>-0.044</td><td></td><td></td><td>{-0.035}</td></t<>					-0.002	-0.046	-0.044			{-0.035}
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	Type of residence: IDP camp (Yes=1)	0.427	0.413	0.406	0.014	0.021	0.007	0.466	0.473	-0.007
Type of residence: Urban (Yes=1) 0.398 0.400 0.404 -0.002 -0.006 -0.004 0.303 0.306 -0.003 Type of residence: Rural (Yes=1) 0.175 0.186 0.190 -0.012 -0.003 0.231 0.221 0.010 Household size (Number) 6.471 6.400 6.514 0.071 -0.042 -0.114 6.432 6.500 -0.068 Household Dietary Diversity Score (0-12) 6.471 6.490 6.478 -0.019 -0.007 0.012 6.106 6.345 -0.239 Total number of types of assets (0-24) 4.313 4.468 4.165 -0.155 0.148 0.039** 3.947 4.434 -0.487** Household expenses last 30 days (USD) 192.835 196.342 190.608 -3.508 2.227 5.734 165.447 175.315 -9.868 Panel C: Parental characteristics - - - -0.023 (0.015] (0.03) - -0.010 Mother has earning activity (Yes=1) 0.596 0.546 0.543 0.050** 0.007 0.407 0.412 -0.005					0.028	0.043	0.015			{-0.015}
$\begin{array}{cccccccccccccccccccccccccccccccccccc$	Type of residence: Urban (Yes=1)	0.398	0.400	0.404	-0.002	-0.006	-0.004	0.303	0.306	-0.003
Type of residence: Rural (Yes=1) 0.175 0.186 0.190 -0.012 -0.015 -0.003 0.231 0.221 0.010 Household size (Number) 6.471 6.400 6.514 0.071 -0.042 -0.114 6.432 6.500 -0.068 Household Dietary Diversity Score (0-12) 6.471 6.490 6.478 -0.019 -0.007 0.012 6.106 6.345 -0.239 Total number of types of assets (0-24) 4.313 4.468 4.165 -0.155 0.148 0.039 (0.004) (-0.078) Household expenses last 30 days (USD) 192.835 196.342 190.608 -3.508 2.227 5.734 165.447 175.315 -9.868 $Panel C: Parental characteristics (0.015) (0.039) (-0.73) (-0.078) (-0.078) Panel C: Parental characteristics (0.021) (0.031) (0.039) (-0.73) (-0.73) Father has earning activity (Yes=1) 0.596 0.546 0.543 0.050^{**} 0.054^{**} 0.004 0.634 0.578 0.056$					-0.004	-0.012	-0.008			{-0.007}
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	Type of residence: Rural (Yes=1)	0.175	0.186	0.190	-0.012	-0.015	-0.003	0.231	0.221	0.010
$\begin{array}{c c c c c c c c c c c c c c c c c c c $					-0.031	-0.040	-0.009			{0.025}
$\begin{array}{c c c c c c c c c c c c c c c c c c c $	Household size (Number)	6.471	6.400	6.514	0.071	-0.042	-0.114	6.432	6.500	-0.068
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $					0.029	-0.017	-0.045			{-0.028}
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	Household Dietary Diversity Score (0-12)	6.471	6.490	6.478	-0.019	-0.007	0.012	6.106	6.345	-0.239
Total number of types of assets (0-24) 4.313 4.468 4.165 -0.155 0.148 0.303^{**} 3.947 4.434 -0.487^{**} Household expenses last 30 days (USD) 192.835 196.342 190.608 -3.508 2.227 5.734 165.447 175.315 -9.868 Panel C: Parental characteristics -0.596 0.546 0.543 0.050^{**} 0.054^{**} 0.004 0.634 0.578 0.056 Mother has earning activity (Yes=1) 0.400 0.388 0.382 0.012 0.018 0.007 0.412 -0.005 Mother has earning activity (Yes=1) 0.400 0.388 0.382 0.012 0.018 0.007 0.412 -0.005 Father has some education (Yes=1) 0.213 0.193 0.214 0.019 -0.002 -0.021 0.265 0.204 0.0154 0.146 0.173 -0.009 Mother has some education (Yes=1) 0.156 0.174 0.165 -0.018 -0.008 0.009 0.164 0.173 -0.009 Mother has some education ({-0.006}	{-0.002}	{0.004}			{-0.078}
$ \begin{array}{cccccccccccccccccccccccccccccccccccc$	Total number of types of assets (0-24)	4.313	4.468	4.165	-0.155	0.148	0.303**	3.947	4.434	-0.487**
Household expenses last 30 days (USD) 192.835 196.342 190.608 -3.508 2.227 5.734 165.447 175.315 -9.868 Panel C: Parental characteristics -					{-0.051}	{0.049}	{0.099}			{-0.173}
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $	Household expenses last 30 days (USD)	192.835	196.342	190.608	-3.508	2.227	5.734	165.447	175.315	-9.868
$ \begin{array}{c c c c c c c c c c c c c c c c c c c $					{-0.023}	{0.015}	{0.039}			{-0.081}
Father has earning activity (Yes=1) 0.596 0.546 0.543 0.050^{**} 0.054^{**} 0.004 0.634 0.578 0.056 Mother has earning activity (Yes=1) 0.400 0.388 0.382 0.012 0.018 0.007 0.407 0.412 -0.005 Father has some education (Yes=1) 0.213 0.193 0.214 0.019 -0.002 -0.021 0.265 0.204 (0.048) Mother has some education (Yes=1) 0.213 0.193 0.214 0.019 -0.002 -0.021 0.265 0.204 0.061^{**} Mother has some education (Yes=1) 0.156 0.174 0.165 -0.018 -0.008 0.009 0.164 0.173 -0.009 Mother has some education (Yes=1) 0.156 0.174 0.165 -0.018 -0.008 0.009 0.164 0.173 -0.009 Mumber of observations 1.068 1.087 1.082 2.55 0.482 0.285 0.225 F-test of joint significance (p-value) 2.155 2.150 2.169	Panel C: Parental characteristics				()		(,			()
$ \begin{array}{c ccccccccccccccccccccccccccccccccccc$	Father has earning activity (Yes=1)	0.596	0.546	0.543	0.050**	0.054**	0.004	0.634	0.578	0.056
Mother has earning activity (Yes=1) 0.400 0.388 0.382 0.012 0.018 0.007 0.407 0.412 -0.005 Father has some education (Yes=1) 0.213 0.193 0.214 0.019 -0.002 -0.021 0.265 0.204 $0.061**$ Mother has some education (Yes=1) 0.156 0.174 0.165 -0.018 -0.008 0.009 0.164 0.173 -0.009 Mother has some education (Yes=1) 0.156 0.174 0.165 -0.018 -0.008 0.009 0.164 0.173 -0.009 Mumber of observations 1.068 1.087 1.082 2.68 578 F-test of joint significance (p-value) 0.557 0.482 0.285 0.225 F-test, number of observations 2.155 2.150 2.169 846	· · · · · · · · · · · · · · · · · · ·				{0.101}	{0.109}	{0.008}			{0.115}
Horizon factor interview (100 17) 0.100 0.000 0.001 0.001 0.001 (-0.010) Father has some education (Yes=1) 0.213 0.193 0.214 0.019 -0.002 -0.021 0.265 0.204 0.061** Mother has some education (Yes=1) 0.156 0.174 0.165 -0.018 -0.008 0.009 0.164 0.173 -0.009 Number of observations 1,068 1,087 1,082 2.057 0.482 0.285 2.169 846	Mother has earning activity (Yes=1)	0 400	0.388	0.382	0.012	0.018	0.007	0 407	0 412	-0.005
Father has some education (Yes=1) 0.213 0.193 0.214 0.019 -0.002 -0.021 0.265 0.204 0.061** Mother has some education (Yes=1) 0.156 0.174 0.165 -0.018 -0.008 0.009 0.164 0.173 -0.009 Number of observations 1,068 1,087 1,082 268 578 F-test of joint significance (p-value) 2.155 2.150 2.169 846		01100	0.000	01002	{0 024}	{0 037}	{0.013}	01107	02	{-0.010}
Number of observations 1,068 1,087 1,082 1,082 0.156 0.174 0.165 0.018 -0.005}{(-0.027)} 0.156 0.174 0.165 0.018 -0.008 0.009 0.164 0.173 -0.009 Number of observations 1,068 1,087 1,082 268 578 F-test of joint significance (p-value) 0.557 0.482 0.285 0.225 F-test, number of observations 2.155 2.150 2.169 846	Father has some education (Yes=1)	0 213	0 1 9 3	0 214	0.019	-0.002	-0.021	0 265	0 204	0.061**
Mother has some education (Yes=1) 0.156 0.174 0.165 0.018 -0.008 0.009 0.164 0.173 -0.009 Number of observations 1,068 1,087 1,082 268 578 F-test of joint significance (p-value) 2.155 2.155 2.150 2.169 846		0.210	0.150	0.211	{0.048}	{-0.005}	{-0.053}	0.200	0.207	{0 146}
Number of observations 1,068 1,087 1,082 1,082 0.174 0.174 0.173 0.005 F-test of joint significance (p-value) 1,068 1,087 1,082 0.255 0.482 0.285 0.225 F-test of joint significance (p-value) 2.155 2.150 2.169 846	Mother has some education (Ves=1)	0 1 5 6	0 174	0 165	-0 018	-0 008	0.000	0 164	0 173	-0 000
Number of observations 1,068 1,087 1,082 268 578 F-test of joint significance (p-value) 0.557 0.482 0.285 0.225 F-test, number of observations 2.155 2.150 2.169 846		0.150	0.174	0.100	{-0 0/7	{-0 022}	{0 025	0.104	0.175	{-U U23/
F-test of joint significance (p-value) 0.000 0.000 0.000 0.000 F-test number of observations 2.155 2.150 2.169 846	Number of observations	1,068	1,087	1.082	(0.047)	ر ۲.۵۷۷	(0.020)	268	578	(0.020)
F-test, number of observations 2.155 2.150 2.169 846	E-test of joint significance (p-value)	1,000	1,007	1,002	0.557	0 482	0.285	200	0/0	0.225
E.100 E.100 E.100 E.100	F-test number of observations				2,155	2,150	2.169			846

Table 1: Randomization Balance Check

 ****, **, and * indicate significance at the 1, 5, and 10 percent critical level.

 Value variables have been winsorized at the 99th percentile. VYA is "Very Young Adolescent" meaning any children with program eligible age of 10-14 years.

 All variables in Table 1 measured during the baseline survey

5.5 Survey Attrition

In Table A9 in the Appendix, we show that survey attrition rates are balanced across treatment and control groups. At midline, 62% of respondents were successfully re-surveyed; and at endline we achieved an 82% survey completion rate. The attrition rate at midline was higher than anticipated due to the COVID-19 pandemic lockdown in Somalia. In fact, the midline data collection was halted before completing the final district, Galdogob. Overall, 89% of the study sample were tracked and resurveyed at either midline or endline. Most of the intended lab-in-the-field sessions were completed (91%), so that 1,000 adolescents participated in a total of 50 games sessions.³⁸

5.6 Treatment Uptake

All intervention activities took place between June 2018 and April 2020. The trainings were rolled out in different time periods across the 6 districts and conducted in two batches within each district.³⁹ Training facilitators were tasked with reporting participants' attendance at the individual level. However, the attendance lists were collected on paper and poorly digitized so it is difficult to decipher which adolescents and adults attended which training sessions. Based on these incomplete training lists we estimate that the average take-up rate is likely to be at least 40-50% among invited adolescents across all districts, with no considerable difference between the two treatment arms. With incomplete information on the individual participation, however, we entirely focus on estimating intention-to-treat (ITT) estimates and refrain from conducting treatment-on-the-treated (ToT) analysis.

³⁸Somalia is a fragile and conflict affected country where operations and research face particular challenges. The project operates in a number of regions of Somalia covering both Somaliland and Puntland and included adolescents living in internally displaced persons (IDP) camps. Data collection efforts in Somalia are difficult especially when tracking young adolescents over a period of several years. Even aside of the COVID-19 pandemic, there are several context-specific hurdles for collecting data, including border tensions between Somaliland and Puntland restricting travel to certain program areas, working in IDP camps, elections, and increased migration due to droughts that were experienced throughout the study. Despite these setbacks this study was successful in collecting data in a context where dire data gaps exist.

³⁹Batch rollout of the training was also randomized.

6 Empirical Strategy

6.1 Survey Data Analysis

Adolescent Outcomes

Our main ITT impact estimations for adolescent outcomes use the following analysis of covariance (ANCOVA) specification:⁴⁰

$$Y_{i,t=1,2} = \beta_0 + \beta_1 CHOICES_i + \beta_2 EMB_i + \beta_3 Y_{i,t=0} + \beta_4 X'_{i,t=0} + \lambda_d + \varepsilon_{it}$$
(1)

Where $Y_{i,t=1,2}$ is the outcome variable for adolescent *i* measured at midline (t = 1) or endline (t = 2), and $Y_{i,t=0}$ is the baseline value of the outcome variable.⁴¹ CHOICES_i is a dummy variable that equals 1 if the adolescent was randomly assigned to receive the *Choices* training program. That is, this indicator equals 1 if adolescent *i* was assigned to either the *Choices* only or the combined *Choices* and *EMB* treatment arms and 0 for the control group. *EMB_i* is a dummy variable that equals 1 if the adolescent also had their household assigned to receive the *EMB* training sessions, i.e. if the household was assigned to the *Choices* and *EMB* treatment arm. $X_{i,t=0}$ is a vector of baseline control variables, λ_d are district-level fixed effects, and $\varepsilon_{i,t}$ is the error term. Standard errors are clustered at the community cluster level.

In Equation 1 the coefficient β_1 measures the impact of the *Choices* training relative to the control group, and β_2 measures the marginal impact of having parents or guardians trained in *EMB*. That is, we are interested in showing the marginal effect of involving parents directly in the program. Throughout the analysis, we control for whether there were multiple adolescents eligible for the program in the household, and the age of the adolescent. We do

 $^{^{40}}$ The ANCOVA estimator has been shown to have more statistical power than a difference-in-differences estimator when outcomes are not strongly autocorrelated (McKenzie, 2012).

⁴¹For those outcome variables that were only collected during follow-up surveys, we estimate equation 1 without controlling for the baseline value of the outcome $(Y_{i,t=0})$.

not pool the two follow-up survey rounds as we are interested in estimating the short-run effects (midline) and longer-run effects (endline) separately. We run the regressions on both the full sample of adolescents as well as conduct heterogeneity analysis by respondent sex.

Our survey instruments include several questions about individual behaviors or beliefs that are closely related. To account for multiple hypothesis testing, we create indices for outcome families following the method used by Kling et al. (2007). More precisely, we aggregate primary outcome variables into an index or composite variable to be the equally weighted average of z-scores, where z-scores are calculated by first subtracting the control group mean of the variable and then dividing by the control group standard deviation before taking the average of these normalized variables. We also present sharpened False Discover Rate q-values, aggregated across outcomes within each main results table, based on a simple method proposed by Benjamini et al. (2006) to calculate the smallest level of significance at which the null hypothesis would be rejected as described in Anderson (2008). These sharpened two-stage q-values are presented in our main regression tables in square parentheses below the standard errors.

We conduct additional robustness checks for our main treatment results in the Appendix. First, we control for a social desirability scale to examine if any treatment effects are driven by responses with a socially desirable answer.⁴² Second, we test whether effects are robust to including additional control variables identified using the double-LASSO-selected controls procedure of Belloni et al. (2014).

Parental Outcomes

When analyzing the responses from adults, we are mainly interested in the ITT impact estimates for those parents who had adolescents assigned to the combined *Choices and EMB* treatment arm, i.e. the treatment arm where the respondents themselves were invited to attend a training. To do so, we estimate Equation 1 where all variables are measured at the

 $^{^{42}}$ We use the 13-item social desirability scale, which Reynolds (1982) shows is a viable substitute for the 33-item Marlowe-Crowne scale.

individual adult level but otherwise have the equivalent definition as before.⁴³ Hence, the coefficient β_1 measures the impact of a child assigned to participate in the *Choices* training on outcomes measured at the parent-level. This is relevant in cases in which adolescents, for example, pass on information from the training to their parents or if they change behavior that is observed by their parents. β_2 measures the impact of parents or guardians trained in *EMB* on their own attitudes or behaviors. Again, throughout the analysis, we control for whether there were multiple adolescents eligible for the program in the sampled household, and the age of the respondent. We also split the sample by father and mother.⁴⁴

6.2 Experimental Games Outcomes

In the analysis of *Choices* program impacts on adolescent outcome variables collected during the lab-in-the-field experiments, we use outcome-specific regression specifications as described below. For the impact analysis of the games outcomes, we focus on estimating the impact of the invitation to participate in the *Choices* program only and do not report the marginal effect for the *EMB* component.

Social Conformity and Perceived Norms

We set up a series of games to capture social influences on attitudes expressed in private and in public. First, participants were randomized into teams of 5 with different gender compositions and then were asked to report their beliefs in private during an initial pregame registration . Adolescents gave their views on three statements measuring gender attitudes.⁴⁵ Next, while in groups, adolescents were asked to give responses to the same three statements but now expressed in public to the other team members. The order in

⁴³Again, for those outcome variables that were only collected during follow-up surveys, we estimate equation 1 without the level of the outcome variable at baseline $(Y_{i,t=0})$.

