Online Appendix to: Empowerment and Restraint – Evidence from Teenage Girls' Clubs in Kenya

Contents

A Additional Information

A.1 Pre-registration

All analysis procedures follow our pre-registration plan. The only exception is that, at the time of pre-registration, we did not yet know that four households in our baseline sample were duplicates. Hence, we did not pre-specify that we would use IPW to deal with the greater probability of these four units being assigned to treatment. We nonetheless adopt this approach, because it preserves the unbiasedness of our estimation. Analyses that have been pre-specified but are not (yet) included in this draft:

- Analysis of cross-household spillovers
- Estimates of complier average causal effects (CACEs)
- Robustness to extreme value bounds
- Analyses of HIV-related outcomes which have been pre-specified as a separate paper

A.2 Explanatory note for regression tables

The unit of analysis is the respondent. Standard errors are heteroskedasticity robust and allow for clustering on the household level for analyses of siblings, brothers, and sisters. As pre-specified, we control for the following nine pre-treatment covariates that were measured at baseline among main respondents:

1. *Age*: How old are you?
    - Integer
2. *Access*: Do you own a mobile phone? If no: Do you have regular access to someone else's mobile phone (e.g., mobile phone of a family member)?
    - 0 = No access
    - 1 = Access but no own phone
    - 2 = Owns phone
3. *Stand Up*: Which of the following two statements comes closest to your view? Statement 1: If my brother argues that some of my friends are not good for me, I would distance myself from them even if I like them. Statement 2: I am quite capable of choosing the friends I want and will continue spending time with my friends even if my brother argues that they are not good for me. If statement 2: Would you confront your brother and ask him to stop questioning your choice of friends or would you rather not discuss your friends with him anymore?
    - 0 = Statement 1 is closest to my view

- 1 = Statement 2 is closes to my view and I would confront my brother and ask him to stop questioning my choice of friends
- 2 = Statement 2 is closest to my view and I would rather not discuss my friends with him anymore.

4. *Hang Out*: Neema is best friends with Brian. A few months ago, Neema started dating Jayson. Neema used to always watch Brian play basketball on Saturday mornings. Jayson is not comfortable with Neema going to see Brian and his basketball friends. Which of the following statements do you agree with more?

   - 1 = Statement 1: Neema should be able to hang out with whomever she wants and however much she pleases.
   - 0 = Statement 2: I understand that Jayson is not comfortable with his girlfriend spending time with another boy and Neema should take that into account

5. *Share Cutlery*: And what about these two statements:

   - 1 = I have no problem sharing cutlery with someone who is HIV positive.
   - 0 = I would not want to share cutlery with HIV positive people.

6. *Any Sky Exposure*: Have you read the following magazines? Have you ever watched any of the following shows?

   - 0 = Respondent has not read the SKY magazine, and not watched the Sista Show, and not watched PAA
   - 1 = Respondent has read the SKY magazine, or watched the Sista Show, or watched PAA

7. *People*: Do you have your own room? If yes: How many people sleep in the room that you sleep in?

   - Integer

8. *Media Usage*: Index of

   - Do you have a TV at home? If yes: How often do you watch TV at home? 0 = Does not have TV, 0.25 = Less than once a month/ Never, 0.5 = Once a month, 0.75 = A few times a week (e.g., only on weekends), 1 = Every day
   - In the past week, how often did you use/visit the following websites/apps? We are interested in how often you used each of these. It does not matter whether you accessed them through a computer or a phone. Facebook, Instagram, WhatsApp, YouTube, TikTok. 0 = Never, 0.33 = Just once or twice, 0.66 = Roughly every other day, 1 = Every day

9. *Enjoy Swimming*: Would you enjoy going swimming? If yes: Would you enjoy it a lot or just a little bit? If no: Would you not enjoy it at all or just a little bit?

   - 0 = Not enjoy it at all
   - 1 = Not enjoy it a little bit
   - 2 = Enjoy it a little bit
   - 3 = Enjoy it a lot

We used the first seven covariates to form blocks, because we deem it likely that they are predictive of potential outcomes. We control for the last two covariates to remove conditional bias which could arise from the fact that these variables show imbalances across treatment and control conditions (see section B.1). For analyses of siblings, we treat these measures as household-level covariates, since we do not have baseline measures for siblings. We impute missing values in covariates using means as pre-registered. For analyses of the outcome *Like Afrobeats*, we additionally control for whether the main respondent in a household reported liking Afrobeats at baseline, as pre-registered.

We weight observations by the inverse of the probability that they are assigned to the condition that they were actually assigned to:

- Respondents from households that were blocked into pairs receive a weight of $\omega_i = 2$.
- Respondents from households that were blocked into a triplet and assigned to treatment receive a weight of $\omega_i = 3$.
- Respondents from households that were blocked into a triplet and assigned to control receive a weight of $\omega_i = \frac{3}{2}$.

See appendix section A.3 for more details on weights of respondents from duplicate households.

The row named "RI p-value" in each table shows the randomization inference p-values calculated by permuting treatment assignment 8,000 times to simulate the sampling distribution under the sharp null hypothesis of no (positive or negative) treatment effect for any unit. The row labeled "Hypothesis" indicates the direction of this hypothesis test.

The row "Joint p-value" shows the p-value from an omnibus test of multiple hypotheses pertaining to the same outcome category, as indicated in the table header. To conduct these tests, we use nonparametric combination (NPC), a randomization-inference-based procedure to conduct an exact test of the global null hypothesis that a set of sharp sub-hypotheses are true against the alternative hypothesis that at least one of the sub-hypotheses is false, taking into account the dependence across hypotheses. NPC proceeds by calculating an RI p-value for each sub-hypothesis, combining these p-values through a combination function and comparing the result to a simulated null distribution of the same summary statistic of p-values under the global null hypothesis. We combine the J p-values

p_j pertaining to one hypothesis into an overall test statistic T using Fisher's production function:

$$T = -2 \sum_{j=1}^{J} \log(p_j).$$

For analyses of spillover effects, the row labelled "N clusters" shows the number of households.

Some tables show interactions between the SKY treatment and other randomized primes. These estimates come from regressions that add to our main specification an indicator for the respondent's assignment to the randomized prime as well as the interaction between this indicator and the indicator for assignment to the SKY intervention. The row named "RI p-value SKY" in such tables shows randomization inference p-values calculated by subsetting the data to respondents who were not assigned to the respective prime and permuting assignment to the SKY treatment 8,000 times to simulate the sampling distribution under the sharp null hypothesis of no (positive or negative) treatment effect of the SKY treatment for any unit not assigned to the respective prime. The row labeled "Hyp. SKY" indicates the direction of this hypothesis test. The row named "RI p-value Prime" in such tables shows randomization inference p-values calculated by subsetting the data to respondents from households that were not assigned to the SKY treatment and permuting assignment to the prime 8,000 times to simulate the sampling distribution under the sharp null hypothesis of no (positive or negative) treatment effect of the prime for any respondent not assigned to the SKY treatment. The row labeled "Hyp. Prime" indicates the direction of this hypothesis test. The row named "RI p-value Diff" in such tables shows parametric p-values that pertain to the weak null hypothesis of no difference in conditional average treatment effects. The row labeled "Hyp. Diff." indicates the direction of this hypothesis test.

We use parametric p-values in such cases, because RI-based tests of hypotheses regarding interactions require the researcher to commit to a sharp null hypothesis which stipulates that treatment effects in both sub-groups are equal to a particular constant and it is not obvious which constant to choose. Since our method for calculating omnibus p-values requires a randomization-inference-based test of all sub-hypotheses, we rely on an RI p-value when conducting an omnibus test where a sub-hypothesis involves a difference between conditional effects. In such cases, we test the sub-hypothesis that the effect in both sub-groups is equal to the estimated average effect in the entire sample.

A.3 Ethical considerations

Well-being during data collection. Following calls for more transparency in research ethics (Baron and Young, 2022), we monitored participants' knowledge of their rights and experience during data collection through telephonic back-checks with a subset of our sample at baseline and endline. We use questions that trauma psychologists developed to measure childrens' reactions to research (Kassam-Adams and Newman, 2002). Reassuringly, very few respondents report being bored by the study or worrying about the privacy of their responses (see Tables A1 to A3). Close to all respondents felt they could freely choose whether to participate, felt well informed about the study, and were interviewed in their preferred language. However, at baseline, over one-third of respondents were unaware of the option to skip questions or halt the survey. Our enumerator training at endline stressed the importance of these rights and the vast majority of respondents report being aware of them at endline. 7 out of a total of 273 backchecked respondents expressed feeling upset or sad as a result of their study participation. These responses triggered a well-defined referral protocol that enumerators were trained to follow whenever they encountered respondents who seemed distressed both in the field and on the phone. Higher-level study staff followed up and, if necessary, would have referred respondents to free counseling services. However, follow-ups suggested respondents' backcheck responses did not reflect a need for assistance but instead were due to conflicts with parents that had already been resolved or misunderstandings about the questions. We also implemented several additional measures that are standard in research with minors. Both baseline and endline enumerators underwent a safeguarding training, female respondents are interviewed by female enumerators and male respondents by male enumerators, enumerators are trained to interview respondents in private, and we obtain both parental consent and teenager assent (for all respondents who are minors) for participation in both the surveys and the intervention.

Well-being during intervention. Many topics covered by SKY – e.g., negotiation in relationships – are sensitive. Hence, the emotional well-being of club participants was a concern. In addition, SKY seeks to empower girls in a context that is gender conservative which may create the risk of harm through negative reactions of girls' social environment. We took a three-pronged approach to these issues. First, we rely heavily on the local expertise of our partners. The SKY program was designed by local experts and is the result of extensive qualitative research. No new SKY content was

developed for this study and even absent our study, SKY content is widely available, e.g. through national TV, and popular in Kenya. All local partners deemed SKY appropriate for teenage girls. The partners have not encountered cases of backlash against SKY from girls' social networks. One reason may be that the program, while focused on teenage girls, also models supportive behavior by teenage boys. Still, any program which seeks to empower a marginalized group may create backlash, and even though we deemed severe backlash that would cause physical or emotional harm unlikely, our research interest was in part whether male peers would react positively or negatively to their sisters' participation. Hence, our second approach was to implement a close-knit monitoring and support system to be able to pick up and take appropriate steps should backlash arise. In addition to the backcheck surveys and referral protocol mentioned above, study staff at the club venue paid close attention to the girls and looked out for signs of distress. Throughout each session, the moderator repeatedly reminded girls that they could approach staff if they felt upset or alternatively call or message 1190, the hotline of One2One, a Kenyan organization that provides free assistance to adolescents and young adults. Neither our endline survey with girls nor that with boys provides evidence of backlash. Third, we ensured that girls and families understood the intervention before consenting to participation and were aware that participation in the club was voluntary. While all local partners agreed on the ethical obligation to compensate families for girls' club attendance because they typically perform economically valuable chores such as child care of younger siblings, the compensation amount chosen was low enough such that the partners deemed it unlikely that the incentives would be coercive. We also sought to make the club intervention as inclusive as possible. To this end, the partners organized childcare services at the club venue, since some of the older girls in the treatment group turned out to have young children of their own. Doing so ensured that the intervention could be inclusive of girls in all life situations without running the risk of altering the club experience or endangering the well-being of small children through their presence during the sessions. A final concern was risks to girls' physical safety that could result from the intervention. Girls were picked up from a network of pickup points and dropped off by buses to and from the venue. The second club session always ended on time, even if the session had to be cut short because of a late start, such that girls would never arrive at the pick-up points late and be in danger when walking home.

Adverse incident. Despite these efforts, an adverse event occurred during the study. One of the study participants could not be located for one night after the close of the second club session

on June 10. To protect the identity of the individuals involved in the adverse incident, we cannot here go into detail about what happened. Yet, we would like to convey that all parties involved in the study took this incident extremely seriously. Study staff worked closely with the family and local authorities to locate the girl. Afterwards, the family was contacted multiple times, including by a local child safeguarding expert, to learn whether the participant had suffered any harm, which, to the best of our knowledge, was not the case. The study partners were able to nonetheless provide the family with support for challenges that the girl had been facing unrelated to the study. All study procedures were halted immediately and we reported the incident to all IRBs that approved this study. After it became clear that the participant hadn't been harmed, all parties involved in the study took time to carefully reflect on whether it would be possible to continue the study in a manner such that potential risks to participants, especially those arising from the transportation procedure, would be outweighed by the benefits of the research. All parties collectively decided that this would be possible given a set of changes, subject to IRB review and approval, especially by the local IRB at Strathmore University which took the lead in reviewing the incident and the new procedures.

Specifically, the responsibility for transportation was handed over to the implementing partner who has extensive experience with transporting large numbers of teenagers to mass events. In retrospect, we should have implemented this division of tasks in the first place, because it better aligns with the expertise of the involved parties. The initial reason to organize the transport the way we did was that we could avoid sharing personally identifiable data like names, addresses and phone numbers collected at baseline among different partners. In the aftermath of the incident, partners worked to ensure that their data infrastructure meets Kenyan data protection regulations and signed a data usage agreement which made this data transfer possible. Once the IRB granted approval for the new protocols, study staff re-visited all households in the treatment group and newly obtained written informed consent from parents and written informed assent from participants who were minors for their participation in the SKY club intervention. During the re-consenting exercise, parents and participants were informed in detail about the new transportation protocol as well as about the fact that their data would be shared. Parents and participants were not given details about the adverse incident – again to protect the identity of the individuals involved. In total, 364 out of 426 participants re-consented. Personally identifiable data were shared among partners only for respondents for whom consent and assent were newly obtained. On August 26, the club sessions resumed (see Table A9).

No further incident occurred during the remaining six sessions.

	No	Maybe	Yes	Don't Know
Being in this study was boring	95% (N = 88)	3% (N = 3)	2% (N = 2)	-
It was my choice if I was in the study.	-	-	100% (N = 93)	-
Being in this study made me feel upset or sad.	96% (N = 89)	1% (N = 1)	3% (N = 3)	-
The things I said will stay private.	4% (N = 4)	2% (N = 2)	94% (N = 87)	-
I was told the truth about the study before it started.	4% (N = 4)	1% (N = 1)	94% (N = 87)	1% (N = 1)
I knew I could skip questions or parts of the study if I wanted to.	39% (N = 36)	-	61% (N = 57)	-
I knew I could stop at any time.	42% (N = 39)	-	58% (N = 54)	-
The interview was conducted in my preferred language.	-	-	100% (N = 93)	-

Table A1: Main respondents' assessment of study experience during baseline backchecks

	No	Maybe	Yes	Don't Know
Being in this study was boring	90% (N = 134)	-	10% (N = 15)	-
It was my choice if I was in the study.	1% (N = 1)	-	99% (N = 148)	-
Being in this study made me feel upset or sad.	99% (N = 147)	-	1% (N = 2)	-
The things I said will stay private.	3% (N = 5)	1% (N = 2)	95% (N = 142)	-
I was told the truth about the study before it started.	2% (N = 3)	1% (N = 2)	97% (N = 144)	-
I knew I could skip questions or parts of the study if I wanted to.	7% (N = 11)	-	93% (N = 138)	-
I knew I could stop at any time.	13% (N = 19)	-	86% (N = 128)	1% (N = 2)
The interview was conducted in my preferred language.	-	-	100% (N = 149)	-

Table A2: Girls' assessment of study experience during endline backchecks (main respondents and sisters)

	No	Maybe	Yes	Don't Know
Being in this study was boring	94% (N = 29)	-	6% (N = 2)	-
It was my choice if I was in the study.	3% (N = 1)	-	97% (N = 30)	-
Being in this study made me feel upset or sad.	97% (N = 30)	-	3% (N = 1)	-
The things I said will stay private.	-	3% (N = 1)	97% (N = 30)	-
I was told the truth about the study before it started.	-	-	100% (N = 31)	-
I knew I could skip questions or parts of the study if I wanted to.	3% (N = 1)	-	97% (N = 30)	-
I knew I could stop at any time.	3% (N = 1)	-	97% (N = 30)	-
The interview was conducted in my preferred language.	-	-	100% (N = 31)	-

Table A3: Brothers' assessment of study experience during endline backchecks

A.4 Sampling strategy

Households. Enumerators worked in pairs or groups of three and were accompanied by local Community Health Workers with knowledge about the location of households with teenagers. Each day, enumerators were assigned starting points that were scattered around the four neighborhoods that we worked in and a direction in which to walk from the starting point. Depending on the population density of the area, enumerators were instructed to skip either 3-4 houses after each interview (in areas dominated by so-called *estates*, i.e., apartment buildings), or 10 houses (in areas with many small single-family houses). This skip pattern was meant to limit spillover effects by reducing the chance that teenagers who enter the sample know each other. For the same reason, we never sampled more than one apartment in an apartment block or more than one household in a yard with multiple dwellings. When sampling in apartment buildings, enumerators were instructed to start looking for a household with an eligible teenage girl either on the top or the bottom floor of the building, and to move along apartment units until they found an eligible respondent. Moreover, enumerators alternated between starting from the top or bottom floor. We used a similar strategy for yards with multiple households where enumerators alternated between start from the household closest to the or furthest from the entrance gate. During the first day of sampling, we enforced that each households needs to have at least two members who are teenage girls of the relevant age range. The goal was to

ensure that every household would contains at least one respondent who could experience spillover from the program (see more below). Since this criterion proved extremely restrictive, we abandoned it after one day of sampling.

N per HH	Sisters	Brothers
0	797	763
1	176	201
2	18	26
3	3	5
4	2	1
Total spillover N	229	272

Table A4: Number of sisters and brothers listed per HH at baseline

Note that these numbers differ slightly from the numbers presented in our PAP, because we have removed four duplicate households that were interviewed twice at baseline.

