

DISCUSSION PAPER SERIES

IZA DP No. 13653

**Measuring Gender Attitudes Using List
Experiments**

M Niaz Asadullah
Elisabetta De Cao
Fathema Zhura Khatoon
Zahra Siddique

SEPTEMBER 2020

DISCUSSION PAPER SERIES

IZA DP No. 13653

Measuring Gender Attitudes Using List Experiments

M Niaz Asadullah

University of Malaya and IZA

Elisabetta De Cao

LSE and IZA

Fathema Zhura Khatoun

BRAC

Zahra Siddique

University of Bristol and IZA

SEPTEMBER 2020

Any opinions expressed in this paper are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but IZA takes no institutional policy positions. The IZA research network is committed to the IZA Guiding Principles of Research Integrity.

The IZA Institute of Labor Economics is an independent economic research institute that conducts research in labor economics and offers evidence-based policy advice on labor market issues. Supported by the Deutsche Post Foundation, IZA runs the world's largest network of economists, whose research aims to provide answers to the global labor market challenges of our time. Our key objective is to build bridges between academic research, policymakers and society.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ISSN: 2365-9793

IZA – Institute of Labor Economics

Schaumburg-Lippe-Straße 5–9
53113 Bonn, Germany

Phone: +49-228-3894-0
Email: publications@iza.org

www.iza.org

ABSTRACT

Measuring Gender Attitudes Using List Experiments*

We elicit adolescent girl's attitudes towards intimate partner violence and child marriage using purposefully collected data from rural Bangladesh. Alongside direct survey questions, we conduct list experiments to elicit true preferences for intimate partner violence and marriage before age eighteen. Responses to direct survey questions suggest that very few adolescent girls in the study accept the practise of intimate partner violence or child marriage (5% and 2%). However, our list experiments reveal significantly higher support for both intimate partner violence and child marriage (at 30% and 24%). We further investigate how numerous variables relate to preferences for egalitarian gender norms in rural Bangladesh.

JEL Classification: I15, O10, C13, C83

Keywords: list experiment, indirect response survey methods, intimate partner violence, child marriage, Bangladesh

Corresponding author:

Dr M Niaz Asadullah
Department of Economics
University of Malaya
Kuala Lumpur 50603
Malaysia
E-mail: m.niaz@um.edu.my

* This study has been funded by BRAC and Institute of Education and Development (IED) of BRAC University. We are indebted to Andrew Jenkins and Erum Mariam for their encouragement and support. We thank Julian Jamison and Giulia La Mattina for useful comments on an earlier version of the paper. In addition, we are grateful to three anonymous referees and the Editor, Klaus F. Zimmermann for their help and guidance in the review process. The usual disclaimers apply.

1. Introduction

Gender disparities in outcomes favouring men within developing countries are generally larger than in the developed world (Jayachandran, 2015). Similar disparities also often exist in attitudes towards gender equality (see for instance Asadullah and Wahhaj, 2019; Borrell-Porta, Mireia, and Philipp, *forthcoming*). Such attitudes, including attitudes supporting domestic violence and child marriage, can have important effects on behaviours, as well as playing a prominent role in explaining observed gender disparities in outcomes. For instance, Dhar et al (2016) find that parent's discriminatory attitudes reduce the aspirations of daughters to pursue schooling beyond secondary school in India. Maertens (2013) finds that parent's perception of ideal age at marriage has an adverse impact on daughter's schooling. Despite the important role that attitudes play in explaining behaviours and outcomes, much of the existing empirical literature relies on survey-based direct questions eliciting attitudes which are likely to suffer from measurement error leading to biased estimates.

One aspect of their lives in which women face a disadvantage is as a victim of violence. Women in South Asia face various forms of violence throughout their lifetime, from early childhood to old age (Solotaroff and Pande, 2014). Surprisingly, some of the global hotspots for child marriage and violence against women are also places that have seen considerable progress in poverty reduction and women's economic participation. One such developing country setting is Bangladesh where female labor force participation for women age 15+ increased from 23% in 1990 to 33% in 2017¹. While the incidence of child marriage has fallen over the recent past, it remained the case that approximately 38% of women in Bangladesh were married by the time they were 15 years of age and 74.7% were married by the time they were 18 years of age in 2014. Support for intimate partner violence is relatively high among both young and old women in Bangladesh and has remained unchanged over the recent past despite improvements in female labor force

¹ World Bank Indicators, <https://data.worldbank.org/indicator/SL.TLF.CACT.FE.ZS?view=chart>

participation (see Appendix A for more details). In 2014, 28.3% of Bangladeshi women agreed that a husband is justified in beating his wife for one of the following reasons: if the wife burns the food, if she argues with the husband, if she goes out without telling him, if she neglects the children or if she refuses to have sexual intercourse with him.² Finally, 22.4% of ever married Bangladeshi women report being a victim of physical and/or sexual violence by their husband/partner in the last 12 months.³ These kinds of disparities, which disproportionately put women at a disadvantage within the household compared to men, are likely to have serious implications for women's welfare.

The objective of this paper is to measure attitudes towards intimate partner violence and child marriage among adolescent girls in rural Bangladesh. We do this by using standard survey questions (direct questions) as well as methods to elicit responses to sensitive survey questions (list experiments, see Section 2 for a description and examples). We find that the method of measurement matters when eliciting responses to sensitive questions. Standard direct survey questions under-estimate support for socially harmful practices -- few adolescent girls in the study accept the practice of intimate partner violence (5%) or child marriage (2%) when asked directly. List experiments reveal substantially higher support for both intimate partner violence (30%) and child marriage (24%). Adolescent girls with lower levels of education under-report their support for child marriage by 16 percentage points compared to adolescent girls with higher education. We also find that girls randomly exposed to village level adolescent clubs set up by the Bangladeshi non-government organization BRAC, that educated them on marital rights and laws, under-report their support for intimate partner violence in comparison with non-exposed adolescent girls. To our knowledge, ours is the first study to use list experiment methods to elicit attitudes towards

² These estimates are based on data from the 2014 Demographic and Health Survey (DHS), <https://dhsprogram.com/topics/gender/index.cfm>.

³ These estimates are based on data from the 2007 Demographic and Health Survey which included the Domestic Violence module, <https://dhsprogram.com/topics/gender/index.cfm>.

child marriage as well as domestic violence; and has greater external validity than similar empirical investigations.

The rest of the paper is organised as follows. Section two provides a literature review and lists the contributions this paper makes to the literature. Section three describes the background of our study and list experiment that generated the data we use in this paper. Section four lays out the empirical analysis that we perform on the data as well as including a discussion of our findings, while section five concludes.

2. Literature review and contribution

An important concern when eliciting gender attitudes in surveys is related to measurement error.⁴ Suppose a survey respondent is queried on whether they consider domestic violence to be acceptable. It is very likely that they would either choose not to respond to the query (leading to systematic item non-response) or respond to the effect that they do not consider domestic violence to be acceptable (misreporting, which might arise from social desirability). In either case, the resulting measurement error will lead to biased estimates when investigating the relationship between gender attitudes and other outcome variables that we might be interested in (Bound et al, 2001).

There are different strategies that have been employed to deal with measurement error when eliciting responses to sensitive questions. One is to rely on administrative data rather than self-reports, even though in developing countries they are usually not systematically collected nor

⁴ A related literature explores the formation of attitudes regarding gender equality in developing country contexts. Beaman et al (2009) find that prior exposure to female leaders increased the chances of women's electoral success in India. They find that changes in voter's gender attitudes (measured using Implicit Association Tests) arising from exposure to female leaders are an important channel through which this change happens. Jensen and Oster (2009) use panel data to find that introduction of cable television in rural India reduced the reported acceptability of domestic violence and son preference. They also find accompanying increases in female autonomy, decreases in fertility and increased school enrolment of young children. Dhar et al (2016) examine how intergenerational transmission plays a role in the formation of gender attitudes in India.

well registered.⁵ Another is to use intensive qualitative fieldwork, as done by Blattman et al (2016), in which local researchers spend several days with a random sub-sample of survey respondents after a survey has taken place. They then obtain verbal confirmation of sensitive behaviours allowing the employment of a validation technique to examine the nature of measurement error in survey responses. Alternatively, one may use quantitative survey methods to examine measurement error in responses to sensitive questions, such as randomized response techniques, endorsement or list experiments.⁶

In this paper we use a list experiment which is also known as the item count or unmatched count technique. In a list experiment, survey respondents are queried on the number of items they agree with on a list, which (randomly) either includes or excludes a sensitive item (Miller, 1984; Imai, 2011). We use list experiments to deal with measurement error in elicited gender attitudes in Bangladesh.

Several recent empirical studies make use of list experiments. Karlan and Zinman (2012) use a list experiment to indirectly elicit how borrowers from microfinance institutions (MFI) use their loan proceeds in Peru and the Philippines. Using the results from the list experiment and comparing them with responses to direct survey questions, they find that direct elicitation underreports the non-enterprise use of the loan proceeds by borrowers from MFIs. List experiments have also been employed to elicit truthful responses to sexual behaviours, such as condom use, number of partners and unfaithfulness in Uganda (Jamison et al, 2013), Colombia (Chong et al, 2013), and Côte d'Ivoire (Chuang et al., 2019); harmful traditional practices against

⁵ Palermo et al (2014) show that administrative data in developing countries capture only a small fraction of women who experienced domestic violence, and this depends on different socio-economic characteristics of the women. See also Jamison et al. (2013) and Moseson et al. (2015).