 $^{^{44}}$ Note that the sample size for fathers during follow-up survey rounds is low with a response rate of 22%, whereas the sample of mothers is relatively more representative.

 $^{^{45}}$ Adolescents were asked if they agreed or disagreed to the following 3 statements: (1) It is more important for a girl to help at home than spend time studying, (2) It is okay for a man to hit his wife if she disagrees with him, and (3) Boys should have more free time than girls.

which adolescents announced their view in the team setting was randomly determined for each round. To measure social conformity we estimate the following equation:

$$Y_{ij} = \beta_0 + \beta_1 CHOICES_i + \beta_2 NumChoices_i + \beta_3 X'_i + \lambda_d + \delta_j + \varepsilon_i \quad (2)$$

CHOICES_i is an indicator variable for whether adolescent *i* was randomly assigned to Choices and X_i a vector of controls measured at baseline; defined as before. NumChoices_i controls for the number of adolescents assigned to the Choices treatment within the team excluding adolescent *i* to account for peer effects. We pool all three gendered statements and analyze the results as a stacked regression and thus include statement fixed effects (δ_j) . λ_d are district-level fixed effects. The standard errors are clustered at the individual level. The regressions are shown for the full sample of adolescents as well as split by gender.

For gender attitudes, the outcome Y_{ij} is a dummy variable that indicates whether the adolescent changed their statement j in a public group setting from what they stated in private.⁴⁶ For example, if an adolescent gave a gender egalitarian response in private and switched to a gender discriminatory response in public the outcome for this individual would be coded as 1. This is our measure of social conformity in the gender domain.

As a robustness check, we also ask adolescents to play two versions of a line match game in private, and then in public to measure social conformity more generally; in an abstract and non-gendered context. For the line match game study participants were asked to choose one line, out of four possible options, that would match a reference line in length. Two of the four options had the correct length. The outcome Y_{ij} is therefore a dummy variable for whether adolescent *i* is classified as displaying either weak conformity or strong conformity. Weak conformity refers to those adolescents who switch from one correctly identified option in private to the other correct option in public. Strong conformity refers to those adolescents who switch from a correctly identified option in private to an incorrect option in public. The sample is restricted to those who identified the correct answer in private.⁴⁷ The standard

 $^{^{46}}$ A switch between private and public responses is coded as 1 if the adolescent changed their answer in response to any one of the three statements.

⁴⁷Note as many as 91% of the study sample correctly matched the line in private.

errors are clustered at the team level.⁴⁸

For the confederates line game, the setup is very similar except that the first two respondents were asked in private to identify one of the incorrect options, and that there was only a single correct that correctly matched the length of the reference line. Again, the sample is restricted to those who answered correctly in private; omitting the leader and deputy, who were the ones prompted to answer incorrectly. Here, the outcome Y_{ij} is one of three dummy variables that capture whether the respondent (i) maintained the correct answer; (ii) named an incorrect option but not the one identified by the leader and deputy; and (iii) conformed and identified the same incorrect options as the leader and deputy. In the regressions we cluster standard errors at the team level as before.

When analyzing social conformity we also examine position effects. That is, we want to examine the effect of prior responses in the public setup. To account for this in the regressions, we include an interaction term between the *Choices* treatment dummy and whether the previous n - 1 or modal response of all previous answers in the public setting is different to the private response of adolescent *i*. For the position analysis the sample is restricted to those seated in positions 2-5 of each public game as their is no information on prior answers for the person in position 1.

$$Y_{1i} = \beta_0 + \beta_1 CHOICES_i + \beta_2 Mode_i + \beta_3 CHOICES \times Mode_i + \beta_4 NumChoices_i + \beta_5 X'_{i,t=0} + \lambda_d + \gamma_t + \varepsilon_i$$
(3)

Where $Mode_i$ is a dummy variable that is defined to either equal 1 if the n-1 response or the modal response was different from the answer respondent i gave in private and 0 otherwise. This setup allows us to understand how previous responses by others in the same team change the way respondents answer publicly and whether this is affected by the program.⁴⁹ NumChoices_i again controls for the number of adolescents assigned to the

 $^{^{48}\}mathrm{Games}$ are described in more detail in the Appendix.

⁴⁹In the line match game when analyzing the outcome for strong conformity, $Mode_i$ is a dummy variable for whether the previous or modal response for those adolescents in the positions before adolescent *i* answer incorrectly (A or C). For the outcome weak conformity, $Mode_i$ is a dummy variable for whether the previous

Choices treatment within the team excluding adolescent i to account for peer effects. As before, we run these regressions for the only boys and only girls samples separately.

Leadership

In the analysis of program impacts on the likelihood of being chosen as a team leader we estimate the following OLS regression:

$$Y_{i} = \beta_{0} + \beta_{1}CHOICES_{i} + \beta_{2}Female_{i} + \beta_{3}CHOICES \times Female_{i} + \beta_{4}NumChoices_{i} + \beta_{5}X'_{i,t=0} + \lambda_{d} + \gamma_{t} + \varepsilon_{i}$$
(4)

 Y_i is the outcome variable for whether adolescent i was selected by their team to be the team leader. As before, $CHOICES_i$ is an indicator variable for whether adolescent i is randomly assigned to receive the Choices training (with or without the parental EMB component). $Female_i$ is a dummy variable equal to 1 if adolescent i is female and $CHOICES * Female_i$ as an interaction term between treatment status and respondent sex. The coefficient β_1 measures the impact of the CHOICES training for boys, relative to the control group; β_2 gives the gender differential among control group adolescents, and β_3 is the treatment differential effect of *Choices* between girls and boys. To account for peer effects we include a control variable $NumChoices_i$ in the regression which is the number of treatment adolescents within a team, excluding adolescent i. X_i is a vector of baseline control variables as was included in the specification for the main survey - multiple adolescents in the household and adolescent age. λ_d are district-level fixed effects, γ_t are team composition controls (indicators for all boys, all girls, or mixed-gender teams), and ε_i is the error term. Standard errors are clustered

at the team level.

or modal response for those adolescents in the positions before adolescent i answer the other correct answer. In the confederates game, $Mode_i$ for the outcome stick to your guns is a dummy variable for whether any adolescent in the positions before adolescent i answered line C (correct answer); for confused, $Mode_i$ is a dummy variable for whether any adolescent in the positions before adolescent i answered line B or D (incorrect answer not primed); and for conform, $Mode_i$ is a dummy variable for whether any adolescent in the positions before adolescent i answered line A (incorrect answer primed).

7 Results

In the following section we present our primary results on a range of adolescent-level outcomes: (1) self-reported gender equitable attitudes, (2) social conformity, and (3) leadership outcomes.⁵⁰ Results at both midline and endline are presented, where applicable.

7.1 Impacts on Gender Equitable Attitudes

Table 2 shows the treatment effect of the *Choices* gender norms training on adolescent gender equitable attitudes. The outcomes in Table 2 include an index that aggregates personal beliefs across 14 gender attitude statements, which we further decompose into 3 sub-indices (see further details on the construction of these indices and statements in Figure 4 and Table A8 in the Appendix). We find that *Choices* led respondents to report more egalitarian gender attitudes. Overall, adolescents assigned to receive the *Choices* treatment have a 0.05 standard deviation (s.d.) higher attitudes index than those in the control group at endline (Panel A), and a 0.07 s.d. higher index at midline (Panel B), relative to the control group as shown in Table 2, column $1.^{51}$ *Choices* adolescents whose parents were also invited to attend the *EMB* training saw minimal marginal impact on attitudes. Columns 2 and 3 present impacts by gender. *Choices* had a positive impact on the gender attitudes of both adolescent boys and girls in the short-run and in the longer-run this effect persists for girls. The effect on the full attitudes index is no longer statistically significant for boys at endline.

Decomposing the full-index into three domains (right to education, women's role and masculinity), however, suggests that the effect on attitudes for boys persists at endline in the education domain. The *Choices* treatment leads to a 0.12 s.d. increase in boys' gender

 $^{^{50}}$ In the Appendix we examine additional impacts on self-reported behavioral outcomes for adolescents, including time spent on household chores, caring behaviors, mental health, and schooling; and the direct *EMB* impacts on parents' gender equitable attitudes and behaviors.

 $^{^{51}}$ To help interpret the magnitude of the *Choices*' effect on attitudes, we can translate the results in terms of changes in the proportion of the sample with gender egalitarian attitudes. The effect size suggests that the share of adolescents with egalitarian attitudes towards gender equality (across all 14 statements) increased from around 50% at baseline to 56% at endline in the *Choices* treatment group.

progressiveness with respect to the right to education at endline (column 4). However, boys' attitudes towards women's role and masculinity norms (columns 6 and 8) are not statistically different from the control group. Girls who were assigned to the *Choices* treatment expressed higher gender equitable attitudes across a greater number of sub-themes. For example, attitudes related to education increase by 0.14 s.d. at endline (0.10 s.d. at midline), and by 0.09 s.d. for attitudes related to women's role at endline (0.12 at midline).⁵²

In Appendix Table A8 we also decompose the index into its individual statements and estimate differences between treatment and control at baseline, midline and endline.⁵³ Overall, our results suggest variation in the types of attitudes that were influenced by the *Choices* intervention. The program was mostly successful in shifting beliefs to be more gender progressive with respect to a girl's right to education and a women's role in the economic sphere.

7.2 Impacts on Social Conformity

In the following we analyze the social dynamics of how an adolescent's views are influenced by others around them using an experiment that measured attitudes in private vs. public. Our measure of conformity considers whether the adolescent changes their response from private to public based on the responses of others and by having to respond publicly. We randomize: (i) the gender composition of the team i.e. all boys, all girls or mixed-gender teams; and (ii) the position that the adolescent responds in public. This allows us to estimate both peer group effects and position effects.⁵⁴ Individual private responses were first taken from each adolescent after they were randomly assigned to their team. Next, the different conformity games were played in teams of 5 adolescents in public. We are interested in perceived normative pressures and whether *Choices* adolescents are influenced by their peers in the same way as adolescents from the control group.

⁵²Masculinity norms appear stickier and do not seem to have changed. However, attitudes towards a girl's right to education appears more malleable.

⁵³Note that the sample size differs between Table 2 and Table A8 as A8 includes the panel respondents.

⁵⁴The team members were randomly seated in a row to sequentially answer three gender attitudes statements and two line match games. In each subsequent game the position of the adolescent was re-randomized.

	O a m d a m E	av site belle	A 44:4		E du a ati a m			la Dala	-	NA			
	Gender E Stand	duitable / lardized l	ndex	Attitu	udes		women Attitu	is Role Jdes		Attit	udes		
	1	- ull-Index		Sub-I	ndex		Sub-I	ndex		Sub-Index			
	All	Boys	Girls	Boys	Girls		Boys	Girls		Boys	Girls		
Donal A: Endlina	(1)	(2)	(3)	(4)	(3)		(0)	(/)		(0)	(9)		
	0 051++	0.000	0 075+++	0 11 5+++	0 1 0 7+++		0.006	0 000++		0.000	0.001		
CHUICES	0.051^^	0.029	0.075^^^	0.115^^^	0.137^^^		0.020	0.089^^		-0.029	0.031		
	(U.UZ)	(0.03)	(0.03)	(0.04)	(0.05)		(0.04)	(0.03)		(0.03)	(0.03)		
	[0.03]	[0.43]	[0.01]	[0.00]	[0.01]		[1.00]	[0.01]		[0.41]	[0.34]		
EMB (marginal)	0.019	-0.016	0.043	-0.011	0.021		-0.052	0.032		-0.005	0.054**		
((0.02)	(0.03)	(0.03)	(0.04)	(0.05)		(0.04)	(0.05)		(0.03)	(0.03)		
	[0.67]	[1 00]	[0 1 9]	[1 00]	[1 00]		[0.26]	[0 99]		[1 00]	[0.05]		
	[0.07]	[1.00]	[0.19]	[1.00]	[1.00]		[0.20]	[0.55]		[1.00]	[0.00]		
Control group mean	-0.00	-0.02	0.01	-0.04	0.02		-0.01	0.01		0.00	0.00		
Adjusted R ²	0.07	0.06	0.08	0.07	0.08		0.04	0.04		0.02	0.02		
Observations	2,643	1,285	1,358	1,285	1,358		1,285	1,358		1,285	1,358		
ANCOVA	Yes	Yes	Yes	Yes	Yes		Yes	Yes		Yes	Yes		
Panel B: Midline													
CHOICES	0.066***	0.064**	0.073**	0.075	0.104*		0.030	0.116***		0.039	0.047		
	(0.02)	(0.03)	(0.03)	(0.05)	(0.06)		(0.04)	(0.04)		(0.03)	(0.03)		
	0.01	[0.04]	[0.02]	[0.17]	0.08		0.89	0.00]		0.20	[0.14]		
EMB (marginal)	-0.002	-0.002	-0.014	-0.019	-0.012		-0.013	0.013		-0.041	-0.026		
	(0.02)	(0.03)	(0.03)	(0.05)	(0.04)		(0.03)	(0.04)		(0.03)	(0.03)		
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]		[1.00]	[1.00]		[0.26]	[0.47]		
Control group mean	-0.01	-0.05	0.03	-0.08	0.09		-0.00	0.02		-0.01	-0.00		
Adjusted P^2	0.06	0.07	0.06	0.00	0.11		0.06	0.06		0.02	0.00		
Aujusteu R Obcorvations	1 0 9 7	0.07	1 0/10	0.10	1 0/10		0.00	1 0/10		0.02	1 0/12		
	1,907	909 Vaa	1,040 Vaa	909	1,040 Maa		909 Vee	1,040		909 Vee	1,040		
ANCOVA	res	res	res	res	res		res	res		res	res		

Table 2: Impact on Adolescent Gender Equitable Attitudes

Notes: * significant at 10% level ** significant at 5% level *** significant at 1% level

(1) CHOICES=1 if adolescent was invited to attend the CHOICES training (assigned to CHOICES only or CHOICES+EMB treatment

groups). EMB=1 if adolescent was in CHOICES+EMB group, effectively giving the marginal impact of the EMB program in addition to the CHOICES program.

(2) ANCOVA estimation controls for the level of the outcome variable. All regressions include district fixed effects and a set of controls measured at baseline: dummy for whether household has multiple adolescents and age of respondent.

(3) Standard errors are clustered at the community level and are reported under the coefficient in parentheses. Sharpened q-values that correct p-values for the false discovery rate (FDR) are in square brackets. A sharpened q-value of 1.00 indicates the null hypothesis is not rejected at any level of FDR.

(4) Outcomes: Gender Equitable Index is a standardized index of 14 gender-progressive statements the respondent agrees with (index standardized as per method from Kling et al. 2007). Higher scores indicate more gender-equitable attitudes. The education sub-index, women's role sub-index, and masculinity norms sub-index are constructed from (mutually exclusive) statements in the Gender Equitable Index.

7.2.1 Gender Attitudes Statements

Tables 3, 4 and 5 present the results for the gender statements where we observe social dynamics specific to gender attitudes. In Table 3 we examine group effects, in Table 4 position effects, and in Table 5 the direction of switching (egalitarian-discriminatory) from private to public. Our sample consists of all adolescents in the lab-in-the-field experiment.⁵⁵

Private vs. Public

We are primarily interested in how the gender composition of their peer group as well as what a peer in the position just ahead of them said influences their responses; and how *Choices* affects that influence.⁵⁶

In Table 3, columns 1-3 present the results for the private responses where the outcome is a dummy variable for whether adolescent i gives a gender egalitarian response to the statement in private. We show that the *Choices* treatment increases the likelihood that the adolescent gives a gender egalitarian response to any of the three gender attitudes; a 6.6 percentage point (pp) increase in egalitarian attitudes.⁵⁷ In columns 2 and 3 we split the sample by respondent sex. The treatment effect is statistically significant for both adolescent boys and girls separately.

⁵⁵The sample size for the experimental sessions is 1,000 respondents with 994 in the analysis as position was not recorded accurately for 6. With 3 different statements this makes 2,982 observations in Table 3.

⁵⁶The reference group in our lab-in-the-field experiments are other adolescent peers randomly assigned to the same team. We concede that those who are non-conformist in the lab setting may conform in other settings where social sanctions or disapproval are a bigger threat. For example, perhaps if the reference group were parents, siblings, teachers, or other adults in the community we might expect different social dynamics as to what is found among peers.

 $^{^{57}}$ Note the control group mean for the three attitudes in the lab experiment in Table 3 column 1 is 75% which is already relatively high and suggests that study participants already held largely egalitarian views on gender.