A.5 Blocking and random assignment to treatment

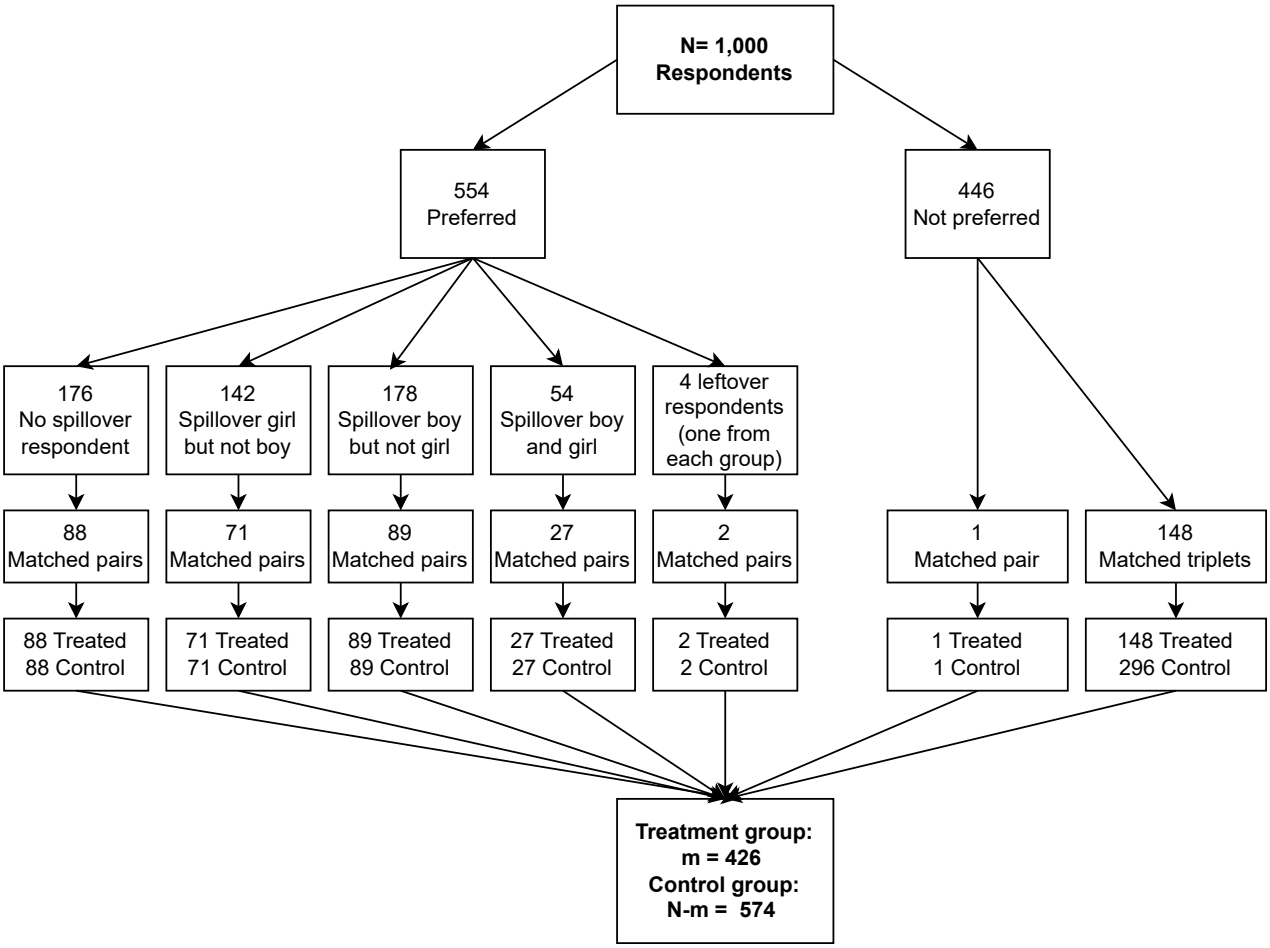

Figure A1: Blocking and random assignment (not accounting for duplicates)

The blocking procedure was designed prior to us discovering that four of the 1,000 households that we sampled at baseline were duplicates and was meant to take into account the possibility that it may not be possible to interview all $N = 1,000$ main respondents at endline. The initial target number of main respondents was 800, 400 of which were meant to be assigned to treatment. We enrolled more respondents at baseline, because the initial sample included fewer respondents in the age range of 14-17 years for which the partner deems its intervention most effective. One option to deal with the fact that we had more main respondents than we needed would have been to remove some respondents outside the preferred age range of 14-17 years from the sample. Yet, this procedure would have meant that the excluded respondents could not have been used to estimate treatment effects even if, for some reason, we would have been able to interview more respondents at endline.

Hence, we instead opted for the following procedure (see Figure A1): First, we divided the sample of $N = 1,000$ main respondents into two sub-samples: 544 *preferred* respondents, and 456 *non-preferred* respondents. We defined as *preferred* all those respondents who fall into the preferred age range of 14-17 years as well as all main respondents who live in households that include at least one spillover respondent (boy or girl). Among *preferred* respondents, we created 272 matched pairs and among *non-preferred* respondents we created 148 matched triplets (and one matched pair since 456 is not divisible by 3). Within each block – whether matched pair or triplet – we randomly assigned one respondent to treatment. This created a treatment group of $m = 426$ respondents.

If we had been able to sample only a subset of the $N = 1,000$ baseline respondents at endline, we would have opted to interview all respondents from matched pairs (and hence all *preferred* respondents). Within matched triples among *on-preferred* respondents, we would have randomly selected only one of the two respondents assigned to the control group to be interviewed. The result would have been a balanced design with a sample of $N = 852$ that would have included all main respondents in the preferred age range and all main respondents from households with spillover units. In other words, our blocking procedure was designed to ensure that we would have satisfied the partner's desire to have the maximum possible number of 14-17 year-olds in the sample while maximizing statistical power for the estimation of within-household spillover effects – even if we could have re-interviewed only a subset of baseline respondents. At the same time, the blocking ensured that we are now able to draw on the entire sample of $N = 1,000$ respondents.

Within the group of *preferred* respondents, we formed matched pairs separately within four groups: respondents from households without spillover respondents, respondents from households with only spillover girls, respondents from households with only spillover boys, and respondents from households with both at least one spillover boy and one spillover girl. The goal was to maximize statistical power for the estimation of treatment effects among, respectively, only spillover boys and only spillover girls. Four respondents could not be matched in pairs within these groups, because these groups were not neatly divisible by 2. These four respondents were re-matched into two pairs.[1] Matched triplets were formed within the entire group of *non-preferred* respondents.[2]

[1] These two pairs were formed by minimizing the within pair Mahalanobis distance of only a subset of our regular blocking covariates: age, one of our measures of assertiveness in relationships and a measure of tolerance towards people who are HIV-positive.

[2] Recall that we formed one matched pair among *non-preferred* respondents, because the number of *non-preferred* respondents was not divisible by 2.

We chose pairs and triplets to minimize the within-block Mahalanobis distance on nine covariates that we deemed likely to be prognostic of potential outcomes: age, two measures of assertiveness in relationships, a measure of tolerance towards people who are HIV-positive, the number of people with whom the respondent sleeps in a room, the respondent's level of access to a mobile phone, whether the respondent has any prior exposure to SKY, as well as latitude and longitude (provided to researchers only as a rotation to preserve privacy of subjects). Blocking on latitude and longitude ensures that the sample is spread out and maximizes our ability to pick up spatial spillover effects across households should they exist.

A.6 Dealing with duplicate observations

	Block 1	**Block 2**	$\mathbf{z_1}$	$\mathbf{z_2}$	**z**	**Prob.**$(\mathbf{z = 1})$	**Prob**$(\mathbf{z = 0})$
1	Pair	Triplet	0	1	1	2/3	1/3
2	Triple	Pair	1	0	1	2/3	1/3
3	Triplet	Triplet	0	1	1	5/9	4/9
4	Pair	Pair	0	0	0	3/4	1/4

Table A5: Blocks, assignments and assignment probabilities of four duplicate households
The first two columns indicate the size of the blocks into which the two instances of the same unit were sorted. The next two columns show the two realized treatment assignments for each instance and the fifth column shows the resulting realized assignment for the household based on the rule that a household is assigned to treatment whenever one of the two instances of this household ended up in the treatment group. The last two columns show the resulting probabilities of assignment to, respectively, treatment and control for the household.

At endline, we discovered that four households in our sample of $1,000$ baseline households were duplicates. In all four cases, the same main respondent had been interviewed twice. When analyzing the data, we assume that a household that was in our sample twice has been assigned to treatment whenever one of the two instances of this household in our sample was assigned to treatment. After all, in this case, the main respondent from this household received an invitation to participate in the SKY club. As can be seen in Table A5, the treatment assignment probabilities for these four households vary depending on whether each of the two instances of this household was part of a matched triplet or a pair. Fortunately, it is not difficult to derive these assignment probabilities. Both our estimation strategy as well as our randomization inference based testing strategy accounts for this variation in assignment probabilities. Finally, we should note that we cannot know whether our

sample contains additional duplicates that we would have discovered had we been able to interview all initially sampled main respondents at baseline. In the absence of additional information, we proceed based on the assumption that the duplicates we discovered are the only existing ones.

A.7 Random assignment to club sessions

The implementing partners organized two separate club sessions every week, because it seemed infeasible to hold a single session with 426 participants. We do not aim to separately identify the effect of participating in one weekly club session or the other. Yet, how girls are sorted into club sessions may nonetheless matter to the degree that the composition of club sessions materially alters their effects. For example, one may think that the effect of participating in the club for a given girl of a given age may be different in a scenario where other girls in the club are close to her in age compared to one in which most other participants are, say, much older. One may counter such concerns with the observation that our estimand is the effect of the SKY club given the composition that endogenously arose in our experiment. Yet, if compositional effects exist and girls sort endogenously into sessions, the nature of the treatment becomes difficult to conceptualize, because it may be materially different across the two sessions. Moreover, our estimand becomes difficult to conceptualize as well, because the quantity we seek to estimate may itself depend on the realized random assignment. This problem is akin to a SUTVA violation, because participants' treated potential outcomes depend on the full vector of random assignments and not just on whether or not they themselves were assigned to treatment. From this perspective, it would seem ideal to randomly assign girls in the treatment group to one of the two weekly sessions, such that both have the same composition in expectation. The problem with this approach is that it may generate non-compliance. We here took a mixed approach. Initially, we randomly assigned girls to the two sessions using simple random assignment. This procedure resulted in 200 of girls being assigned to the first and 226 being assigned to the second session. We only realized late that it would be important to randomly assign girls to sessions. Since there was little time left, IPA Kenya staff conducted the random assignment using Excel by generating a random number between 0 and 1 for all respondents in the treatment group and assigning all those with a number less than or equal to 0.5 to the first session. This procedure is far from ideal, since it is not replicable. Yet, at least we have access to both the random number as well as the initial assignment for each participant. Tables A7 and A8 display covariate balance across session

assignments. IPA staff non-randomly moved 15 respondents from one session assignment to the other in order to have an equal number of participants in both groups. We count these 15 respondents as having switched groups when we analyze the rates at which switching occurred. Other than that, girls were invited to their assigned session, but were subsequently allowed to switch sessions.

A.8 Compliance

	N Attended	Exposed TV or Magazine	Knowledge Club	SKY Exposure
SKY	5.835***	0.287***	4.440***	2.007***
	(0.154)	(0.023)	(0.108)	(0.113)
Outcome range	0 – 13	0 – 1	0 – 7	0 – 9
Control Mean	0.193	0.682	0.259	1.803
Control SD	0.959	0.466	1.043	1.530
Hypothesis	upr	upr	upr	upr
Num. obs.	829	829	829	829

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A6: Effects of SKY on girls' exposure to SKY

Models include pre-registered covariates listed in Section A.2. Stars indicate significance based on RI p-values from test of sharp null of no (positive/negative) treatment effect for any unit. Row "Hypothesis" shows the direction of the test. A test of the hypothesis that all null hypotheses for the tests shown in this table is $p = 0.001$. Details on model specification and testing are in appendix section A.2 and outcome wording and coding in section D.7.

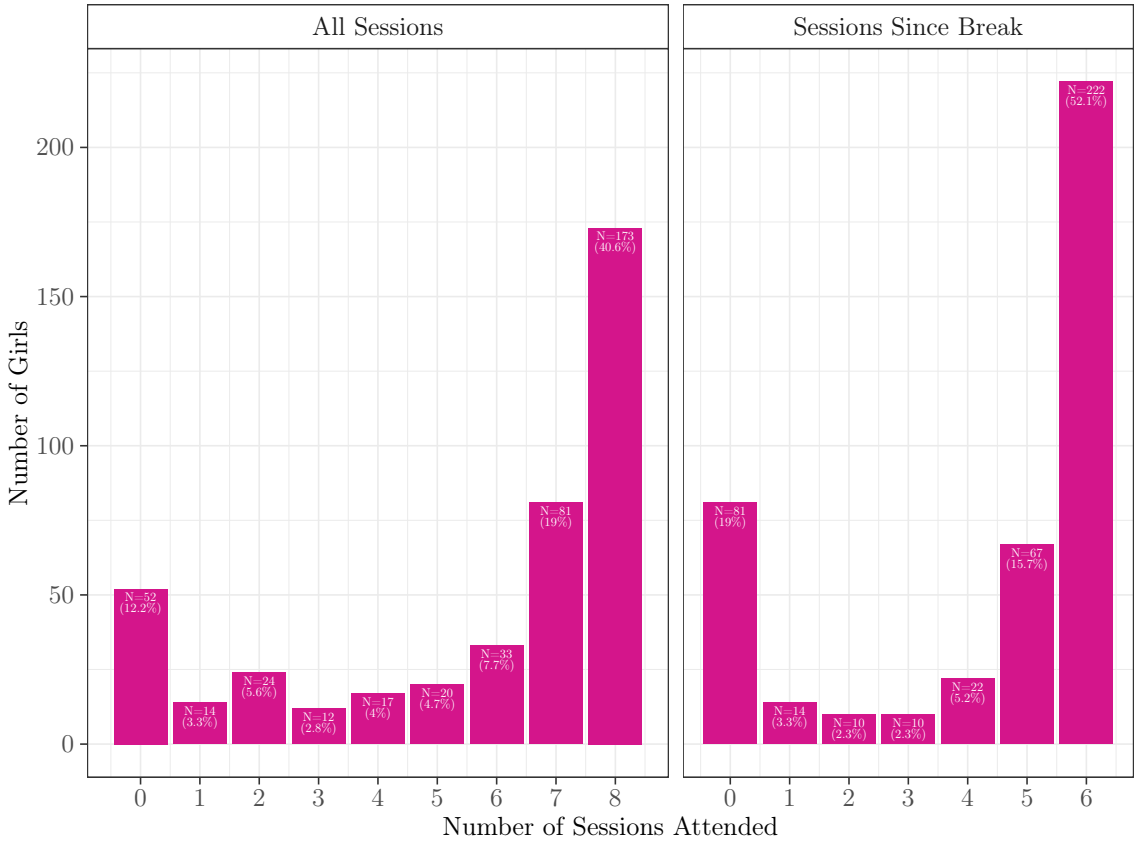

Figure A2: SKY Club Attendance

A.9 Covariate balance across club sessions

Covariate	Min	Max	Week 1		Week 2		Week 3		Week 4	
			Session 1	Session 2	Session 1	Session 2	Session 1	Session 2	Session 1	Session 2
enrolled_school	0	1	0.62	0.66	0.65	0.64	0.63	0.65	0.62	0.66
day	0	3	2.4	2.39	2.42	2.37	2.38	2.4	2.4	2.39
age	12	19	15.78	15.59	15.68	15.7	15.75	15.63	15.86	15.55
access	0	2	1.16	1.01	1.09	1.09	1.15	1.03	1.13	1.05
media_usage	0	1	0.31	0.28	0.31	0.28	0.31	0.28	0.3	0.29
stand_up	0	2	0.54	0.48	0.51	0.51	0.54	0.47	0.53	0.49
program	0	2	1.24	1.31	1.26	1.29	1.3	1.26	1.29	1.27
hang_out	0	1	0.65	0.61	0.66	0.61	0.65	0.62	0.63	0.63
tattoo_stylish	0	3	0.59	0.66	0.6	0.65	0.61	0.64	0.61	0.64
piercing_stylish	0	3	0.75	0.84	0.78	0.81	0.76	0.83	0.74	0.84
enjoy_swimming	0	3	2.38	2.36	2.4	2.35	2.37	2.36	2.33	2.4
like_afrobeats	0	1	0.48	0.36	0.47	0.38	0.49	0.36	0.49	0.37
like_bongo	0	1	0.44	0.47	0.47	0.44	0.45	0.46	0.44	0.47
like_gospel	0	1	0.66	0.76	0.67	0.74	0.66	0.76	0.66	0.75
share_cutlery	0	1	0.54	0.54	0.53	0.56	0.53	0.55	0.53	0.55
any_sky_exposure	0	1	0.48	0.39	0.49	0.39	0.5	0.38	0.47	0.41
assets	0	1	0.54	0.58	0.55	0.57	0.54	0.58	0.54	0.58
own_house	0	1	0.29	0.29	0.29	0.29	0.29	0.29	0.28	0.31
people	0	11	2.69	2.42	2.65	2.48	2.69	2.42	2.72	2.42
block_house	0	1	0.86	0.88	0.86	0.87	0.86	0.87	0.87	0.86
hhsize	1	15	4.88	4.63	4.76	4.75	4.81	4.7	4.77	4.74
kikuyu	0	1	0.42	0.38	0.43	0.37	0.42	0.38	0.43	0.37
luo	0	1	0.19	0.16	0.17	0.17	0.18	0.16	0.18	0.16
swahili	0	1	0.8	0.75	0.79	0.77	0.8	0.75	0.8	0.76
N participants			209	217	192	234	212	214	191	235

Table A7: Covariate balance among actual SKY club participants across weeks and session

Covariate	Min	Max	Week 5		Week 6		Week 7		Week 8	
			Session 1	Session 2	Session 1	Session 2	Session 1	Session 2	Session 1	Session 2
enrolled_school	0	1	0.6	0.68	0.61	0.67	0.6	0.67	0.61	0.67
day	0	3	2.37	2.41	2.37	2.41	2.39	2.4	2.4	2.39
age	12	19	15.92	15.51	15.94	15.5	15.96	15.49	15.95	15.48
access	0	2	1.16	1.03	1.15	1.04	1.17	1.03	1.16	1.03
media_usage	0	1	0.31	0.28	0.3	0.29	0.31	0.28	0.31	0.28
stand_up	0	2	0.56	0.46	0.57	0.46	0.55	0.48	0.53	0.49
program	0	2	1.28	1.28	1.29	1.27	1.26	1.29	1.28	1.28
hang_out	0	1	0.66	0.61	0.66	0.61	0.64	0.62	0.63	0.63
tattoo_stylish	0	3	0.65	0.61	0.62	0.63	0.66	0.6	0.64	0.62
piercing_stylish	0	3	0.77	0.82	0.77	0.82	0.79	0.8	0.77	0.81
enjoy_swimming	0	3	2.36	2.38	2.35	2.39	2.34	2.39	2.35	2.38
like_afrobeats	0	1	0.49	0.37	0.49	0.37	0.5	0.36	0.49	0.37
like_bongo	0	1	0.45	0.46	0.45	0.46	0.45	0.46	0.47	0.44
like_gospel	0	1	0.65	0.76	0.64	0.76	0.64	0.76	0.66	0.76
share_cutlery	0	1	0.54	0.55	0.52	0.56	0.53	0.55	0.52	0.56
any_sky_exposure	0	1	0.48	0.4	0.47	0.41	0.46	0.42	0.47	0.41
assets	0	1	0.54	0.58	0.54	0.58	0.54	0.58	0.54	0.58
own_house	0	1	0.3	0.29	0.31	0.28	0.3	0.29	0.29	0.3
people	0	11	2.77	2.39	2.71	2.44	2.74	2.42	2.74	2.41
block_house	0	1	0.87	0.87	0.86	0.87	0.86	0.87	0.85	0.88
hhsize	1	15	4.8	4.72	4.79	4.73	4.79	4.73	4.78	4.73
kikuyu	0	1	0.45	0.35	0.44	0.36	0.45	0.36	0.45	0.35
luo	0	1	0.17	0.17	0.17	0.17	0.18	0.17	0.17	0.17
swahili	0	1	0.82	0.75	0.8	0.76	0.8	0.76	0.79	0.76
N participants			186	240	180	246	179	247	189	237

Table A8: Covariate balance among actual SKY club participants across weeks and session continued

A.10 Additional figures

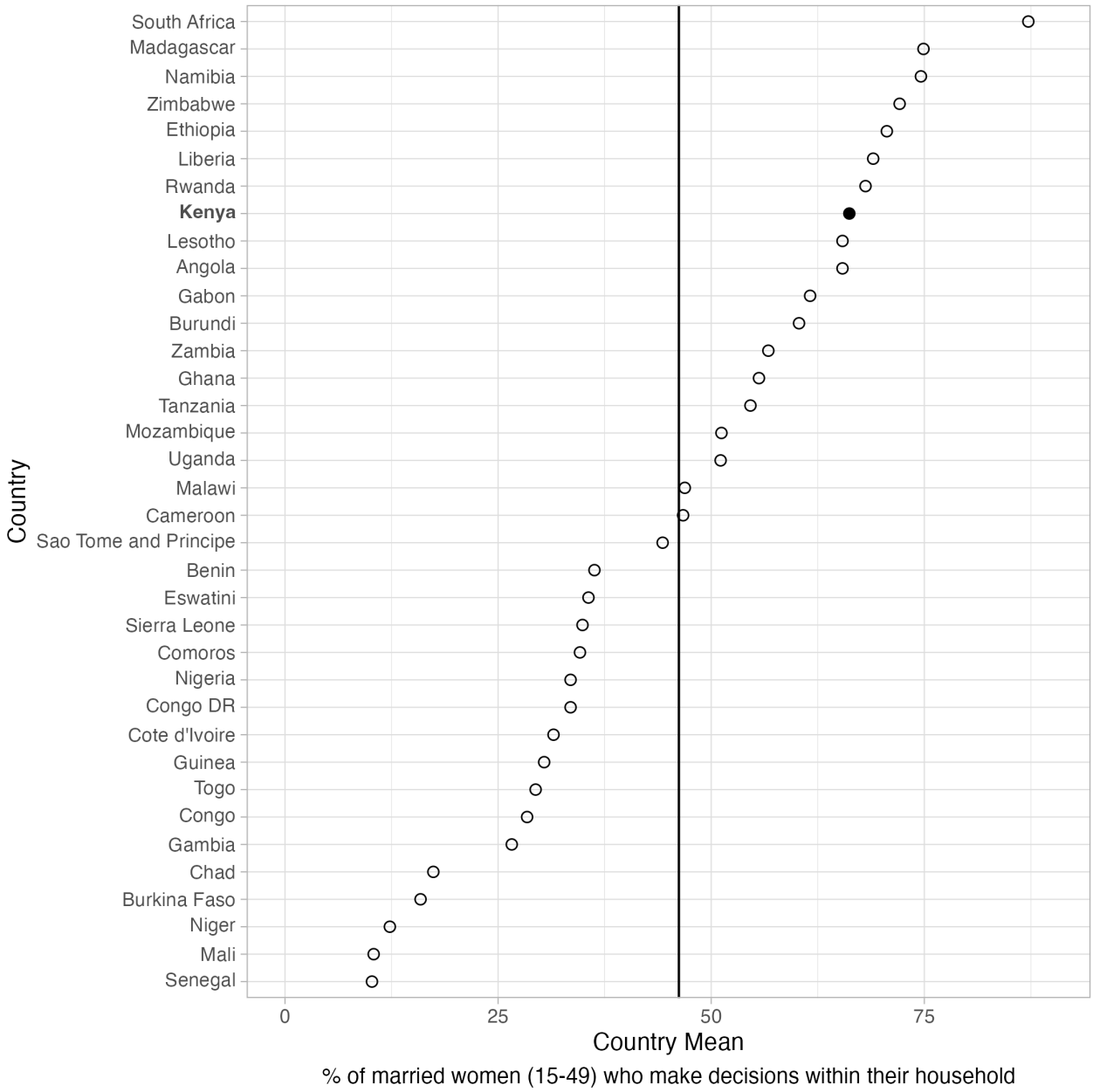

Figure A3: Country mean of currently married women and girls aged 15-49 who make decisions about various issues within their household either by themselves or jointly with their husband.