⁶ In a randomized response technique respondents are asked to use a randomization device such as a dice or coin whose outcome is unknown to the enumerator (Warner, 1965). In an endorsement experiment, randomly selected survey respondents are asked for their support of policies which have been endorsed by a socially sensitive actor whilst other survey respondents are asked for their support for the same policies without the endorsement. If endorsement increases support for the policies then this is taken as evidence of support for the socially sensitive actor (Bullock et al, 2011).

women in Ethiopia (De Cao et al., 2017, De Cao and Lutz, 2018, Gibson et al., 2018) and anti-gay sentiment in the US (Coffman et al., 2016).

A recent study examining gender attitudes (specifically those related to female genital mutilation/cutting) in Ethiopia finds underreporting in direct attitude questions of 10 percentage points (De Cao and Lutz, 2018). This study also provides suggestive evidence that underreporting is more pronounced among uneducated women and among women who were targeted by a non-government organization (NGO) intervention to strengthen the health system as well as sexual and reproductive health knowledge.

A few recent studies have also used list experiments to examine measurement error in domestic violence reporting and prevalence. Peterman et al (2016) use a list experiment combined with an unconditional cash transfer given to female caregivers of children younger than 5 in rural Zambia. They find that 15% of the women had experienced physical intimate partner violence in the last 12 months. They also find no effect of the cash transfer on intimate partner violence four years after the program. Since direct questions are not asked, it is not possible to examine the direction or magnitude of the measurement error in such questions. Joseph et al (2017) use a list experiment in Kerala, India to find that the level of under reporting of domestic violence is over 9 percentage points, while being negligible for physical harassment on buses. They analyse the list experiment using difference-in-means across sub-groups of the population. Unlike Peterman et al (2016) and Joseph et al (2017), Agüero and Frisáncho (2018) follow WHO guidelines and protocol to ask female respondents direct questions on violence (which are comparable to the widely used domestic violence questions asked in the Demographic and Household Surveys) and compare these with a list experiment used to elicit information on experiences of physical and sexual intimate partner violence. They use a sample of female clients of a micro credit organisation operating in urban areas of Lima, Peru. Agüero and Frisáncho (2018) find that more educated women systematically underreport violence more often, but that there is no underreporting by women who have less education. They also describe a low-cost

solution to correct for bias in a setting in which there is non-classical measurement error in the dependent variable (for instance the dependent variable could be intimate partner violence as in Agüero and Frisanchó, 2018), there is no measurement error in the independent variable and endogeneity is present. This solution involves using estimates generated from list experiments carried alongside other survey instruments.

Our work contributes to this literature in several important ways. It is the first study to use a list experiment to elicit attitudes towards child marriage, at the same time being among the first to develop a list experiment for domestic violence alongside Peterman et al (2016), Joseph et al (2017) and Agüero and Frisanchó (2018). Since our sample comprises a third of all districts of Bangladesh, it has greater external validity than many of the other empirical investigations in this area (e.g., urban Lima in Agüero and Frisanchó (2018)). Ours is also the first study that makes use of an RCT (non-formal education intervention) to analyse how support for domestic violence and child marriage changes with the intervention while also using a list experiment. We analyse our list experiments by using regression techniques that allow us to investigate how the probability of supporting the sensitive question varies as a function of respondent's characteristics (as in Coffman et al., 2016, Agüero and Frisanchó, 2018, De Cao and Lutz, 2018), improving on earlier papers that only compute difference-in-means across sub-groups of the population (e.g., Karlan and Zinman, 2012, Joseph et al., 2017). Finally, we discuss the validity of our list experiments in relation to recent criticisms raised by Chuang et al. (2019).

3. Data and study design

3.1 Survey design

Data on list experiments eliciting attitudes towards domestic violence and child marriage used in this study was collected in February 2017 as part of an end line survey to evaluate the Adolescent Development Programme (ADP), a Randomized Control Trial (RCT) intervention implemented by, BRAC, the largest NGO in Bangladesh. The ADP intervention introduced village level random

variation in adolescent girls' exposure to non-formal education on marital rights and laws. The program design of the intervention is described in detail in Appendix B. The baseline survey design considered twenty-seven BRAC branch offices located in the nineteen poorest districts⁷ where BRAC was about to implement and scale up the ADP scheme by the end of 2012. Of a total 216 villages in the sample under twenty-seven BRAC branches, half were assigned to the program and the remaining served as non-program villages. Randomization was done at the village level where BRAC branches were considered as clusters i.e. it's a clustered RCT. Within the catchment area of each sample village, 20 adolescents (ages 11-16), of whom fifteen were girls and five boys, were interviewed. A total of 4,320 adolescents (3240 females) ages 11-16 years were interviewed across all villages as part of the baseline survey in June 2012.⁸ The same adolescents were interviewed in the end-line survey in February 2017, with 2,732 (or 63% of the baseline respondents) being successfully re-interviewed (of which 2020 are females).⁹ Appendix Table A1 gives a comparison of characteristics across the sample of 3,240 adolescent girls and the sub-sample of 2,020 who completed the end-line survey. An important concern is related to potentially selective attrition of subjects from baseline to end-line. However, observable characteristics of adolescent girls which are related to age, religion and education as well as their mother's age, education and empowerment measures at baseline for the complete sample and the sub-sample successfully re-interviewed at end-line are very similar, making selective attrition unlikely in this setting.

3.2 List experiments to elicit gender attitudes

The list experiment question used to elicit attitudes towards domestic violence included the following items: (1) If the father is too busy with outside work, this has a negative impact on children's education; (2) It is not acceptable to use contraceptives to avoid pregnancy; (3) In a

⁷ These were selected based on national poverty ranking.

⁸ The adolescent sample size was considered sufficient with 80 percent power and 95 percent confidence level for a 20 percent effect size (Khatoon et al, 2018).

⁹ Given random program placement at the village level, all adolescent girls in the program area had an equal probability of participation in the ADP intervention. As such, a comparison of mean outcomes across ADP program and control villages yield intention-to-treat (ITT) estimate of the program effect.

marriage both husband and wife should decide on how many children to have; (4) A wife can be hit, slapped, kicked or physically hurt by the husband under any circumstances.¹⁰

The list experiment question used to elicit attitudes towards child marriage included the following items: (1) It is important for girls to attend school (2). Birth of a girl brings as much happiness to a family as birth of a boy does (3) Literate mothers can take care of their children better than illiterate mothers (4) A girl should be married off before 18.

We randomly divided our respondents in two groups, A and B, that acted as either control or treatment for the first or second list experiment. This allowed us to reduce bias in the answers, given that only one list experiment with one sensitive item was asked from each respondent. In both list experiments, the sensitive item is the last item.¹¹ For each list experiment the control group was asked the list experiment with only items (1)-(2)-(3). We carefully selected our non-sensitive questions after discussions with BRAC, and although the items 1-3 in both list experiments seem sensitive too, in the local setting they fit as non-sensitive items given the illegality of the sensitive one.¹² Moreover, recent research shows that non-sensitive items more closely related to the sensitive one perform better because they make the sensitive item less salient (Chuang et al., 2019).

¹⁰ The instructions given the respondent before the list experiment module are as follows: “Now I will read out a number of statements to you and request you to anonymously give me your answer indicating how many statements you agree with. I’ll give you 4 stones which I’ll request you to keep in your right hand. Both of your hands will have to be kept behind you so that they’re not visible to me. If you agree with the statement I read out, transfer one stone from your right to your left hand. Please don’t show this to me or tell me verbally your answer or whether you transferred any stone from right to left hand. If you don’t agree with a sentence, do not transfer any stone from your right hand. At the end of this exercise, tell me the total number of stones in your left hand.”

¹¹ List experiments should not be too short to avoid the ceiling and floor effect, and usually include a 3-item or 4-item list (Kuklinski et al., 1997). We did not have enough power to randomize the order of the items, but it is common practice to have the sensitive item at the very end.

¹² There is a variety of punishments against perpetrators under the following relevant laws: the Domestic Violence Prevention and Protection Act of 2010; the Prevention of Oppression against Women and Children Act of 2000; the Child Marriage Restraint Act of 1929 (later revised as “the Child Marriage Restraint Act of 2017”).

In early January 2017, BRAC researchers *piloted* the list experiment questions in the Karail slums in Dhaka district with 10 female adolescents participating in the pilot. The primary objective was to verify the adequacy of the list experiment statements, and to assess the appropriateness of using stones (marbles) by interviewees to describe their responses. Stones were used to avoid numeracy-related bias in responses (as in De Cao and Lutz, 2018). The survey work was conducted by a team of 50 enumerators who received an intensive week-long training at the BRAC head office which was directly supervised by study team members as well as field management trainers from BRAC's Research and Evaluation Division (RED). The majority of enumerators (38 out of 50) were females keeping in mind the study population (where 75% adolescent respondents were female). In total, the enumerators were organized in 15 teams so that each team in a sample site had enough female enumerators available to interview a female adolescent.

The list experiment questions were asked in the last page of a long questionnaire (about forty pages long). Direct questions phrased in the same way as the sensitive item in the list experiments were asked from all respondents but at around page twenty of the questionnaire. We have no reason to believe that respondents were cognisant of this design and that the format influenced the list experiment results as so many different issues were dealt with during the interview. However, when we analyse the direct questions, we focus on the sample corresponding to the list experiment control group. See also Section 4.4 for further discussion on the validity of our list experiments.