		Private	(Egalitarian (All Pos	Gender Att sitions)	titudes)	Public (Egalitarian Gender Attitudes) (All Positions)							
· · · · · · · · · · · · · · · · · · ·	All	Boys	Girls	All	Boys	Girls	All	Boys	Girls	All	Boys	Girls	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	
CHOICES (0/1)	0.066***	0.060**	0.074***	0.056*	0.087**	0.068**	0.011	0.003	0.021	0.021	0.031	0.024	
	(0.02)	(0.03)	(0.02)	(0.03)	(0.04)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)	(0.04)	(0.03)	
Number of CHOICES peers in team (0-4)	-0.004	-0.003	0.007	-0.004	-0.003	0.007	0.018**	0.023**	0.015	0.019**	0.023**	0.015	
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	
Mixed-gender teams	-0.066***	0.065***	-0.044**	-0.072**	0.102**	-0.052	-0.014	0.044*	-0.012	0.001	0.080*	-0.007	
	(0.02)	(0.03)	(0.02)	(0.03)	(0.05)	(0.04)	(0.02)	(0.03)	(0.02)	(0.03)	(0.04)	(0.03)	
Boys only team	-0.144***			-0.159***			-0.079***			-0.086**			
	(0.02)			(0.05)			(0.02)			(0.04)			
CHOICES*Mixed-gender teams				0.009	-0.051	0.013				-0.022	-0.052	-0.006	
				(0.04)	(0.06)	(0.04)				(0.04)	(0.05)	(0.04)	
CHOICES*Boys only team				0.021						0.009			
				(0.05)						(0.05)			
Control group mean	0.75	0.72	0.78	0.75	0.72	0.78	0.80	0.78	0.81	0.80	0.78	0.81	
Adjusted R-Squared	0.11	0.11	0.12	0.11	0.12	0.12	0.09	0.12	0.06	0.09	0.12	0.06	
Observations	2982	1473	1509	2982	1473	1509	2937	1446	1491	2937	1446	1491	

Table 3: Impact on Social Conformity Outcomes (Group Conformity Effects) Gender Statement Game

Notes: * significant at 10% level ** significant at 5% level *** significant at 1% level

(1) CHOICES=1 if adolescent was assigned to CHOICES only or CHOICES+EMB treatment groups; 0 if in control group.

(2) Standard errors are clustered at the individual level and are reported under the coefficient in parentheses. All OLS regressions includes district fixed effects and controls for whether household has multiple adolescents, the age of respondent, and peer effects given by the number of CHOICES adolescents in the adolescent's team excluding the respondent (0-4). Regressions control for randomized team gender composition (whether team is mixed-gender, boys only, and girls only) where girls only is the omitted category in the regression, statement fixed effects and enumerator-gender fixed effects.

(3) Outcomes: Private Egalitarian is a dummy variable for whether the adolescent reported an egalitarian response to the gender statements in private during registration before the start of the lab experimental games. Public Egalitarian is a dummy variable for whether the adolescent reported an egalitarian response to the gender statements in a public group setting. Sample size in private (994 respondents) is slightly larger than in public (979 respondents) since the public responses of 3 teams (15 respondents) were missing.

(4) Table 7A columns (4)-(6) and (10)-(12) analyzes the group dynamics by treatment that interacts the Choices treatment dummy with the gender composition of the team.

Table 3 also shows the group effects; namely the coefficients on *mixed-gender teams* and *boys only team*. In column 1, the privately expressed attitudes of individuals in the girls-only team (the omitted category) are significantly more gender egalitarian than those expressed by individuals in the mixed-gender or boys-only teams.⁵⁸ Decomposing this by the sex of the respondent (columns 2 and 3) we show the group effects for boys and girls point in opposite directions. That is, when boys are in a mixed team they are more likely to state privately gender egalitarian attitudes (6.5pp) than in a team with all boys. When girls are in a mixed team they are more likely to state privately discriminatory response (-4.4pp) than in a team with all girls. Our findings indicate significant social influence based on the sex-composition of the group. Even without communication, we find adolescents altering what attitudes are expressed in private. Boys respond to the presence of girls by giving more egalitarian responses and the opposite is true for girls. This may mean that girls and boys experience different types of social influence and that social desirability bias might go in opposite directions. In columns 4-6 we also interact the team composition dummy variables with the *Choices* treatment dummy and do not find a differential impact.

In Table 3 columns 7-12 we turn to the responses given in public where the outcome is a dummy variable for whether adolescent i gives a gender egalitarian response in the public group setting. On average, we find no evidence that *Choices* adolescents have greater egalitarian attitudes in public, relative to the control group. For boys, in Table 3 column 8 the public response is positively influenced by the number of other *Choices* peers in their team. The group-conformity effects based on the gender composition of the team seen in private still matter, but less so in public. Boys in mixed-gender teams are more gender egalitarian than boys in the all-boys teams in public. As was seen in private, we find that *Choices* does not differentially impact the extent of group conformity in public.

In summary, when answering in private, boys and girls give more egalitarian responses if they participated in *Choices*. The contexts of (i) *Choices* peers and (ii) mixed-gender teams

 $^{^{58}82\%}$ of responses for girls in the control group who are randomized into the all girls team are egalitarian, compared with 67% of control group boys allocated to the all boys team.

both encourage boys to give more egalitarian responses. In public, boys give more egalitarian responses if there are more *Choices* peers in their group. Perhaps boys with more *Choices* peers in their group are more confident that their peers will give egalitarian responses or more confident that they will not be sanctioned for giving egalitarian responses themselves. Whereas for girls, the presence of *Choices* peers has minimal impact and mixed-gender teams encourage less egalitarian responses. This suggests that normative pressures from peers are different for boys and girls.⁵⁹ Taken together these results suggest that group dynamics are important and that there is some attitudinal change occurring between private and public.

Position Effects

In Table 4 the outcome indicates whether there was any switch in response from private to public, i.e. dummy variable equals 1 if the adolescent switched from being egalitarian in private to discriminatory in public, or 1 if the adolescent switched from being discriminatory in private and egalitarian in public; and 0 if the adolescent gave the same response in public as in private.⁶⁰ When adolescents are asked their attitudes publicly, some adjust their responses to conform to the peers around them (24% of the control group switch their responses in public from what they reported in private). The negative coefficient in Table 4 column 1 on the *Choices* treatment dummy suggests the program reduces the likelihood that an adolescent changes their response when publicly stating a gender attitude in the presence of a group of randomly selected peers. We show that *Choices* adolescents are, on average, more likely to be egalitarian in private and, more importantly, they are more likely to stick to those egalitarian beliefs when asked to state them in public.

In the following we examine position effects i.e. the influence of hearing the attitudes of others in public in the positions before you.⁶¹ In Table 4 columns 4-9 we examine the

 $^{^{59}}$ Comparing the control group means in Table 3 suggests that the private vs. public mean for boys is statistically significantly different.

 $^{^{60}}$ In Table 4 we simply interpret a switch as indicative of conformist behavior, and in Table 5 we further explore the direction of the switch.

 $^{^{\}hat{6}1}$ In the analysis for position effects we include controls for the team sex-composition and do not interact with treatment.

Switch in Public outcome for the adolescents' seated in the 2nd to 5th positions to explore whether adolescents change their opinion from private to public based on the responses of adolescents in the preceding position(s).⁶² The negative coefficient on *Choices* in column 4 suggests that results hold after restricting the sample to adolescents in positions 2-5 only. We find *Choices* adolescents randomly positioned in seats 2-5 are less likely to change their response when publicly stating gender attitudes, relative to the control group. In columns 7-9 we test if the adolescents are more or less likely to change their attitudes if the public response of the person in the preceding position to adolescent was different from their own. The coefficient on the position variable is positive and significant which suggests that there is a greater likelihood that the adolescent changes their response in public, i.e. there is a greater tendency to conform to the responses of those who are positioned before you.⁶³ In columns 8 and 9 we show the results disaggregated by sex. *Choices* boys in positions 2-5 are less likely to switch their response in public, especially when the response of those in the preceding position is different to what they privately said (the interaction term is significant at 99% confidence). The influence of a different prior response is diminished for *Choices* boys - they are more likely to stick to their personal beliefs even when the person positioned before them disagrees with them. *Choices* girls are also less likely to change their response in public than the control group. However, this does not seem to be influenced by the responses of those in the preceding positions. That is, *Choices* girls are less likely to switch their response in public irrespective of what was said in the position before. In mixed-gender teams, however, girls are more likely to switch their response in public.

 $^{^{62}}$ Those in position 1 are not affected by hearing the responses given by peers in their group but might be affected by having their own response being heard in public. Perhaps those who change their response in public in position 1 may be inclined to do so because of social desirability, for example, by claiming a more gender egalitarian/discriminatory response to a statement in a group vs what was said in private.

⁶³In the Appendix Table A11 we show results are robust to considering the responses of the majority in the preceding positions being different to i's private response i.e. the mode of responses. Results in Table 4 columns 7-9 are robust to using the modal response, i.e. when 50 percent or more of the responses in the preceding positions in the row are different to adolescent i's private response. Note for the adolescents seated in odd positions the mode was taken to be n-1 when exactly 50 percent of the public responses in the preceding positions are different to i's private response.

	Switch in Public (All Positions)			Switch in Public (Positions 2-5)			Switch in Public (Positions 2-5 and response in positic before interacted with treatment)		
	All	Boys	Girls	All	Boys	Girls	All	Boys	Girls
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
CHOICES (0/1)	-0.045**	-0.044	-0.049**	-0.043**	-0.031	-0.056**	-0.016	0.008	-0.041*
	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.02)
Number of CHOICES peers in team (0-4)	-0.003	0.003	-0.010	0.003	0.018	-0.013	0.003	0.018	-0.012
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Mixed-gender teams	0.085***	0.000	0.082***	0.076***	-0.007	0.070***	0.039*	0.007	0.040*
	(0.02)	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
Boys only team	0.086**			0.083**			0.039		
	(0.03)			(0.04)			(0.03)		
Response of position before different to i's private response							0.513***	0.613***	0.420***
							(0.04)	(0.05)	(0.06)
CHOICES*Response of position before different							-0.078*	-0.202***	0.038
							(0.05)	(0.06)	(0.07)
Control group mean	0.24	0.28	0.21	0.25	0.27	0.23	0.25	0.27	0.23
Adjusted R-Squared	0.10	0.07	0.11	0.10	0.07	0.13	0.30	0.29	0.31
Observations	2937	1446	1491	2352	1151	1201	2352	1151	1201

Table 4: Impact on Social Conformity Outcomes (Position Effects) Gender Statement Game

Notes: * significant at 10% level ** significant at 5% level *** significant at 1% level

(1) CHOICES=1 if adolescent was assigned to CHOICES only or CHOICES+EMB treatment groups; 0 if in control group.

(2) Standard errors are clustered at the individual level and are reported under the coefficient in parentheses. All OLS regressions includes district fixed effects and controls for whether household has multiple adolescents, the age of respondent, and peer effects given by the number of CHOICES adolescents in the adolescent's team excluding the respondent (0-4). Regressions control for randomized team gender composition (whether team is mixed-gender, boys only, and girls only) where girls only is the omitted category in the regression, statement fixed effects and enumerator-gender fixed effects.

(3) Outcomes: Switch in Public is whether the adolescent changed her privately held belief when in a public group setting (=1 if adolescent changed their private response in public; 0 if stuck with their private response). A negative coefficient of being less likely to change in public is indicative of demonstrating greater non-conformist behavior.

(4) Table 3B columns (7) to (9) analyzes the influence of position that you play the game by interacting the CHOICES treatment dummy variable with a dummy variable that equals 1 if the public response of the individual seated in the position before adolescent i is different to i's private response.

Direction of Switching in Public

In Table 5 we unpack the direction of the *Switch in Public* outcome presented before.⁶⁴ In columns 1-3 the outcome *Always Egalitarian* is a dummy variable equal to 1 if the adolescent gave a gender egalitarian attitude in private *and* in public; and in columns 4-6 *Always Discriminatory* is a discriminatory attitude in private and public. Both *Choices* boys and girls have a higher probability of being in the *Always Egalitarian* category compared to adolescents in the control group (see grey bars in Figure 7 multinomial logit in the Appendix). *Choices* boys are less likely to be influenced by a gender discriminatory prior response than boys in the control group where *Choices* boys are more likely to be *Always Egalitarian* outcome in column 2 shows a positive and significant coefficient on CHOICES interacted with the prior response, and conversely for *Always Discriminatory* in column 5 this coefficient is negative and significant). *Choices* girls are more likely to be *Always Egalitarian* and are not differently influenced by the responses of those in the preceding positions than control group girls.

In terms of the switchers, in Table 5 columns 7-12 we condition the analysis on being privately egalitarian (7-9) or privately discriminatory (10-12). We show a higher likelihood to switch and conform to a discriminatory view in public if the person before you gives a discriminatory response, and a lower likelihood to switch to an egalitarian view if the person before you gives a discriminatory response. Among boys, column 8 suggests that *Choices* boys (but not girls) are less influenced by a prior response to switch from egalitarian to discriminatory; and column 11 suggests the *Choices* peer effects point to an egalitarian shift - boys are more likely to switch from privately discriminatory to publicly egalitarian when they had a greater number of *Choices* adolescents in their team. Table 5 column 12 suggests that *Choices* girls who were discriminatory in private are even less likely to switch to an

 $^{^{64}}$ Figure 7 in the Appendix shows the predicted probabilities from a multinomial logistic regression which is a useful visualization of the dynamics of what is occurring in private vs. public.

egalitarian attitude in public when the prior response was discriminatory than the control group.

In summary, Table 5 suggests that having participated in *Choices* encouraged boys to stick with an egalitarian response in the face of opposing discriminatory views, presumably because *Choices* made boys think that expressing such views would lead to greater social approval or lower chances of disapproval. Another way of saying this is that the presence of *Choices* peers and the presence of girls make boys more likely to think that their peers are egalitarian. For girls, the effect is driven by *Choices* both encouraging girls to stick with an egalitarian private view in public, and affirming girls who gave a discriminatory response in private to not switch to egalitarian in public especially when the prior response was discriminatory. In other words, *Choices* convinced some girls with discriminatory attitudes that they could defend those attitudes in public. While *Choices* encouraged more egalitarian private attitudes among girls and they stick to those convictions, there is not a lot of evidence that it convinced (more) girls that egalitarian responses are the popular ones among these peers. Perhaps there are social influences that are pushing girls in opposing directions.

We consider a number of potential mechanisms through which the program could influence an individual's response: personal attitudinal change, change in perceived egalitarianism of peers, change in perceived likelihood of positive/negative sanctions for expressing egalitarian ideals, or change in willingness to endure sanctions for expressing egalitarian ideals. We see evidence of the personal attitudinal change for boys and girls. There is evidence that *Choices* changed some aspect of perceived norms for boys, but minimal evidence of this channel for girls. In Table A2 we present some explicit measures of perceived community norms that asks "out of 10 of your neighbors how many would believe..." (related to women's work out of the home and child marriage). We show that while *Choices* girls perceive slightly more support for women's work out of the home from their community, there are limited changes in their overall perceptions of norms among the community.⁶⁵

⁶⁵Girls' perceptions of community norms are, on average, similar to parents' perceptions of the same norm.

	Always Egalitarian in Private and in Public (Positions 2-5)			Alway in Priv (P	Always Discriminatory in Private and in Public (Positions 2-5)			Egalitarian to Discriminatory Switch in Public (Positions 2-5)			Discriminatory to Egalitarian Switch in Public (Positions 2-5)		
	All	Boys	Girls	All	Boys	Girls	All	Boys	Girls	All	Boys	Girls	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	
CHOICES (0/1)	0.031	0.013	0.054**	0.000	0.016	-0.017	-0.004	0.004	-0.015	-0.031	-0.102	0.073	
	(0.02)	(0.03)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.05)	(0.07)	(0.09)	
Number of CHOICES peers in team (0-4)	-0.004	-0.017	0.014	-0.001	-0.004	-0.002	-0.008	-0.001	-0.013	0.031	0.054**	-0.006	
	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)	(0.03)	(0.03)	
Mixed-gender teams	-0.043*	0.021	-0.041*	-0.027*	-0.031*	-0.024	0.016	0.019	0.018	0.154**	0.028	0.128*	
	(0.02)	(0.03)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.02)	(0.02)	(0.07)	(0.07)	(0.08)	
Boys only team	-0.063*			0.002			0.003			0.125			
	(0.04)			(0.02)			(0.03)			(0.09)			
Discriminatory response in position before	-0.446***	-0.464***	-0.425***	0.241***	0.299***	0.187***	0.421***	0.535***	0.365***	-0.333***	-0.411***	-0.199*	
	(0.04)	(0.06)	(0.06)	(0.04)	(0.06)	(0.05)	(0.06)	(0.10)	(0.08)	(0.07)	(0.10)	(0.11)	
CHOICES*Discriminatory response in position before	0.102**	0.201***	0.001	-0.056	-0.178**	0.056	-0.119*	-0.265**	-0.029	0.014	0.186	-0.215*	
	(0.05)	(0.07)	(0.08)	(0.05)	(0.07)	(0.06)	(0.07)	(0.11)	(0.10)	(0.09)	(0.12)	(0.13)	
Control group mean	0.66	0.62	0.69	0.09	0.10	0.08	0.13	0.14	0.12	0.61	0.62	0.60	
Adjusted R-Squared	0.25	0.22	0.28	0.14	0.15	0.15	0.21	0.19	0.23	0.18	0.18	0.20	
Observations	2352	1151	1201	2352	1151	1201	1861	870	991	491	281	210	

Table 5: Impact on Social Conformity Outcomes (Direction of Switching) Gender Statement Game

Notes: * significant at 10% level ** significant at 5% level *** significant at 1% level

(1) CHOICES=1 if adolescent was assigned to CHOICES only or CHOICES+EMB treatment groups; 0 if in control group.

(2) Standard errors are clustered at the individual level and are reported under the coefficient in parentheses. All OLS regressions includes district fixed effects and controls for whether household has multiple adolescents, the age of respondent, and peer effects given by the number of CHOICES adolescents in the adolescent's team excluding the respondent (0-4). Regressions control for randomized team gender composition (whether team is mixed-gender, boys only, and girls only) where girls only is the omitted category in the regression, statement fixed effects and enumerator-gender fixed effects.