Each point represents the country-level share of married women who usually makes decisions about various issues within their household, comparing Kenya (bold) to other Sub-Saharan African countries. The decisions include on their own health care, daily and/or large household purchases, and visits to family or relatives. The data is compiled from the final reports of the latest Demographic & Health Survey for each country. The solid vertical line represents the average across countries.

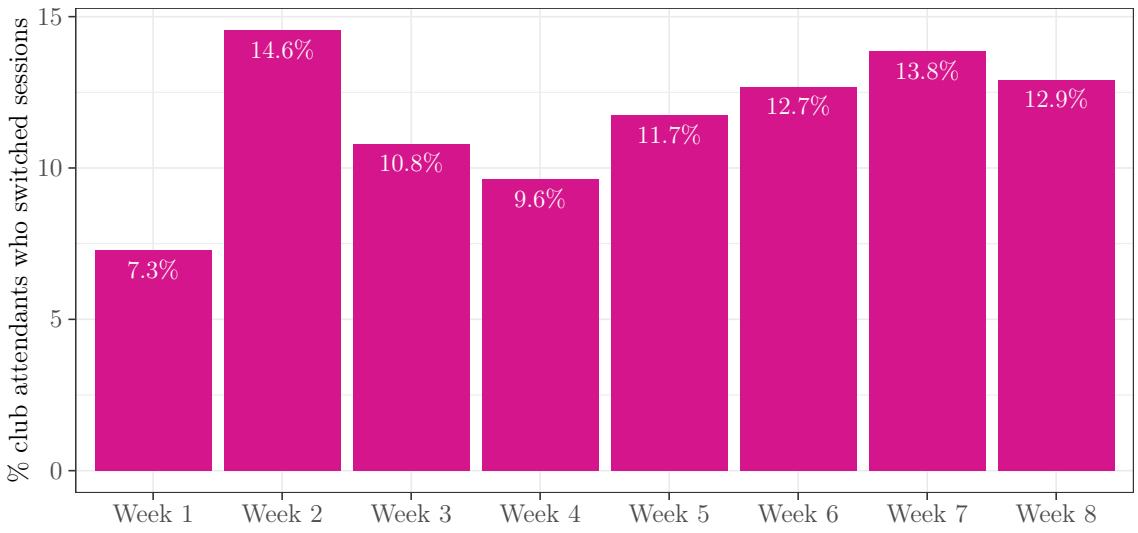

Figure A4: Share of club attendants who switched to the session to which they had not been assigned

Figure A5: The SKY club flyer for the last 6 sessions

Figure A6: SKY Club membership card.

May 1-24	June 3 - 10	June 10	June 11 - Aug 25	Aug 26 - Sep 30	Oct 9 - Jan 18
Baseline survey	First 2 SKY club sessions	Incident	Clubs paused and reconsenting	Remaining 6 SKY club sessions	Endline survey

Table A9: Study timeline

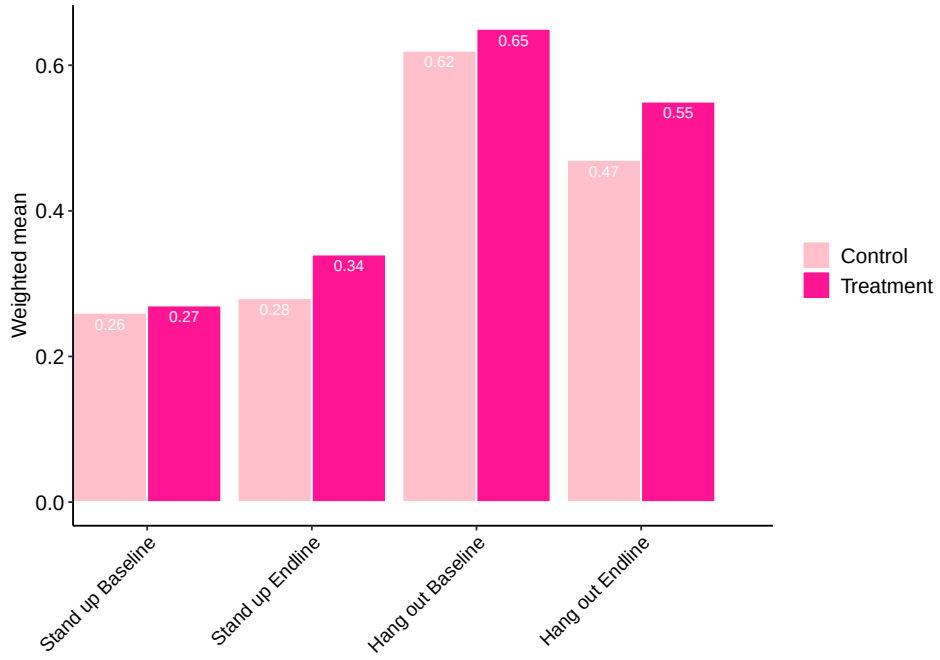

Figure A7: Main respondents' assertive behavior and attitudes by survey wave and treatment assignment

Each bar indicates the weighted average among the main girls (light pink) in the control group and main girls (dark pink) in the treatment group for two outcomes measuring assertive behavior (*Stand up*) and attitudes (*Hang out*) at baseline and endline. We weight by the inverse of each respondent's probability of being assigned to the condition they were assigned to. All outcomes were standardized to range from 0 to 1 by dividing them by their maximum possible value. Question wording: *Stand up*: Which of the following two statements comes closest to your view? If Statement 2: Would you confront your brother and ask him to stop questioning your choice of friends, or would you rather not discuss your friends with him anymore? 0 = Statement 1: If my brother argues that some of my friends are not good for me, I would distance myself from them even if I like them. 0.5 = I am quite capable of choosing the friends I want and will continue spending time with my friends even if my brother argues that they are not good for me. AND I would rather not discuss my friends with him anymore. 1 = I am quite capable of choosing the friends I want and will continue spending time with my friends even if my brother argues that they are not good for me. AND I would confront my brother and ask him to stop questioning my choice of friends. *Hang out*: Let's consider another scenario: Neema is best friends with Brian. A few months ago, Neema started dating Jayson. Neema used to always watch Brian play basketball on Saturday mornings. Jayson is not comfortable with Neema going to see Brian and his basketball friends. Which of the following statements do you agree with more? 1 = Statement 1: Neema should be able to hang out with whomever she wants and however much she pleases. 0 = I understand that Jayson is not comfortable with his girlfriend spending time with another boy and Neema should take that into account.

A.11 Descriptive statistics

Variable	N	Mean	SD	Min	P25	P50	P75	Max
Baseline survey (Total N = 996)								
Age	996.00	15.69	2.59	12.00	13.00	16.00	18.00	19.00
Enrolled in school	996.00	0.63	0.48	0.00	0.00	1.00	1.00	1.00
HH size	996.00	4.73	1.81	1.00	4.00	5.00	6.00	15.00
Numb. of ppl sleeping in same room	996.00	2.91	1.30	1.00	2.00	2.00	3.00	11.00
Phone access	996.00	0.74	0.44	0.00	0.00	1.00	1.00	1.00
Main language Swahili	996.00	0.80	0.40	0.00	1.00	1.00	1.00	1.00
Ethnic group: Kikuyu	996.00	0.39	0.49	0.00	0.00	0.00	1.00	1.00
Ethnic group: Luo	996.00	0.17	0.38	0.00	0.00	0.00	0.00	1.00
Endline survey (Total N = 829)								
Age	829.00	15.90	2.69	12.00	13.00	16.00	18.00	22.00
Numb. of sisters	829.00	0.28	0.56	0.00	0.00	0.00	0.00	4.00
Numb. of brothers	829.00	0.31	0.60	0.00	0.00	0.00	0.00	5.00

Table A10: Main girl sample descriptives of baseline and endline survey

The variables listed in the first column are covariates measured for each main respondent during the baseline and endline survey. The second column is the number of observations for each variable. The third and fourth columns show the mean and standard deviation of each variable among the main respondents who got interviewed at baseline and endline. Columns 5-9 indicate the minimum, 25^{th}, 50^{th}, and 75^{th} percentile, and the maximum of each variable among the main respondents who got interviewed at baseline and endline.

Variable	N	Mean	SD	Min	P25	P50	P75	Max
Endline survey (Total N = 829)								
Age	255.00	15.70	2.79	12.00	13.00	15.00	18.00	22.00
Older than main girl	255.00	0.43	0.50	0.00	0.00	0.00	1.00	1.00
Age difference to main girl	255.00	0.27	3.92	-10.00	-3.00	1.00	3.00	9.00

Table A11: Spillover boy sample descriptives of endline survey

The variables listed in the first column are covariates measured for each brother of a main respondent during the endline survey. The second column is the number of observations for each variable. The third and fourth columns show the mean and standard deviation of each variable among the brothers who got interviewed at endline. Columns 5-9 indicate the minimum, 25^{th}, 50^{th}, and 75^{th} percentile, and the maximum of each variable among the brothers who got interviewed at endline.

Variable	N	Mean	SD	Min	P25	P50	P75	Max
Endline survey (Total N = 829)								
Age	235.00	16.09	2.85	12.00	13.00	16.00	18.00	22.00
Older than main girl	235.00	0.46	0.50	0.00	0.00	0.00	1.00	1.00
Age difference to main girl	235.00	0.03	3.85	-9.00	-3.00	0.00	3.00	9.00

Table A12: Spillover girl sample descriptives of endline survey

The variables listed in the first column are covariates measured for each sister of a main respondent during the endline survey. The second column is the number of observations for each variable. The third and fourth columns show the mean and standard deviation of each variable among the sisters who got interviewed at endline. Columns 5-9 indicate the minimum, 25^{th}, 50^{th}, and 75^{th} percentile, and the maximum of each variable among the sisters who got interviewed at endline.

A.12 SKY club schedule and giveaways

Table A13: SKY Club Schedule

Session	Date	Plan
Session 1	3-Jun	**SKY Intro + Mag 8 engagement** MT/NMT discussion Mag 8 Distribution - discussion around selected articles Mention what will happen in session 2
Session 2	10-Jun	**SISTA Show Screening** Hosted by Owendi Segments: - Celeb corner with the PAA girls (relationship and smoking) - Pink Couch - the guys talk about relationship pressure - SKY advice with the hosts - Doreen's Pink Couch Tease the next session by showing mag 9 (Doreen + SISTA show hosts featured)
Session 3	26-Aug	**SKY @ 3 & Mag 11** Article discussion - SKY advice, SKY guy SKY @ 3 birthday celebration with Kendi Q
Session 4	2-Sep	**Mag 9 Engagement with Doreen** Article discussion - SKY Sesh: Am 1 allowed to say no. How to stay true to yourself Engagement with Doreen (HIV activist)
Session 5	9-Sep	**PAA Screening (Ep 1-3** 55 min screening + 35 min discussion Guided discussion at the end Tease part 2 of screening in session 6 with PAA cast
Session 6	16-Sep	**PAA Screening (ep 4-6)** 55 min screening + 35 min discussion Guided discussion at the end Tease Collo Blue attendance in session 7
Session 7	23-Sep	**Mag 10 & Collo Blue** Articles - I am confident Engagement with Collo Blue - dance & confidence discussion
Session 8	30-Sep	**Summary + Vision Board workshop** Vision board co-creation with girls Giveaways based on attendance

Note: These are the weekly activities for the SKY club sessions.

Table A14: Weekly giveaways to SKY club participants

Session	Date	Giveaway
Session 1	3-Jun	Giveaways: 400 Mag 8 Wristbands for those who pledge 100 chocolate bars for participants (50 per group)
Session 2	10-Jun	Giveaways: 400 Mag 9 Wristbands for those who pledge 100 chocolate bars for participants (50 per group)
Session 3	26-Aug	Giveaways: 400 Mag 11 60 body lotions (30 per group)
Session 4	2-Sep	Giveaways: 400 pens 40 pencil pouches (20 per group) 20 T-shirts (10 per group)
Session 5	9-Sep	Giveaways: 400 button pins 30 pads (15 per group) 30 socks (15 per group)
Session 6	16-Sep	Giveaways: 400 PAA posters 60 socks (30 per group)
Session 7	23-Sep	Giveaways: 400 Mag 10 60 T-shirts (30 per group)
Session 8	30-Sep	Giveaways: 400 notebooks 400 Vision boards + stickers Reward for overall attendance

Note: These items are given to the participants at the end of the weekly SKY club sessions

Table A15: Giveaways to SKY club participants

Prize	Condition	Items
Grand Hamper	Attended all 8 sessions	T-shirt Pencil Pouch Fancy Wristband Chocolate Pads Body lotion
Runner up 1	Attended 5-7 sessions	Pencil Pouch Pads Chocolate
Runner up 2	Attended 1-4 sessions	Pencil Pouch Chocolate

Note: These items are given to the participants at the end of all the SKY club sessions

A.13 Qualitative interviews and FGDs

Qualitative interviews and focus group discussions (FGDs) with a selection of girls from our main study were conducted in April 2025. All interviews were set up and completed in collaboration with Innovations for Poverty Action (IPA) Kenya. Interviews and FGDs were led by one of the PIs and a female enumerator from IPA who also completed the consenting process with the parents and the girl. Questions were asked in English if the girl confirmed that she felt comfortable speaking English and in Swahili otherwise. Girls could answer in English or Swahili. All interviews and FGDs were audio-recorded conditional on parental and the girls' consent. IPA transcribed and anonymised all interviews and discussions, and translated Swahili versions to English. For this qualitative research, we received IRB approval from Strathmore University Institutional Ethics Review Committe (protocol *Social Norms and Interactions Among Teenage Girls*).

Sampling strategy. We selected 40 girls out of the 829 girls who participated in our endline survey. The selected girls' ages ranged from 12-19 at the time of their endline interview. 18 of the girls were part of our main study's control group and 22 were part of the treatment group. The plan was to do two FGDs with 10 girls from each group and hold individual interviews with 8 girls from the control and 12 girls from the treatment group. For the focus groups, we selected girls from the same neighbourhood to minimize their travel times to the FGD venue. The individual interviews were conducted at the girls' homes.

Individual interviews. We ended up conducting 21 individual interviews – 13 with girls from the treatment and 8 with girls from the control group. This was because one girl showed up late to her FGD and agreed to do an interview instead. Interviews lasted approximately 20-40 minutes and followed a guide of questions about girls' daily social interactions as well as their experiences with and concerns about executing assertive behavior. Girls in the treatment group were also asked about their SKY club experience.

Focus groups. In the control group FGD, 3 young women between the ages of 17-21 attended. In the FGD with the treatment group girls, 7 girls between the ages of 14-16 participated. FGDs

lasted about an hour and were held in an empty classroom at a local school in one of the four Nairobi neighbourhoods sampled from in our main study. Like the interviews, the FGDs followed a guide of discussion questions asking the girls to share experiences where they considered asserting themselves in social interactions and what the actual or anticipated responses to their assertions were. In the treatment group FGD, girls also shared their SKY club experiences.