3.3 Estimation sample

Given that the targets of the NGO intervention were primarily adolescent girls, we restrict our estimation sample to adolescent girls who responded to the list experiment questions asked in the end-line survey. This gives us an estimation sample of 2,020 adolescent girls. Table 1 reports

descriptive statistics for this sample. Half of the sample was exposed to the ADP program.¹³ 42.5% of the respondents had less than nine years of schooling or had (at most) completed junior secondary education while the rest had either secondary or tertiary education. The average age while completing the end-line survey was 17.5 years. 28% of the adolescent girls were married by the time they completed the end-line survey, of whom 71% were married before age 18.

When directly asked about gender attitudes only 2% of the adolescent girls agreed that a girl should be married off before age 18. Similarly, only 5% agreed that a wife can be hit, slapped, kicked or physically hurt by the husband under any circumstances. This is striking because of the high prevalence of early marriage as well as violence against women in the study area. For example, turning to maternal characteristics, in 15% of the cases, respondents' mothers reported being beaten at least once by their husband in the last twelve months. Moreover, about 46% of the girls' mothers in the estimation sample were pregnant before age 18. Female respondents in rural Bangladesh are also subject to patriarchal social norms – 89% of the mothers of adolescent respondents reportedly practiced Purdah¹⁴ when they went out.

4. Empirical Strategy and Results

4.1 Empirical strategy

In a standard list experiment design a sample of respondents (N) is randomly divided in two groups: control and treatment. Each respondent in the control group ($T_i = 0$, where i indicates the individual) receives a list of J non-sensitive, yes/no items, and is asked to provide the total number of items he/she agrees on. The same applies to each respondent in the treatment group ($T_i = 1$) where the list is increased by one item to include the sensitive item ($J + 1$ items). Let us assume Z_{ij}^* to be the respondent i 's truthful preference to the j th item, $j = 0, 1, \dots, J$ (Imai, 2011).

¹³ Balance tests on the ADP intervention are given in Appendix Table A2.

¹⁴ The word “Purdah” refers to female seclusion or dressing conservatively in presence of a non-related male member or when venturing outside the house. Among our female adolescent respondents (i.e. the main study population), 91% practise Purdah.

We have that $Z_{ij}(T)$ is one if the answer to the j th item is one, and zero otherwise. The econometrician only observes $Y_i = Y_i(T_i)$ where $Y_i(0) = \sum_{j=1}^J Z_{ij}(0)$ or $Y_i(1) = \sum_{j=1}^{J+1} Z_{ij}(1)$.

A list experiment is valid (Imai, 2011; Blair and Imai, 2012) if: (a) the *randomization* is good, meaning that for each respondent $\{\{Z_{ij}(0), Z_{ij}(1)\}_{j=1}^J, Z_{i,J+1}(1)\} \perp T_i$; (b) there are *no design effects* meaning that the inclusion of the sensitive item does not change the sum of affirmative answers to the non-sensitive items ($\sum_{j=1}^J Z_{ij}(0) = \sum_{j=1}^J Z_{ij}(1)$); (c) there are *no liars* meaning that the respondent replies truthfully to the sensitive item ($Z_{i,J+1}(1) = Z_{i,J+1}^*$). Assumption (c) is also called ceiling and floor effects. Ceiling effects occur when a respondent in the treatment group gives the answer $Y_i = J$ even if he/she would have replied $Y_i = J + 1$. Floor effects occur instead when a respondent in the treatment group answers $Y_i = 1$ even if he/she would have replied $Y_i = 0$.

If the list experiment satisfies (a), (b) and (c) then support for the sensitive item can be obtained by simply using a difference-in-means estimator:

$$\hat{\rho} = \frac{1}{N_1} \sum_{i=1}^N T_i Y_i - \frac{1}{N_0} \sum_{i=1}^N (1 - T_i) Y_i \quad (1)$$

Where $N_1 = \sum_{i=1}^N T_i$ is the treatment group size, and N_0 the control group size. To investigate how preferences over the sensitive item change with changes in respondent's characteristics, a multivariate regression model can be used.¹⁵ In particular, the following equation can be estimated:

$$Y_i = X_i^T \gamma + T_i X_i^T \delta + \varepsilon_i \quad (2)$$

Where X_i are the respondent's characteristics, and (γ, δ) are the parameters to estimate. We can estimate (γ, δ) using ordinary least squares (OLS).

4.2 Estimation results

¹⁵ See Imai (2011) for the methodological contribution, and De Cao and Lutz (2018) for a recent application.

In Table 2 we present the distribution of responses to our two list experiments (LE). The proportion of women in favour of domestic violence (DV) and child marriage (CM) is computed using the difference-in-means estimator and is respectively 30% (SE=0.028) and 24% (SE=0.026).¹⁶

Table 3 reports the analysis when we run a linear regression model, as in equation (2). The first four columns report the results where the outcome is the list experiment outcome for domestic violence (LE DV), while the remaining four refer to the list experiment outcome for child marriage (LE CM). Columns (1) and (5) report regressions where the list experiment outcomes are regressed only on a list experiment indicator (T_i); these correspond to the difference-in-means estimate from equation (1). The next columns in Table 3 add the most important individual characteristics. The coefficients of interest are the ones interacted with the list experiment dummy (δ). Column (2) provides the effect of being exposed to ADP on LE DV and finds a surprisingly positive effect indicating an increase in support for domestic violence. Column (4) also includes age, primary education and marital status, and it shows that ADP exposed adolescent girls are 11.4 percentage points (p-value=0.005) more likely to be in favour of domestic violence than girls not exposed to the ADP intervention even after controlling for other individual characteristics. The results of the list experiment outcome for child marriage, instead, reveal an interesting effect of education. Column (8) shows that less educated girls, who have at most completed primary education, are 16.2 percentage points (p-value=0.012) more likely to support child marriage compared to more educated girls.

We report and use robust standard errors when interpreting our results in the previous paragraph. We also compute and report p-values computed using the wild bootstrap when

¹⁶ The average answer in the control group for the LE DV question is 2.19 (SE=0.019), while in the treated it is 2.48 (SE=0.020); the difference-in-means is then 0.297 (SE=0.028). The average answer in the control group for the LE CM question is 2.48 (SE=0.019), while in the treated it is 2.71 (SE=0.019); the difference-in-means is then 0.238 (SE=0.026).

clustering at the NGO branch level (since there are only 27 NGO branches). Our results remain robust to the use of clustered standard errors at the NGO branch level.

4.3 Social desirability bias

In this section we examine social desirability bias by comparing attitudes towards domestic violence and child marriage measured via a list experiment with the same attitudes measured via a standard direct survey question (DQ). Table 1 reports that only 5% and 2% of respondents support domestic violence and child marriage when asked directly. When considering the direct question on domestic violence (DQ DV), we restrict our sample to the control group in the list experiment for domestic violence. Similarly, when considering the direct question on child marriage (DQ CM), we restrict our sample to the control group in the list experiment for child marriage. This implies that when we compare the direct question response with the list experiment, each respondent would have answered the sensitive question only once. In Table 4 we estimate linear probability models, using an indicator variable taking the value one if a girl supports domestic violence in columns (1)-(4) as the dependent variable of interest.¹⁷ An indicator variable taking the value one if a girl supports child marriage is used as the dependent variable of interest in columns (5)-(8). Explanatory variables include a girl's main characteristics (ADP exposure, age, marital status and education). While being exposed to ADP has no effect on adolescent girl's attitudes, primary education is positively associated with the probability of supporting domestic violence, while age is negatively associated with the probability of supporting child marriage. Lower educated girls (with at most primary education) are about 3 percentage points more likely to support domestic violence; while being a year older decreases the probability that child marriage is supported by 0.5 percentage points.¹⁸

¹⁷ Results are very similar if we instead compute average marginal effects from a non-linear probit model (see Table A3).

¹⁸ The R-squared is low for DQ in specifications that include an intercept only or an intercept and ADP exposure only (columns (1)-(2) and (5)-(6), Table 4); these specifications were estimated to facilitate comparison with the LE results (Table 3). We also do not find a good model fit for these specifications when using a non-linear probit model (Appendix Table A3). Inclusion of additional variables (such as education, age and whether married) improve model fit by increasing the R-squared. We interpret this as

Next, we empirically test whether there are statistically significant differences between the estimates obtained using list experiments vs. direct questions eliciting gender attitudes. This difference tells us how much the true support for domestic violence or child marriage is under-reported. The underlying assumption is that true support for domestic violence or child marriage is measured using the list experiment. A second assumption is that the measurement error in the direct questions and list experiments has the same sign. Formally, let us define $Z_{i,J+1}(0)$ as the respondent i 's potential answer to the sensitive item when asked directly (Blair and Imai, 2012). Then social desirability bias is:

$$S(x) = \Pr(Z_{i,J+1}^* = 1 | X_i = x) - \Pr(Z_{i,J+1}(0) = 1 | X_i = x) \quad (3)$$

The first term can be estimated as in equation (2), while the second can be estimated with a linear probability model regressing the observed value of $Z_{i,J+1}(0)$ on X_i . Given that the LE only allows to identify the total number of items the respondent agrees on but not which ones (e.g., $Z_{i,J+1}^*$ cannot be identified), we cannot study the social desirability bias at the individual level, but at aggregate level.