(3) Outcomes: Always Egalitarian is a dummy variable for whether the adolescent reported an egalitarian response to the gender statements in private and in public. Always Discriminatory is whether the adolescent gave a discriminatory response in private and in public. Egalitarian to Discriminatory Switch in Public is whether the adolescent changed her privately held belief when in a public group setting from an egalitarian view to discriminatory to Egalitarian Switch in Public is whether the adolescent changed her privately held belief when in a public group setting from a egalitarian view (=1 if adolescent changed their private response). Sample size in columns 7-12 is lower than k in coulmns 1-6 since we condition the analysis on being egalitarian (7-9) or discriminatory (10-12) in private and then examine where there is a switch in public.

(4) Sample is restricted to those adolescents in positions 2-5 to analyze position effects. The influence of position that you play the game is analyzed by interacting the CHOICES treatment dummy variable with a dummy variable that equals 1 if the public response of the individual seated in the position before adolescent i is discriminatory.
Taken together, this means that, when adolescents are asked their attitudes publicly, they are inclined to adjust their responses to conform to those in the positions before them (see positive coefficient on "Response of position before different to i's private response" in Table 4). In public there are higher average gender egalitarian attitudes reported by everyone largely driven by *Choices* adolescents being more likely to be privately egalitarian and sticking with that view in public. Control group adolescents are more likely to be egalitarian in public than in private, which suggests a shift towards a more egalitarian group norm perhaps since these became the dominant attitudes set within the lab.⁶⁶ We refer to *Choices* adolescents as "non-conformist" in reference to sticking to a privately held view in public, even when faced with opposing views. In the subsequent line match conformity games we provide evidence that our results do not reflect general adolescent rebellion as the treatment effects on non-conformity are only found in the gender domain.

7.2.2 Non-gender Conformity Experiments

Line Match Game

Table 6 presents the results from line match game designed to replicate the Asch (1952) conformity experiments.⁶⁷ In the line match game we do not prime subjects (i.e. actors or confederates) but modify the game to have 2 lines of equal size to create a 50-50 chance of picking one of the correct answers (lines B and D). In columns 1-3 the outcome is a dummy variable for whether the adolescent was privately correct when matching the line. Column 1 suggests that adolescents were correct in their response 91% of the time i.e. when responding to an enumerator in private, about 9% make a mistake and wrongly match the line. We do not find a *Choices* treatment effect on the likelihood the adolescent gave a correct response to the line match game in private.⁶⁸

 $^{^{66}}$ This suggests that it may be possible to alter the attitudes of a few key actors in small close-knit groups to leverage positive feedback effects from social interactions.

⁶⁷See the Appendix for further details on the design of the conformity games.

 $^{^{68}\}mathrm{This}$ is reassuring since we would not expect to find a treatment effect on accuracy.

To examine conformity we define two outcomes: weak and strong conformity. Where *weak conformity* is a dummy variable that indicates the adolescent changed their response in public from what was said in private but that response was still correct (i.e. a change from B to D, or D to B). Here we examine whether adolescents in preceding positions in the line-up influence responses. We restrict the sample to those adolescents in positions 2-5 as before and interact treatment with the modal public response of those in the preceding positions to adolescent i being the opposite correct answer to i's private answer.⁶⁹ We label this weak conformity as we consider a change to the responses of those in the preceding positions still conformist behavior, even if both answers are correct. *Strong conformity* is a dummy variable that indicates the adolescent changed their response in public from what was said in private and their public response was incorrect (i.e. change from B or D in private to C or A in public). For the strong conformity outcome we similarly restrict the sample to adolescents in positions 2-5 and interact the treatment dummy with the modal response of those in the preceding positions to adolescents in positions 2-5 and interact the treatment dummy with the modal response of those in the preceding positions to adolescents in positions 2-5 and interact the treatment dummy with the modal public response of those in the preceding positions to adolescents in positions 2-5 and interact the treatment dummy with the modal response of those in the preceding positions to adolescent i being an incorrect answer.

The results suggest that 26% of the control group demonstrate weak conformist behavior and 20% demonstrate strong conformist behavior. We find no evidence of a *Choices* treatment effect on weak or strong conformity for the line match 1 game. Similar to the Asch experimental findings, we find participants are more likely to respond incorrectly when asked to respond in public i.e. 9% incorrectly matched the line in private, whereas in public 20% incorrectly matched the line. Interestingly, we find that the modal response of what adolescents say in the preceding positions is positive and significant for the girls sample (see column 9). That is, adolescent girls were more influenced by the responses of adolescents in the preceding positions and were more likely to change their response to an incorrect answer when most of the responses were incorrect in the positions before them in the random lineup. Since girls are more influenced by the prior responses, this indicates that girls could be more sensitive to social pressure in this game (and possibly more broadly).

 $^{^{69}}$ We exclude those adolescents who gave an incorrect response in private as their public response may be measuring something different to conformity.

Social Conformity - Line Match Game 1	Private Correct Line Match (B or D)			V Change in F (P	Veak Conforn Public from B Positions 2-5	nity to D or D to B only)	Strong Conformity Change in Public from B or D to incorrect (Positions 2-5 only)		
	All	Boys	Girls	All	Boys	Girls	All	Boys	Girls
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
CHOICES	-0.019	0.002	-0.038	0.024	0.028	0.023	-0.004	0.004	-0.023
	(0.02)	(0.03)	(0.03)	(0.04)	(0.06)	(0.05)	(0.03)	(0.03)	(0.04)
Number of CHOICES peers in team (0-4)	-0.020	-0.052***	0.003	-0.002	-0.013	0.012	0.010	-0.009	0.033
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)	(0.02)
Mixed-gender teams	-0.005	0.021	-0.005	-0.100**	0.056	-0.107**	-0.023	0.074	-0.020
	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Boys only team	-0.045			-0.158**			-0.077		
	(0.06)			(0.07)			(0.07)		
Response of position before is other correct answer				0.226***	0.184*	0.271***			
				(0.07)	(0.10)	(0.10)			
CHOICES # Response of position before is other correct answer				0.010	0.112	-0.091			
				(0.08)	(0.12)	(0.13)			
Response of position before is an incorrect answer							0.331***	0.131	0.481***
							(0.10)	(0.14)	(0.12)
CHOICES # Response of positions before is an incorrect answer							-0.004	0.195	-0.172
							(0.10)	(0.15)	(0.14)
Control group mean	0.90	0.90	0.91	0.26	0.22	0.29	0.21	0.16	0.24
Adjusted R-Squared	0.02	0.05	0.00	0.08	0.08	0.07	0.18	0.22	0.18
Observations	994	491	503	697	351	346	697	351	346

Table 6: Impact on Social Conformity Outcomes (Group and Position Effects) – Asch Line Match 1 Game

Notes: * significant at 10% level ** significant at 5% level *** significant at 1% level

(1) CHOICES=1 if adolescent was assigned to CHOICES only or CHOICES+EMB treatment groups; 0 if in control group.

(2) Standard errors are clustered at the session-team level and are reported under the coefficient in parentheses. All OLS regressions includes district fixed effects and controls for whether household has multiple adolescents, the age of respondent, and peer effects given by the number of CHOICES adolescents in the adolescent's team excluding the respondent (0-4). Regressions control for randomized team gender composition (mixed-gender, boys only, and girls only where girls only is the omitted category), and enumerator-gender fixed effects.

(3) Outcomes Weak Conformity and Strong Conformity are restricted to the adolescents in positions 2-5 who correctly matched the line when responding in private. 90% were correct in private.
(4) Position effects for the outcome "Weak conformity" in columns (4) to (6) are estimated by the influence of the response in the position before i being the other correct answer (i.e. B if you said D in private, or D if you said B in private). "Strong conformity" position effects in (7) to (9) are estimated by the response of the previous position being an incorrect answer (A or C).

Table 7: Impact on Social Conformity Outcomes (Group and Position Effects) – Asch Confederates Line Match Game

Social Conformity _ Confederates Game	Private Correct Line Match (C)			Stick to your guns in public (C in private and C in public)			Confused in Public (C in private and incorrect B or D in public)			Conform in Public (C in private and incorrect A in public)		
	All	Boys	Girls	All	Boys	Girls	All	Boys	Girls	All	Boys	Girls
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
CHOICES	0.039	0.072	0.009	-0.020	-0.062	0.007	-0.004	-0.005	-0.021	0.025	0.067	0.013
	(0.03)	(0.04)	(0.04)	(0.05)	(0.08)	(0.07)	(0.05)	(0.08)	(0.06)	(0.06)	(0.08)	(0.08)
Number of CHOICES peers in team (0-4)	-0.011	-0.002	-0.021	0.009	-0.010	0.033	-0.023	-0.032	-0.036	0.014	0.042	0.003
	(0.02)	(0.02)	(0.02)	(0.02)	(0.04)	(0.03)	(0.02)	(0.04)	(0.02)	(0.02)	(0.04)	(0.03)
Mixed-gender teams	0.001	0.056	-0.005	-0.067	0.045	-0.082	0.033	0.006	0.066	0.033	-0.051	0.016
	(0.04)	(0.05)	(0.04)	(0.06)	(0.06)	(0.06)	(0.05)	(0.07)	(0.05)	(0.07)	(0.08)	(0.07)
Boys only team	-0.061			-0.128			0.039			0.089		
	(0.07)			(0.08)			(0.08)			(0.10)		
Any response in positions before correct answer C				0.455***	0.433***	0.464***	-0.177***	-0.247**	-0.136*	-0.278***	-0.185*	-0.329***
				(0.08)	(0.13)	(0.11)	(0.07)	(0.11)	(0.08)	(0.07)	(0.11)	(0.10)
CHOICES # Any response in positions before correct answer C				-0.076	-0.032	-0.097	0.051	0.041	0.094	0.026	-0.009	0.004
				(0.08)	(0.14)	(0.12)	(0.07)	(0.12)	(0.08)	(0.08)	(0.13)	(0.12)
Control group mean	0.80	0.77	0.83	0.43	0.42	0.45	0.32	0.38	0.28	0.24	0.20	0.28
Adjusted R-Squared	0.03	0.06	0.00	0.29	0.28	0.29	0.30	0.28	0.32	0.06	0.03	0.09
Observations	994	491	503	469	221	248	469	221	248	469	221	248

Notes: * significant at 10% level ** significant at 5% level *** significant at 1% level

(1) CHOICES=1 if adolescent was assigned to CHOICES only or CHOICES+EMB treatment groups; 0 if in control group.

(2) Standard errors are clustered at the session-team level and are reported under the coefficient in parentheses. All OLS regressions includes district fixed effects and controls for whether household has multiple adolescents, the age of respondent, and peer effects given by the number of CHOICES adolescents in the adolescent's team excluding the respondent (0-4). Regressions control for randomized team gender composition (mixed-gender, boys only, and girls only where girls only is the omitted category), and enumerator-gender fixed effects.

(3) Outcomes Stick to your guns, Confused in Public and Conform in Public are restricted to the adolescents in positions 3-5 who correctly matched the line when responding in private. 80% were correct in private. The leader and deputy sitting in positions 1 and 2 who were prompted to say A in public are excluded from the analysis.

(4) For the outcomes "Stick to your guns", "Confused in Public" and "Conform in Public" the position effects are estimated by considering if any of the adolescents in the positions before adolescent i answered correctly in public (i.e. they answered C).

Confederates Game

Table 7 presents the results from the line match confederates game. In this game there is only one correct answer (line C). However, each team's leader and deputy were instructed to act as confederates and state an incorrect answer (A) when playing the game with their team in public.⁷⁰ The leader and deputy were asked to sit in positions 1 and 2, respectively, with the remaining three team members seated in a random order.⁷¹

The results in column 1 suggests that 80% of the control group correctly matched the line for this game in private with no treatment differences. In the remaining columns in Table 5 we define 3 outcomes: *stick to your guns* if correct in public (C), *confused* if incorrect in public but was not the primed response (B or D); and *conform* if incorrect in public and responded with the primed response (A).

When answering in private only 20% incorrectly matched the line, whereas when in public as many as 57% incorrectly matched the line. In the following we exclude the responses by the leader and deputy who were primed to answer incorrectly and show results for adolescents in positions 3-5 only. In column 4 we show that 43% stick to the correct answer in the public game. While we find no *Choices* treatment difference, the coefficient on the variable *Any response in positions before correct answer* C is positive and significant. This suggests that the adolescents were more inclined to give the correct answer in public when any adolescent before them in the line-up gave a correct response. In column 7 we show that 32% of the control group were perhaps confused in public. They give an incorrect answer in public but it was not what the leader and deputy said. Similarly, in column 10 we show that 24% of the control group switch to conform and answer the primed response (A) in the public game. For both these outcomes we find no evidence of a *Choices* treatment effect. The coefficients on the variable *Any response in positions before correct answer* C for the *confused* and *conform*

 $^{^{70}}$ Note - the line match confederates game was played at the very end of the session so that the leader and deputy could be primed outside before playing the final game in public.

⁷¹Asch (1952) showed that conformity increases when other members of the group are of a higher social status so we may expect conformity by the adolescents to be higher after the priming the leader and deputy.

outcomes in columns 7-12 are negative and significant. This suggests that the adolescents (especially for girls in columns 9 and 12) were less likely to give an incorrect response in public when an adolescent before them in the line-up gave a correct response.

Overall, the results in Tables 6 and 7 provide evidence of conformity among adolescent peers as was found in the Asch experiments. However, *Choices* participation does not affect how adolescents play the line match games and therefore *Choices* does not appear to affect general levels of conformity. We also find no *Choices* treatment effect in the response to the survey statement: "There have been times when I felt like rebelling against people in authority even though I knew they were right" (not shown here). The *Choices* program did affect private gender attitudes and public reports of gender attitudes in ways that suggest that adolescents' perceptions of gender norms (or acceptable responses) may have shifted in interesting ways that could theoretically be linked directly to the program content.

7.3 Impacts on Leadership

In Table 8 we present the *Choices* impacts on the likelihood an adolescent was chosen for a leadership position in their team where the outcome is a dummy variable for whether the adolescent was selected as a leader.⁷² Columns 1 and 2 present the results for all teams combined.⁷³ While in column 1 the coefficient on *Choices* treatment is positive, only in column 2 that examines heterogeneous results by gender the result for *Choices* boys is statistically significant. That is, column 2 shows that the *Choices* treatment increases the likelihood that boys in the treatment group lead with no impact for girls. In columns 3 to 6 we split the sample into the mixed-gender teams (3-4) and single-gender teams (5-6). We find the *Choices* impact for boys is only statistically significant for treated boys in the all boys team.⁷⁴ In the mixed-gender teams we find no evidence of a treatment impact; however, overall girls are less likely to lead their team than boys (see coefficient on *Girl* is negative).

 $^{^{72}}$ We find no impact of *Choices* on being selected for the deputy position (not shown).

⁷³The regression analysis is first run for the unrestricted, full sample, then we restrict the analysis to mixed-gender teams only, and single-gender teams: the boys-only and girls-only teams.

⁷⁴Possibly, *Choices* boys are better able to negotiate that they should be made the leader.

Table 8: Impact on Leadership

			<u>Child is Lea</u>	<u>ader (Yes=1)</u>		
	All t	ams	Mixed-gen	der teams	Boys only	Girls only
			Mixed gen		team	team
	(1)	(2)	(3)	(4)	(5)	(6)
CHOICES	0.040	0.079**	0.054	0.061	0.103*	-0.042
	(0.03)	(0.04)	(0.04)	(0.06)	(0.06)	(0.06)
Girl		-0.064		-0.103*		
		(0.05)		(0.06)		
CHOICES # Girl		-0.076		-0.017		
		(0.05)		(0.07)		
Number of CHOICES peers in team (0-4)	-0.010	-0.010	-0.013	-0.014	-0.026*	0.011
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Control group mean	0.17	0.17	0.17	0.17	0.13	0.23
Adjusted R ²	-0.01	0.00	-0.01	0.01	-0.01	-0.03
Observations	1,000	1,000	495	495	250	255

Notes: * significant at 10% level ** significant at 5% level *** significant at 1% level

(1) CHOICES=1 if adolescent was assigned to CHOICES only or CHOICES+EMB treatment groups.

(2) Standard errors are clustered at the community session level and are reported under the coefficient in parentheses. All OLS regressions includes district fixed effects and controls for whether household has multiple adolescents, the age of respondent, and peer effects given by the number of CHOICES adolescents in the adolescent's team excluding the respondent themself (0-4). Regressions in Table 6 columns (1) and (2) also include a control for team composition (whether team is mixed-gender, all boys, all girls).

(3) Outcomes: Dummy variable for whether the adolescent was chosen to be leader by their team

(4) Table 6 Columns (2) and (4) present gender differences by including dummy variables for gender of the adolescent "girl" and an interaction term between treatment and gender (CHOICES # girl).

8 Conclusion

We evaluate a gender norms program in Somalia that tries to encourage the development of positive, mutually-respectful gender attitudes and behaviors to promote greater gender equality. The program is targeted to young adolescents aged 10-14, an age preceding many important schooling and work decisions. We find evidence that the program led to a shift towards more gender egalitarian attitudes for both girls and boys. The adolescents were also more likely to stand by these gender egalitarian views in public when they were surrounded by a group of peers. We also find changes in perceived gender norms for boys, but not for girls which suggests normative pressures from peers may be different for boys and girls. We find no evidence of a program effect on the desire to conform, proxied in our experimental setup through a line match game, that supports the notion that what we observe are social dynamics specific to gender attitudes/norms taught by the program; not conformity in general or outright rebellion. Our findings have potential broader implications for understanding various forms of social influence more generally. We encourage future research that further explores social dynamics with alternative reference groups.