B Identification

B.1 Covariate balance

	coefficient	std. error	RI p-value	param. p-value	N	mean control	std. dev. control	mean treatment	std. dev. treatment
Enrolled school	0.01	0.03	0.63	0.70	829	0.67	0.47	0.68	0.47
Day	-0.00	0.05	1.00	1.00	829	2.39	0.68	2.39	0.61
Age	-0.03	0.18	0.79	0.88	829	15.52	2.57	15.49	2.56
Access	-0.03	0.06	0.32	0.55	829	1.11	0.81	1.07	0.79
Media usage	-0.02	0.02	0.17	0.23	829	0.32	0.24	0.30	0.21
Stand up	0.01	0.05	0.70	0.82	829	0.53	0.76	0.54	0.77
Program	-0.02	0.05	0.70	0.72	829	1.29	0.75	1.27	0.76
Hang out	0.03	0.03	0.08	0.36	829	0.62	0.48	0.65	0.48
Tattoo stylish	-0.10	0.07	0.15	0.14	829	0.68	1.02	0.58	0.97
Piercing stylish	-0.11	0.08	0.13	0.15	829	0.88	1.16	0.77	1.06
Enjoy swimming	-0.09	0.07	0.15	0.16	829	2.48	0.86	2.39	0.93
Like Afrobeats	-0.06	0.04	0.06	0.07	829	0.48	0.50	0.42	0.49
Like Bongo	-0.02	0.04	0.63	0.64	829	0.45	0.50	0.43	0.50
Like Gospel	0.04	0.03	0.21	0.22	829	0.67	0.47	0.71	0.46
Share cutlery	0.02	0.04	0.33	0.63	829	0.52	0.50	0.54	0.50
Any SKY exposure	-0.02	0.04	0.25	0.54	829	0.49	0.50	0.46	0.50
Assets	0.00	0.02	0.90	0.90	829	0.56	0.29	0.56	0.28
Own house	0.05	0.03	0.14	0.15	829	0.27	0.45	0.32	0.47
People	-0.06	0.11	0.44	0.59	829	2.65	1.55	2.59	1.60
Block house	-0.02	0.02	0.35	0.39	829	0.89	0.31	0.87	0.34
Household size	-0.17	0.13	0.15	0.20	829	4.86	1.85	4.69	1.78
Kikuyu	-0.01	0.04	0.76	0.76	829	0.41	0.49	0.40	0.49
Luo	0.00	0.03	0.91	0.91	829	0.17	0.38	0.18	0.38
Swahili	-0.04	0.03	0.19	0.20	829	0.82	0.38	0.79	0.41

Table A16: Balance of baseline covariates across treatment and control group interviewed at endline

The outcomes listed in the first column are covariates measured for each respondent during the baseline survey. The first two columns are the coefficient and clustered standard error from a weighted regression of each outcome on the treatment indicator. The "RI p-value" displays the p-value from a test of the sharp null hypothesis of no treatment effect for any unit. The "param. p-value" displays the parametric p-value from the same regression pertaining to the weak null hypothesis of no average treatment effect. "N" is the number of observations. "Mean control" is the covariate mean among the girls in the control group who got interviewed at endline, and "st. dev. control" the weighted standard deviation. "Mean treatment" is the weighted covariate mean among the girls in the treatment group who got interviewed at endline, and "st. dev. treatment" the weighted standard deviation.

	coefficient	std. error	RI p-value	param. p-value	N	mean control	std. dev. control	mean treatment	std. dev. treatment
Enrolled school	-0.00	0.03	0.80	0.90	996	0.63	0.48	0.63	0.48
Day	0.00	0.04	1.00	0.99	996	2.40	0.66	2.40	0.61
Age	0.06	0.17	0.30	0.72	996	15.67	2.61	15.73	2.56
Access	-0.06	0.05	0.00	0.24	996	1.17	0.80	1.11	0.80
Media usage	-0.03	0.01	0.00	0.08	996	0.32	0.24	0.30	0.22
Stand up	-0.02	0.05	0.30	0.66	996	0.54	0.77	0.52	0.77
Program	0.01	0.05	0.70	0.89	996	1.27	0.75	1.28	0.75
Hang out	0.00	0.03	0.70	0.90	996	0.63	0.48	0.63	0.48
Tattoo stylish	-0.07	0.07	0.20	0.31	996	0.69	1.03	0.62	1.00
Piercing stylish	-0.11	0.07	0.00	0.13	996	0.90	1.16	0.79	1.08
Enjoy swimming	-0.15	0.06	0.00	0.01	996	2.51	0.85	2.36	0.96
Like Afrobeats	-0.04	0.03	0.20	0.18	996	0.47	0.50	0.42	0.49
Like Bongo	-0.02	0.03	0.70	0.61	996	0.46	0.50	0.44	0.50
Like Gospel	0.05	0.03	0.10	0.13	996	0.67	0.47	0.72	0.45
Share cutlery	0.02	0.03	0.00	0.49	996	0.52	0.50	0.55	0.50
Any SKY exposure	-0.02	0.03	0.00	0.47	996	0.47	0.50	0.44	0.50
Assets	0.01	0.02	0.50	0.56	996	0.56	0.29	0.57	0.28
Own house	0.03	0.03	0.10	0.31	996	0.27	0.44	0.30	0.46
People	-0.08	0.10	0.30	0.45	996	2.60	1.53	2.53	1.54
Block house	-0.02	0.02	0.40	0.37	996	0.89	0.31	0.87	0.33
Household size	-0.11	0.12	0.20	0.33	996	4.77	1.83	4.66	1.80
Kikuyu	0.00	0.03	1.00	0.97	996	0.39	0.49	0.39	0.49
Luo	-0.00	0.02	0.90	0.94	996	0.17	0.38	0.17	0.38
Swahili	-0.05	0.03	0.10	0.08	996	0.82	0.38	0.78	0.42

Table A17: Balance of baseline covariates across treatment and control group interviewed at baseline

The outcomes listed in the first column are covariates measured for each respondent during the baseline survey. The first two columns are the coefficient and clustered standard error from a weighted regression of each outcome on the treatment indicator. The "RI p-value" displays the p-value from a test of the sharp null hypothesis of no treatment effect for any unit. The "param. p-value" displays the parametric p-value from the same regression pertaining to the weak null hypothesis of no average treatment effect. "N" is the number of observations. "Mean control" is the weighted covariate mean among the girls in the control group who got interviewed at baseline, and "st. dev. control" the weighted standard deviation. "Mean treatment" is the weighted covariate mean among the girls in the treatment group who got interviewed at baseline, and "st. dev. treatment" the weighted standard deviation.

B.2 Attrition

Outcome	Sample	Control	Treatment	p-value	N
Attrited	Main girls	16.9%	16.2%	0.759	996
Attrited	Spillover respondents listed at baseline	39%	33.1%	0.250	501
Any new respondent	All study households	12.7%	16.5%	0.096	996

Table A18: Attrition rates across experimental conditions

The outcome in the first two rows is an indicator for whether a respondent attrited. The outcome in the third row is an indicator for whether a new respondent was interviewed in a study household. The columns labeled "Control" and "Treatment" show the weighted averages of outcomes in, respectively, the treatment and control groups. Weights account for varying treatment assignment probabilities across blocks as well as duplicate observations. Two-tailed p-values stem from a regression of the outcome on the treatment assignment indicator and are calculated using randomization inference by permuting treatment assignment 8,000 times to generate the distribution of the regression coefficient under the sharp null hypothesis of no effect of treatment on attrition or the presence of new respondents for any unit.

Outcome	Sample	p-value	N
Attrited	Main girls	0.084	996
Attrited	Spillover respondents listed at baseline	0.756	501
Any new respondent	All study households	0.957	996

Table A19: F-test of treatment-by-covariate interactions in models of attrition

The outcome in the first two rows is an indicator for whether a respondent attrited. The outcome in the third row is an indicator for whether a new respondent was interviewed in a study household. p-values come from F-tests that compare the following two models. The full model regresses the outcome on an indicator for treatment assignment and all treatment-by-covariate interactions using eight pre-registered baseline covariates. The nested model restricts all interaction terms to zero. p-values have been calculated using randomization inference by permuting treatment assignment 8,000 times.

C Additional Analyses

C.1 Distribution of prisoner's dilemma outcomes in control group

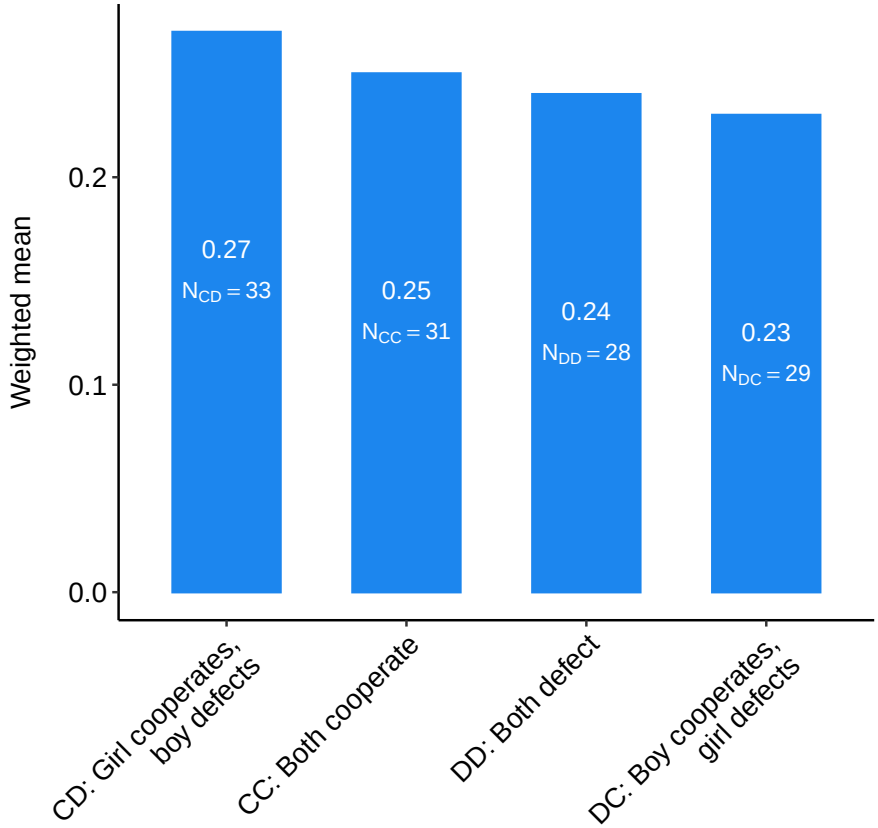

Figure A8: Distribution of prisoner's dilemma outcomes in control group

Bars show weighted averages of outcome occurrence in the prisoner's dilemma game (PD) between main girls and brothers from control group households. The first bar displays the share of PD outcomes in which the main girl chose the *Split* card – equivalent to cooperating – and her brother chose the *Keep* card – equivalent to defecting – in the game. The second bar shows the share of PD outcomes in which both the main girl and her brother chose the *Split* card, and the third column the share of outcomes in which both chose the *Keep* card. The fourth bar indicates the share of PD outcomes in which the main girl selected the *Keep* card and her brother selected the *Split* card. The number of occurrences of each outcome is displayed inside the bars. Black lines indicate parametric 95% confidence intervals. Weights correspond to the inverse of the probability of being assigned to control. Weighting ensures means are unbiased estimates of average control potential outcomes across the experimental sample. See appendix section D.1 for question wording.

C.2 Comparing high-cost interactions of main girl with brothers vs. sisters

	Quarrel w/ Brother	Quarrel w/ Brother	Quarrel w/ Sister	Quarrel w/ Sister
SKY	0.186	0.173*	0.142	0.083
	(0.115)	(0.107)	(0.110)	(0.105)
Sample	Main w/ interv. brother	Main w/ brother	Main w/ interv. sister	Main w/ sister
Outcome range	$0-2$	$0-2$	$0-2$	$0-2$
Control Mean	0.508	0.497	0.582	0.616
Control SD	0.765	0.756	0.746	0.748
Hypothesis	two	two	two	two
Num. obs.	205	232	198	218

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A20: Effects of SKY on girls' proclivity to report having recently quarreled with their brothers/sisters
Appendix section A.2 provides details on model specification and testing and section D.8 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect of assignment to the SKY intervention for any unit and the row "Hypothesis" displays the direction of this test.

	Defect vs. Brother	Defect vs. Sister
SKY	−0.019	0.030
	(0.068)	(0.070)
Outcome range	$0-1$	$0-1$
Control Mean	0.479	0.465
Control SD	0.480	0.490
Hypothesis	two	two
Num. obs.	207	198

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A21: Effects of SKY on girls' proclivity to defect against their brothers/sisters in prisoner's dilemma
Appendix section A.2 provides details on model specification and testing and section D.9 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect of assignment to the SKY intervention for any unit and the row "Hypothesis" displays the direction of this test. The outcome *Defect vs. Brother* is the average defection rate against all brothers of a main respondent, and *Defect vs. Sister* the average defection rate against all sisters.

C.3 Comparing "Assert Brother" measure between girls with and without siblings

	Assert Brother	Assert Brother	Assert Brother	Assert Brother
SKY	0.154*	0.116	0.081	0.108*
	(0.092)	(0.087)	(0.083)	(0.088)
Sample	Main w/ interv. sibling	Main w/ sibling	Main w/o interv. sibling	Main w/o sibling
Outcome range	$0-2$	$0-2$	$0-2$	$0-2$
Control Mean	0.523	0.538	0.572	0.565
Control SD	0.831	0.838	0.829	0.824
Hypothesis	upr	upr	upr	upr
Num. obs.	355	394	474	435

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A22: Effect of SKY on girls' proclivity to assert themselves towards their brother

Appendix section A.2 provides details on model specification and testing and section **??** on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect for any unit and the row "Hypothesis" the direction of this test. The analysis in column 1 is based on all main respondents with a sibling, column 2 on all main respondents with an older sibling, and column 3 on all main respondents without a sibling.

	Assert Brother	Assert Brother	Assert Brother	Assert Brother
SKY	0.133	0.105	0.110*	0.120*
	(0.120)	(0.113)	(0.071)	(0.073)
Sample	Main w/ interv. brother	Main w/ brother	Main w/o interv. brother	Main w/o brother
Outcome range	$0-2$	$0-2$	$0-2$	$0-2$
Control Mean	0.495	0.503	0.57	0.571
Control SD	0.799	0.809	0.838	0.837
Hypothesis	upr	upr	upr	upr
Num. obs.	207	233	622	596

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A23: Effect of SKY on girls' proclivity to assert themselves towards their brother

Appendix section A.2 provides details on model specification and testing and section D.11 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect for any unit and the row "Hypothesis" the direction of this test. The analysis in column 1 is based on all main respondents with a brother, column 2 on all main respondents with an older brother, and column 3 on all main respondents without a brother.

	Assert Brother	Assert Brother	Assert Brother	Assert Brother
SKY	0.187*	0.162	0.084	0.089*
	(0.124)	(0.116)	(0.070)	(0.072)
Sample	Main w/ interv. sister	Main w/ sister	Main w/o interv. sister	Main w/o sister
Outcome range	$0-2$	$0-2$	$0-2$	$0-2$
Control Mean	0.538	0.532	0.557	0.56
Control SD	0.848	0.843	0.825	0.826
Hypothesis	upr	upr	upr	upr
Num. obs.	198	218	631	611

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A24: Effect of SKY on girls' proclivity to assert themselves towards their brother

Appendix section A.2 provides details on model specification and testing and section D.11 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect for any unit and the row "Hypothesis" the direction of this test. The analysis in column 1 is based on all main respondents with a sister, column 2 on all main respondents with an older sister, and column 3 on all main respondents without a sister.

C.4 Within-household negotiation outcomes for main girl with and without siblings

	Chores	Chores	Chores	Fun Activities	Fun Activities	Fun Activities
SKY	−0.153*	0.040	0.039	0.068	0.011	0.006
	(0.091)	(0.073)	(0.072)	(0.121)	(0.113)	(0.114)
Has Sibling			0.015			0.141
			(0.075)			(0.109)
SKY × Has Sibling			−0.189			0.063
			(0.117)			(0.165)
Sample	Main w/ interv. sibling	Main w/o interv. sibling	All Main	Main w/ interv. sibling	Main w/o interv. sibling	All Main
Outcome range	$0-4$	$0-4$	$0-4$	$0-6$	$0-6$	$0-6$
Control Mean	3.435	3.384	3.404	2.404	2.307	2.345
Control SD	0.770	0.764	0.767	1.176	1.238	1.215
Hyp. SKY	two	two	two	upr	upr	upr
Hyp. Has Sib.			two			two
Hyp. Diff.			two			two
Num. obs.	355	474	829	355	474	829

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A25: Effects of SKY on number of chores and fun activities done by girls with and without interviewed siblings

Appendix section A.2 provides details on model specification and testing and section D.12 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars for the coefficients of "SKY" indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect of assignment to the SKY intervention for any unit and the row "Hyp. SKY" the direction of this test. Stars for the coefficients of "Has Sibling" denote the parametric p-value from a test of the weak null hypothesis of no difference in the outcome between those who have and have no siblings and "Hyp. Has Sib." the direction of this test. Stars for the coefficients of "SKY x Has Sibling" indicate the parametric p-value from a test of the weak null hypothesis of no difference between the effect of the treatment among those who have and have no siblings and "Hyp. Diff" the direction of this test. Analyses in columns 1 and 4 are based on all main respondents who have at least one sibling who was interviewed at endline and analyses in column 2 and 5 on all main respondents who have no siblings who were interviewed at endline.

	Chores	Chores	Chores	Fun Activities	Fun Activities	Fun Activities
SKY	−0.110	0.005	0.012	0.070	0.016	0.009
	(0.085)	(0.077)	(0.077)	(0.115)	(0.120)	(0.120)
Has Sibling			−0.019			0.097
			(0.073)			(0.107)
SKY × Has Sibling			−0.117			0.059
			(0.115)			(0.166)
Sample	Main w/ sibling	Main w/o sibling	All Main	Main w/ sibling	Main w/o sibling	All Main
Outcome range	0 − 4	0 − 4	0 − 4	0 − 6	0 − 6	0 − 6
Control Mean	3.416	3.395	3.404	2.373	2.323	2.345
Control SD	0.785	0.752	0.767	1.193	1.233	1.215
Hyp. SKY	two	two	two	upr	upr	upr
Hyp. Has Sib.			two			two
Hyp. Diff.			two			two
Num. obs.	394	435	829	394	435	829

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A26: Effects of SKY on number of chores and fun activities done by girls with and without siblings

Appendix section A.2 provides details on model specification and testing and section D.12 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars for the coefficients of "SKY" indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect of assignment to the SKY intervention for any unit and the row "Hyp. SKY" the direction of this test. Stars for the coefficients of "Has Sibling" denote the parametric p-value from a test of the weak null hypothesis of no difference in the outcome between those who have and have no siblings and "Hyp. Has Sib." the direction of this test. Stars for the coefficients of "SKY x Has Sibling" indicate the parametric p-value from a test of the weak null hypothesis of no difference between the effect of the treatment among those who have and have no siblings and "Hyp. Diff" the direction of this test. Analyses in columns 1 and 4 are based on all main respondents who have at least one sibling and analyses in column 2 and 5 on all main respondents who have no siblings.

C.5 Brothers' and sisters' interactions with main girl

	Perceptions	Behavior			
	Assert Brother	Quarrel w/ Main Girl	Defect vs. Main Girl	Chores	Fun Activities
SKY	0.022	−0.049	−0.054	−0.018	−0.029
	(0.062)	(0.097)	(0.064)	(0.143)	(0.182)
Outcome range	0 − 1	0 − 2	0 − 1	0 − 4	0 − 6
Control Mean	0.325	0.580	0.514	2.829	2.976
Control SD	0.470	0.730	0.501	1.058	1.333
Hypothesis	upr	two	two	two	two
Num. obs.	255	254	255	255	255
N Clusters	207	206	207	207	207

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A27: Effect of SKY on brothers' perceptions and behavior related to main girl

Appendix section A.2 provides details on model specification and testing and section D.13 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect for any unit and the row "Hypothesis" the direction of this test. Analyses are based on all brothers in the households of the main respondents and are clustered at the level of the main respondent.

	Perceptions	Behavior			
	Assert Brother	Quarrel w/ Main Girl	Defect vs. Main Girl	Chores	Fun Activities
SKY	0.050	0.195*	0.051	0.049	−0.080
	(0.064)	(0.116)	(0.065)	(0.119)	(0.156)
Outcome range	0 – 1	0 – 2	0 – 1	0 – 4	0 – 5
Control Mean	0.347	0.705	0.433	3.311	2.714
Control SD	0.477	0.869	0.497	0.945	1.157
Hypothesis	upr	two	two	two	two
Num. obs.	235	235	235	235	235
N Clusters	198	198	198	198	198

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A28: Effect of SKY on sisters' perceptions and behavior related to main girl

Appendix section A.2 provides details on model specification and testing and section D.13 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect for any unit and the row "Hypothesis" the direction of this test. Analyses are based on all sisters in the households of the main respondents and are clustered at the level of the main respondent.

C.6 Brothers' and sisters' perceptions of main girl's assertiveness

	Perceptions				Knowledge	
	Assert Brother	Assert Brother	Empowered	Empowered	Correct Guesses abt. Main Girl	Correct Guesses abt. Main Girl
SKY	0.022	0.050	0.152*	−0.081	−0.182	−0.005
	(0.062)	(0.064)	(0.103)	(0.101)	(0.166)	(0.149)
Sample	Brothers	Sisters	Brothers	Sisters	Brothers	Sisters
Outcome range	0 – 1	0 – 1	0 – 3	0 – 3	0 – 6	0 – 5
Control Mean	0.325	0.347	0.792	0.909	2.563	2.789
Control SD	0.470	0.477	0.752	0.796	1.208	1.101
Hypothesis	upr	upr	upr	upr	upr	upr
Num. obs.	255	235	255	235	255	235
N Clusters	207	198	207	198	207	198

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A29: Effect of SKY on siblings' perceptions of main girl's assertiveness

Appendix section A.2 provides details on model specification and testing and section D.14 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect for any unit and the row "Hypothesis" the direction of this test. Analyses in columns 1 and 4 are based on all siblings in the households of the main respondents, analyses in columns 2 and 5 on all sisters, and columns 3 and 6 on all brothers. All analyses are clustered at the level of the main respondent.