Table 5 reports the differences in the estimated proportion of girls answering the sensitive item in the affirmative when using the list experiment or the direct question by socio-demographic characteristics.¹⁹ The direct question estimates correspond to the list experiment control group sub-samples. The first row of Table 5 shows the unconditional results; these reveal a large difference of 24 percentage points in support for domestic violence, and 22 percentage points in support for child marriage. In the following rows of Table 5 we report differences in gender attitudes elicited using list experiments and direct questions by examining the estimated proportions for different groups whilst controlling for all other characteristics. All differences are

showing that it is difficult to predict individual responses to direct questions with much accuracy using the models at hand (either linear OLS or non-linear probit), at least in specifications where a full set of controls is not used.

¹⁹ Results are similar if we use a non-linear probit rather than a linear probability model to estimate the second term in equation (3), see Appendix Table A4.

highly statistically significant and between 15 and 30 percentage points. When indirectly questioned, girls seem to be much more in favour of both domestic violence and child marriage than when asked directly. Which girls under-report their support the most? By taking differences again between groups (e.g., married versus non-married, primary educated versus secondary/tertiary educated, etc) from columns (5) and (6), we find two interesting results. First, girls exposed to the ADP intervention are more likely to under-report their support (by 12 percentage points) for domestic violence compared to girls who are not exposed to the intervention (p-value=0.044). Second, less educated girls (i.e. primary schooling or below) are 15 percentage points more likely to under-report their support for child marriage than the higher educated girls (p-value=0.008).

We also estimate the proportions for the DQ using probit models. Table A6 shows the social desirability bias when the DQ predictions and their standard errors (columns (3)-(4)) come from the probit models used in Table A5. Reassuringly, the results are very similar to Table 5.

4.4 Validity of the list experiments

In Section 4.1 we discussed the conditions for list experiments to be valid. Here we discuss the validity of each of them in the context of the list experiments that we implemented. The balance tests for the *randomization* of the list experiments can be seen in Table 6. Column (5) reports the p-value of the t-test statistic where each main variable in the control group is compared with the one in the treatment group. None of the differences are statistically significant, indicating that our list experiment randomization is good.

To test if there is a violation of the design effects assumption, Blair and Imai (2012) developed a statistical test. The null hypothesis of this test indicates no design effects, and we fail to reject it.²⁰ This indicates that the inclusion of the sensitive item did not change the responses to the non-sensitive items.

²⁰ Results of this test are reported in Appendix Table A5. For technical details we refer to Blair and Imai (2012, pp. 64-65).

The third requirement for a valid list experiment is the absence of ceiling or floor effects. This assumption called *no liars* cannot be statistically tested (with the linear model used in this paper, Blair and Imai, 2012), but we can analyse the distribution of responses to our list experiments (see Table 2 and Figure 1). As can be seen, responses to the list experiments are well distributed, being mainly concentrated around 2 and 3. None of the respondents responded zero to either list experiment, but floor effects are expected to play a minor role. Table 2 shows that 4% and 5% of the respondents have no problems in revealing their support for domestic violence and child marriage, but there are quite a few girls who replied “3” to both list experiments, particularly the one on child marriage. In Table 8 we run different regressions to analyse floor and ceiling effects. We create an outcome called Floor LE DV (Floor LE CM) that takes the value one if the LE DV (LE CM) is equal to one and zero otherwise; and an outcome Ceiling LE DV (Ceiling LE CM) that takes the value one if the LE DV (LE CM) is equal to three and zero otherwise. We regress these outcomes on the main respondent characteristics for the list experiment control group. This allows us to see who is most likely hitting the floor or the ceiling and may thus be over- or under-reporting her support for domestic violence or child marriage, not reporting “0” or “4”. We find no statistically significant effect of any of those characteristics on the outcomes, except for primary education on ceiling effects for the list experiment on child marriage. This result could indicate a ceiling effect for less educated girls. Bearing in mind this limitation, it has been shown that when there are ceiling (or floor) effects, the true support for the sensitive item is underestimated (Blair and Imai, 2012).

Given that our list experiments show heterogenous effects by ADP exposure for domestic violence, and by education for child marriage, we examine the distribution of responses to the list experiments by these characteristics in Figures 2 and 3. This is not a formal test, but the idea behind these Figures is to try to understand if these girls (less educated and ADP exposed) have understood the mechanism behind the list experiment and have manipulated their results. For the sake of comparison, we report the distribution of the list experiments also for the highly educated

girls and girls not exposed to ADP. In Figure 2, we can see how responses to the list experiment on domestic violence in the different groups is well distributed, with only a small number of cases at the extremes. In Figure 3, the distribution of responses to the list experiment on child marriage shows that some girls gave the response 3, which might indicate the presence of a ceiling effect. In this case we might have an underestimate of the true support for child marriage. Tests for design effects run on the subsample of low educated, high educated, ADP exposed and ADP non-exposed adolescent girls always fail to reject the null hypothesis of no design effects (results available upon request).

In a recent paper, Chuang et al. (2019) critically examine the usefulness of indirect survey methods such as list experiments and randomized response techniques. They implement a large number of double list experiments within a single survey taken by respondents in Côte d'Ivoire where groups A and B acted as treatment and controls for the same sexual or reproductive health sensitive behaviour; in this design the non-sensitive items for groups A and B need to be different by construction. Use of double list experiments allows the generation of two difference-in-means estimators that can be compared, which to date had only been used to reduce the variance compared to a single list experiment (Droitcourt et al., 1991; Glynn, 2013). For most sensitive behaviours Chuang et al. (2019) find statistically significant differences in the two difference-in-means estimates obtained for every list experiment; they conclude by suggesting that such comparisons (which can only be carried out using double list experiments) be used to check the internal consistency of the list experiment technique.

Since we did not implement double list experiments, we cannot carry out the tests proposed in Chuang et al. (2019). We preferred to ask our respondents direct questions on attitudes to compare them with the indirect list experiment questions. We believe this method is better suited to our objective to examine social desirability bias since in a double list experiment everyone

is asked about the sensitive item twice, both directly or via the list experiment.²¹ An important exercise in Chuang et al. (2019)'s work is the variation in the type of non-sensitive items ranging from innocuous items to items related to the sensitive item. The authors find that non-sensitive items more closely related to the sensitive one perform better. None of the non-sensitive items in our list experiments are innocuous, making the sensitive item less salient, and supporting the validity of our design.

Tables A6 and A7 report similar analysis to respectively Tables 3 and 4 but with additional controls. We included controls for the following measures of maternal empowerment: if the respondent's mother has been beaten by her husband; was married early; became pregnant early; and if she practices purdah. None of these additional variables are statistically significant, except if the mother practices purdah which increases the likelihood of supporting child marriage when asked directly. Nonetheless, adding these variables does not change our main findings.

4.5 Discussion

Our findings show that measurement error is important when examining attitudes towards sensitive issues such as domestic violence or child marriage. Under-reporting can be quite high. We find that only 5.4% of adolescent girls support domestic violence when questioned directly, but 29.7% support domestic violence when questioned indirectly via a list experiment. Similar results are shown for child marriage, where 2.1% of the respondents think a girl should be married off by age 18 when asked a direct question, but support increases to 23.9% when asked via a list experiment.

Interestingly, we find girls who have lower education under-report their support for child marriage compared to girls who have higher education. To the best of our knowledge, this is the first study which implements a list experiment to examine attitudes towards child marriage,

²¹ Ideally, the direct question should only be asked from the list experiment control group to avoid potential underreporting (see our discussion on how we deal with this in our list experiment, subsection 4.1).

therefore, we cannot compare this result with existing studies. There are no heterogeneous effects by education when looking at attitudes towards domestic violence. In contrast, Aguero and Frisancho (2018) use a list experiment to study domestic violence experiences in urban Lima (Peru), and find high underreporting among the most educated respondents. This difference could be related to the different contexts or to the fact that we aim at measuring attitudes, while Aguero and Frisancho focus on behaviours. Our survey asks girls if “a wife can be hit, slapped, kicked or physically hurt by the husband under any circumstances”. In our context, girls with at most primary education might have more to lose if they do not support domestic violence, while educated girls might have better outside options (e.g., better jobs) and depend less on their husband. De Cao and Lutz (2018) examine attitudes towards female genital cutting in Ethiopia and find, similarly to us, that uneducated women are less willing to share their support for the practice.

Finally, we find suggestive evidence that the social desirability bias for domestic violence is larger among adolescent girls exposed to ADP. ADP is a random intervention, hence we can interpret its effect to be causal, even if only marginally statistically significant.²² The intervention focuses on the change in traditional attitudes through non-formal training and dissemination of information regarding sexual health, gender rights and legal provisions for violence against women including child marriage. It is certainly possible that respondents in ADP exposed areas conform to the expectations of those providing the program treatment. The ADP campaign aims at changing the local customs and this may increase social pressures around gender attitudes resulting in a stronger incentive to reveal a biased answer. We provide a more detailed comparison of the ADP program with other similar programs in developing countries in Appendix B.

5. Conclusion

²² Similar findings were found by De Cao and Lutz (2018) that study attitudes towards female genital cutting in Ethiopia. The intervention they consider, however, is not random and prevents them from claiming causal effects.