Program impacts occurred in a relatively short timeframe and were maintained two years post-intervention. Attitudes towards equal gender rights to education and a woman's role in the economic sphere were more malleable in the Somali context, whereas attitudes around masculinity were difficult to shift. We show that the norms intervention can be effective when delivered through an after-school program for both in- and out-of-school adolescents at an early age. Government policy that embeds gender norms programming in school curricula or through school management committees can be a pathway to scale. While we find a complementary gender norms program to parents also resulted in more egalitarian attitudes among mothers and fathers, we find limited additional impact on adolescent attitudes or behaviors. Reassuringly, adolescents do not report experiencing any immediate backlash from their parents or the community. Programs that tackle restrictive gender norms may be better targeted at both adolescent girls and boys in tandem. By allowing space for critical reflection on gender equality, especially during an age where these views are being reinforced and internalized by both sexes, can help to address constraints on future economic outcomes and wellbeing. Adolescence is a period of great developmental change that leads to social and emotional maturity. In this phase, the program led to a decrease in the emotional and behavioral problems faced by both boys and girls.

Though we are unable to link these changes in attitudes and behavior to longer run education or labor market outcomes, the fact that we find increased support for gender egalitarian ideals in public, even when faced with opposing discriminatory views, is important. It suggests that the program can help to build the foundation on which adolescents will be able to negotiate important future life decisions related to education, employment, marriage, and fertility more effectively. For when they come up against societal expectations, they will need the wherewithal to stand up for what they personally believe in and resist societal pressures to conform.

References

- Aizer, Anna, "The Gender Wage Gap and Domestic Violence," American Economic Review, September 2010, 100 (4), 1847–59.
- Akerlof, G.A. and R.E. Kranton, "Economics and Identity," Quarterly Journal of Economics, 2000, 115 (3), 715–753.
- and _, "Identity and Schooling: Some Lessons for the Economics of Education," Journal of Economic Literature, 2002, 40 (3), 1167–1201.
- _ and _ , "Identity Economics," New Jersey: Princeton University Press., 2010.
- Akerlof, George A., "Social Distance and Social Decisions," *Econometrica*, 1997, 65 (5), 1005–1027.
- Alesina, A, P. Giuliano, and N. Nunn, "On the Origins of Gender Roles: Women and the Plough," *Quarterly Journal of Economics*, 2013, 128 (2), 469–530.
- Anderson, M.L., "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects.," *Journal of the American Statistical Association*, 2008, 103 (484), 1481–1495.
- Andrew, Alison, Sonya Krutikova, Gabriela Smarrelli, and Hemlata Verma, "Gender norms, violence and adolescent girls' trajectories: evidence from a field experiment in India.," *Economics Series Working Papers 984, University of Oxford, Department of Economics.*, 2022.
- Asch, Solomon, "Effects of group pressure on the modification and distortion of judgments. Groups, Leadership and Men: Research in Human Relations.," *Carnegie Press*, 1951, p. 177–190.
- Asch, Solomon E., "Social psychology.," Englewood Cliffs, NJ:Prentice Hall, 1952.
- Austen-Smith, David and Roland G. Fryer, "An Economic Analysis of "Acting White"," The Quarterly Journal of Economics, 2005, 120 (2), 551–583.
- Baird, Sarah, Zulfiqar A. Bhutta, Bassam Abu Hamad, Joan Hamory Hicks, Nicola Jones, and Jennifer Muz, "Do restrictive gender attitudes and norms influence physical and mental health during very young Adolescence? Evidence from Bangladesh and Ethiopia," SSM - Population Health, 2019, 9.
- Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen, "Inference on treatment effects after selection among high-dimensional controls," *The Review of Economic Studies*, 2014, *81* (2), 608–650.
- Benjamini, Yoav, Abba M. Krieger, and Daniel Yekutieli, "Adaptive linear step-up procedures that control the false discovery rate," *Biometrika*, 2006, *93* (3), 491–507.

- Bernhardt, Arielle, Erica Field, Rohini Pande, Natalia Rigol, Simone Schaner, and Charity Troyer-Moore, "Male Social Status and Women's Work," *AEA Papers* and Proceedings, May 2018, 108, 363–67.
- Bernheim, Douglas B., "A Theory of Conformity.," Journal of Political Economy, 1994, 102 (5), 841–877.
- Bertrand, Marianne., "Gender in the Twenty-First Century," AEA Papers and Proceedings, 2020, 110, 1–24.
- Bicchieri, Cristina, "Norms in the Wild: How to Diagnose, Measure, and Change Social Norms," Oxford University Press, 2016.
- Bursztyn, L., T. Fujiwara, and A. Pallais, "Acting Wife: Marriage Market Incentives and Labor Market Investments," *American Economic Review*, 2017, 107 (11), 3288–3319.
- Bursztyn, Leonardo, Alessandra L. Gonzalez, and David Yanagizawa-Drott, "Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia," *American Economic Review*, October 2020, 110 (10), 2997–3029.
- _ , Alexander W Cappelen, Bertil Tungodden, Alessandra Voena, and David H Yanagizawa-Drott, "How Are Gender Norms Perceived?," Working Paper 31049, National Bureau of Economic Research March 2023.
- and Robert Jensen, "How Does Peer Pressure Affect Educational Investments?," The Quarterly Journal of Economics, 2015, 130 (3), 1329–1367.
- _, Georgy Egorov, and Robert Jensen, "Cool to be Smart or Smart to be Cool? Understanding Peer Pressure in Education," The Review of Economic Studies, 2019, 86 (4), 1487–1526.
- Chetty, Raj and Nathaniel Hendren, "The Impacts of Neighborhoods on Intergenerational Mobility I: Childhood Exposure Effects.," *The Quarterly Journal of Economics*, 2018, 133 (3), 1107–1162.
- Cialdini, Robert B., Carl A. Kallgren, and Raymond R. Reno, "A Focus Theory of Normative Conduct: A Theoretical Refinement and Reevaluation of the Role of Norms in Human Behavior.," *Advances in Experimental Social Psychology*, 1991.
- Dahl, Gordon B, Cristina Felfe, Paul Frijters, and Helmut Rainer, "Caught between Cultures: Unintended Consequences of Improving Opportunity for Immigrant Girls.," National Bureau of Economic Research (NBER) Working Paper, 2020, 26674.
- Dhar, Diva, Tarun Jain, and Seema Jayachandran, "Intergenerational Transmission of Gender Attitudes: Evidence from India," *The Journal of Development Studies*, 2019, 55 (12), 2572–2592.
- _ , _ , and _ , "Reshaping Adolescents' Gender Attitudes: Evidence from a School-Based Experiment in India," *American Economic Review*, March 2022, 112 (3), 899–927.

- **Duflo, Esther**, "Women Empowerment and Economic Development," Journal of Economic Literature, 2012, 50 (4), 1051–1079.
- Eagly, A. H. and L. L. Carli, "Through the labyrinth: The truth about how women become leaders.," *Harvard Business School Press.*, 2007.
- Elster, Jon, "Social Norms and Economic Theory," Journal of Economic Perspectives, December 1989, 3 (4), 99–117.
- Field, Erica, Rohini Pande, Natalia Rigol, Simone Schaner, and Charity Troyer Moore, "On Her Own Account: How Strengthening Women's Financial Control Impacts Labor Supply and Gender Norms," *American Economic Review*, 2021, 111 (7), 2342–75.
- Goodman, Robert, "The Strengths and Difficulties Questionnaire: A Research Note," Journal of Child Psychology and Psychiatry, 1997, 38, 581–586.
- Heinicke, Franziska, Christian König-Kersting, and Robert Schmidt, "Injunctive vs. descriptive social norms and reference group dependence.," *Journal of Economic Behavior Organization*, 2022.
- ILO, "Female labor force participation statistic.," ILOSTAT database., 2019.
- Jayachandran, Seema, "Social Norms as a Barrier to Women's Employment in Developing Countries," *IMF Econ Rev*, 2021, 69 (5), 576–595.
- Jones, Stephen R.G., "The Economics of Conformism.," Oxford: Basil Blackwell, 1984.
- Kling, J.R., J.B. Liebman, and L.F. Katz, "Experimental Analysis of Neighborhood Effects.," *Econometrica*, 2007, 75 (1), 83–119.
- Lundgren, Rebecka, Miranda Beckman, Surendra Prasad Chaurasiya, Bhawna Subhedi, and Brad Kerner, "Whose turn to do the dishes? Transforming gender attitudes and behaviours among very young adolescents in Nepal," Gender Development, 2013, 21 (1), 127–145.
- _, Susannah Gibbs, and Brad Kerner, "Does it take a village? Fostering gender equity among early adolescents in Nepal," International Journal of Adolescent Medicine and Health, 2020, 32 (4).
- McKenzie, David, "Beyond baseline and follow-up: The case for more T in experiments.," Journal of Development Economics, 2012, 99 (2), 210–221.
- Perez-Truglia, R., "Political Conformity: Event-Study Evidence from the United States," *Review of Economics and Statistics*, 2018, 100 (1), 14–28.
- Reitan, H. T. and M. E Shaw, "Group membership, sex-composition of the group, and conformity behavior.," *The Journal of Social Psychology*, 1964.
- Reynolds, W. M., "Development of reliable and valid short forms of the Marlowe-Crowne Social Desirability Scale.," *Journal of Clinical Psychology*, 1982, 38 (1), 119–125.

- Schwarzer, R. and M. Jerusalem, "Generalized Self-Efficacy scale," In J. Weinman, S. Wright, and M. Johnston, Measures in health psychology: A user's portfolio., 1995, pp. 35–37.
- Snyder, C. R., B. Hoza, W. E. Pelham, M. Rapoff, L. Ware, M. Danovsky, and K. J. Stahl, "The development and validation of the Children's Hope Scale," *Journal of Pediatric Psychology*, 1997, 22 (3), 399–421.
- Sunstein, Cass R., "Conformity: The Power of Social Influences.," New York: New York University Press., 2019.
- Tankard, M. E. and E. L Paluck, "Norm perception as a vehicle for social change," *Social Issues and Policy Review*, 2016, 10 (1), 181–211.
- UNDP, "Human Development Report.," New York: Palgrave Macmillan., 2012.
- Vaillant, J, E Koussoubé, D Roth, R Pierotti, M Hossain, and KL Falb, "Engaging men to transform inequitable gender attitudes and prevent intimate partner violence: a cluster RCT in Democratic Republic of Congo," BMJ Glob Health, 2020, 5 (5).
- Young, H. Peyton, "The Evolution of Social Norms," Annual Review of Economics, 2015, 7 (1), 359–387.

Appendix

Details on lab-in-the-field experimental games

There were a total of 50 sessions held. For each session, 20 adolescents from the *Choices* impact evaluation sample were invited to a predetermined site (community center or local school). The session had 10 boys and 10 girls from a mix of *Choices* treatment and control groups.⁷⁵ As soon as participants entered the session room, enumerators checked their name, gender and age in order to validate that boys and girls in the session were study participants from a list of potential participants. There were 4 enumerators per session.

Team determination round

Adolescents were first randomly assigned into one of 4 teams (red, yellow, blue, green). There were 2 sacks named B and G. Sack B contained 5 red balls, 3 green balls and 2 blue balls. Sack G contained 5 yellow balls, 2 green balls and 3 blue balls. Boys picked a ball from sack B and girls picked a ball from sack G. The color of the ball the adolescent picked is the team they were in for the entire session. Within each team, adolescents were asked to discuss among their team of 5, and decide who should be made the role of leader and deputy from their team before proceeding with a private interview.

Registration and Private Responses

Each enumerator was assigned one team and started by conducting a private interview with each of the 5 adolescents. In private, adolescents were asked to choose if they agree or disagree with 3 statements aimed at capturing their gender attitudes, and to play 2 line matches. The two line match games asked the adolescent in private to correctly identify the correct length of a line drawn on paper. The private interview was done with an enumerator asking questions that are typical of self-reported survey questions. The gender attitudes statements include:

- Statement 1: It is more important for a girl to help at home than spend time studying.
- Statement 2: It is okay for a man to hit his wife if she disagrees with him.
- Statement 3: Boys should have more free time than girls.

Gender Attitudes Statement game - Public Responses

The 5 adolescents in each team sat in a row of chairs in random order generated automatically on a tablet. They were then asked whether they agree or disagree with each gender statement read out by the enumerator. This time they were asked to say their response out loud in

 $^{^{75}}$ We decided to mix treatment and control adolescents in a session to address potential criticism that *Choices* adolescents could know each other better after attending a training together and thereby exhibit peer effects.

public in front of their group. They were asked to change seating position after each gender statement game (3 games in total) and sit in a new order generated by the tablet for each round. The adolescents were instructed to answer quickly without discussing answers with other team members. The same gender statements that were asked in private were again asked in public. In our experiment, we examine conformity as a behavioral outcome that measures an adolescent's likelihood of changing their privately-held personal attitudes when grouped with peers of the same or opposite sex and when stating their attitudes in public. In our analysis, we also consider the effect of the person's response in the position before (position effects) and potential peer effects that may come from teammates who had attended the gender norms program (Choices peer effects). Figure 1 outlines the random assignments used to examine these effects.



Figure 1: Effects measured in the lab-in-the-field experiment

Line match game 1

Again the 5 adolescents in each team sat in a row of chairs in random order to play the line match 1 game. In the random order, adolescents were asked one after the other: "Which line matches the line on the left? Choose one from A, B, C or D?"

Confederates line match game

• The leader and deputy from all teams were asked to leave the room, where they were instructed to give answer A (an incorrect answer) to a new line match game that they

would play with their teams at the end of the session.

- The leader and deputy returned to their teams and were seated in position 1 and 2 in the row of seats with the three other team members seated in a random order following them. The order of seating for the three members will be automatically randomized on the tablet.
- In this order, adolescents were asked one after the other: "Which line matches the line on the left? Choose one from A, B, C or D?" Image of the lines is in Figure 3.

Images of the lines used for line match 1 game is in Figure 2 and for the confederates game is in Figure 3.



Figure 2: Lines used for line match game 1



Figure 3: Lines used for confederates line game

Baseline Attitudes and Behaviors

In Figure 4 we present the percentage of adolescent girls, boys, mothers and fathers who personally agree with a number of gender attitudes statements reported at baseline.⁷⁶ For example, from the baseline data, we find over 50% of fathers, mothers, girls and boys personally think it is more important for a girl to help at home than to spend time studying; and over 50% think boys should have more free time than girls. For the statement related to intimate partner violence (IPV) we find approximately 30% would personally agree that it is okay for a man to hit his wife if she disagrees with him. It is interesting to note that even as young as 10-14 years, adolescents hold a similar aggregate level of gender discriminatory beliefs as their parents.



Gender Discriminatory Views at Baseline

Note: Personal attitudes of adolescent girls and boys and their parents at baseline. Includes the 14 statements that make up full gender attitudes index.

Figure 4: Gender Discriminatory Attitudes at Baseline

⁷⁶The attitudes are those 14 included in the full index.

In Figure 5 we indicate the percentage of respondents who reported they were engaged in any of the listed activities in the past 7 days at baseline: cooking, washing clothes, fetching firewood, cleaning, washing dishes, buying food items, tending family livestock, and farming activities. In general, across all listed activities, girls are more likely than boys to have conducted chore activities at baseline. Even in young adolescence, girls are involved in domestic activities at a rate similar to their mothers. For example, at baseline 73% of girls reported they were involved in cooking in the last week, compared with 80% of mothers, 31% of boys, and 33% of fathers. Figure 6 suggests that the aggregate mean for adolescent boys is similar to fathers and the mean for adolescent girls is similar to mothers. In terms of time use, we show that conditional on spending time on an activity - girls spend more time on care and domestic tasks than boys, and less time on leisure and studying out of school.





Figure 5: Gender Differences in Unpaid Work and Time Use at Baseline

Gender Statements Game Directional Analysis

In Somalia, social norms are viewed as gender discriminatory (see Figure 4 in previous section for views of adolescents and their parents at baseline). However, the actual beliefs among adolescents in the lab setting should be considered overall more gender progressive than alternative reference groups in Somali society. For example, perhaps if the reference group were parents or other adults in their communities we may expect the social cost of non-conformity to be higher than a reference group of peers. Alternatively, behavior among a group of friends might differ to a random group of peers as was used in the lab experiment. Nevertheless, it is still informative to examine the direction of any switching that occurs in the public lab setting to understand the dynamics of norms and conformity. In the following analysis we further scrutinize the treatment effects on social conformity by unpacking the direction of the effects i.e. were any conforming adolescents more likely to switch to a discriminatory or egalitarian view in public? We also examine how adolescents are influenced by the responses of those in positions before them in Table 5.



Figure 6: Predicted Probabilities Multinomial Logistic Regression on Conformity

In the multinomial logistic model we show the impact of *Choices* by gender and use the Always Egalitarian category as the base case to estimate the predicted probabilities in each outcome category reported as average partial effects. Figure 6 shows the results from a multinomial logistic regression with an outcome that has 4 distinct categories" Always Egalitarian (private egal – public egal), Always Discriminatory (private disc – public disc), Switch Egal-

itarian in Public (private disc – public egal), and Switch Discriminatory in Public (private egal – public disc). The notation "egal" means egalitarian and "disc" mean discriminatory. *Choices* boys and girls have a higher probability of being in the Always Egalitarian (private egal – public egal) category compared to adolescents in the control group (i.e. the solid grey bar in Figure 5 is higher for *Choices* adolescents than the control group). We show, ceteris paribus, 73% of *Choices* adolescent girls are private egal – public egal, compared to 66% of control girls i.e. a 6.9 percentage point difference as shown in Table 7C.