C.7 Behavioral measures with enumerator fixed effects

	Assert Enumerator	Assert Enumerator
SKY	0.036*	0.018
	(0.025)	(0.050)
Sample	Main	Sisters
Outcome range	$0-1$	$0-1$
Control Mean	0.472	0.526
Control SD	0.362	0.372
Hypothesis	upr	upr
Num. obs.	829	235
N Clusters		198

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A30: Effects of SKY on girls' confidence in their interaction with enumerators including enumerator fixed effects

Appendix section A.2 provides details on model specification and testing and section D.15 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect for any unit and the row "Hypothesis" the direction of this test. Analysis in column 1 is based on all main respondents and analysis in column 2 on all sisters. All analyses include enumerator fixed effects and the analysis in column 2 is clustered at the level of the main respondent.

C.8 Alternative mechanisms

	Gender Salience	Peer Pressure
SKY	-0.015	-0.109^{**}
	(0.021)	(0.058)
Outcome range	$0-1$	$0-2$
Control Mean	0.506	0.687
Control SD	0.287	0.832
Hypothesis	upr	lwr
Num. obs.	828	829

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A31: Effect of SKY on gender salience and perceived pressure to be in a relationship

Appendix section A.2 provides details on model specification and testing and section D.16 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect for any unit and the row "Hypothesis" the direction of this test.

C.9 Dictator game with prompt for anonymous girl not being a SKY girl

	Kept Pens Girl	Kept Pens Girl
SKY	0.138*	0.105
	(0.083)	(0.094)
Not in Study Prompt		−0.124
		(0.127)
SKY × Not in Study Prompt		0.134
		(0.201)
Outcome range	0 − 5	0 − 5
Control Mean	2.760	2.760
Control SD	1.127	1.127
Hyp. SKY	two	two
Hyp. Prompt		two
Hyp. Diff.		upr
Num. obs.	829	829

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A32: Interaction effect of SKY and randomized prompt informing that anonymous girl is new to the study on number of pens kept by girl in dictator game

Appendix section A.2 provides details on model specification and testing and section D.17 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars for the coefficients of "SKY" indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect of assignment to the SKY intervention for any unit and the row "Hyp. SKY" the direction of this test. Stars for the coefficients of "Not in Study Prompt" denote the RI p-value from a test of the sharp null hypothesis of no (positive) treatment effect of the Not-in-study prompt for any unit and the row "Hyp. Prompt" the direction of this test. Stars for the coefficients of "SKY × Not in Study Prompt" indicate the parametric p-value from a test of the weak null hypothesis of no difference between the effect of the treatment among those who were and were not assigned to the Not-in-study prompt and "Hyp. Diff." the direction of this test.

C.10 Analyses without covariates

	N Attended	Exposed TV or Magazine	Knowledge Club	SKY Exposure
SKY	5.835***	0.283***	4.425***	1.973***
	(0.154)	(0.024)	(0.109)	(0.119)
Outcome range	$0-13$	$0-1$	$0-7$	$0-9$
Control Mean	0.193	0.682	0.259	1.803
Control SD	0.959	0.466	1.043	1.530
Hypothesis	upr	upr	upr	upr
Num. obs.	829	829	829	829

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A33: Effects of SKY on girls' exposure to SKY

Appendix section A.2 provides details on model specification and testing and section D.18 on outcome question wording and coding. Models include no covariates. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect for any unit and the row "Hypothesis" the direction of this test. The p-value from a randomization inference-based test of the global null hypothesis that all null hypotheses for the tests shown in this table are correct is $p = 0.000$.

	Assert (Boy)friend	Assert Enumerator	Kept Pens Girl	Kept Pens Boy
SKY	0.046***	0.039*	0.136*	0.152**
	(0.016)	(0.026)	(0.083)	(0.082)
Outcome range	$0-1$	$0-1$	$0-5$	$0-5$
Control Mean	0.644	0.472	2.760	2.843
Control SD	0.230	0.362	1.127	1.142
Hypothesis	upr	upr	two	upr
Num. obs.	825	829	829	829

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A34: Effects of SKY on girls' proclivity to behave empowered in low-cost interactions

Appendix section A.2 provides details on model specification and testing and section D.18 on outcome question wording and coding. Models include no covariates. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect of assignment to the SKY intervention for any unit and the row "Hypothesis" displays the direction of this test. The p-value from a randomization inference-based test of the global null hypothesis that all null hypotheses pertaining to the tests shown in columns 1-4 of this table are correct is $p = 0.001$.

	Assert Brother	Quarrel w/ Sibling	Defect vs. Sibling	Chores	Fun Activities
SKY	0.128**	0.157*	0.013	−0.053	0.052
	(0.062)	(0.082)	(0.050)	(0.059)	(0.088)
Outcome range	0 – 2	0 – 2	0 – 1	0 – 4	0 – 6
Control Mean	0.553	0.566	0.467	3.404	2.345
Control SD	0.830	0.740	0.463	0.767	1.215
Hypothesis	upr	two	two	two	upr
Num. obs.	829	355	355	829	829

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A35: Effects of SKY on girls' proclivity to behave empowered in high-cost interactions

Appendix section A.2 provides details on model specification and testing and section D.18 on outcome question wording and coding. Models include no covariates. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect of assignment to the SKY intervention for any unit and the row "Hypothesis" displays the direction of this test. The p-value from a randomization inference-based test of the global null hypothesis that all null hypotheses pertaining to the tests shown in columns 1-5 of this table are correct is $p = 0.043$.

	Perceptions		Knowledge	Behavior			
	Assert Brother	Empowered	Correct Guesses abt. Main Girl	Quarrel w/ Main Girl	Defect vs. Main Girl	Chores	Fun Activities
SKY	0.028	0.049	−0.069	0.078	−0.000	0.038	−0.057
	(0.045)	(0.072)	(0.115)	(0.077)	(0.047)	(0.091)	(0.112)
Outcome range	0 – 1	0 – 3	0 – 6	0 – 2	0 – 1	0 – 4	0 – 6
Control Mean	0.336	0.848	2.671	0.640	0.475	3.059	2.851
Control SD	0.473	0.775	1.162	0.801	0.500	1.033	1.258
Hypothesis	upr	upr	upr	two	two	two	two
Num. obs.	490	490	490	489	490	490	490
N Clusters	355	355	355	354	355	355	355

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A36: Effect of SKY on siblings' interactions with main girl

Appendix section A.2 provides details on model specification and testing and section D.18 on outcome question wording and coding. Models include no covariates. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect for any unit and the row "Hypothesis" the direction of this test. The p-value from a randomization inference-based test of the global null hypothesis that all null hypotheses pertaining to the tests shown in columns 1-7 of this table are correct is $p = 0.9$.

	Social Connections			Norm Perceptions	
	Girl Friends	New Friend	Turn To	Descriptive	Prescriptive
SKY	−0.110	0.104***	0.102*	0.086***	0.093***
	(0.140)	(0.034)	(0.069)	(0.025)	(0.025)
Outcome range	0 – 10	0 – 1	0 – 3	0 – 1	0 – 1
Control Mean	2.877	0.582	2.186	0.559	0.506
Control SD	2.073	0.494	0.989	0.348	0.346
Hypothesis	upr	upr	upr	upr	upr
Num. obs.	829	829	828	827	827

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

Table A37: Effect of SKY on network ties and perceptions of social norm

Appendix section A.2 provides details on model specification and testing and section D.18 on outcome question wording and coding. Models include no covariates. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect for any unit and the row "Hypothesis" the direction of this test. The p-value from a randomization inference-based test of the global null hypothesis that all null hypotheses regarding outcomes in the *Social Connections* (*Norm Perceptions*) category are true is $p = 0.004$ ($p = 0.000$).

	Attitudes		Beliefs About Self	
	Should Assert	Healthy Rel.	Confident	Empowered
SKY	0.054***	−0.027*	0.002	0.074
	(0.021)	(0.015)	(0.012)	(0.053)
Outcome range	0 – 1	0 – 1	0 – 1	0 – 3
Control Mean	0.688	0.514	0.840	0.954
Control SD	0.296	0.219	0.170	0.743
Hypothesis	upr	upr	upr	upr
Num. obs.	829	817	829	829

***$p < 0.01$; **$p < 0.05$; *$p < 0.1$

Table A38: Effect of SKY on assertive attitudes and self-confidence

Appendix section A.2 provides details on model specification and testing and section D.18 on outcome question wording and coding. Models include no covariates. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect for any unit and the row "Hypothesis" the direction of this test. The p-value from a randomization inference-based test of the global null hypothesis that all null hypotheses regarding outcomes in the *Attitudes* (*Beliefs About Self*) category are true is $p = 0.02$ ($p = 0.15$).

	Gender Salience	Peer Pressure
SKY	−0.018	−0.110*
	(0.021)	(0.058)
Outcome range	0 − 1	0 − 2
Control Mean	0.506	0.687
Control SD	0.287	0.832
Hypothesis	upr	lwr
Num. obs.	828	829

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A39: Effect of SKY on gender salience and perceived pressure to be in a relationship

Appendix section A.2 provides details on model specification and testing and section D.18 on outcome question wording and coding. Models include no covariates. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect for any unit and the row "Hypothesis" the direction of this test. We do not report a joint RI p-value for this table, because the outcomes in columns 1-2 were not pre-registered as sub-hypotheses to one global hypothesis.

C.11 Disaggregating indices

	Assert (Boy)friend	Voice Feeling	Enjoy Movie	Intervene Tickling
SKY	0.048***	0.097***	0.045	0.057*
	(0.017)	(0.034)	(0.087)	(0.040)
Outcome range	$0-1$	$0-1$	$0-3$	$0-2$
Control Mean	0.553	0.513	1.892	1.583
Control SD	0.202	0.500	1.235	0.582
Hypothesis	upr	upr	upr	upr
Num. obs.	825	829	825	829

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A40: Effects of SKY on girls' proclivity to assert themselves in low-cost relationships

Appendix section A.2 provides details on model specification and testing and section D.18 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect of assignment to the SKY intervention for any unit and the row "Hypothesis" displays the direction of this test.

	Assert Enumerator	Correct Date	Complain Gift
SKY	0.045**	0.044	0.045*
	(0.026)	(0.036)	(0.035)
Outcome range	$0-1$	$0-1$	$0-1$
Control Mean	0.472	0.409	0.535
Control SD	0.362	0.492	0.499
Hypothesis	upr	upr	upr
Num. obs.	829	829	829

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A41: Effects of SKY on girls' proclivity to assert themselves toward the enumerator

Appendix section A.2 provides details on model specification and testing and section D.18 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect of assignment to the SKY intervention for any unit and the row "Hypothesis" displays the direction of this test.

	Correct Guesses abt. Main Girl	Ear Piercings Correct	Activity Correct	TV Show Correct	Sport Correct	Dream Job Correct	Stand Up Correct
SKY	−0.083	0.008	0.008	−0.015	−0.011	−0.006	−0.067
	(0.114)	(0.046)	(0.047)	(0.051)	(0.045)	(0.051)	(0.049)
Outcome range	$0-6$	$0-1$	$0-1$	$0-1$	$0-1$	$0-1$	$0-1$
Control Mean	2.671	0.565	0.364	0.416	0.308	0.430	0.589
Control SD	1.162	0.496	0.482	0.493	0.462	0.496	0.492
Hypothesis	upr	upr	upr	upr	upr	upr	upr
Num. obs.	490	490	490	490	490	490	490
N Clusters	355	355	355	355	355	355	355

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A42: Effects of SKY on the number of correct guesses siblings make about the main girl

Appendix section A.2 provides details on model specification and testing and section D.18 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect of assignment to the SKY intervention for any unit and the row "Hypothesis" displays the direction of this test.

	Norm Perceptions		
	Descriptive	Not Come Over	Intervene Tickling Other Girls
SKY	0.086***	0.119***	0.097
	(0.025)	(0.035)	(0.058)
Outcome range	$0-1$	$0-1$	$0-2$
Control Mean	0.559	0.508	1.223
Control SD	0.348	0.500	0.862
Hypothesis	upr	upr	upr
Num. obs.	827	828	828

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A43: Effects of SKY on girls' perceptions of descriptive norms about assertive behavior among other girls

Appendix section A.2 provides details on model specification and testing and section D.18 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect of assignment to the SKY intervention for any unit and the row "Hypothesis" displays the direction of this test.

	Norm Perceptions		
	Prescriptive	Hang Out Other Girls	Take Time Other Girls
SKY	0.089^{***}	0.070^{***}	0.105^{***}
	(0.025)	(0.035)	(0.033)
Outcome range	$0-1$	$0-1$	$0-1$
Control Mean	0.506	0.415	0.599
Control SD	0.346	0.493	0.490
Hypothesis	upr	upr	upr
Num. obs.	827	829	827

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A44: Effects of SKY on girls' perceptions of prescriptive norms about assertive behavior among other girls

Appendix section A.2 provides details on model specification and testing and section D.18 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect of assignment to the SKY intervention for any unit and the row "Hypothesis" displays the direction of this test.

	Should Assert	Hang Out	Take Time
SKY	0.048^{***}	0.073^{***}	0.024^{***}
	(0.020)	(0.035)	(0.019)
Outcome range	$0-1$	$0-1$	$0-1$
Control Mean	0.688	0.466	0.910
Control SD	0.296	0.499	0.286
Hypothesis	upr	upr	upr
Num. obs.	829	829	829

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A45: Effects of SKY on girls' attitudes about assertive behavior in low-cost relationships

Appendix section A.2 provides details on model specification and testing and section D.18 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect of assignment to the SKY intervention for any unit and the row "Hypothesis" displays the direction of this test.

	Healthy Rel.	Around All Time	Encourage Friends	Wear Heels	Expensive Gifts	Movie With Friends	Asking Love	Texting All Time
SKY	−0.030	0.001	0.010	−0.010	−0.068	−0.037	−0.039	−0.075
	(0.015)	(0.036)	(0.028)	(0.036)	(0.030)	(0.033)	(0.035)	(0.035)
Outcome range	0 − 1	0 − 1	0 − 1	0 − 1	0 − 1	0 − 1	0 − 1	0 − 1
Control Mean	0.514	0.514	0.792	0.491	0.282	0.350	0.625	0.552
Control SD	0.219	0.500	0.406	0.500	0.450	0.477	0.484	0.498
Hypothesis	upr	upr	upr	upr	upr	upr	upr	upr
Num. obs.	817	827	827	826	826	827	828	829

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A46: Effects of SKY on girls' attitudes about healthy relationship behaviors

Appendix section A.2 provides details on model specification and testing and section D.18 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect of assignment to the SKY intervention for any unit and the row "Hypothesis" displays the direction of this test.

	Confident	Good Qualities	Deal Challenges
SKY	0.003	0.101**	−0.085
	(0.012)	(0.052)	(0.050)
Outcome range	0 − 1	0 − 3	0 − 3
Control Mean	0.840	2.422	2.620
Control SD	0.170	0.750	0.662
Hypothesis	upr	upr	upr
Num. obs.	829	829	829

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A47: Effects of SKY on girls' belief in their own qualities and abilities

Appendix section A.2 provides details on model specification and testing and section D.18 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect of assignment to the SKY intervention for any unit and the row "Hypothesis" displays the direction of this test.

	Gender Salience	Girl Identity	Girl Identity Important
SKY	−0.015	−0.010	−0.038
	(0.021)	(0.034)	(0.043)
Outcome range	0 – 1	0 – 1	0 – 2
Control Mean	0.506	0.339	1.344
Control SD	0.287	0.474	0.584
Hypothesis	upr	upr	upr
Num. obs.	828	829	828

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A48: Effects of SKY on girls' attitudes about healthy relationship behaviors

Appendix section A.2 provides details on model specification and testing and section D.18 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect of assignment to the SKY intervention for any unit and the row "Hypothesis" displays the direction of this test.

C.12 Effects on sisters

	Attendance	Awareness of SKY Club		Exposure to SKY Content	
	N Attended	Main Respondent Attended	SKY is for Girls	Exposed TV or Magazine	SKY Exposure
SKY	0.137	0.795***	0.102	0.230***	0.298*
	(0.109)	(0.041)	(0.128)	(0.059)	(0.232)
Outcome range	0 – 8	0 – 1	0 – 6	0 – 1	0 – 8
Control Mean	0.087	0.133	0.197	0.555	1.721
Control SD	0.422	0.341	0.834	0.498	1.806
Hypothesis	two	upr	upr	upr	upr
Num. obs.	235	235	235	235	235
N Clusters	198	198	198	198	198

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A49: Spillover effects of SKY on sisters' exposure to SKY

Appendix section A.2 provides details on model specification and testing and section D.19 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars indicate significance based on the RI *p*-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect for any unit and the row "Hypothesis" the direction of this test. The *p*-value from a randomization inference-based test of the global null hypothesis that all null hypotheses regarding outcomes in the *Awareness of SKY Club* (*Exposure to SKY Content*) category are true is $p = 0.000$ ($p = 0.000$). We do not report such a *p*-value for *Attendance*, because only one outcome falls into this category.

	Assert (Boy)friend	Assert Enumerator	Kept Pens Girl	Kept Pens Boy
SKY	−0.032	0.063*	−0.194	−0.050
	(0.031)	(0.047)	(0.157)	(0.174)
Outcome range	0 – 1	0 – 1	0 – 5	0 – 5
Control Mean	0.682	0.526	3.482	3.735
Control SD	0.222	0.372	1.216	1.135
Hypothesis	upr	upr	lwr	two
Num. obs.	232	235	235	235
N Clusters	196	198	198	198

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A50: Spillover effects of SKY on sisters' proclivity to behave empowered in low-cost interactions

Appendix section A.2 provides details on model specification and testing and section D.20 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars indicate significance based on the RI *p*-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect for any unit and the row "Hypothesis" the direction of this test. The *p*-value from a randomization inference-based test of the global null hypothesis that all null hypotheses pertaining to the tests shown in columns 1-4 of this table are correct is $p = 0.67$.

	Assert Brother	Quarrel w/ Main Girl	Defect vs. Main Girl	Chores	Fun Activities
SKY	0.068	0.195*	0.051	0.049	−0.080
	(0.061)	(0.116)	(0.066)	(0.122)	(0.160)
Outcome range	0 – 1	0 – 2	0 – 1	0 – 4	0 – 5
Control Mean	0.302	0.705	0.433	3.311	2.714
Control SD	0.460	0.869	0.497	0.945	1.157
Hypothesis	upr	two	two	two	upr
Num. obs.	235	235	235	235	235

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A51: Spillover effects of SKY on sisters' proclivity to behave empowered in high-cost interactions

Appendix section A.2 provides details on model specification and testing and section D.20 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect for any unit and the row "Hypothesis" the direction of this test. The p-value from a randomization inference-based test of the global null hypothesis that all null hypotheses pertaining to the tests shown in columns 1-5 of this table are correct is $p = 0.29$.

	Social Connections			Norm Perceptions	
	Girl Friends	New Friend	Turn To	Descriptive	Prescriptive
SKY	−0.473	−0.000	0.082	−0.005	−0.034
	(0.270)	(0.065)	(0.137)	(0.047)	(0.049)
Outcome range	0 – 10	0 – 1	0 – 3	0 – 1	0 – 1
Control Mean	2.988	0.569	2.152	0.589	0.560
Control SD	2.012	0.496	1.068	0.351	0.384
Hypothesis	upr	upr	upr	upr	upr
Num. obs.	235	235	235	235	234
N Clusters	198	198	198	198	197

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A52: Spillover effects of SKY on network ties and perceptions of social norm among sisters

Appendix section A.2 provides details on model specification and testing and section D.20 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect for any unit and the row "Hypothesis" the direction of this test. The p-value from a randomization inference-based test of the global null hypothesis that all null hypotheses regarding outcomes in the *Social Connections* (*Norm Perceptions*) category are true is $p = 0.63$ ($p = 0.27$).