Traditional “gender attitudes” or beliefs regarding the appropriateness and/or acceptability of gender-specific roles and behaviour in society are considered important drivers of women’s well-being. While measures of gender attitudes are now included in many representative international and national surveys, they suffer from potential measurement error, limiting their usefulness in empirical research. Using a unique data set from Bangladesh, we confirm that subjective responses to sensitive direct questions under-estimate support for regressive social practices such as wife beating and child marriage. We find that girls with higher education are more supportive of egalitarian gender norms pertaining to child marriage. While we do not claim this to be a causal relationship, our finding is supportive of expanding access to education to young girls in developing country settings. We also find that exposure to a program that disseminated knowledge on gender empowerment led girls to hide their true support for domestic violence. This indicates that (at least in the short term) programs like the ADP might not have the desired effects on gender attitudes. We also find that different individual characteristics are associated with under-reporting of different aspects of gender attitudes. For instance, education matters for under-reporting of attitudes which pertain to child marriage, while ADP exposure matters for under-reporting of attitudes regarding domestic violence. This indicates that there are no simple prescriptions or general rules that apply across all aspects of gender attitudes. Our research suggests that survey methods matter in eliciting attitudes towards gendered violence and child marriage. The evidence presented in this paper also highlights the difficulty in permanently shifting gender attitudes exclusively through social empowerment programs even in a setting where girls’ schooling and economic opportunities have improved considerably in recent decades.

Our results confirm the relevance of potential bias in responses to standard direct questions when the outcome of interest is sensitive. We suggest practitioners to measure each sensitive outcome using different survey methodologies to test if there is indeed under- or over-reporting. We believe this is particularly important in the context of policy impact evaluations

where gathering complementing evidence about the effectiveness of a program or intervention is crucial when attitudes or behaviour concern sensitive topics.

Compliance with Ethical Standards

Funding: This study was funded by BRAC (Research and Evaluation Department).

Conflict of Interest: The corresponding author has received research grants from BRAC. Author C is an employee of BRAC.

References

- Aguero J, Frisancho V (2018) Misreporting of Intimate Partner Violence in Developing Countries: Measurement error and new strategies to minimize it. *Working paper*.
- Asadullah MN, Wahhaj Z (2019) "Early Marriage, Social Networks and the Transmission of Norms," *Economica*. 86(344), 801-831.
- Amin S, Saha J and Ahmed J (2018) "Skills-building programs to reduce child marriage in Bangladesh: A randomized controlled trial," *Journal of Adolescent Health*, 63(3):293-300.
- Bandiera O, Buehren N, Burgess R, Goldstein M, Gulesci S, Rasul I, Sulaiman M (2018) "Women's Empowerment in Action: Evidence from a Randomized Control Trial in Africa," CEPR Discussion Papers 13386, C.E.P.R. Discussion Papers.
- Beaman, L., R. Chattopadhyay, E. Duflo, R. Pande, and P. Topalova (2009). Powerful women: Does exposure reduce bias? *Quarterly Journal of Economics* 124(4), 1497– 1540.
- Blair, G. and K. Imai (2012) Statistical analysis of list experiments. *Polit. Anal.* 20 (1), 47–77.
- Blattman, C., J. Jamison, T. Koroknay-Palicz, K. Rodrigues and M. Sheridan (2016) Measuring the measurement error: A method to qualitatively validate survey data. *Journal of Development Economics*. 120. 99–112.
- Bound, J., Brown, C. and N. Mathiowetz (2001) Measurement error in survey data. *Handb. Econ.* 2, 3705–3843.
- Buchmann, N., E. Field, R. Glennerster, S. Nazneen, S. Pimkina, and I. Sen (2018) Power vs. money: Alternative approaches to reducing child marriage in Bangladesh, A randomized control trial. working paper. <https://www.povertyactionlab.org/sites/default/files/publications/Power-vs-Money-Working-Paper.pdf>
- Bullock, W., Imai, K., Shapiro, J.N. (2011) Statistical analysis of endorsement experiments: measuring support for militant groups in Pakistan. *Polit. Anal.* 19 (4), 363–384.
- Borrell-Porta, Mireia, Joan Costa-Font and Julia Philipp (2019) The 'mighty girl' effect: does parenting daughters alter attitudes towards gender norms?, *Oxford Economic Papers*, 71(1), 25–46.
- Chong, A., Gonzales-Navarro, M., Karlan, D., & Valdivia, M. (2013). Effectiveness and spillovers of online sex education: Evidence from a randomized evaluation in Colombian public schools. *NBER Working Paper 18776*.
- Chuang, E., Dupas P., Huillery, E. and Seban, J. (2019) Sex, lies and measurement: Do indirect response survey methods work? *Mimeo*.
- Coffman, K. B., Coffman, L. C., & Ericson, K. M. M. (2016). The size of the LGBT population and the magnitude of antigay sentiment are substantially underestimated. *Management Science*, 63(10), 3168-3186.

De Cao, E., Huis, M., Jemaneh, S. and Lensink, R. (2017). 'Community conversations as a strategy to change harmful traditional practices against women', *Applied Economics Letters*, Vol. 24, pp. 72–74.

De Cao, E., & Lutz, C. (2018). Sensitive survey questions: Measuring attitudes regarding female genital mutilation through a list experiment. *Oxford Bulletin of Economics and Statistics*, doi: 10.1111/obes.12228.

Dhar, D., T. Jan and S. Jayachandran (2016). Intergenerational Transmission of Gender Attitudes: Evidence from India. Working paper.

Dhar, D., Jain, Tarun and Jayachandran, Seema (2018) "Reshaping Adolescents' Gender Attitudes: Evidence from a School-Based Experiment in India".

Droitcour, Judith, Rachel A. Caspar, Michael L. Hubbard, Teresa L. Parsley, Wendy Visscher, and Trena M. Ezzati (1991) "The Item-Count Technique as a Method of Indirect Questioning: A Review of Its Development and a Case Study Application." In *Measurement Errors in Surveys*, edited by P. P. Biemer, R. M. Groves, L. E. Lyberg, N. A. Mathiowetz, and S. Sudman, 185–210. New York: John Wiley & Sons.

Gibson, Mhairi A. & Eshetu Gurm & Beatriz Cobo & María M. Rueda & Isabel M. Scott (2018) Indirect questioning method reveals hidden support for female genital cutting in South Central Ethiopia, PLOS ONE, <https://doi.org/10.1371/journal.pone.0193985>

Glynn, A. N. (2013) What can we learn with statistical truth serum? Design and analysis of the list experiment. *Public Opinion Quarterly*, 77(S1), 159-172.

Imai, K. (2011). Multivariate regression analysis for the item count technique. *Journal of the American Statistical Association*, 106(494), 407-416.

Jamison, J. C., Karlan, D., & Raffler, P. (2013) Mixed-method evaluation of a passive m Health sexual information texting Service in Uganda. *Information Technologies and International Development*, 9(3), 1–28.

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

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

Joseph, G., Usman Javaid, S., Andres, L. A., Chellaraj, G., Solotaro, J. L. and Rajan, S. I. (2017) Underreporting of Gender-Based Violence in Kerala, India: An Application of the List Randomization Method, Technical report, *Policy Research Working Paper N. 8044*, World Bank.

Karlan, D.S., Zinman, J. (2012) List randomization for sensitive behavior: an application for measuring use of loan proceeds. *J. Dev. Econ.* 98 (1), 71–75.

Kuklinski, J. H., Cobb, M. D. and Gilens, M. (1997). 'Racial attitudes and the "New South"', *The Journal of Politics*, Vol. 59, pp. 323–349.

Khatoon, Fathema Zhura, Khan, Safayet, Khan, Tahsina Naz and Alim, Md. Abdul (2018) “The changing status of adolescents' knowledge, attitude and practice regarding reproductive and health and social issues: An Impact Assessment of APON intervention of Adolescent Development Programme of BRAC BEP”. RED working paper.

Maertens, Annemie (2013) Social norms and aspirations: age of marriage and education in rural India. *World Development*, 47. pp. 1-15.

Miller, J. D. (1984), “A New Survey Technique for Studying Deviant Behavior,” Ph.D. thesis, The George Washington University.

Moseson, H., Massaquoi, M., Dehlendorf, C., Bawo, L., Dahn, B., Zolia, Y., Vittinghoff, E., Hiatt, R. A. and Gerds, C. (2015) ‘Reducing under-reporting of stigmatized health events using the list experiment: results from a randomized, population-based study of abortion in Liberia’, *International journal of epidemiology* 44(6), 1951–1958.

Palermo, T., Bleck, J. and Peterman, A. (2014) ‘Tip of the Iceberg: Reporting and Gender-based Violence in Developing Countries’, *American Journal of Epidemiology* 179(5), 602– 612.

Peterman, A., Palermo, T., Handa, S. and Seidenfeld, D. (2017) ‘List randomization for soliciting experience of intimate partner violence: Application to the evaluation of Zambia's unconditional child grant program’, *Health Economics Letter* pp. 1-7.

Warner, S. L. (1965) ‘Randomized response: a survey technique for eliminating evasive answer bias’, *Journal of the American Statistical Association*, Vol. 60, pp. 63–69.