In Figure 6 we also demonstrate that the control group adolescents are more likely to switch to an egalitarian view in public (private disc – public egal). Since we know that the *Choices* program leads to *Choices* adolescents stating greater egalitarian views both in private and public, then it is perhaps unsurprising that this in turn influences control group adolescents to switch to report a more egalitarian view in public. The conformity seen among control group adolescents is in the direction of the dominant attitudes within the lab.⁷⁷ The hope is that the dynamics shown here among adolescents will play out among other reference groups confronted in society such that societal norms can eventually be shifted to be more gender egalitarian. This can occur if: 1) the *Choices* adolescents who have greater gender egalitarian personal attitudes do not conform and go against societal norms by sticking to their beliefs; and 2) adolescents who continue to hold gender discriminatory personal attitudes begin to conform to a new progressive norm that is shaped by the greater egalitarian beliefs of their peer group.

⁷⁷Note the response in the lab setting was more egalitarian than the endline survey that was conducted at a similar time. Among control group respondents the private response during the games for the right to free time was 40% discriminatory (52% at endline), for IPV it was 12% discriminatory (21% at endline), and for the importance of home work versus education it was 24% discriminatory (48% at endline). In the lab set-up respondents were already put into teams so there was some group influence (see Table 3) plus being around other adolescent peers in general might have affected their response.

Modules of *Choices* Intervention

The *Choices* training curriculum has been adapted for Somalia and while the adaption of the curriculum is still ongoing and therefore subject to change, the twelve module objectives are:

- 1. Raise awareness about society's perspectives and expectations of boys and girls and resulting lifestyle differences.
- 2. Raise awareness of time use differences between boys and girls and of broader societal expectations regarding gender roles and the challenges they pose.
- 3. Enable children to explore their thoughts about their own future life options and about future life options for the opposite sex.
- 4. Teach children the importance of having hope and encouraging hope in others.
- 5. Understanding how mutual respect and sharing are essential foundations for a home where people feel cared for, loved, and are productive members.
- 6. Make children aware of each other's hopes and dreams and empower children to be mutually encouraging in the pursuit of their hopes and dreams. Understand how selfesteem affects the pursuit of hopes and dreams.
- 7. Raise awareness of societal expectations that create conformity and enable gender inequality. Teach children how to take a stand against situations fostering gender inequality.
- 8. Identify characteristics of respectable, model males and females. Understand that respect is earned and not given.
- 9. Instilling the importance of actively showing caring actions to others as a means of encouragement and companionship.
- 10. Encouraging kids to think about their life options beyond what is traditionally accepted. Helping kids to think through what steps they would take to make some life changes.
- 11. Clarify misconceptions about social expectations and individual abilities and practice talking through difficult issues and situations.
- 12. Validate participants' new commitment to each other and support for specific behavior.

Modules of EMB training intervention

The EMB curriculum involves a series of nine community dialogues sessions, including one final session to agree on actions. The nine module objectives are:

- 1. Gender Norms: To identity the differences between rules of behavior for men and for women and to understand how profoundly these gender rules affect their lives.
- 2. Fatherhood: To reflect upon the influence that fathers or other male authority figures have had on the participants while they were growing up.
- 3. Power: To explore how power can be used in positive and negative ways, to learn how to use power positively, and to reflect on how power is divided between men and women. The session intends to create awareness about gender and social expectations.
- 4. Household work: To become aware of the different tasks and roles men and women do/share in the household and discuss if and how those roles could be exchanged.
- 5. Violence 1: To discuss how violence affects our everyday lives and how to stop the cycle of violence in our lives and our communities.
- 6. Violence 2: To better understand the many ways in which women's (and men's) lives are limited by male violence and/or the threat of men's violence, especially sexual violence.
- 7. Child, early and forced marriage: To explore and discuss the community belief towards early marriage and its relationship with girls' education and future opportunities in life.
- 8. Female genital cutting: To reflect on the practice of FGC, what are the consequences of this practice for girls and women and why it still exists.
- 9. Action planning: To mobilize the community into action against harmful social norms that affect them. To support the group to develop an action plan that will be re-visited after 6 months.

Statistical power calculations

We use power analysis to identify the minimum sample size necessary to detect meaningful effects of the gender norms interventions, and to maximize the chance of finding an effect (conversely minimizing the chance of not finding it) when it actually exists. Based on a sample of 1,400 children in each arm, and assuming 15% attrition, the minimal detectable effect (MDE) is 0.133 standard deviations (SD) at 90% power and 0.115 SD at 80% for comparing any two groups. Estimating proportions, as a number of attitude- and behavior-level outcomes are likely to be binary variables, this sample size is able to capture 6.6 percentage point (pp) changes from a 50% level at 90% power.⁷⁸ These power calculations are for intention to treat (ITT) effects by assuming 100% take-up. However, we anticipated lower take-up rates, and at 80% take-up achieving the 6.6pp difference in ITT would require a real change of 8.3pp.⁷⁹ However, the study also examines the impact by gender of the adolescent. If we take half of the sample for this sub-group analysis, the MDEs are 0.188 SD or 9.4 pp at 90% power. The suggested study sample size is designed to be conservative and allows for some reduction in take-up without affecting power too much.

Examining Possible Spillover Effects

One of the concerns for an individual level randomization design is the possibility of spillover effects. In this study we use household level randomization to ensure that adolescents within a household (i.e. siblings or other relatives) were not assigned to different treatments within a single household to minimize possible spillover or contamination effects. It was not possible to design a cluster-level or community-level randomization given the limited sample size in terms of the number of communities targeted for implementation of the *Choices* program. If spillovers occur between treated and control households then our results are likely biased downward and should be considered a lower bound estimate of the true impact of the program. In order to examine the extent of spillovers in Table A8 in the Appendix we show that, among a panel of adolescents surveyed in all 3 survey rounds, the average gender attitudes among the control group are fairly stable over time from baseline, midline to endline. This suggests there is limited evidence of spillovers to the control group as adolescents hold similar views on aggregate for most measured gender attitudes over time. An alternative approach to assess the spillover effects is by studying individual control group exposure to *Choices* using the natural variation in the number of adolescents in the community treated to analyze the correlation between the number of treated and the change in control group attitudes over time (from baseline to midline and baseline to endline). In the 33 study community clusters, the number of adolescents treated ranges from 47 to 97. We find that the number of treated adolescents in the community is *not* correlated with a change in the overall gender attitudes index of the control group over time (results available on request).⁸⁰

 $^{^{78}}$ Power calculations estimated in Stata using the following command - power two proportions .5, test(lrchi2) power(0.9) nratio(1) n(2380).

⁷⁹These estimates do not account for multiple survey rounds where MDE will be lower.

⁸⁰Note, social network data directly linking control group children to networks of treatment group children was difficult to collect and match; and we were unable to obtain school networks data to examine spillovers.

Appendix tables

Impact on Parents' Gender Attitudes

Table A1 presents the impacts on gender equitable attitudes of parents that replicate the outcomes shown in Table 2 for the adolescent sample. Since parents in T2 (*Choices* + EMB) were assigned to receive the EMB training directly, in Table A1 the coefficient on EMB should be read as a direct effect. Here we show responses from the father and mother separately in columns 2 and 3 for the full index. At midline, we find evidence of an EMB treatment impact on overall gender equitable attitudes of parents (both mother and father have approximately 0.10 standard deviation higher attitudes full index than those in the control group at midline (Panel B). In the longer run, however, at endline, we find the EMB effect on the full index dissipates. However, fathers continue to hold greater gender egalitarian attitudes in specific domains i.e. toward the role of women (including a woman's right to work) and education attitudes (both mother and father).

It is interesting to observe that mothers with adolescents who were assigned to T1 (*Choices* Only) appear to also report of greater gender equitable attitudes at midline. The *Choices* coefficient is positive and statistically significant suggesting that the *Choices* has positive spillovers on the attitudes of mothers (and fathers in their attitudes toward the role of women). However, by endline, any spillover effects from the *Choices* treatment dissipate with only attitudes towards education held by mother significantly different to the control group (at 90% confidence).

Impact on Perceived Community Norms

In Table A2 we show the responses of adolescents (columns 1-4) and their parents (columns 5-12) on the perceptions of community norms. Perceived community norms are estimated by asking "out of 10 of your neighbors how many do you think agree..." and we rescaled their response to be a proportion out of 100%. Among adolescents we ask perceptions on community agreement with women's work outside the home (columns 1, 2, 5, and 6) and child marriage (columns 3, 4, 7, and 8). Among parents we ask the same but also measure perceptions of community acceptance towards intimate partner violence (IPV) and girl circumcision (FGC). On average, at endline, fathers perceive that 42% of their community would speak badly of a woman who works out of the home (mothers 37%, boys 36% and girls 37%). On child marriage, mothers and fathers perceive that 40% of their community think it is important for a girl to be married before she is 18 years old (boys and girls perceive this norm to be closer to 35%). In terms of IPV and FGC, as mothers and fathers give similar responses: they perceive that 55% of their community believes a woman should tolerate violence to keep her family together; and 75% think it is important for girls to be circumcised.

Overall, we find limited treatment impacts of the EMB program or *Choices* program on the perception of community norms by parents. That is, parents do not update their perceptions about what their community thinks after attending a training themselves. However, among adolescent boys and girls there is some updating of their perceived norms. For example, in column 2 we show that at endline *Choices* girls perceive more support for woman's work out of the home (a 4% reduction in the perceived community norm on woman's work); and *Choices* boys perceive less acceptance of child marriage by their communities (a 3% reduction in the perceived community norm on child marriage). These results suggest that the *Choices* program did not misinform participants about the level of support for gender equality in the community. The perceptions about what other people in the community thinks about woman's work outside the home and child marriage appears to be fairly consistent across the genders and generations.

Impact on Adolescents' Relationship with Parents

In Appendix Table A3 we also show the *Choices* treatment has a positive impact on the relationship of adolescents with their mother at midline which dissipates by endline. This results supports that there were no obvious backlash effects towards adolescents from attending the program, at least by their parents.

Impacts on Household Chores and Care

The *Choices* curriculum includes modules that look to raise awareness of time use differences between boys and girls and of broader societal expectations regarding gender roles. It also attempts to instill the importance of actively showing caring actions to others. Table 6 therefore explores any *Choices* treatment impact on self-reported behaviors for boys and girls separately.⁸¹

The outcome in Table A4 columns 1 and 2 is whether the adolescent did any household chores activities in the past week whereas in columns 3 and 4 is the number of hours spent on domestic and care activities in the past week.⁸² Results suggest no evidence of an impact on the likelihood that girls do less household chores. However, we find *Choices* boys increase the likelihood they do any household chores in the past week. In the short-run, at midline, this result is significant for the boys who also had parents assigned to *EMB* - that is, the marginal impact of the *EMB* treatment is positive and significant at 95% confidence. This result might suggest that adolescents do not have much say in the chores allocation within their home and parents might be a constraint. In the longer run, at endline, the result is less robust where *Choices* boys in the control group. It is reasuring that these effects go in the direction we would hypothesize for boys. However, since *Choices* girls do not do less at the extensive or intensive margin this is a puzzling result if *Choices* is expected to lead to a redistribution of the tasks between boys and girls in the home.

In columns 5 and 6 we also present a standardized index of caring behaviors towards a sibling of the opposite sex. Results suggest that *Choices* boys and girls report more caring

⁸¹Here we may expect effects to go in opposite directions for boys and girls, so we do not not show the combined impact. That is, for boys to do more household chores and girls to do fewer household chores as a result of the program.

⁸²The activities included here are cooking, washing clothes, cleaning the house and washing dishes which are domestic tasks; and can also be considered traditional female tasks since at least 2 out of 3 of the girls in the sample performed these activities at baseline. An index measure of the number of chores was also analyzed where we find no treatment differences.

behavior toward siblings of the opposite sex.⁸³ Among boys, the *Choices* treatment leads to a 0.21 standard deviation increase in caring behaviors towards sisters index at midline, and a 0.12 s.d. increase at endline, relative to the control group. That is, boys were more likely to have advocated with their parents for their sisters' wellbeing and/or demonstrated in some way that they care for their sister in the past month. Among girls the *Choices* treatment leads to a 0.20 s.d. increase in caring behaviors towards brothers at midline, and a 0.11 s.d. increase at endline, relative to control group girls (the magnitude of the results for girls are the same as boys but the results are less precise).

Impacts on Adolescent Mental Health and Hope

The *Choices* training curriculum tried to teach children the importance of having hope and encouraging hope in others, and make children understand how self-esteem affects the pursuit of hopes and dreams.

Table A5 presents treatment impacts on measures of mental health and hope. Columns 1 to 3 uses the Strengths and Difficulties Questionnaire (SDQ) which is a behavioral screening questionnaire validated for adolescents that consists of 25 statements: 10 attributes as strengths, 14 as difficulties, and 1 neutral that was developed by Goodman (1997). The statements are used to construct a total number of difficulties index measure that can be decomposed into 4 sub-scales: emotional problems, conduct problems, hyperactivity/inattention, and peer relationship problems.⁸⁴ Additionally, in columns 4 to 6 the outcome is a sub-scale from the SDQ that captures pro-social behavior which are behaviors that are intended to help others. The *Choices* treatment leads to a decrease in the emotional and behavioral problems faced by boys who report that they experience a 1.061 standard deviation decrease in the total difficulties index at midline, and a 1.338 standard deviation decrease at endline, relative to boys in the control group. The direction of the treatment effects are similar for girls; however, these effects are marginally insignificant (column 3, sharpened q-value 0.12). A decomposition of the total index shows that *Choices* treatment effect for boys is significant across all sub-scales at endline: boys face fewer emotional symptoms, conduct problems, problems with inattention, and problems with peers; while girls face fewer emotional symptoms.

Finally, the outcome in columns 7 to 9 is the Children's Hope Scale (CHS) which is a six-item scale developed by Snyder et al. (1997) that measures a child's perception that their goals can be met. The CHS total score is presented in Table 7 and the scale can be decomposed into Agency Thinking and Pathway Thinking sub-scales.⁸⁵ We find evidence of a positive treatment impact on the adolescent's perception that their goals can be met in the short-run. However, impacts are not sustained in the longer-run.

In addition to the SDQ and CHS, at endline we also measured the generalized selfefficacy (GSE) index scale as developed by Schwarzer and Jerusalem (1995) (not shown).⁸⁶

 $^{^{83}}$ Note 99% of the sample have siblings and 90% have siblings of the opposite sex at baseline.

 $^{^{84}}$ The sub-scales are not shown in Table A5 but are available on request.

⁸⁵Agency Thinking captures the child's perception on their ability to take action towards a goal, and Pathway Thinking is the child's perceived capacity to find a way towards that goal. Note the sub-scales are not shown in Table A5.

⁸⁶Self-efficacy has been found to be negatively correlated with anxiety, stress and depression (Schwarzer

The average GSE index score (ranging from 1-4) at endline is 2.83 and we find no significant difference in the average GSE scores of adolescent boys and girls in the sample. We find no evidence of a treatment impact on the self-efficacy outcome.

It is noteworthy that the improvements in mental health outcomes as a result of the *Choices* program are experienced by both boys and girls. This is consistent with increasing evidence that gender inequality is undesirable for everyone (Baird et al., 2019).

Impacts on Education Outcomes

In Table A6 we examine outcomes related to education: whether the child is currently enrolled in school in columns 1 to 3 and hours spent in school or studying in the past week, conditional on enrollment (columns 4 to 6). While the *Choices* program was not explicitly designed to keep adolescents in school we examine education outcomes since this is one of the major activities that the adolescent were involved in. The program targeted adolescents both in and out of school and at baseline 80% of the adolescents were enrolled in school (83% of boys and 78% of girls). The *Choices* treatment had limited impacts on school enrollment rates. The control group mean at midline for school enrollment was 83% and at endline 86%. Similarly, hours spent in school or studying remains unchanged by the program.

Adolescents also report on their aspirations for the future across a number of domains, including aspirations for professional work, education, marriage, and fertility. At baseline most of the aspirations were fairly high e.g., 76% of boys and 75% of girls aspire to study up to university at baseline. By the endline survey, 90% of the control group aspired to study up to university. However, even with such high aspirations, we show there is a small impact of the *Choices* program on girls' aspirations to attend university.

Note, marital status and fertility outcomes were also examined at endline. However, at the time of the endline survey, a large part of the sample were still too young for these outcomes to be relevant. At endline we find less than 1% of the sample who are married and 0.5% of the sample have children with no estimated treatment difference.

Accounting for Experimenter Demand Effects

In Table A7 we use the Marlowe-Crowne social desirability scale as proposed in Dhar et al. (2022) to measure an individual's propensity to give socially desirable answers (Crowne and Marlowe 1960). We included the module in the follow-up survey rounds (i.e. at midline and endline) to explore whether effects are biased by responses to an enumerator rather than their actual beliefs and behaviors. We find that respondents with a higher score on the social desirability index report greater support for gender equality overall, but we find no difference between the treatment and control groups. Therefore, the positive treatment effects on selfreported attitudes and behavior are similar in magnitude for respondents with a low versus high propensity for social desirability bias. We view this analysis as an important robustness check on the validity of our results.

and Jerusalem, 1995).