	Beliefs About Self		Attitudes	
	Confident	Empowered	Should Assert	Healthy Rel.
SKY	0.045**	0.136*	0.091**	−0.040
	(0.022)	(0.098)	(0.040)	(0.028)
Outcome range	0 − 1	0 − 3	0 − 1	0 − 1
Control Mean	0.836	0.885	0.606	0.529
Control SD	0.187	0.799	0.326	0.227
Hypothesis	upr	upr	upr	upr
Num. obs.	235	235	234	234
N Clusters	198	198	197	198

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A53: Spillover effects of SKY on assertive attitudes, self-confidence, and salience of gender identity among sisters

Appendix section A.2 provides details on model specification and testing and section D.20 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect for any unit and the row "Hypothesis" the direction of this test. The p-value from a randomization inference-based test of the global null hypothesis that all null hypotheses regarding outcomes in the *Beliefs About Self* (*Attitudes*) category are true is $p = 0.015$ ($p = 0.87$).

	Gender Salience	Peer Pressure
SKY	−0.022	−0.002
	(0.040)	(0.103)
Outcome range	0 − 1	0 − 2
Control Mean	0.513	0.604
Control SD	0.309	0.803
Hypothesis	upr	lwr
Num. obs.	235	234
N Clusters	198	197

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A54: Effect of SKY on gender salience and perceived pressure to be in a relationship among sisters

Appendix section A.2 provides details on model specification and testing and section D.20 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect for any unit and the row "Hypothesis" the direction of this test. We do not report a joint RI p-value for this table, because the outcomes in columns 1-2 were not pre-registered as sub-hypotheses to one global hypothesis.

C.13 Effects on brothers

	Awareness of SKY Club		Exposure to SKY Content	
	Main Respondent Attended	SKY is for Girls	Exposed TV or Magazine	SKY Exposure
SKY	0.447***	0.102**	0.108**	0.322**
	(0.058)	(0.057)	(0.063)	(0.148)
Outcome range	0 – 1	0 – 1	0 – 1	0 – 6
Control Mean	0.332	0.191	0.385	0.906
Control SD	0.472	0.394	0.488	1.246
Hypothesis	upr	upr	upr	upr
Num. obs.	255	255	255	255
N Clusters	207	207	207	207

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A55: Spillover effects of SKY on brothers' exposure to SKY

Appendix section A.2 provides details on model specification and testing and section D.21 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect for any unit and the row "Hypothesis" the direction of this test. The p-value from a randomization inference-based test of the global null hypothesis that all null hypotheses regarding outcomes in the *Awareness of SKY Club* (*Exposure to SKY Content*) category are true is $p = 0.012$ ($p = 0.000$).

	Norm Perceptions			Stereotypes	Beliefs About
	Prescriptive (Girls)	Prescriptive (Boys)	Express Emotion (Girls)	Gender Roles	Inequality
SKY	−0.056	−0.003	−0.108	−0.026	−0.016
	(0.045)	(0.047)	(0.111)	(0.032)	(0.031)
Outcome range	0 – 1	0 – 1	0 – 2	0 – 1	0 – 1
Control Mean	0.673	0.464	1.188	0.370	0.587
Control SD	0.364	0.331	0.789	0.251	0.230
Hypothesis	two	two	two	two	two
Num. obs.	255	255	255	255	252
N Clusters	207	207	207	207	205

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A56: Spillover effects of SKY on brothers

Appendix section A.2 provides details on model specification and testing and section D.22 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect for any unit and the row "Hypothesis" the direction of this test. The p-value from a randomization inference-based test of the global null hypothesis that all null hypotheses regarding outcomes in the *Norm Perceptions* category are true is $p = 0.51$. We do not report such a p-value for *Stereotypes* and *Beliefs About*, because only one outcome falls into those categories.

	Attitudes		Identity	Masculinity		Supportive Behavior
	Should Assert	Healthy Rel.	Gender Salience	Empowered	Express Emotion	Support Girls
SKY	0.013	0.038	0.001	−0.039	0.042	−0.013
	(0.038)	(0.028)	(0.032)	(0.099)	(0.096)	(0.023)
Outcome range	0 − 1	0 − 1	0 − 1	0 − 3	0 − 2	0 − 1
Control Mean	0.698	0.486	0.310	1.218	1.135	0.647
Control SD	0.293	0.191	0.256	0.767	0.757	0.185
Hypothesis	two	two	two	two	two	two
Num. obs.	255	255	255	255	255	255
N Clusters	207	207	207	207	207	207

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A57: Spillover effects of SKY on brothers (continued)

Appendix section A.2 provides details on model specification and testing and section D.23 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect for any unit and the row "Hypothesis" the direction of this test. The p-value from a randomization inference-based test of the global null hypothesis that all null hypotheses regarding outcomes in the *Attitudes* (*Masculinity*) category are true is $p = 0.34$ ($p = 0.83$). We do not report such a p-value for *Identity* and *Supportive Behavior*, because only one outcome falls into those categories.

	Kept Pens Girl	Kept Pens Boy	Discrimination Pens
SKY	−0.356*	−0.072	0.284
	(0.198)	(0.177)	(0.194)
Outcome range	0 − 5	0 − 5	−4 − 5
Control Mean	3.133	3.051	−0.081
Control SD	1.375	1.372	1.229
Hypothesis	two	two	two
Num. obs.	255	255	255
N Clusters	207	207	207

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A58: Spillover effects of SKY on brothers' behavior in dictator game

Appendix section A.2 provides details on model specification and testing and section D.24 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect for any unit and the row "Hypothesis" the direction of this test. We do not report a joint RI p-value for this table, because the outcomes in columns 1-3 were not pre-registered as sub-hypotheses to one global hypothesis.

C.14 Analyses with siblings from baseline

Outcome	Range	Sample	N	Hypothesis	RI p-value	Estimate
Chores	0 - 4	Siblings from baseline	320	two	0.997	0.001
Fun Activities	0 - 6	Siblings from baseline	320	two	0.637	0.065
Quarrel w/ Main Girl	0 - 2	Siblings from baseline	319	two	0.562	0.055
Defect vs. Main Girl	0 - 1	Siblings from baseline	320	two	0.267	0.060
Correct Guesses abt. Main Girl	0 - 5	Siblings from baseline	320	upr	0.912	-0.208
Prescriptive (Girls)	0 - 1	Brothers from baseline	169	two	0.826	-0.014
Prescriptive (Boys)	0 - 1	Brothers from baseline	169	two	0.694	0.024
Express Emotion (Girls)	0 - 2	Brothers from baseline	169	two	0.577	-0.070
Gender Roles	0 - 1	Brothers from baseline	169	two	0.213	-0.050
Inequality	0 - 1	Brothers from baseline	168	two	0.843	-0.008
Should Assert	0 - 1	Brothers from baseline	169	two	0.610	0.027
Healthy Rel.	0 - 1	Brothers from baseline	169	two	0.105	0.050
Gender Salience	0 - 1	Brothers from baseline	169	two	0.838	0.009
Empowered	0 - 3	Brothers from baseline	169	two	0.389	0.110
Express Emotion	0 - 2	Brothers from baseline	169	two	0.674	0.047
Support Girls	0 - 1	Brothers from baseline	169	two	0.818	-0.007
Kept Pens Girl	0 - 5	Brothers from baseline	169	two	0.564	-0.147
Kept Pens Boy	0 - 5	Brothers from baseline	169	two	0.796	-0.058
Discrimination Pens	-4 - 5	Brothers from baseline	169	two	0.717	0.088
Assert\\(Boy)friend	0 - 1	Sisters from baseline	149	upr	0.882	-0.044
Assert Enumerator	0 - 1	Sisters from baseline	151	upr	0.053	0.095
Kept Pens Girl	0 - 5	Sisters from baseline	151	two	0.686	-0.089
Kept Pens Boy	0 - 5	Sisters from baseline	151	lwr	0.207	-0.171
Turn To	0 - 3	Sisters from baseline	151	upr	0.852	-0.188
Girl Friends	0 - 10	Sisters from baseline	151	upr	0.980	-0.704
New Friend	0 - 1	Sisters from baseline	151	upr	0.773	-0.059
Descriptive	0 - 1	Sisters from baseline	151	upr	0.390	0.014
Prescriptive	0 - 1	Sisters from baseline	151	upr	0.781	-0.046
Pressure	0 - 2	Sisters from baseline	150	lwr	0.508	0.002
Confident	0 - 1	Sisters from baseline	151	upr	0.188	0.026
Empowered	0 - 3	Sisters from baseline	151	upr	0.393	0.039
Should Assert	0 - 1	Sisters from baseline	151	upr	0.020	0.108
Healthy Rel.	0 - 1	Sisters from baseline	150	upr	0.990	-0.087
Gender Salience	0 - 1	Sisters from baseline	151	upr	0.568	-0.009

Table A59: Effect of SKY on siblings listed for each household during the baseline survey

Appendix section A.2 provides details on model specification and testing and section D.25 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. "Outcome" lists the dependent variable of each regression, and "Range" its range. "Sample" indicates the subsample of respondents used. "Hypothesis" reports the direction of a test of the sharp null hypothesis of no (positive/negative) treatment effect for any unit. "RI p-value" displays the randomization inference p-value from this test, and "Estimate" is the treatment coefficient.

C.15 Pre-registered main analyses

Our PAP distinguishes between "main outcomes" that we expected to be affected by the SKY intervention and "equilibrium outcomes" for which we were a priori less certain that they would be affected, since they have to do with social interactions within the household. Broadly speaking, we pre-registered one-tailed tests for the former and two-tailed tests for the latter. The analyses that we present in the main text follow this logic. However, they deviate slightly from what we pre-specified, since we realized that some outcomes that we pre-registered as "main outcomes" really measure within-household interactions and that one outcome is not as closely tied to our theory as we previously thought. In this section, we present our results for both "main outcomes" (hypotheses 1.1 and 1.8 in the PAP) and "equilibrium outcomes" (hypothesis 1.9 in the PAP) exactly as pre-specified and show that our conclusions do not change.

	Assert Relation	Assert Enumerator	Like Afrobeats	Kept Pens Boy	Kept Pens Girl	Fun Activities
SKY	0.050***	0.045**	0.098	0.164**	0.138*	0.047
	(0.014)	(0.026)	(0.097)	(0.081)	(0.083)	(0.082)
Afrobeats Prime			−0.217***			
			(0.091)			
SKY × Prime			−0.076			
			(0.142)			
Outcome range	0 – 1	0 – 1	0 – 3	0 – 5	0 – 5	0 – 6
Control Mean	0.553	0.472	1.792	2.843	2.760	2.345
Control SD	0.202	0.362	1.037	1.142	1.127	1.215
Hyp. SKY	upr	upr	two	upr	two	upr
Hyp. Prime			lwr			
Hyp. Diff.			upr			
Num. obs.	825	829	824	829	829	829

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A60: Effects of SKY on girls' proclivity to behave in empowered ways

Appendix section A.2 provides details on model specification and testing and section D.26 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars for the coefficients of "SKY" indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect of assignment to the SKY intervention for any unit and the row "Hyp. SKY" the direction of this test. Stars for the coefficients of "Afrobeats Prime" denote the RI p-value from a test of the sharp null hypothesis of no (positive) treatment effect of the Afrobeats prime for any unit and the row "Hyp. Prime" the direction of this test. Stars for the coefficients of "SKY × Prime" indicate the parametric p-value from a test of the weak null hypothesis of no difference between the effect of the treatment among those who were and were not assigned to the Afrobeats prime and "Hyp. Diff." the direction of this test. The p-value from a randomization inference-based test of the global null hypothesis that all null hypotheses pertaining to the tests shown in columns 1-4, and 6 of this table are correct is $p = 0.001$. We did not include the outcome *Given Pens Girl* in this global test, because there are competing logics about whether this outcome should be affected.

Table A60 shows estimated treatment effects on all outcomes that we pre-specified as "main outcomes." The table differs from Table 2 in the following ways:

First, the outcome in column 1 is an index that combines the three items that form the index *Assert (Boy)friend* in column 1 of Table 2 with the outcome *Assert Brother* in column 1 of Table 3. We separated the item *Assert Brother* from the other three items and included it in the analysis of high cost outcomes, since it clearly pertains to within-household behavior.

Second, we include the analysis in column 3 which was designed to test whether the SKY intervention causes girls to assert their opinions even in the face of peer pressure. At baseline, we asked respondents whether they like a variety of music genres, among them the well-known genre Afrobeats. 55.4% of respondents reported they do not like Afrobeats. At endline, we asked respondents again how much they like Afrobeats, but used simple random assignment to assign about half of our sample to first be told that a majority of girls at baseline stated they do not like Afrobeats. We expected the prompt to induce peer pressure to express a negative opinion about Afrobeats. The estimate in the second row of Table A60 indeed suggests the prime reduced the stated preference for Afrobeats among girls who were not assigned to the SKY intervention by about one fifth of a control group standard deviation ($p < 0.01$). The interaction term provides no evidence that the SKY intervention reduced girls' tendency to give in to this pressure. If anything, the negative effect of the prime is stronger among those assigned to the SKY club. While this result, at first sight, runs counter to the notion that the SKY intervention increases girls' proclivity to act on their preferences, this result is in line with the idea that SKY club participants bonded with each other. In other words, the SKY campaign may have induced a desire to conform to the views of girls interviewed at baseline, since these girls are also potential SKY club participants. This apparent effect is in line with the rest of our evidence on the mechanisms through which the intervention may have unfolded its effects (see section 4 in the main text). Because of these multiple possible interpretations, we exclude this result from our main outcomes table for simplicity but discuss it briefly in our section on mechanisms.[3]

Third, the table includes the outcome *Fun Activities* in the last column. We highlight in footnote 8 on p.22 of the PAP that we thought about this outcome as unilaterally chosen by the girls in our sample rather than as an outcome of intra-household bargaining with parents and siblings, because we assumed that girls' time budget constraint is not fully binding. However, our qualitative evidence strongly suggests that all of girls' time usage is an outcome of negotiations with parents and sometimes

[3]We do not analyze the interaction between assignment to the SKY intervention and the randomized Afrobeats prime among sisters, because the Afrobeats prime does not appear to shift stated preferences for Afrobeats even among sisters of control group girls. This strategy follows our pre-registration.

with siblings. Hence, we now see this outcome as one that results from interactions in which the social costs of asserting oneself are high and include the outcome in Table 3 accordingly.

We pre-registered a test of the joint significance of the estimates shown in columns 1, 2, 4 and 6 of the table and the interaction term in column 3. We excluded the analysis of dictator game behavior towards anonymous girls from this global test, since there are countervailing logics (greater in-group bonding and identification) that may have offset any empowerment-related effect on this outcome. Note that this exclusion ends up being conservative, since we do find an effect on this outcome. The p-value that results from the joint test is 0.001. Hence, our conclusion that the SKY intervention encouraged empowered behavior in situations where the social costs of asserting oneself are low survives even if we closely follow our PAP.

Conversely, Table A61 shows all analyses that we registered under hypothesis 1.9 (General Equilibrium Outcome). The p-value that results from the pre-registered joint significance test for this set of outcomes is $p = 0.5$. Our finding that the SKY intervention does not appear to affect within-household outcomes and behavior is even stronger if we stick to our pre-registered analyses.

	Quarrel w/ Sister	Quarrel w/ Brother	Chores	Defect vs. Sister	Defect vs. Brother
SKY	0.142	0.186	−0.050	0.030	−0.019
	(0.110)	(0.115)	(0.057)	(0.070)	(0.068)
Outcome range	0 − 2	0 − 2	0 − 4	0 − 1	0 − 1
Control Mean	0.582	0.508	3.404	0.465	0.479
Control SD	0.746	0.765	0.767	0.490	0.480
Hypothesis	two	two	two	two	two
Num. obs.	198	205	829	198	207

$^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Table A61: Effects of SKY on grils' within-household interactions

Appendix section A.2 provides details on model specification and testing and section D.27 on outcome question wording and coding. Models include a set of pre-registered covariates, detailed in section A.2. Stars for the coefficients of "SKY" indicate significance based on the RI p-value from a test of the sharp null hypothesis of no (positive/negative) treatment effect of assignment to the SKY intervention for any unit and the row "Hypothesis" the direction of this test. The p-value from a randomization inference-based test of the global null hypothesis that all null hypotheses pertaining to the tests shown in columns 1-5 of this table are correct is $p = 0.5$.

D Question Wording

The responses "refuse to answer" or "refused to play" are coded as missing. Additionally, the reponse "don't know" is coded as missing unless otherwise indicated. Indices are constructed by dividing individual items by their maximum possible value and then averaging across them.

D.1 Figure 2

Low-cost Settings Plot

Bars 1-2 - Complain Gift: This measure is a single item based on the following question:

- Thank you very much for all your responses so far. We have come almost to the end of the interview. There is only one section left. Before we do this last section, I want to tell you that I will later give a small gift to you as a token of gratitude for your participation in our study. You can choose between a handkerchief and ankle socks.

  At the end of the interview: ENUMERATOR: *Please give the respondent the XXXX [TABLET TO SELECT THE GIFT THAT THE RESPONDENT DID NOT CHOOSE] and say "here is your gift".*

  - Did the respondent tell you that you gave her the wrong gift? 0 = No, 1 = Yes

Bars 3-4 - Correct Date: This measure is an integer based on the following question:

- ENUMERATOR: *Read slowly as if you are thinking about it: "I first need to input the date here on the tablet. Give me a moment. Today is XXX [TABLET TO WRITE TODAY'S DAY], the [TABLE TO WRITE CORRECT DATE, E.G. 13th of May], **2022**."*

  - Did the respondent intervene to correct the date? 0 = No, 1 = Yes

Bars 5-6 - Kept Pens Girl: This measure is a single item based on a dictator game played with an anonymous girl:

- For this game, we have selected a girl who also participates in our study to be your partner. She is not someone from your household. Here are 5 pens. You can decide how many of these pens you would like to give to the girl and how many you would like to keep for yourself. You can give all pens, or no pens, or any number in between to the girl. We will deliver the pens

that you don't keep to the girl and you get to keep the rest. Please take the pens you would like to keep.

  - ENUMERATOR: How many pens did the respondent keep?
    * 5 = Respondent kept 5, gave back 0
    * 4 = Respondent kept 4, gave back 1
    * 3 = Respondent kept 3, gave back 2
    * 2 = Respondent kept 2, gave back 3
    * 1 = Respondent kept 1, gave back 4
    * 0 = Respondent kept 0, gave back 5

Bars 7-8 - Kept Pens Boy: This measure is a single item based on a dictator game played with an anonymous boy:

- For this game, we have selected a boy who also participates in our study to be your partner. He is not someone from your household. Here are 5 pens. You can decide how many of these pens you would like to give to the boy and how many you would like to keep for yourself. You can give all pens, or no pens, or any number in between to the boy. We will deliver the pens that you don't keep to the boy and you get to keep the rest. Please take the pens you would like to keep.