TABLES AND FIGURES

Table 1. Descriptive statistics

Variables	(1) Mean	(2) SD
<u>Adolescent's characteristics</u>		
Residing in ADP exposed village (0 = ADP non-exposed, 1 = ADP exposed)	0.497	0.500
≤ Primary education (0 = secondary/tertiary, 1 = primary schooling or below)	0.425	0.494
Age	17.500	1.728
Married	0.276	0.447
Attitude DV (1 = a wife can be hit, slapped, kicked or physically hurt by the husband under any circumstances)	0.054	0.227
Attitude CM (1 = a girl should be married off before 18)	0.021	0.143
<u>Mother's characteristics (measured at baseline)</u>		
Beaten by her husband at least once in the last 12 months	0.152	0.359
Early marriage (i.e. married before age 18)	0.709	0.454
Early pregnancy (i.e. pregnant before age 18)	0.456	0.498
Practises purdah when going out	0.885	0.319
N	2,020	

Notes: Mother and adolescent specific variables (except Domestic Violence or DV and Child Marriage or CM attitudes) are measured at baseline.

Table 2. Distribution of responses to the list experiments

	Control group		Treatment group	
	Freq. (1)	Percent (2)	Freq. (3)	Percent (4)
<u>Panel A: Outcome LE DV</u>				
1	108	10.53	42	4.23
2	618	60.23	465	46.78
3	300	29.24	451	45.37
4			36	3.62
Total	1,026	100	994	100
<u>Panel B: Outcome LE CM</u>				
1	50	5.03	24	2.34
2	421	42.35	297	28.95
3	523	52.62	653	63.65
4			52	5.07
Total	994	100	1,026	100

Notes: LE refers to List Experiment, DV to Domestic Violence and CM to Child Marriage.

Table 3. Linear regression estimates for responses to the list experiments

Variables	(1) LE DV	(2) LE DV	(3) LE DV	(4) LE DV	(5) LE CM	(6) LE CM	(7) LE CM	(8) LE CM
LE	0.297 (0.028) <i>0</i>	0.245 (0.039) <i>0</i>	0.426 (0.286) <i>0.196</i>	0.324 (0.299) <i>0.321</i>	0.239 (0.026) <i>0</i>	0.233 (0.038) <i>0</i>	0.434 (0.278) <i>0.243</i>	0.382 (0.286) <i>0.316</i>
ADP exposure		-0.057 (0.038) <i>0.123</i>	-0.063 (0.038) <i>0.0917</i>	-0.063 (0.038) <i>0.0896</i>		-0.034 (0.038) <i>0.530</i>	-0.037 (0.038) <i>0.480</i>	-0.037 (0.038) <i>0.475</i>
≤ Primary education			-0.029 (0.038) <i>0.587</i>	-0.045 (0.040) <i>0.386</i>			-0.116 (0.038) <i>0.00980</i>	-0.109 (0.040) <i>0.0108</i>
Age			0.018 (0.011) <i>0.129</i>	0.014 (0.012) <i>0.233</i>			0.007 (0.011) <i>0.573</i>	0.009 (0.011) <i>0.499</i>
Married				0.052 (0.045) <i>0.408</i>				-0.026 (0.046) <i>0.536</i>
ADP exposure x LE		0.104 (0.055) <i>0.00780</i>	0.114 (0.056) <i>0.00410</i>	0.114 (0.056) <i>0.00540</i>		0.009 (0.053) <i>0.864</i>	0.005 (0.053) <i>0.929</i>	0.005 (0.053) <i>0.930</i>
≤ Primary education x LE			0.043 (0.057) <i>0.432</i>	0.066 (0.059) <i>0.256</i>			0.149 (0.054) <i>0.0170</i>	0.162 (0.057) <i>0.0118</i>
Age x LE			-0.012 (0.016) <i>0.502</i>	-0.005 (0.017) <i>0.780</i>			-0.015 (0.016) <i>0.434</i>	-0.012 (0.016) <i>0.558</i>
Married x LE				-0.085 (0.067) <i>0.235</i>				-0.041 (0.065) <i>0.431</i>
Constant	2.187 (0.019) <i>0</i>	2.215 (0.026) <i>0</i>	1.909 (0.196) <i>0</i>	1.972 (0.204) <i>0</i>	2.476 (0.019) <i>0</i>	2.493 (0.028) <i>0</i>	2.432 (0.192) <i>0</i>	2.402 (0.199) <i>0</i>
Mean of outcome	2.333	2.333	2.331	2.331	2.597	2.597	2.599	2.599
Observations	2,020	2,020	2,004	2,004	2,020	2,020	2,004	2,004
R-squared	0.054	0.056	0.057	0.058	0.039	0.040	0.045	0.046

Notes: LE refers to List Experiment, DV to Domestic Violence and CM to Child Marriage. The dependent variable is the response to the list experiment questions. It is either 0, 1, 2, 3 for respondents in the control group or 0, 1, 2, 3, 4 for respondents in the treatment group. Data on education is missing for 16 observations. Estimated coefficients are from the item count technique linear regression model (see Imai 2011). Robust standard errors in parentheses, and wild bootstrap p-values clustered at the NGO branch level in italics.

Table 4. Linear regression estimates for responses to the direct attitude questions

Variables	(1) DQ DV	(2) DQ DV	(3) DQ DV	(4) DQ DV	(5) DQ CM	(6) DQ CM	(7) DQ CM	(8) DQ CM
ADP exposure		-0.003 (0.015)	-0.002 (0.015)	-0.002 (0.015)		-0.002 (0.009)	-0.003 (0.009)	-0.003 (0.009)
		<i>0.866</i>	<i>0.890</i>	<i>0.893</i>		<i>0.778</i>	<i>0.752</i>	<i>0.756</i>
≤ Primary education			0.034 (0.016)	0.028 (0.015)			0.015 (0.009)	0.014 (0.010)
			<i>0.0490</i>	<i>0.126</i>			<i>0.0976</i>	<i>0.174</i>
Age			-0.001 (0.004)	-0.002 (0.005)			-0.005 (0.002)	-0.005 (0.002)
			<i>0.880</i>	<i>0.734</i>			<i>0.0169</i>	<i>0.00810</i>
Married				0.018 (0.019)				0.004 (0.011)
				<i>0.419</i>				<i>0.728</i>
Constant	0.058 (0.007)	0.059 (0.010)	0.060 (0.078)	0.083 (0.083)	0.019 (0.004)	0.02 (0.006)	0.096 (0.044)	0.101 (0.045)
	<i>0</i>	<i>0</i>	<i>0.601</i>	<i>0.518</i>	<i>0</i>	<i>0</i>	<i>0.0101</i>	<i>0.00400</i>
Mean of outcome	0.0575	0.0575	0.0580	0.0580	0.0191	0.0191	0.0193	0.0193
Predictions outside [0,1]	0	0	0	0	0	0	5.09%	3.74%
Observations	1,026	1,026	1,018	1,018	994	994	986	986
R-squared	0.000	0.000	0.005	0.006	0.000	0.000	0.002	0.003

Notes: The dependent variables are dummies equal to one if the respondent is in favour of Domestic Violence (DV) or Child Marriage (CM). DQ refers to Direct Question. The sample used for the DQ DV (DQ CM) corresponds to the list experiment DV control group (list experiment CM control group). Data on education is missing for 16 observations. Robust standard errors in parentheses, and wild bootstrap p-values clustered at the NGO branch level in italics.

Table 5. Social desirability bias.

	LE		DQ		Differences		Test	P-value
	Est.	SE	Est.	SE	Est.	SE		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: DV								
No covariates	0.297	0.028	0.058	0.007	0.239	0.029	8.384	0.000
Covariates	0.296	0.028	0.058	0.007	0.238	0.029	8.294	0.000
ADP non-exposed	0.240	0.040	0.059	0.010	0.181	0.041	4.393	0.000
ADP exposed	0.354	0.039	0.057	0.010	0.297	0.040	7.338	0.000
Secondary/tertiary education	0.268	0.037	0.046	0.008	0.222	0.037	5.924	0.000
Primary education (or below)	0.335	0.045	0.074	0.013	0.260	0.047	5.540	0.000
Not married	0.320	0.034	0.053	0.008	0.267	0.035	7.614	0.000
Married	0.235	0.055	0.071	0.016	0.164	0.058	2.836	0.005
Panel B: CM								
No covariates	0.239	0.026	0.019	0.004	0.219	0.027	8.203	0.000
Covariates	0.235	0.027	0.019	0.004	0.215	0.027	8.017	0.000
ADP non-exposed	0.232	0.038	0.021	0.006	0.211	0.038	5.494	0.000
ADP exposed	0.237	0.037	0.018	0.006	0.219	0.038	5.839	0.000
Secondary/tertiary education	0.166	0.035	0.013	0.005	0.152	0.035	4.295	0.000
Primary education (or below)	0.328	0.043	0.027	0.008	0.301	0.044	6.875	0.000
Not married	0.246	0.032	0.018	0.005	0.228	0.033	6.985	0.000
Married	0.205	0.054	0.022	0.010	0.183	0.054	3.354	0.001

Notes: LE refers to List Experiment, DQ to Direct Question, DV to Domestic Violence and CM to Child Marriage. Predictions are based on the linear probability model for the direct question (columns (4) and (8) of Table 4), and on the linear model for the indirect question (columns (4) and (8) of Table 3). The results are averaged over the sample distribution of covariates. Standard errors are robust.