Gender Equitable Attitudes Index Breakdown

In Table A8 we present the individual statements in the attitudes index for adolescents and show the mean differences between treatment and control groups for the panel surveyed across three survey rounds.

Survey Attrition Rates

In Table A9 we show that survey attrition rates are balanced across treatment and control groups.

Robustness Check - Selecting Controls Using LASSO

As a robustness check, we estimate the main treatment impacts using the post-doubleselection (PDS) methodology by (Belloni et al., 2014) which uses the LASSO (Least Absolute Shrinkage and Selection Operator) estimator to identify a set of control variables from a long list of baseline covariates. We show in Table A10 that our results are not sensitive to the method of selection of controls.

Conformity Position Effects - Modal Response

In Table A11 we rerun the analysis for the outcome from Table 6 to examine the effects of the mode of the public responses in the preceding position(s) for boys and girls. In the regression $ModalPosition_i$ is a dummy variable equal to 1 when the mode (50 percent or more) of the public responses in the preceding positions in the row are different to adolescent *i*'s private response. We show results are robust when using the modal response - *Choices* adolescents are more likely to stick to their privately held response in public, even when the majority of peers responding before i has an opposing view.

Comparing Endline Survey and Games Responses

In Table A12 we show there is a positive treatment effect at midline and endline on a standardized index that is composed of the three gender statements that were also included in the lab-in-the-field experiment. We run analysis in Table A12 to compare effect sizes for endline survey data to the lab-in-the-field games data. The private responses in the lab context were overall more gender egalitarian than the average responses captured during the endline survey and treatment effect sizes are higher in magnitude. This may reflect a certain comfort that the adolescents had within the lab session held around peers, peer pressure, or for other reasons. Importantly, since the private and public responses were measured within the same lab session, results should be considered internally valid.

		Attitudes:		Att	itudes:	Attit	udes:	Attit	udes:
	Gende	r Equitable	e Index	Edu	ucation	Wome	n's Role	Masculin	ity Norms
		Full-Index		Sul	o-Index	Sub-	Index	Sub-	Index
	All	Father	Mother	Father	Mother	Father	Mother	Father	Mother
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Panel A: Endline									
CHOICES	0.020	-0.005	0.027	-0.058	0.089*	0.036	0.040	0.015	0.023
	(0.03)	(0.04)	(0.03)	(0.08)	(0.05)	(0.08)	(0.04)	(0.05)	(0.03)
	[0.91]	[1.00]	[0.63]	[0.86]	[0.09]	[1.00]	[0.42]	[1.00]	[0.95]
EMB (direct effect)	0.039	0.055	0.033	0.132*	0.101*	0.124**	0.041	0.001	0.002
· · · ·	(0.03)	(0.04)	(0.03)	(0.07)	(0.05)	(0.05)	(0.04)	(0.04)	(0.03)
	[0.19]	[0.17]	[0.33]	[0.08]	[0.08]	[0.03]	[0.36]	[1.00]	[1.00]
Control group mean	-0.00	-0.01	0.00	0.03	-0.02	-0.06	0.02	-0.02	0.01
Adjusted R ²	0.12	0.19	0.11	0.20	0.11	0.14	0.08	0.06	0.04
Observations	1733	381	1352	381	1352	381	1352	381	1352
ANCOVA	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Midline									
CHOICES	0.065**	0.069	0.065**	0.060	0.110	0.126**	0.106**	0.016	0.017
	(0.03)	(0.05)	(0.03)	(0.07)	(0.07)	(0.06)	(0.04)	(0.05)	(0.03)
	[0.03]	[0.25]	[0.05]	[0.69]	[0.12]	[0.05]	[0.02]	[1.00]	[1.00]
EMB (direct effect)	0.097***	0.081*	0.100***	0.057	0.158**	0.118	0.142***	0.084	0.044
· · · · ·	(0.03)	(0.05)	(0.03)	(0.10)	(0.06)	(0.07)	(0.04)	(0.05)	(0.03)
	0.00]	0.11	0.00	[1.00]	[0.02]	[0.12]	0.00]	[0.12]	0.18
Control group mean	-0.00	0.02	-0.01	0.09	-0.02	0.01	-0.01	0.00	-0.01
Adjusted R ²	0.06	0.03	0.06	0.05	0.13	0.02	0.05	0.02	0.04
Observations	1695	327	1368	327	1368	327	1368	327	1368
ANCOVA	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table A1: Impact on Parents' Gender Equitable Attitudes

Notes: * significant at 10% level ** significant at 5% level *** significant at 1% level

(1) CHOICES=1 if adolescent was assigned to CHOICES only group where parents were not invited to a gender norms training. EMB=1 if adolescent was assigned to the CHOICES+EMB group and parent was invited to the EMB gender norms training (EMB measures the direct effect).

(2) ANCOVA estimation controls for the level of the outcome variable. All regressions include district fixed effects and a set of controls measured tbaseline: dummy for whether household has multiple adolescents and age of respondent.

(3) Standard errors are clustered at the community level and are reported under the coefficient in parentheses. Sharpened q-values that correct p-values for the false discovery rate (FDR) are in square brackets. A sharpened q-value of 1.00 indicates the null hypothesis is not rejected at any level of FDR.

(4) Outcomes: Gender Equitable Index is a standardized index of 14 gender-progressive statements the respondent agrees with (index standardized as per method from Kling et al. 2007). Higher scores indicate more gender-equitable attitudes. The education sub-index, women's role sub-index, and masculinity norms sub-index are constructed from (mutually exclusive) statements in the Gender Equitable Index.

		Adole	scents		Parents							
	Speak b woman w outside th (Women	adly of a rho works ne home? 's Work)	Think it is important for a girl to be married before she is 18years old? (Child Marriage)		Speak badly of a woman who works outside of the home? (Women's Work)		Think it is important for a girl to be married before she is 18years old? (Child Marriage)		Believe a woman should tolerate violence to keep her family together? (IPV)		Think it is important for girls to be circumcised? (FGM)	
	Boy	Girl	Boy	Girl	Father	Mother	Father	Mother	Father	Mother	Father	Mother
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Panel A: Endline												
CHOICES	-0.126	-3.974**	-2.865**	-1.435	-5.243	-0.149	0.748	0.112	-2.091	-1.213	-0.568	-1.918
	-1.41	-1.57	-1.31	-1.53	(3.36)	(1.67)	(1.78)	(1.49)	(2.91)	(1.51)	(2.45)	(1.29)
	[1.00]	[0.02]	[0.04]	[0.55]	[0.15]	[1.00]	[1.00]	[1.00]	[0.91]	[0.75]	[1.00]	[0.17]
EMB (marginal effect)	-0.583	1.314	3.188*	1.19	1.972	-1.350	-0.924	-1.132	4.375	1.296	0.459	0.224
	-1.48	-1.33	-1.85	-1.48	(2.73)	(1.47)	(2.13)	(1.67)	(3.46)	(1.87)	(2.54)	(1.21)
	[1.00]	[0.49]	[0.10]	[0.75]	[0.91]	[0.58]	[1.00]	[1.00]	[0.27]	[0.97]	[1.00]	[1.00]
Control group mean	35.68	37.44	34.76	34.56	42.14	37.45	40.79	39.89	54.93	53.23	76.19	74.53
Adjusted R ²	0.04	0.06	0.16	0.15	0.07	0.04	0.23	0.14	0.06	0.14	0.21	0.19
Observations	1373	1431	1219	1279	422	1422	424	1446	416	1405	426	1446
ANCOVA	No	No	No	No	No	No	No	No	No	No	No	No
Panel B: Midline												
CHOICES	-4.394**	-0.028	0.262	-1.894	0.960	-1.048	-0.676	-1.264				
	(1.75)	(2.47)	(2.05)	(2.00)	(2.73)	(2.13)	(2.11)	(1.50)				
	[0.02]	[1.00]	[1.00]	[0.54]	[1.00]	[1.00]	[1.00]	[0.68]				
EMB (marginal effect)	-0.668	1.282	-2.894*	-0.539	-2.564	-1.500	-4.199*	-0.285				
	(1.70)	(2.32)	(1.55)	(1.99)	(2.57)	(1.89)	(2.09)	(1.24)				
	[1.00]	[1.00]	[0.08]	[1.00]	[0.49]	[0.77]	[0.06]	[1.00]				
Control group mean	40.44	36.19	37.63	39.29	37.20	40.54	40.83	36.89				
Adjusted R ²	0.11	0.11	0.10	0.11	0.18	0.09	0.10	0.14				
Observations	917	978	651	724	356	1417	379	1478				
ANCOVA	No	No	No	No	No	No	No	No	No	No	No	No

Table A2: Impact on Perceived Norms (Adolescents and Parents)

Notes: * significant at 10% level ** significant at 5% level *** significant at 1% level

(1) CHOICES=1 if adolescent was assigned to CHOICES only group where parents were not invited to a gender norms training. EMB shown as a marginal effect.

(2) All regressions include district fixed effects and a set of controls measured at baseline: dummy for whether household has multiple adolescents and age of respondent.

(3) Standard errors are clustered at the community level and are reported under the coefficient in parentheses. Sharpened q-values that correct p-values for the false discovery rate (FDR) are in square brackets. A sharpened q-value of 1.00 indicates the null hypothesis is not rejected at any level of FDR.

	Relationsh	nip Quality w	ith Mother	Relations	ship Quality wi	th Father
	Sta	ndardized In	dex	Sta	andardized Inc	lex
	All	Boys	Girls	All	Boys	Girls
	(1)	(2)	(3)	 (4)	(5)	(6)
Panel A: Endline						
CHOICES	0.002	0.001	0.001	0.000	0.002	-0.004
	(0.01)	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)
	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]	[1.00]
EMB (marginal)	-0.008	0.009	-0.022	-0.020	-0.002	-0.035
	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.02)
	[1.00]	[1.00]	[0.63]	[0.35]	[1.00]	[0.13]
Control group mean	0.00	-0.00	0.00	0.00	-0.01	0.01
Adjusted R ²	0.08	0.08	0.08	0.09	0.09	0.08
Observations	2,532	1,235	1,297	2,223	1,082	1,141
ANCOVA	Yes	Yes	Yes	Yes	Yes	Yes
Panel B: Midline						
CHOICES	0.050**	0.066**	0.036	0.018	0.004	0.029
	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)	(0.03)
	[0.03]	[0.04]	[0.18]	[0.46]	[1.00]	[0.37]
EMB (marginal)	-0.029	-0.028	-0.028	-0.018	-0.000	-0.030
	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)	(0.02)
	[0.12]	[0.24]	[0.38]	[0.45]	[1.00]	[0.22]
Control group mean	-0.00	-0.00	0.00	0.00	0.00	-0.00
Adjusted R ²	0.04	0.04	0.03	0.05	0.06	0.06
Observations	1,909	905	1,004	1,751	826	925
ANCOVA	Yes	Yes	Yes	Yes	Yes	Yes

Table A3: Impact on Adolescents' Relationship with Parents

Notes: * significant at 10% level ** significant at 5% level *** significant at 1% level

(1) CHOICES=1 if adolescent was invited to attend the CHOICES training (assigned to CHOICES only or CHOICES+EMB treatment groups). EMB=1 if adolescent was in CHOICES+EMB group, effectively giving the marginal impact of the EMB program in addition to the CHOICES program.

(2) ANCOVA estimation controls for the level of the outcome variable at baseline. All regressions include district fixed effects and a set of controls measured at baseline: dummy for whether household has multiple adolescents and age of respondent.

(3) Standard errors are clustered at the community level and are reported under the coefficient in parentheses. Sharpened q-values that correct p-values for the false discovery rate (FDR) are in square brackets. A sharpened q-value of 1.00 indicates the null hypothesis is not rejected at any level of FDR.

(4) Outcomes: Indices for relationship with mother and father are a standardized index of the number of statements (of 14) in which the adolescent indicates they perceive their relationship with their parent to be positve (index standardized as per method from Kling et al. 2007). Higher scores indicate higher-quality relationships.

	Completed any domestic tasks in past week (Yes=1) Boys Girls		Hours s domestic average week	pent on tasks in day last (0-24)	Index: Ca sibling of gender	aring for opposite · (0-1)
	Boys (1)	Girls (2)	Boys (3)	Girls (4)	Boys (5)	Girls (6)
Panel A: Endline						
CHOICES	0.070*	0.012	-0.027	-0.150	0.117**	0.114
	(0.04)	(0.02)	(0.13)	(0.20)	(0.06)	(0.08)
	[0.07]	[0.83]	[1.00]	[0.88]	[0.04]	[0.17]
EMB (marginal)	-0.013	0.015	-0.102	-0.019	0.007	-0.044
	(0.04)	(0.01)	(0.17)	(0.14)	(0.05)	(0.06)
	[1.00]	[0.22]	[1.00]	[1.00]	[1.00]	[0.97]
Control group mean	0.50	0.94	3.61	5.08	0.01	-0.01
Adjusted R ²	0.04	0.02	0.02	0.01	0.11	0.09
Observations	1,285	1,358	1,285	1,358	1,287	1,271
ANCOVA	Yes	Yes	Yes	Yes	No	No
Panel B: Midline						
CHOICES	0.008	-0.015	-0.149	0.046	0.214***	0.202*
	(0.03)	(0.01)	(0.15)	(0.16)	(0.07)	(0.12)
	[1.00]	[0.31]	[0.47]	[1.00]	[0.00]	[0.11]
EMB (marginal)	0.087**	-0.007	0.305*	-0.075	-0.037	-0.078
	(0.04)	(0.02)	(0.16)	(0.14)	(0.09)	(0.09)
	[0.05]	[1.00]	[0.08]	[1.00]	[1.00]	[0.71]
Control group mean	0.53	0.96	3.24	5.03	0.00	-0.02
Adjusted R ²	0.03	0.07	0.03	0.07	0.10	0.14
Observations	939	1,048	939	1,048	967	1,026
ANCOVA	Yes	Yes	Yes	Yes	No	No

Table A4: Impact on Adolescent Behaviors - Chores and Care

Notes: * significant at 10% level ** significant at 5% level *** significant at 1% level

(1) CHOICES=1 if adolescent was assigned to CHOICES only or CHOICES+EMB treatment groups. EMB=1 if adolescent was assigned to CHOICES+EMB group, effectively giving the marginal impact of the EMB program in addition to the CHOICES program.

(2) ANCOVA estimation controls for the level of the outcome variable. All regressions include district fixed effects and a set of controls measured at baseline: dummy for whether household has multiple adolescents and age of respondent.

(3) Standard errors are clustered at the community level and are reported under the coefficient in parentheses. Sharpened q-values that correct p-values for the false discovery rate (FDR) are in square brackets. A sharpened q-value of 1.00 indicates the null hypothesis is not rejected at any level of FDR.

(4) Outcomes: "Completed any domestic tasks in the past week" =1 if adolescent completed any of the following household chores activities in the past 7 days: cooking, washing clothes, cleaning the house, or washing dishes. "Hours spent on domestic tasks in average day last week (0-24)" = number of hours spent on care for others and domestic tasks on an average day in the past week. "Index: Caring for sibling of opposite gender (0-1)" is based on whether the child mentions one thing they did in the past week to let their sibling know they care for them, and whether they have spoken to their parents about their sibling's wellbeing during the last month (Index standardized using method by Kling et al. 2007).

ale: dized)
Girls (9)
0.182
(0.43)
[1.00]
-0.145
(0.26)
Ì1 00Ĵ
[]
-0.03
0.04
1.358
,
0 452
(0.35)
[0.00]
[0.27]
-0.018
(0.36)
[1 00]
[1.00]
-0.04
0.02
1.048
Yes

Table A5: Impact on Adolescent Mental Health and Hope

Notes: * significant at 10% level ** significant at 5% level *** significant at 1% level

(1) CHOICES=1 if adolescent was assigned to CHOICES only or CHOICES+EMB treatment groups. EMB=1 if adolescent was assigned to CHOICES+EMB group, effectively giving the marginal impact of the EMB program in addition to the CHOICES program.

(2) ANCOVA estimation controls for the level of the outcome variable. All regressions include district fixed effects and a set of controls measured at baseline: dummy for whether household has multiple adolescents and age of respondent.

(3) Standard errors are clustered at the community level and are reported under the coefficient in parentheses. Sharpened q-values that correct p-values for the false discovery rate (FDR) are in square brackets. A sharpened q-value of 1.00 indicates the null hypothesis is not rejected at any level of FDR.

(4) Outcomes in Table 4 columns 1 and 2 are created from the Strengths and Difficulties Questionnaire (SDQ) which is a behavioral screening questionnaire consisting of 25 statements developed by Goodman (1997) that aggregates 4 sub-scales: emotional problems, conduct problems, hyperactivity/inattention, peer problems into an index of "Total Difficulties Scale" (0-20). The sub-scale that measures prosocial behavior is also shown. The Children's Hope Scale (CHS) is a six-item scale developed by Snyder et al. (1997) that measures a child's perception that their goals can be met. All index measures have been standardized as per the method from Kling et al. 2007.