  - ENUMERATOR: How many pens did the respondent keep?
    * 5 = Respondent kept 5, gave back 0
    * 4 = Respondent kept 4, gave back 1
    * 3 = Respondent kept 3, gave back 2
    * 2 = Respondent kept 2, gave back 3
    * 1 = Respondent kept 1, gave back 4
    * 0 = Respondent kept 0, gave back 5

High-cost Settings Plot

Bar 1 - Quarrel w/ Main: This measure is a single item based on the following survey question, asked by the enumerator to the brother(s) of the main respondent:

- Now I would like you to think about the last two months. Thinking about this time frame, can you remember getting into a quarrel with [THE MAIN RESPONDENT IN THIS HOUSEHOLD]? 0 = No, Yes

  - **If Yes:** Thinking back, would you say this happened at least once a day or less than that? 2 = At least once a day, 1 = Less than that

Bar 2 - Quarrel w/ Brother: This measure is a single item based on the following survey question, asked by the enumerator to the main respondent:

- **If at least one boy in household:** Thinking about the same time frame [the last two months], can you remember getting into a quarrel with your brother/a teenage boy in the household? 0 = No, Yes

  - **If Yes:** Thinking back, would you say this happened at least once a day or less than that? 2 = At least once a day, 1 = Less than that

Bar 3 - Defect vs. Main: This measure is a single item based on the following prisoner's dilemma game played by the brother(s) with the main respondent:

- Now, we would like you to play a game with [NAME OF THE MAIN RESPONDENT]. The game is about four of these notebooks [ENUMERATOR: *Show the notebooks*]. You receive a card with the word **split** written on it and another card with the word **keep** written on it. [ENUMERATOR: *Give respondents the two cards.*] When we interview [NAME OF THE MAIN RESPONDENT], she will also receive the same two cards. Both of you can play either the **split** card or **keep** card. If you both play the **split** card, you **split** the four notebooks, and you receive two notebooks each. If one of you chooses the **keep** card and the other chooses the **split** card, whoever chooses the **keep** card receives all four notebooks and whoever chooses the split card receives nothing. But if you both choose the **keep** card, I will take two of the four notebooks away, and each of you will receive only one notebook. We will come back next week to distribute the notebooks depending on how you both played.

  - Do you want to play the **split** or **keep** card? 0 = Split, 1 = Keep.

Bar 4 - Defect vs. Brother: This measure is a single item based on the following prisoner's dilemma game played by the main respondent with her brother(s):

- **If one teenager in the household:** Now, we would like you to play a game with [NAME OF THE OTHER TEENAGER]. The game is about four of these notebooks [ENUMERATOR: *Show the notebooks*]. You receive a card with the word **split** written on it and another card with the word **keep** written on it. [ENUMERATOR: *Give respondents the two cards.*] When we interview [NAME OF THE OTHER TEENAGER], [HE/SHE] will also receive the same two cards. Both of you can play either the **split** card or **keep** card. If you both play the split card, you split the four notebooks, and you receive two notebooks each. If one of you chooses the **keep** card and the other chooses the **split** card, whoever chooses the **keep** card receives all four notebooks and whoever chooses the split card receives nothing. But if you both choose the keep card, I will take two of the four notebooks away, and each of you will receive only one notebook. We will come back next week to distribute the notebooks depending on how you both played.

  - Do you want to play the **split** or **keep** card? 0 = Split, 1 = Keep.

- **If more than one teenager in the household:** Now, we would like you to play a game separately with each of the other teenagers in the household. The game is about four of these notebooks as follows. [ENUMERATOR: *Show the notebooks*]. You receive a card with the word **split** written on it and another card with the word **keep** written on it. [ENUMERATOR: *Give respondents the two cards.*] When we interview your game partner, he or she will also receive the same two cards. Both of you can play either the **split** card or **keep** card. If you both play the **split** card, you **split** the four notebooks, and you receive two notebooks each. If one of you chooses the **keep** card and the other chooses the **split** card, whoever chooses the **keep** card receives all four notebooks and whoever chooses the **split** card receives nothing. But if you both choose the **keep** card, I will take two of the four notebooks away, and each of you will receive only one notebook. You will play the game with all the youths in the household. We will come back next week to distribute the notebooks according to how you and the other youths interviewed in this household played.

Randomize order of siblings in household.

- First, you are playing with [TEENAGER 1]. Do you want to play the **split** or **keep** card? 0 = Split, 1 = Keep.
- Second, you are playing with [TEENAGER 2]. Do you want to play the **split** or **keep** card? 0 = Split, 1 = Keep

Continues until main respondent has played against all siblings in her household.

This variable is the proportion of times that the main respondent chose to defect across all prisoners' dilemma games she played against a brother in the household.

Bars 5-6 - Chores: This measure is a single item based on the following survey question:

- Now, I would like you think about the past week. Please tell me which of the following you can remember doing in the past week.
  - 1 = Shopping for food
  - 1 = Cooking
  - 1 = Cleaning the house
  - 1 = Washing clothes

The variable reports the number of chores that respondent recalls doing.

Bars 7-8 - Fun Activities: This measure is a single item based on the following survey question:

- Now, I would like you to think about the past week. Please tell me which of the following you can remember doing in the past week.
  - 1 = Going to the movies
  - 1 = Hanging out at the mall
  - 1 = Doing sports
  - 1 = Going to a party
  - 1 = Reading a novel for fun
  - 1 = Studying

The variable reports the number of fun activities that respondent recalls doing.

D.2 Table 2

Column 1 - Assert (Boy)friend: This measure is an index of three items:

- Suppose your boyfriend trains to become an athlete every afternoon. He is very good at it and hopes to win a scholarship to become a professional athlete. You feel like his sport is always his top priority and that makes you sad. You would like to spend more time with him, at least on the weekends. Which of the following will you do?
  - 0 = I would not tell my boyfriend about my feelings, because I don't want to interfere with his professional dream
  - 1 = I would tell my boyfriend how I feel
- Now imagine a different situation: Your girlfriends are going to the movies tonight and you would like to join them. However, your boyfriend wants you to watch the same movie with him instead. Which of the following would you rather do?
  - 0 = I would go watch the movie with my boyfriend because I don't want him to be upset
  - I want to spend time with my girlfriends and so I will watch the movie with them
    - **If I want to spend time with my girlfriends and so I will watch the movie with them:** Will saying no to your boyfriend reduce how much you enjoy the evening? 0.33 = No, not at all, 0.67 = Yes, by a little, 1 = Yes, by a lot.
- Suppose you are hanging out with some of your friends at a party at someone's house. One of your guy friends is tickling one of your girl friends and she says "Stop tickling me, I don't like it", to which he says "We are just playing, what is wrong with you?" Which of the following are you most likely to do?
  - 0 = I would tell my friend that the boy is just having fun and doesn't mean any harm
  - 1 = I would not confront the boy, but would diffuse the situation by asking my friend to join me in the kitchen to get something to eat
  - 2 = I would intervene and tell the boy to not touch my friend if she doesn't like it

Column 2 - Assert Enumerator: This measure is an index of two items:

- ENUMERATOR: *Read slowly as if you are thinking about it: "I first need to input the date here on the tablet. Give me a moment. Today is XXX [TABLET TO WRITE TODAY'S DAY], the [TABLE TO WRITE CORRECT DATE, E.G. 13th of May],* ***2022****."*
  - Did the respondent intervene to correct the date? 0 = No, 1 = Yes
- Thank you very much for all your responses so far. We have come almost to the end of the interview. There is only one section left. Before we do this last section, I want to tell you that I will later give a small gift to you as a token of gratitude for your participation in our study. You can choose between a handkerchief and ankle socks.

  At the end of the interview: ENUMERATOR: *Please give the respondent the XXXX [TABLET TO SELECT THE GIFT THAT THE RESPONDENT DID NOT CHOOSE] and say "here is your gift".*
  - Did the respondent tell you that you gave her the wrong gift? 0 = No, 1 = Yes

Columns 3-4: See section D.1.

D.3 Table 3

Column 1 - Assert Brother: This measure is a single item based on the following question:

- Which of the following two statements comes closest to your view:
  - 0 = *Statement 1:* If my brother argues that some of my friends are not good for me, I would distance myself from them even if I like them.
  - *Statement 2:* I am quite capable of choosing the friends I want and will continue to spend time with my friends even if my brother argues that they are not good for me.
    - **If Statement 2:** Would you confront your brother and ask him to stop questioning your choice of friends or would you rather not discuss your friends with him anymore? 0.5 = I would rather not discuss my friends with him anymore, 1 = I would confront my brother and ask him to stop questioning my choice of friends.

Column 2 - Quarrel w/ Sibling: This measure is an average of the following items:

- **If more than one girl in household:** Thinking about the same time frame [the last two months], can you remember getting into a quarrel with your sister/another teenage girl in the household?
  0 = No, Yes
  - **If Yes:** Thinking back, would you say this happened at least once a day or less than that?
    2 = At least once a day, 1 = Less than that

- **If at least one boy in household:** Thinking about the same time frame [the last two months], can you remember getting into a quarrel with your brother/a teenage boy in the household?
  0 = No, Yes
  - **If Yes:** Thinking back, would you say this happened at least once a day or less than that?
    2 = At least once a day, 1 = Less than that

  The variable is an average of the responses to both items if the main respondent has both sisters and brothers. If the respondent has only at least one sister (brother) but no sisters (brothers), then the variable is coded as the response to the first (second) item only.

Column 3 - Defect vs. Sibling: For question wording, see section D.1. This variable is the proportion of times that the main respondent chose to defect across all prisoners' dilemma games she played against a sibling in the household.
Columns 4-5 - See section D.1.

D.4 Table 4

Column 1 - Perceptions: Assert Brother: This measure is a single item based on the following survey question, that captures the number of chosen adjectives that are not stereotypically female:

- Finally, we asked [NAME OF MAIN RESPONDENT] which of the following two statements comes closest to her view:
  - Statement 1: If my brother argues that some of my friends are not good for me, I would distance myself from them even if I like them.
  - Statement 2: I am quite capable of choosing the friends I want and will continue spending time with my friends even if my brother argues that they are not good for me.

Which statement do you believe [NAME OF MAIN RESPONDENT] chose? 0 = Statement 1, 1 = Statement 2

Column 2 - Perceptions: Empowered: This measure is a single item based on the following survey question:

- I would like you to take a look at these words. Please tell me which three words you think best describe [NAME OF MAIN RESPONDENT], that captures the number of chosen adjectives that are not stereotypically female:
  - 1 = Ambitious (male)
  - 0 = Caring (female)
  - 0 = Helpful (female)
  - 1 = Assertive (male)
  - 0 = Sweet (female)
  - 1 = Charismatic (male)
  - 1 = Strong (male)
  - 1 = Adventurous (male)
  - 0 = Understanding (female)
  - 0 = Emotional (female)

Column 3 - Knowledge: Correct Guesses abt. Main Girl: This measure is a single item based on the sum derived from the following survey questions:

- We asked [NAME OF MAIN RESPONDENT] whether she thinks that multiple ear piercings are stylish? Do you think she said yes or no? 1 = if answer matches main respondent's answer in baseline, 0 = if otherwise including don't knows
- We asked [NAME OF MAIN RESPONDENT] which of the following activities she would rather do if she had to choose one of them:
  - Going to the mall

  - Watching movies
  - Listening to music
  - Reading books
  - Doing sports

  Which do you think [NAME OF MAIN RESPONDENT] picked?

  - Going to the mall
  - Watching movies
  - Listening to music
  - Reading books
  - Doing sports
  - Don't know

  1 = if answer matches the main respondent's answer in endline, 0 = if otherwise including don't knows.

- We also asked [NAME OF MAIN RESPONDENT] "Do you have a favorite TV show?" Do you think [NAME OF MAIN RESPONDENT] said yes or no?

  - If yes: Which is [NAME OF MAIN RESPONDENT]'s favorite TV show?

  1 = if answer matches the main respondent's answer in endline: whether respondent has favorite show and if yes, which TV show, 0 = if otherwise including don't knows.

- We also asked [NAME OF MAIN RESPONDENT] "Do you have a favorite sport?" Do you think [NAME OF MAIN RESPONDENT] said yes or no?

  - If yes: Which is [NAME OF MAIN RESPONDENT]'s favorite sport?

  1 = if answer matches the main respondent's answer in endline: whether respondent has favorite sport and if yes, which sport, 0 = if otherwise including don't knows.

- We also asked [NAME OF MAIN RESPONDENT] what her dream job is. What do you think she said? 1 = if answer matches main respondent's answer in endline, 0 = if otherwise including don't knows.

- Finally, we asked [NAME OF MAIN RESPONDENT] which of the following two statements comes closest to her view:
  - Statement 1: If my brother argues that some of my friends are not good for me, I would distance myself from them even if I like them.
  - Statement 2: I am quite capable of choosing the friends I want and will continue spending time with my friends even if my brother argues that they are not good for me.

  1 = if answer matches main respondent's answer in endline, 0 = if otherwise including don't knows.

Columns 4-7 - See section D.1.

D.5 Table 5

Column 1 - Attitudes: Should Assert: This measure is an index of two survey items:

- Let's consider another scenario: Neema is best friends with Brian. A few months ago, Neema started dating Jayson. Neema used to always watch Brian play basketball on Saturday mornings. Jayson is not comfortable with Neema going to see Brian and his basketball friends. Which of the following statements do you agree with more?
  - 1 = *Statement 1:* Neema should be able to hang out with whomever she wants and however much she pleases.
  - 0 = *Statement 2:* I understand that Jayson is not comfortable with his girlfriend spending time with another boy and Neema should take that into account.
- Let's consider another scenario. Peter is 17 years old and Maya is 16. The two have been together for a year. They love and respect each other. Lately, Peter has been suggesting to Maya that he thinks they should have sex. Maya is unsure if she feels ready yet. Which of the following statements comes closest to your view?
  - 1 = *Statement 1:* It is important that Maya takes her time to figure out whether she is ready to have sex with Peter.
  - 0 = *Statement 2:* I don't see any reason why the couple should not have sex now given that they love and respect each other.

Column 2 - Attitudes: Healthy Rel.: This measure is an index of seven survey items:

- Suppose your close friend is in a relationship. I will read you a list of things that her boyfriend may do. For each behavior, please tell me whether you think this behavior can be a sign of a toxic relationship.
  - He wants to be around her all the time. 1 = Yes, 0 = No
  - He encourages her to spend time with her friends. 0 = Yes, 1 = No
  - He tells her all the time that she would look amazing in heels. 1 = Yes, 0 = No
  - He buys her very expensive gifts. 1 = Yes, 0 = No
- Still thinking about your friend and her boyfriend. I will now read you a list of things that your friend may do. For each behavior, please tell me whether you think it would contribute to a healthy relationship.
  - Telling him that she cannot hang out with him because she has to watch a movie with her friends. 1 = Yes, 0 = No
  - Asking him every day whether he loves her. 0 = Yes, 1 = No
  - Texting him all the time about where he is. 0 = Yes, 1 = No

Column 3 - Beliefs About Self: Confident: This measure is an index of two items:

- **Now, please tell me whether you agree or disagree with the following statement.** "I feel that I have a number of good qualities". Do you agree or disagree? Agree OR Disagree
  - **If Agree:** Do you agree a lot with this statement or do you agree just a little? 3 = Agree a lot, 2 = Agree just a little
  - **If Disagree:** Do you disagree a lot with this statement or do you disagree just a little? 1 = Disagree just a little, 0 = Disagree a lot
- And finally, what about this statement? I am able to deal with life's challenges at least as well as most other people in my age group. Agree or Disagree
  - **If Agree:** Do you agree a lot with this statement or do you agree just a little? 3 = Agree a lot, 2 = Agree just a little

  - **If Disagree:** Do you disagree a lot with this statement or do you disagree just a little? 1 = Disagree just a little, 0 = Disagree a lot

Column 4 - Beliefs About Self: Empowered: This measure is a single item based on the following survey question, that captures the number of chosen adjectives that are not stereotypically female:

- Next, I would like to ask you a few questions about how you see yourself. Remember that there are no right or wrong answers - we are just interested in your views. To begin with, I would like you to take a look at these words. Please select the three words that best describe you.
  - 1 = Ambitious (male)
  - 0 = Caring (female)
  - 0 = Helpful (female)
  - 1 = Assertive (male)
  - 0 = Sweet (female)
  - 1 = Charismatic (male)
  - 1 = Strong (male)
  - 1 = Adventurous (male)
  - 0 = Understanding (female)
  - 0 = Emotional (female)

D.6 Table 6

Column 1 - Social Connections: Girl Friends: This measure is an integer that stores the number of a girl's close female friends, and is based on the following survey questions:

- I would like you to think about the friends you spend time with. Could you tell me the first name of the friend you spend most time with?
  - Is this friend a boy or a girl?
- Do you spend time with another friend? Yes/No
  - **If Yes:** Please tell me the first name of the second friend you spend time with.

    - Is this friend a boy or a girl?

  This question is repeated until the enumerator has asked about all the girls' close friends - up to ten.

Column 2 - Social Connections: New Friend: This measure is a single item based on the following survey question:

- Now I would like you to think about the last two months. Can you remember making a new friend during this time? 1 = Yes, 0 = No

Column 3 - Social Connections: Turn To: This measure is a single item based on the following survey question:

- Now, I would like you to imagine that you are very upset, perhaps because a friend has badly disappointed you. In a situation like this, how likely is it that you will find a supportive person to reach out to? 3 = Very likely, 2 = Somewhat likely, 1 = Not very likely, 0 = Not likely at all.

Column 4 - Norm Perceptions: Descriptive: This measure is an index of the following two items:

- Imagine the following situation: Otieno and Njeri have been together for 6 months. Otieno is constantly asking Njeri to come over to his house to hang out. Njeri is not comfortable going to Otieno's house and has been refusing. Think about most other teenage girls. If they were in Njeri's shows, which of the following would they rather do? 1 = Most teenage girls would definitely not go to Otieno's house if they don't feel comfortable doing so, 0 = Most teenage girls would maybe consider going to Otieno's house if they really liked him.

- Suppose you are hanging out with some of your friends at a party at someone's house. One of your guy friends is tickling one of your girl friends and she says "Stop tickling me, I don't like it", to which he says "We are just playing, what is wrong with you?" What would most other girls from your community do if they were in your situation? 1 = Most other girls would intervene and tell the boy to not touch their friend if she doesn't like it. 0.5 = Most other girls would not confront the boy but would diffuse the situation by asking their friend to join them

in the kitchen to get something to eat. 0 = Most other girls would tell their friend that the boy is just having fun and doesn't mean any harm.

Column 5 - Norm Perceptions: Prescriptive: This measure is an index of two survey items:

- Let's consider another scenario: Neema is best friends with Brian. A few months ago, Neema started dating Jayson. Neema used to always watch Brian play basketball on Saturday mornings. Jayson is not comfortable with Neema going to see Brian and his basketball friends. What would most girls in your community think about this question? 1 = Most girls would think Neema should be able to hang out with whomever she wants and however much she pleases. 0 = Most girls would understand that Jayson is not comfortable with his girlfriend spending time with another boy and Neema should take that into account.
- Let's consider another scenario. Peter is 17 years old and Maya is 16. The two have been together for a year. They love and respect each other. Lately, Peter has been suggesting to Maya that he thinks they should have sex. Maya is unsure if she feels ready yet. What do most other girls in your community think about this issue? 1 = Most other girls in my community think that it is important that Maya takes her time to figure out whether she is ready to have sex with Peter, 0 = Most other girls in my community don't see any reason why the couple should not have sex now, given that they love and respect each other.

D.7 Table A6

Column 1 - N Attended: This measure is an integer based on two questions:

- Think back over the last couple of weeks. Have you attended any sessions of the SKY club, which is a program for teenage girls? 0 = No, Yes
  - **If Yes to having attended**: How many sessions do you remember attending? The integer is coded as 0 if the respondent has not attended any session.

Column 2 - Exposed TV or Magazine: This measure is an index of two items:

- Have you ever watched any of the following shows? 0 = Monica, 0 = Gameshakers, 1 = PAA, 1 = SISTA show

- Have you read the following magazines? 0 = The Pulse, 0 = Buzz, 0 = Shujaaz, 1 = SKY magazine

Column 3 - Knowledge Club: This measure is the sum of three items:

- Think back over the last couple of weeks. Have you attended any sessions of the SKY club, which is a program for teenage girls? 0 = No, Yes
  - **If Yes to having attended:** Which of the following do you recall happening in the club? 1 = Dancing, 1 = Talking about relationships, 1 = Watching the SISTA show, 0 = Watching Gameshakers, 0 = Learning about cooking, 0 = Reading Buzz.