Table 6: Tests of randomization for the list experiments

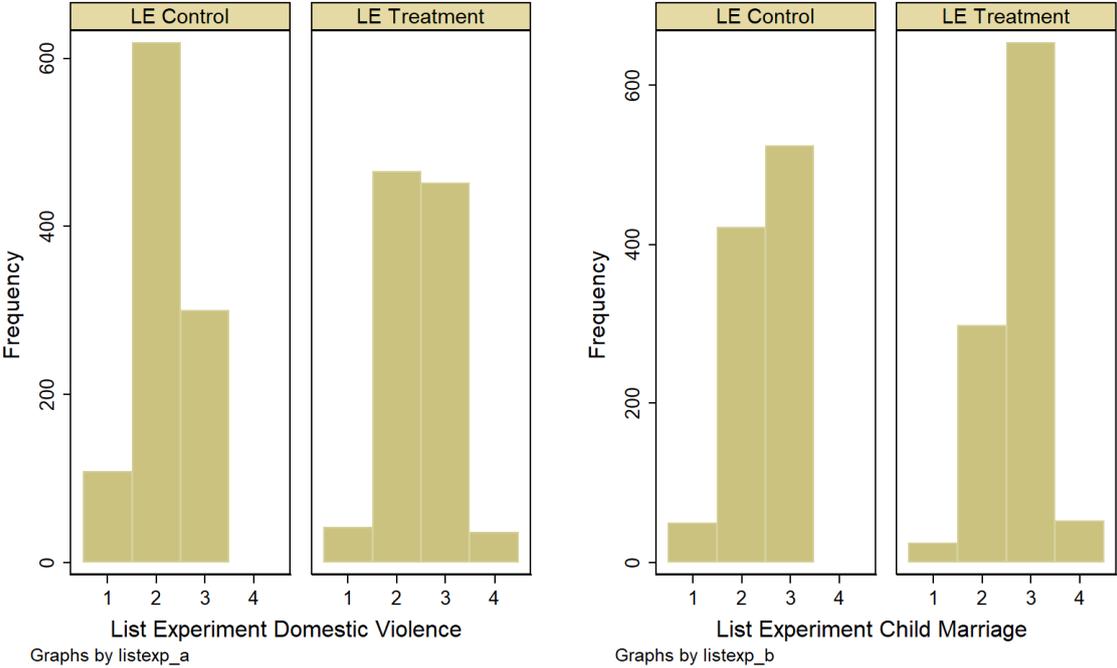
Variables	Control group		Treatment group		Diff p value
	Mean	SD	Mean	SD	
ADP exposure	0.506	0.500	0.487	0.500	0.401
≤ Primary education	0.431	0.495	0.418	0.494	0.570
Age	17.489	1.728	17.510	1.729	0.787
Married	0.283	0.451	0.270	0.444	0.523
Attitude DV (1=a wife can be hit, slapped, kicked or physically hurt by the husband under any circumstances)	0.051	0.221	0.058	0.233	0.540
Attitude CM (1=a girl should be married off before 18)	0.019	0.137	0.022	0.148	0.603
Mother beaten by her husband at least once in the last 12 months	0.142	0.349	0.162	0.369	0.203
Mother was married before age 18	0.697	0.460	0.720	0.449	0.249
Mother was pregnant before age 18	0.439	0.497	0.472	0.499	0.144
Mother practises purdah when going out	0.894	0.309	0.877	0.329	0.238

Table 7: Floor and ceiling effects

VARIABLES	(1)	(2)	(3)	(4)
	Floor LE DV	Floor LE CM	Ceiling LE DV	Ceiling LE CM
ADP exposure	0.023 (0.019) <i>0.208</i>	-0.020 (0.014) <i>0.227</i>	-0.040 (0.029) <i>0.143</i>	-0.058* (0.032) <i>0.184</i>
≤ Primary education	0.004 (0.020) <i>0.859</i>	0.018 (0.016) <i>0.269</i>	-0.040 (0.031) <i>0.269</i>	-0.091*** (0.033) <i>0.00460</i>
Age	-0.010 (0.006) <i>0.0919</i>	-0.002 (0.004) <i>0.479</i>	0.004 (0.008) <i>0.657</i>	0.006 (0.010) <i>0.576</i>
Married	-0.030 (0.020) <i>0.194</i>	-0.007 (0.017) <i>0.643</i>	0.022 (0.037) <i>0.653</i>	-0.033 (0.038) <i>0.365</i>
Constant	0.272** (0.111) <i>0.0213</i>	0.094 (0.070) <i>0.0996</i>	0.245* (0.148) <i>0.151</i>	0.496*** (0.170) <i>0.0196</i>
Mean of outcome	0.106	0.0497	0.292	0.528
N	1,018	986	1,018	986
R-squared	0.007	0.004	0.004	0.014

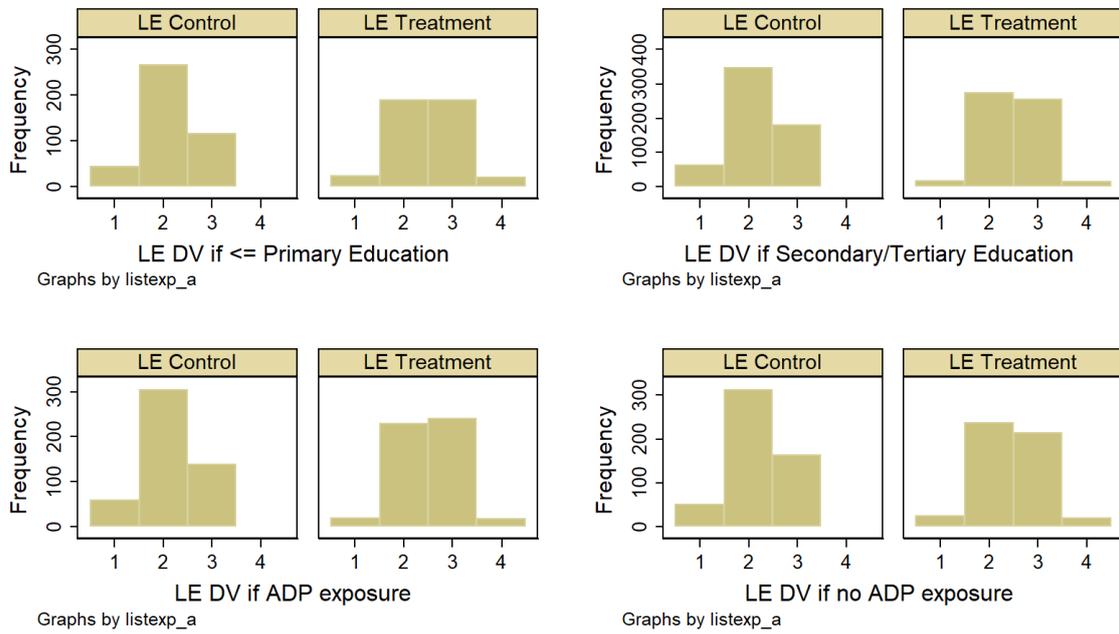
Notes: LE refers to List Experiment, DV to Domestic Violence and CM to Child Marriage. The outcome Floor LE DV (CM) is a dummy that takes value one if the LE DV (CM) is equal to one and zero otherwise. The outcome Ceiling LE DV (CM) is a dummy that takes value one if the LE DV (CM) is equal to three and zero otherwise. Robust standard errors in parentheses, and wild bootstrap p-values clustered at the NGO branch level in italics.

Figure 1: Distribution of List Experiment responses



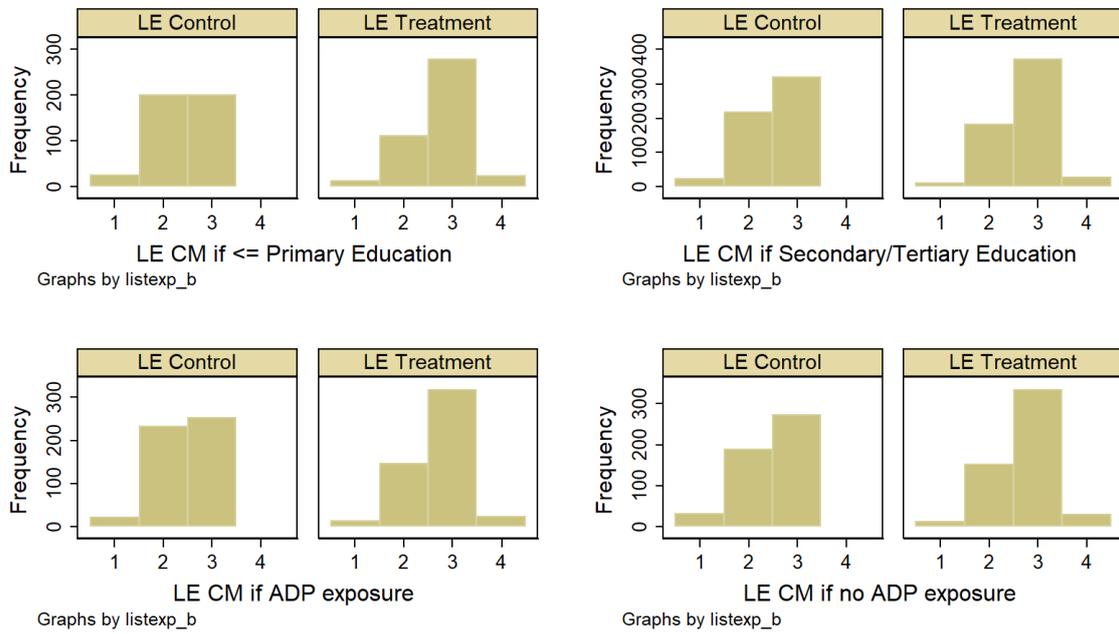
Notes: LE refers to List Experiment.