	Enr	olled in Sch (Yes = 1)	iool	Hours per day spent in school or studying last week (0-24) (conditional on enrolment) Aspiration to stu				tion to stuc versity (Yes	dy up to s = 1)		
	All (1)	Boys (2)	Girls (3)	A (4	. 4)	Boys (5)	Girls (6)		All (7)	Boys (8)	Girls (9)
Panel A: Endline				`					<u> </u>		
CHOICES	-0.004 (0.02) [1.00]	0.017 (0.02) [0.89]	-0.024 (0.03) [0.69]	0.1 (0. [0.8	08 14) 83]	0.133 (0.18) [0.83]	0.054 (0.24) [1.00]	0.0 (0 [0	025* .01) .09]	0.021 (0.02) [0.34]	0.029* (0.02) [0.08]
EMB (marginal)	0.008 (0.02) [1.00]	-0.025 (0.03) [0.67]	0.042* (0.02) [0.08]	0.1 (0.1 [0.4	50 15) 49]	0.157 (0.24) [1.00]	0.180 (0.21) [0.64]	-0 (0 [1	.004 .02) .00]	-0.005 (0.02) [1.00]	-0.003 (0.02) [1.00]
Control group mean	0.86	0.88	0.84	6.	63	6.92	6.35	0	.90	0.91	0.90
Adjusted R ²	0.07	0.03	0.11	0.	03	0.03	0.02	0	.01	0.02	0.00
Observations	2.643	1.285	1.358	2.6	643	1.285	1.358	2	506	1.229	1.277
Panel B: Midline			<u>·</u>			,	<u> </u>				•
CHOICES	0.005 (0.02) [1.00]	-0.016 (0.03) [1.00]	0.024 (0.03) [0.71]	-0.((0. [1.))69 16) 00]	-0.096 (0.24) [1.00]	-0.041 (0.21) [1.00]	-0 (0 [1	.014 .02) .00]	-0.041 (0.03) [0.15]	0.010 (0.03) [1.00]
EMB (marginal)	0.003 (0.02) [1.00]	0.011 (0.02) [1.00]	0.003 (0.03) [1.00]	0.0 (0.1 [1.0)22 11) 00]	-0.056 (0.22) [1.00]	0.165 (0.16) [0.48]	0. (0 [0	034 .02) .14]	0.041 (0.03) [0.21]	0.029 (0.03) [0.42]
Control group mean	0.83	0.87	0.79	6.3	33	6.81	5.89	0	.87	0.89	0.85
Adjusted R ²	0.21	0.14	0.27	0.	14	0.10	0.17	0	.07	0.07	0.06
Observations	1,987	939	1,048	1,9	87	939	1,048	1,	886	908	978
ANCOVA	Yes	Yes	Yes	Ye	es	Yes	Yes	١	′es	Yes	Yes

Table A6: Impact on Schooling Outcomes

Notes: * significant at 10% level ** significant at 5% level *** significant at 1% level

(1) CHOICES=1 if adolescent was assigned to CHOICES only or CHOICES+EMB treatment groups. EMB=1 if adolescent was assigned to CHOICES+EMB group, effectively giving the marginal impact of the EMB program in addition to the CHOICES program.

(2) ANCOVA estimation controls for the level of the outcome variable. All regressions include district fixed effects and a set of controls measured at baseline: dummy for whether household has multiple adolescents and age of respondent.

(3) Standard errors are clustered at the community level and are reported under the coefficient in parentheses. Sharpened q-values that correct p-values for the false discovery rate (FDR) are in square brackets. A sharpened q-value of 1.00 indicates the null hypothesis is not rejected at any level of FDR.

	Gender Equitable Attitudes (Standardized Index) Full Index All	Completed any domestic tasks in past week (Yes=1) Boys	Strengths and Difficulties (SDQ) Scale: (Standardized Index) Total difficulties index All
	(1)	(2)	(3)
Panel A: Endline CHOICES	0.047** (0.02)	0.060* (0.04)	-0.830** (0.38)
Social Desirability Score	0.010*** (0.00)	-0.008* (0.00)	-0.827*** (0.05)
CHOICES x Social Desirability Score	0.007** (0.00)	0.000 (0.01)	0.036 (0.07)
EMB (marginal)	0.013 (0.02)	-0.004 (0.03)	0.439 (0.42)
EMB x Social Desirability Score	-0.001 (0.00)	0.004 (0.01)	-0.119* (0.07)
Control group mean	-0.00	0.49	-0.00
Adjusted R ²	0.11	0.10	0.32
Observations	2,643	1,285	2,643
Coefficient: CHOICES + CHOICES × Soc Desirability = 0	0.054	0.061	-0.794
p-value: CHOICES + CHOICES × Soc Desirability = 0	0.009	0.094	0.033
	Yes	Yes	Yes
Panel B: Midline	0 0EZ***	0.021	0.602*
CHOICES	(0.02)	(0.03)	(0.35)
Social Desirability Score	0.006	-0.010*	-0.760***
	(0.00)	(0.01)	(0.08)
CHOICES x Social Desirability Score	0.009** (0.00)	-0.003 (0.01)	0.155** (0.07)
EMB (marginal)	0.004 (0.02)	0.075* (0.04)	0.414 (0.35)
EMB x Social Desirability Score	-0.003 (0.00)	0.026*** (0.01)	-0.133** (0.06)
Control group mean	0.00	0.51	0.00
Adjusted R^2	0.11	0.04	0.35
Observations	1987	939	1987
Coefficient: CHOICES + CHOICES × Soc Desirability = 0	0.066	0.018	-0.448
p-value: CHOICES + CHOICES × Soc Desirability = 0	0.003	0.585	0.233
ANCOVA	Yes	Yes	Yes

Table A7: Robustness Check for Social Desirability Bias

Notes: * significant at 10% level ** significant at 5% level *** significant at 1% level

(1) CHOICES=1 if adolescent was assigned to CHOICES only or CHOICES+EMB treatment groups. EMB=1 if adolescent was assigned to CHOICES+EMB group, effectively giving the marginal impact of the EMB program in addition to the CHOICES program. Social Desirability Score is a continuous standardized index measure of the propensity to give socially desirable answers measured at the time of the follow-up surveys.

(2) ANCOVA estimation controls for the level of the outcome variable. All regressions include district fixed effects and a set of controls measured at baseline: dummy for whether household has multiple adolescents and age of respondent.

(3) Standard errors are clustered at the community level and are reported under the coefficient in parentheses.

(4) All index measures have been standardized as per the method from Kling et al. 2007.

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
			Baseline			Midline			Endline	
		Control	CHOICES Treatment	t-test Difference	Control	CHOICES Treatment	t-test Difference	Control	CHOICES Treatment	t-test Difference
	Variable (coded 1 = gender egalitarian)	Mean/SE	Mean/SE	(1)-(2)	Mean/SE	Mean/SE	(4)-(5)	Mean/SE	Mean/SE	(7)-(8)
1	In a family, girls and boys should get the same amount to eat no matter how much food there is.	0.882 [0.014]	0.884 [0.009]	-0.001	0.808 [0.017]	0.830 [0.011]	-0.021	0.863 [0.015]	0.903 [0.009]	-0.040**
2	Girls should be honored with bride price rather than investing in her education.	0.468 [0.021]	0.463 [0.015]	0.005	0.506 [0.021]	0.580 [0.015]	-0.074***	0.485 [0.021]	0.558 [0.015]	-0.073***
3	The most important aspiration for a girl is to be a mother and take care of her family.	0.318 [0.020]	0.324 [0.014]	-0.006	0.336 [0.020]	0.410 [0.015]	-0.074***	0.338 [0.020]	0.388 [0.014]	-0.050**
4	The most important aspiration for a boy is to do well in school and succeed professionally.	0.174 [0.016]	0.184 [0.011]	-0.010	0.188 [0.017]	0.201 [0.012]	-0.013	0.172 [0.016]	0.168 [0.011]	0.004
5	The more successful the boy is in his profession, the more he has to pay as bride price.	0.231 [0.018]	0.213 [0.012]	0.018	0.244 [0.018]	0.260 [0.013]	-0.016	0.226 [0.018]	0.226 [0.012]	0.001
6	Boys as well as girls should be responsible for carrying wood and water.	0.761 [0.018]	0.734 [0.013]	0.027	0.776 [0.018]	0.765 [0.013]	0.011	0.803 [0.017]	0.834 [0.011]	-0.031
7	It is okay for a man to hit his wife if she disagrees with him.	0.736 [0.019]	0.712 [0.013]	0.024	0.747 [0.019]	0.784 [0.012]	-0.037*	0.808 [0.017]	0.837 [0.011]	-0.028
8	A woman should be obedient to her husband in order to keep her family together.	0.244 [0.018]	0.281 [0.013]	-0.037	0.241 [0.018]	0.259 [0.013]	-0.018	0.168 [0.016]	0.177 [0.011]	-0.009
9	If a family can only afford for one child to go to school it should be the boy.	0.436 [0.021]	0.464 [0.015]	-0.028	0.354 [0.020]	0.386 [0.014]	-0.032	0.356 [0.020]	0.440 [0.015]	-0.083***
10	Dowry is more important for the family's esteem rather than having educated daughter.	0.535 [0.021]	0.554 [0.015]	-0.019	0.552 [0.021]	0.568 [0.015]	-0.017	0.608 [0.021]	0.687 [0.014]	-0.079***
11	Girls who spend their time studying rather than helping their mothers will not be the ideal wives.	0.524 [0.021]	0.534 [0.015]	-0.010	0.600 [0.021]	0.602 [0.014]	-0.002	0.647 [0.020]	0.698 [0.014]	-0.051**
12	A boy who shows his affection to his sister is weak.	0.821 [0.016]	0.815 [0.011]	0.006	0.844 [0.015]	0.863 [0.010]	-0.018	0.866 [0.014]	0.877 [0.010]	-0.011
13	Boys should have more free time than girls.	0.439 [0.021]	0.460 [0.015]	-0.020	0.416 [0.021]	0.461 [0.015]	-0.045*	0.427 [0.021]	0.481 [0.015]	-0.054**
14	It is more important for a girl to help at home and learn household activities than to spend time studying.	0.458 [0.021]	0.456 [0.015]	0.001	0.485 [0.021]	0.536 [0.015]	-0.051**	0.515 [0.021]	0.582 [0.015]	-0.067***
	Number of Observations (Panel)	553	1144		553	1144		553	1144	

Table A8: Gender Attitudes - Individual Statements

Note: ***, **, and * indicate significance at the 1, 5, and 10 percent critical level. The value displayed for t-tests are the differences in the means across the groups at baseline, midline and endline. Means are presented for the panel i.e. the sample of respondents who were surveyed across all 3 survey rounds.
Table A9: Survey Attrition Rates

	Attrition Rate Midline	Attrition Rate Endline
	(1)	(2)
CHOICES	-0.015	-0.015
	(0.019)	(0.020)
CHOICES + EMB	0.028 (0.018)	-0.006 (0.018)
Mean attrition rate of control group	0.383	0.198
Number of observations (households)	3237	3237

Notes: * significant at 10% level, ** significant at 5% level, *** significant at 1% level

(1) CHOICES=1 if adolescent was assigned to CHOICES only or CHOICES+EMB treatment groups. CHOICES+EMB=1 if adolescent was assigned to CHOICES+EMB treatment group.

(2) Columns (1) and (2) display results of OLS regressions of attrition on each treatment arm with district fixed effects. Attrition=1 if child attrits between baseline and midline(or endline), attrition=0 otherwise. Standard errors are clustered at the community level.

	Gender Equitable Attitudes (Standardized Index) Full Index	Completed any domestic tasks in past week (Yes=1)	Strengths and Difficulties (SDQ) Scale: (Standardized Index) Total difficulties index
	All	Boys	All
	(1)	(2)	(3)
Panel A: Endline			
CHOICES	0.054**	0.071*	-1.338***
	(0.02)	(0.04)	(0.44)
EMB (marginal)	0.015	-0.012	0.484
	-0.02	(0.04)	(0.44)
Control group mean	-0.00	0.50	-0.11
Observations	2,643	1,285	2,643
ANCOVA	Yes	Yes	Yes
Panel B: Midline			
CHOICES	0.071***	0.008	-1.061**
	(0.02)	(0.03)	(0.42)
EMB (marginal)	-0.007	0.095**	0.707*
	(0.02)	(0.04)	(0.38)
Control group mean	-0.01	0.53	-0.07
Observations	1,987	939	1,987
ANCOVA	Yes	Yes	Yes

Table A10: Robustness Check - LASSO selected controls

Notes: * significant at 10% level ** significant at 5% level *** significant at 1% level

(1) CHOICES=1 if adolescent was assigned to CHOICES only or CHOICES+EMB treatment groups. EMB=1 if adolescent was assigned to CHOICES+EMB group, effectively giving the marginal impact of the EMB program in addition to the CHOICES program.

(2) ANCOVA estimation controls for the level of the outcome variable. All regressions include district fixed effects and a set of controls measured at baseline using a double-LASSO-selected controls procedure of Belloni, Chernozhukov, and Hansen (2014).

(3) Standard errors are clustered at the community level and are reported under the coefficient in parentheses.

(4) All index measures have been standardized as per the method from Kling et al. 2007.

Table A11: Robustness Check - Conformity Position Effects - Modal Responses

	Switch in Public (Positions 2-5)		
	All	Boys	Girls
	(1)	(2)	(3)
CHOICES (0/1)	-0.017	0.006	-0.042*
	(0.02)	(0.02)	(0.02)
Number of CHOICES peers in team (0-4)	0.001	0.011	-0.011
	(0.01)	(0.01)	(0.01)
Mixed-gender teams	0.044**	0.000	0.045**
	(0.02)	(0.02)	(0.02)
Boys only team	0.048		
	(0.03)		
Modal response of positions before different to i's private response	0.525***	0.616***	0.436***
	(0.04)	(0.05)	(0.06)
CHOICES*Modal response of positions before different	-0.075	-0.178***	0.025
	(0.05)	(0.06)	(0.07)
Control group mean	0.25	0.27	0.23
Adjusted R-Squared	0.31	0.30	0.32
Observations	2352	1151	1201

Notes: * significant at 10% level ** significant at 5% level *** significant at 1% level

(1) CHOICES=1 if adolescent was assigned to CHOICES only or CHOICES+EMB treatment groups; 0 if in control group.

(2) Standard errors are clustered at the individual level and are reported under the coefficient in parentheses. All OLS regressions includes district fixed effects and controls for whether household has multiple adolescents, the age of respondent, and peer effects given by the number of CHOICES adolescents in the adolescent's team excluding the respondent (0-4). Regressions control for randomized team gender composition (whether team is mixed-gender, boys only, and girls only) where girls only is the omitted category in the regression, statement fixed effects and enumerator-gender fixed effects.

(3) Outcome: Switch in Public is whether the adolescent changed her privately held belief when in a public group setting (=1 if adolescent changed their private response in public; 0 if stuck with their private response).

(4) CHOICES treatment dummy variable is interacted with a dummy variable that equals 1 if the mode of the public responses of the individuals seated in the position before adolescent i is different to i's private response i.e. most of the individuals responding before had an opposing view to adolescent i.

	Endline Survey		Lab-in-the-Field Games Sample		
	Gender Equitable Attitudes Standardized Index	Gender Equitable Attitudes Standardized Index	Gender Equitable Attitudes Standardized Index	Gender Equitable Attitudes Standardized Index	
	Index with all 14 statements as shown in Table 2	Index with 3 statements endline response	Index with 3 statements games private response	Index with 3 statements games private response (plus team gender composition controls)	
	(1)	(2)	(3)	(4)	
Panel A: Endline					
CHOICES	0.056**	0.073*	0.157***	0.170***	
	(0.02)	(0.04)	(0.05)	(0.05)	
	[0.02]	[0.08]	[0.00]	[0.00]	
Number of CHOICES peers in team (0-4)				-0.012	
				(0.03)	
				-0.145**	
Mixed-gender teams				(0.06)	
				-0.316***	
Boys only team				(0.08)	
Control group mean	-0.00	-0.00	0.00	0.00	
Adjusted R ²	0.08	0.12	0.07	0.09	
Observations	2643	2643	994	994	
ANCOVA	Yes	Yes	No	No	
Panel B: Midline					
CHOICES	0.065***	0.092***			
	(0.02)	(0.03)			
	[0.01]	[0.01]			
Control group mean	-0.01	-0.01			
Adjusted R ²	0.08	0.04			
Observations	1987	1987			
ANCOVA	Yes	Yes			

Table A12: Standardized Attitudes Index - Survey versus Lab-in-the-Field Responses

Notes: * significant at 10% level ** significant at 5% level *** significant at 1% level

(1) CHOICES=1 if adolescent was invited to attend the CHOICES training (assigned to CHOICES only or CHOICES+EMB treatment groups).

(2) ANCOVA estimation controls for the level of the outcome variable at baseline. All regressions include district fixed effects and a set of controls measured at baseline: dummy for whether household has multiple adolescents and age of respondent.

(3) Standard errors are clustered at the community level and are reported under the coefficient in parentheses. Sharpened q-values that correct p-values for the false discovery rate (FDR) are in square brackets.

(4) Table A9 Column 1 Gender Equitable Index is a standardized index of 14 gender-progressive statements the respondent agrees with (index standardized as per method from Kling et al. 2007) as was presented in Table 2 column 1. Higher scores indicate more gender-equitable attitudes. Table A9 columns 2, 3 and 4 Gender Equitable Index is a standardized index of 3 gender statements the respondent agrees with (index standardized as per method from Kling et al. 2007). The 3 statement standardized index is shown to have a comparable measure of attitudes at endline as was included in the lab-in-the-field. The 3 statements in the index are: (1) It is more important for a girl to help at home than spend time studying, (2) It is okay for a man to hit his wife if she disagrees with him, and (3) Boys should have more free time than girls.