  This measure is coded as the number of activities that the respondent correctly identified as either happening or not happening in the club, coded as zero if the respondent said she did not attend the club.

- Finally, imagine that your friend has a new boyfriend and is going through a difficult time in that relationship. She feels like her boyfriend is putting a lot of pressure on her and turns to you for help. You think the situation sounds quite serious and you worry about your friend. You believe it would be a good idea if your friend spoke to a counselor about the problem. Do you know an organization that your friend could call for free to get help from a counselor? 0 = No, Yes
  - **If yes:** Please tell me the phone number. 1 = if the respondent said 1190, 0 = if the respondent said another number or said she doesn't know a number, or says she doesn't know a number or organization.
  - **If yes:** Do you know the name of the organization that your friend could call for free to get help from a counselor? 0 = No, Yes
    - **If yes:** Please tell me the name of the organization. 1 = if the respondent said One2One, and 0 = if the respondent said another organization or said she doesn't know an organization, or says she doesn't know a number or organization.

Column 4 - SKY Exposure: This measure is an integer based on the following question:

- **If the respondent has participated in the club or has otherwise heard about SKY:** Please tell me whether you have ever done the following things.
  - 1 = Visited the SKY Facebook page
  - 1 = Visited the SKY page on Instagram
  - 1 = Watched SKY-related content on TikTok
  - 1 = Took the SKY pledge
  - 1 = Visited the SKY hub
  - 1 = Visited the SKY website
  - 1 = Talked about SKY with friends
  - 1 = Listened to a SKY song
  - 1 = Went to a SKY school activation
  - 0 = None

  This measure is coded as the number of SKY activities the respondent has performed. 0 = if the respondent has never heard of SKY.

D.8 Table A20

Columns 1-2 - See section D.1.

Columns 3-4 - Quarrel w/ Sister: This measure is a single item based on the following survey question:

- **If more than one girl in household:** Thinking about the same time frame [the last two months], can you remember getting into a quarrel with your sister/another teenage girl in the household?

  0 = No, Yes
  - **If Yes:** Thinking back, would you say this happened at least once a day or less than that?

    2 = At least once a day, 1 = Less than that

D.9 Table A21

Column 1 - Defect vs. Sister: For question wording, see section D.1. This variable is the proportion of times that the main respondent chose to defect across all prisoners' dilemma games she played against a sister in the household.

Column 2 - Defect vs. Brother: For question wording, see section D.1. This variable is the proportion of times that the main respondent chose to defect across all prisoners' dilemma games she played against a brother in the household.

D.10 Tables A22-A24

Columns 1-4 - See section D.3.

D.11 Tables A22-A24

Columns 1-4 - See section D.3.

D.12 Tables A25-A26

Columns 1-6 - See section D.1.

D.13 Tables A27-A28

Columns 1-5 - See sections D.1 and D.3.

D.14 Table A29

Columns 1-7 - See section D.4.

D.15 Table A30

Columns 1-2 - See section D.3.

D.16 Table A31

Column 1 - Gender Salience: This measure is an index of two items:

- People think about themselves in many different ways. Please tell me which of the following categories best describes the way you think of yourself:
  - 0 = Student
  - 0 = My tribe
  - 0 = Kenyan
  - 1 = Boy/Man
  - 0 = Youth
- Now, I will read you three statements. Please tell me which of the following statements comes closest to your view.

- 2 = *Statement 1:* Being a boy or man is more important to me than being a youth.
- 1 = *Statement 2:* Being a boy or man is equally important to me as being a youth.
- 0 = *Statement 3:* Being a youth is more important to me than being a boy or man.

Column 2 - Peer Pressure: This measure is a single item based on the following survey questions:

- Which of the following statements comes closest to your view?
  - 2 = *Statement 1:* Many teenage girls seem to be in a relationship and I feel pressured to be in a relationship too.
  - *Statement 2:* I don't feel pressured to be in a relationship at this point in time.
    - **If Statement 2:** Do you not feel any pressure at all or do you feel a little pressure? 0 = No pressure at all, 1 = A little pressure,

D.17 Table A32

Columns 1-2 - See section D.1.

D.18 Tables A33-A48

See sections D.1, D.2, D.3, D.7, D.5, and D.16.

D.19 Table A49

Column 1 - N Attended: This measure is an integer based on two questions:

- Think back over the last couple of weeks. Have you attended any sessions of the SKY club, which is a program for teenage girls? 0 = No, Yes
  - **If Yes to having attended**: How many sessions do you remember attending? The integer is coded as 0 if the respondent has not attended any session.

Column 2 - Awareness of SKY Club: Main Respondent Attended: This measure is an integer based on the following question asked to the brothers:

- Think back over the last couple of weeks. Do you recall whether [NAME OF MAIN RESPONDENT IN THE HOUSEHOLD] attended any sessions of the SKY club? 0 = No, 0 = Don't know, 1 = Yes

Column 3 - Awareness of SKY Club: Knowledge Club: This measure is the sum of three items:

- Think back over the last couple of weeks. Have you attended any sessions of the SKY club, which is a program for teenage girls? 0 = No, Yes
  - **If Yes to having attended:** Which of the following do you recall happening in the club? 1 = Dancing, 1 = Talking about relationships, 1 = Watching the SISTA show, 0 = Watching Gameshakers, 0 = Learning about cooking, 0 = Reading Buzz.

  This measure is coded as the number of activities that the respondent correctly identified as either happening or not happening in the club, coded as zero if the respondent said she did not attend the club.

- Finally, imagine that your friend has a new boyfriend and is going through a difficult time in that relationship. She feels like her boyfriend is putting a lot of pressure on her and turns to you for help. You think the situation sounds quite serious and you worry about your friend. You believe it would be a good idea if your friend spoke to a counselor about the problem. Do you know an organization that your friend could call for free to get help from a counselor? 0 = No, Yes
  - **If yes:** Please tell me the phone number. 1 = if the respondent said 1190, 0 = if the respondent said another number or said she doesn't know a number, or says she doesn't know a number or organization.
  - **If yes:** Do you know the name of the organization that your friend could call for free to get help from a counselor? 0 = No, Yes
    - **If yes:** Please tell me the name of the organization. 1 = if the respondent said One2One, and 0 = if the respondent said another organization or said she doesn't know an organization, or says she doesn't know a number or organization.

Column 4 - Exposure to SKY Content: Exposed TV or Magazine: This measure is an index of two items:

- Have you ever watched any of the following shows? 0 = Monica, 0 = Gameshakers, 1 = PAA, 1 = SISTA show

- Have you read the following magazines? 0 = The Pulse, 0 = Buzz, 0 = Shujaaz, 1 = SKY magazine

Column 5 - Exposure to SKY Content: SKY Exposure: This measure is an integer based on the following question:

- **If the respondent has participated in the club or has otherwise heard about SKY:** Please tell me whether you have ever done the following things.
  - 1 = Visited the SKY Facebook page
  - 1 = Visited the SKY page on Instagram
  - 1 = Watched SKY-related content on TikTok
  - 1 = Took the SKY pledge
  - 1 = Visited the SKY hub
  - 1 = Visited the SKY website
  - 1 = Talked about SKY with friends
  - 1 = Listened to a SKY song
  - 1 = Went to a SKY school activation
  - 0 = None

  This measure is coded as the number of SKY activities that the respondent has done. 0 = if the respondent has never heard about SKY.

D.20 Tables A50-A54

See sections D.1, D.2, D.3, D.7, and D.16.

D.21 Table A55

Columns 1-4 - See section D.19.

D.22 Table A56

Column 1 - Norm Perceptions: Prescriptive (Girls): This measure is an index of two items:

- And what about the following situation: Otieno and Njeri have been together for 6 months. Otieno is constantly asking Njeri to come over to his house to hang out. Njeri is not comfortable going to Otieno's house and has been refusing. Otieno is complaining about Njeri's behavior to you and your other male friends.
  - Think about most other teenage girls. Which of the following would they rather tell Otieno?
    - 1 = Most girls would tell Otieno that he should stop putting pressure on Njeri.
    - 0 = Most girls would help Otieno to think about ways to convince Njeri.
- Let's consider another scenario. Peter is 17 years old and Maya is 16. The two have been together for a year. They love and respect each other. Lately, Peter has been suggesting to Maya that he thinks they should have sex. Maya is unsure if she feels ready yet.
  - What do most other girls in your community think about this issue?
    - 1 = Most other girls in my community think it is important that Maya takes her time to figure out whether she is ready to have sex with Peter.
    - 0 = Most other girls in my community don't see any reason why the couple should not have sex now given that they love and respect each other.

Column 2 - Norm Perceptions: Prescriptive (Boys): This measure is an index of two items:

- And what about the following situation: Otieno and Njeri have been together for 6 months. Otieno is constantly asking Njeri to come over to his house to hang out. Njeri is not comfortable going to Otieno's house and has been refusing. Otieno is complaining about Njeri's behavior to you and your other male friends.
  - Think about most other teenage boys. Which of the following would they rather tell Otieno?
    - 1 = Most boys would tell Otieno that he should stop putting pressure on Njeri.
    - 0 = Most boys would help Otieno to think about ways to convince Njeri.

- Let's consider another scenario: Neema is best friends with Brian. A few months ago, Neema started dating Jayson. Neema used to always watch Brian play basketball on Saturday mornings. Jayson is not comfortable with Neema going to see Brian and his basketball friends.
  - What would most boys within your community think about this question?
    - 1 = Most boys would think Neema should be able to hang out with whomever she wants and however much she pleases.
    - 0 = Most boys would understand that Jayson is not comfortable with his girlfriend spending time with another boy and Neema should take that into account.

Column 3 - Norm Perceptions: Express Emotion (Girls): This measure is a single item based on the following question:

- Next, suppose you and your friends are joking around. One of your friends makes a joke that you find hurtful.
  - What do most girls you think you should do in this situation?
    - 2 = Most girls think I should tell my friend that he hurt me and ask him to not say such things anymore.
    - 1 = Most girls think I should laugh along with the others in order to not spoil the fun, but tell my friend that he hurt me in private later.
    - 0 = Most girls think I should not tell my friend that he hurt me - this is just something that happens among guys and there is no need to make a big deal out of it.

Column 4 - Stereotypes: Gender Roles: This measure is an index of four items:

- And what about these two statements?
  - 0 = *Statement 1:* Boys tend to be more ambitious than girls.
  - 1 = *Statement 2:* Girls tend to be more ambitious than boys.
- Which of the following statements comes closest to your view?
  - 1 = *Statement 1:* Boys tend to be more emotional than girls.

  - 0 = *Statement 2:* Girls tend to be more emotional than boys.

- A man should have the final word about decisions in his home.
  - 0 = Agree
  - 1 = Disagree

- Boys make better community leaders than girls.
  - 0 = Agree
  - 1 = Disagree

Column 5 - Beliefs About: Inequality: This measure is an index of four items:

- Think about yourself and [NAME OF MAIN RESPONDENT]. Between the two of you, who would say is more likely to work in a job that they love in the future?
  - 1 = Yourself
  - 0 = [NAME OF MAIN RESPONDENT]

- And again between you and [NAME OF MAIN RESPONDENT], who would you say is most likely to earn a lot of money?
  - 1 = Yourself
  - 0 = [NAME OF MAIN RESPONDENT]

- Now I will read you three statements. Please tell me which of them you most agree with.
  - 1 = *Statement 1:* In today's society, it is better to be a boy or man, because boys and men tend to have more opportunities than women and girls.
  - 0 = *Statement 2:* In today's society, it is better to be a girl or woman, because girls and women tend to have more opportunities than boys and men.

- And what about the following two statements? Remember, we are just interested in which statement is closest to your view - even if you don't agree with any of the statements completely.

- 0 = *Statement 1:* Girl children have a lot of support today. We need a greater focus on boys.
- 1 = *Statement 2:* Boy children have a lot of support today. We need a greater focus on girls.

D.23 Table A57

Column 1-3 - See sections D.5 and D.16.

Column 4 - Masculinity: Empowered: This measure is a single item based on the following question, and captures the number of chosen adjectives that are not stereotypically female:

- Next, I would like to ask you a few questions about how you see yourself. Remember that there are no right or wrong answers - we are just interested in your views. To begin with, I would like you to take a look at these words. Please select the three words that best describe you.
  - 1 = Ambitious (male)
  - 0 = Caring (female)
  - 0 = Helpful (female)
  - 1 = Assertive (male)
  - 0 = Sweet (female)
  - 1 = Charismatic (male)
  - 1 = Strong (male)
  - 1 = Adventurous (male)
  - 0 = Understanding (female)
  - 0 = Emotional (female)

Column 5 - Masculinity: Express Emotion: This measure is a single item based on the following question:

- Next, suppose you and your friends are joking around. One of your friends makes a joke that you find hurtful. Which of the following are you must likely to do?
  - 2 = I would tell my friend that he hurt me and ask him to not say such things anymore.

  - 1 = I would laugh along with the others in order to not spoil the fun, but tell my friend that he hurt me in private later.
  - 0 = I would not tell my friend that he hurt me - this is just something that happens among guys and there is no need to make a big deal out of it.

Column 6 - Supportive Behavior: Support Girls: This measure is an index of five items:

- Now, suppose one of your female cousins tells you that she is very stressed, because final exams are coming up in her school, but she doesn't have enough time to study. The reason is that she must take care of her younger siblings every weekend, because her mother is sick. Which of the following would you rather do?
  - 1 = I would offer my cousin to help her out with her siblings on the next two Saturdays so that she can study.
  - 0 = I would promise my cousin to ask my sister whether she can help, because girls are better at taking care of kids.
- Now, I would like you to think about a different situation. Imagine that your sister or female cousin is a big football fan and wants to play on the school team. Yet, most of the other students that play are boys, and your family thinks that it would be better if she played a sport like Volleyball that is more suitable for girls. Again, which of the following would you rather do?
  - 1 = I would help my sister convince my family to let her be on the football team.
  - 0 = I agree with my family and would try to explain to my sister/female cousin that she will have more fun playing Volleyball with the other girls in her school.
- Next, let's think about another situation: Your little sister is about to go to the mall with her girl friends and you notice that she is wearing a skirt that is very short. Which of the following would you rather do?
  - 2 = I would not say anything to my sister, because my sister should be able to wear whatever she wants.
  - 1 = I would suggest to my sister that she should change her outfit before going out.

  - 0 = I would tell my sister to change her outfit and will not let her leave if she refuses.

- And what about the following situation: Otieno and Njeri have been together for 6 months. Otieno is constantly asking Njeri to come over to his house to hang out. Njeri is not comfortable going to Otieno's house and has been refusing. Otieno is complaining about Njeri's behavior to you and your other male friends. Which of the following would you rather tell Otieno?

  - 1 = I would tell Otieno that he should stop putting pressure on Njeri. She should not go to Otieno's house if she does not feel comfortable doing so.
  - 0 = I would tell Otieno that I understand his frustration and help him think about ways of how to convince Njeri.

- Suppose you are hanging out with some of your friends at a party at someone's house. One of your guy friends is tickling one of your girl friends and she says "Stop tickling me, I don't like it" to which he says "We are just playing, what is wrong with you?" Which of the following are you most likely to do?

  - 2 = I would intervene and tell the boy to not touch my friend if she doesn't like it.
  - 1 = I would not confront the boy, but would diffuse the situation by asking my friend to join me in the kitchen to get something to eat.
  - 0 = I would tell my friend that the boy is just having fun and doesn't mean any harm.

D.24 Table A58

Columns 1-2 - See section D.1

Column 3 - Dictator Game: Discrimination: This measure is a single item which measures the degree to which boys discriminate against girls when they allocate pens in dictator games with anonymous partners:

- Dictator Game: Given Pens Girl - Dictator Game: Given Pens Boy

D.25 Table A59

See sections D.1, D.2, D.3, D.7, D.16, and D.22.

D.26 Table A60

Column 1 - Assert Relation: This measure is an index of four items:

- Which of the following two statements comes closest to your view:
  - 0 = *Statement 1:* If my brother argues that some of my friends are not good for me, I would distance myself from them even if I like them.
  - *Statement 2:* I am quite capable of choosing the friends I want and will continue to spend time with my friends even if my brother argues that they are not good for me.
    - **If Statement 2:** Would you confront your brother and ask him to stop questioning your choice of friends or would you rather not discuss your friends with him anymore? 0.5 = I would rather not discuss my friends with him anymore, 1 = I would confront my brother and ask him to stop questioning my choice of friends.
- Suppose your boyfriend trains to become an athlete every afternoon. He is very good at it and hopes to win a scholarship to become a professional athlete. You feel like his sport is always his top priority and that makes you sad. You would like to spend more time with him, at least on the weekends. Which of the following will you do? 0 = I would not tell my boyfriend about my feelings, because I don't want to interfere with his professional dream, 1 = I would tell my boyfriend how I feel.
- Now imagine a different situation: Your girlfriends are going to the movies tonight and you would like to join them. However, your boyfriend wants you to watch the same movie with him instead. Which of the following would you rather do?
  - 0 = I would go watch the movie with my boyfriend because I don't want him to be upset
  - I want to spend time with my girlfriends and so I will watch the movie with them
    - If **I want to spend time with my girlfriends and so I will watch the movie with them:** Will saying no to your boyfriend reduce how much you enjoy the evening? 0.33 = No, not at all, 0.67 = Yes, by a little, 1 = Yes, by a lot.
- Suppose you are hanging out with some of your friends at a party at someone's house. One of your guy friends is tickling one of your girl friends and she says "Stop tickling me, I don't like

it", to which he says "We are just playing, what is wrong with you?" Which of the following are you most likely to do? 0 = I would tell my friend that the boy is just having fun and doesn't mean any harm, 1 = I would not confront the boy, but would diffuse the situation by asking my friend to join me in the kitchen to get something to eat, 2 = I would intervene and tell the boy to not touch my friend if she doesn't like it.

Column 2 - *Act Enumerator*: See section D.1.

Column 3 - *Like Afrobeats*: This measure is a single item based on the following survey questions:

- Do you like Afrobeats - meaning artists such as Kizz Daniel or Burna Boy? Yes or No
  - **If Yes:** Do you like Afrobeats a lot or little? 3 = Like a lot, 2 = Like it a little
  - **If No:** Do you not like Afrobeats at all or do you not like it a little? 1 = Don't like it a little, 0 = Don't like it at all

  For half of the sample, they are randomly shown the following statement: During the first round of surveys, we interviewed a total of 1,000 girls like you from your area of Nairobi. The majority of them told us that they do NOT like Afrobeats - meaning artists such as Kizz Daniel or Burna Boy.

Columns 4-6 - See section D.1.

D.27 Table A61

Columns 1-3 - See section D.1 and D.8.

Column 4 - *Defect vs. Sister*: For question wording, see section D.1. This variable is the proportion of times that the main respondent chose to defect across all prisoners' dilemma games she played against a sister in the household.

Column 5 - *Defect vs. Brother*: For question wording, see section D.1. This variable is the proportion of times that the main respondent chose to defect across all prisoners' dilemma games she played against a brother in the household.

E References

References

Baron, Hannah and Lauren E Young. 2022. "From principles to practice: Methods to increase the transparency of research ethics in violent contexts." *Political Science Research and Methods* 10(4):840–847.

Kassam-Adams, Nancy and Elana Newman. 2002. "The reactions to research participation questionnaires for children and for parents (RRPQ-C and RRPQ-P)." *General Hospital Psychiatry* 24(5):336–342.

URL: *https://www.sciencedirect.com/science/article/pii/S0163834302002001*