Figure 2: List Experiment responses by education and ADP exposure for Domestic Violence



Notes: LE refers to List Experiment and DV to Domestic Violence.

Figure 3: List Experiment responses by education and ADP exposure for Child Marriage



Notes: LE refers to List Experiment and CM to Child Marriage.

APPENDIX

Appendix A: Female labor force participation and gender attitudes using the Bangladesh Demographic and Health Surveys (DHS) (*for online publication only*)

To examine differences in female labor force participation, as well as the incidence of and attitudes towards child marriage and domestic violence across cohorts and over time we use data from the 2007 and 2014 Demographic and Health Surveys (DHS) for Bangladesh. These are nationally representative surveys that interview repeated cross-sections of Bangladeshi households. For the following discussion we make use of responses to the women's questionnaire from the 2007 and 2014 surveys where the respondents were ever married women from these households between the ages of 15 and 49.

Figure A1 shows the fraction within different age groups of women who report that they are currently working. While the fraction of women currently working is less than 40% within all age groups in both the 2007 and 2014 surveys, these fractions have increased over time if we compare the 2007 and 2014 respondents. Among the 2014 respondents, the fraction of currently working women has particularly increased within the older age groups of 35-39, 40-44 and 45-49 in comparison with the 2007 respondents.

Figure A2 provides the average age at first cohabitation (or marriage) within different age groups for both the 2007 and 2014 respondents. This provides us with information on the incidence of child marriage.²³ As may be seen in Figure A2, there is an increase in average age at first cohabitation over time within all age groups, and particularly among the oldest age groups of 40-44 and 45-49 year old women. For both the 2007 and 2014 respondents, average age at first cohabitation is lowest for the youngest age group 15-19 (largely because the average is over the few women who are already cohabiting at this age), and then for older women belonging to age groups 30-34, 40-44 and 45-49.

Questions on incidence of domestic violence were only asked in the 2007 DHS. Figure A3 shows the fraction of women within different age groups who experienced either less severe violence (left panel) or severe violence (right panel). While relatively high fractions of women experience less severe violence (>40% for all age groups) and severe violence (>10% for all age groups) there does not seem to be much variation in incidence across age groups. In other words, younger and older women seem to be equally likely to experience either less severe or severe violence from an intimate partner.

Next, we turn to attitudes towards domestic violence as given in Figures A4 and A5. These are constructed from a set of questions asking women whether they agree that wife beating is justified in the following situations (for 2007 respondents):

1. If the wife goes out without telling husband
2. If the wife neglects the children
3. If the wife argues with the husband
4. If the wife refuses to have sex with the husband

²³ Note that there are no questions that explicitly ask women on attitudes towards child marriage, so we cannot examine how attitudes towards child marriage vary across cohorts and over time.

In the 2014 survey, female respondents are asked if they agree with the above four statements, and, also, whether they agree that wife beating is justified:

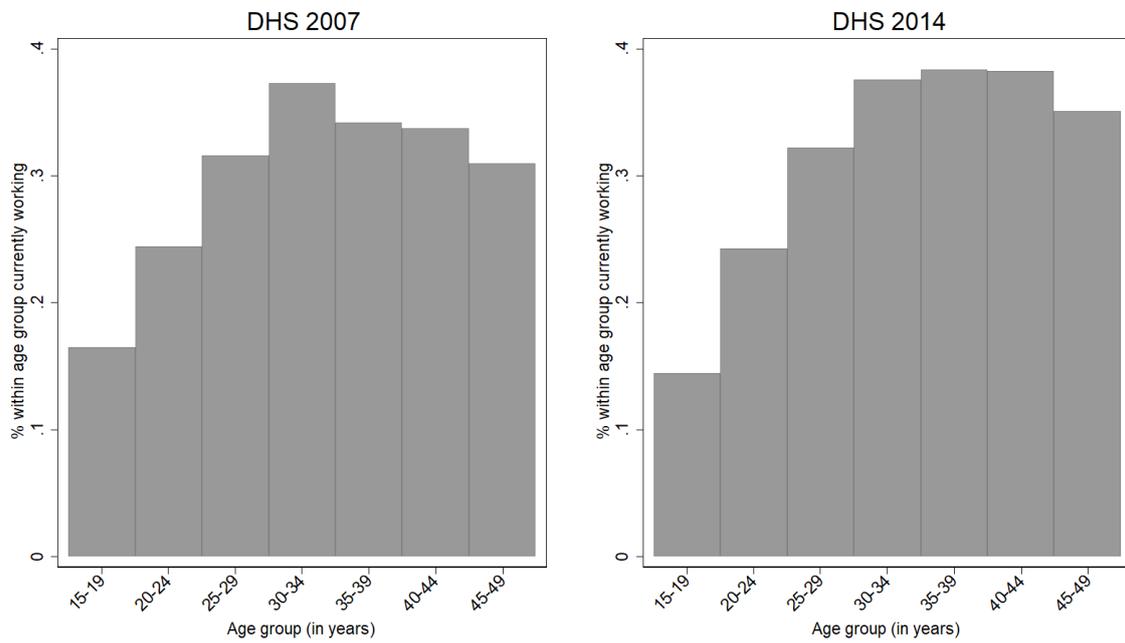
5. If the wife burns the food

Figure A4 provides a summary of responses for the 2007 respondents by age group and Figure A5 provides this summary for the 2014 respondents. Firstly, despite potential measurement error in these responses a relatively large fraction of women agree that wife beating is justified in these situations. Approximately 20% of women support wife beating in situations 1-3 and approximately 10% in situations 4-5.²⁴ Second, there is very little variation in support for wife beating across age groups within the 2007 respondents or within the 2014 respondents. In other words, younger women seem as likely to support wife beating as older women, and this is true in 2007 as well as in 2014. Finally, from a comparison of Figures A4 and A5, it does not seem that attitudes towards wife beating have changed over time since approximately the same fraction of women support wife beating in 2007 and 2014. This is despite the improvements in labor force participation over this period that we discussed earlier as shown in Figure A1.

Comparing over the recent past, Bangladesh has seen an improvement in female force participation particularly among older women and a reduction in the incidence of child marriage. However, gender attitudes specifically towards domestic violence remain unchanged. While we have not ruled out potential confounds in this descriptive discussion, the patterns we have shown indicate that improvements in female labor force participation could have driven reductions in the incidence of child marriage, but that it is *unlikely* that changes in female labor force participation led to changes in gender attitudes (at least those related to domestic violence) in Bangladesh.

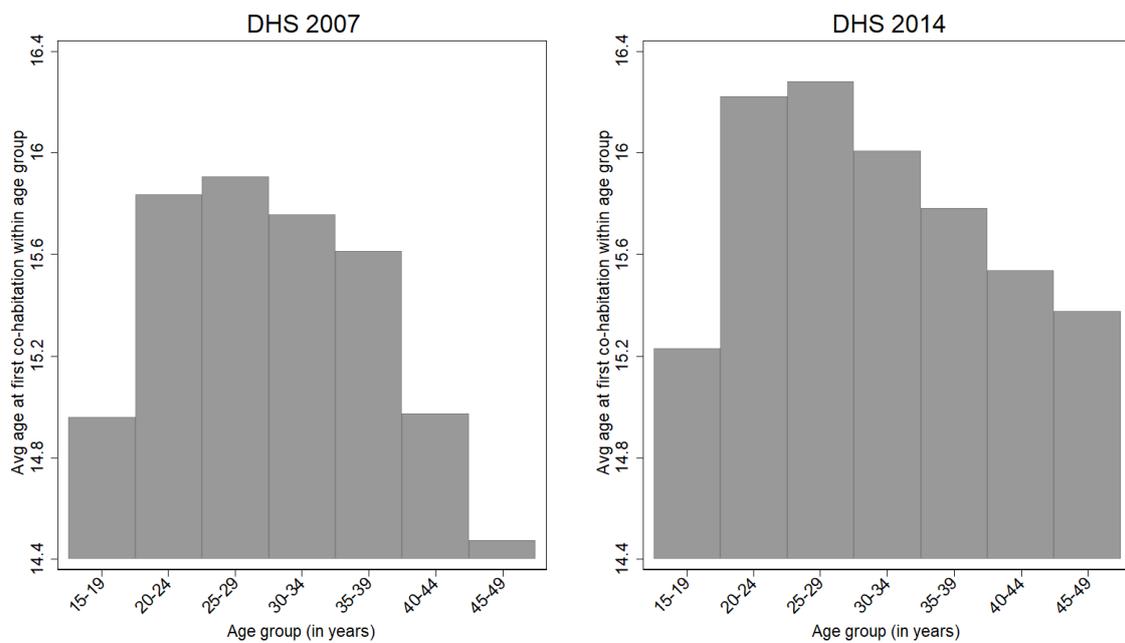
²⁴ We find a lower fraction of respondents support intimate partner violence when asked directly in our survey but our direct question is very different, asking if wife beating is justified under any rather than specific circumstances.

Figure A1: Women currently working



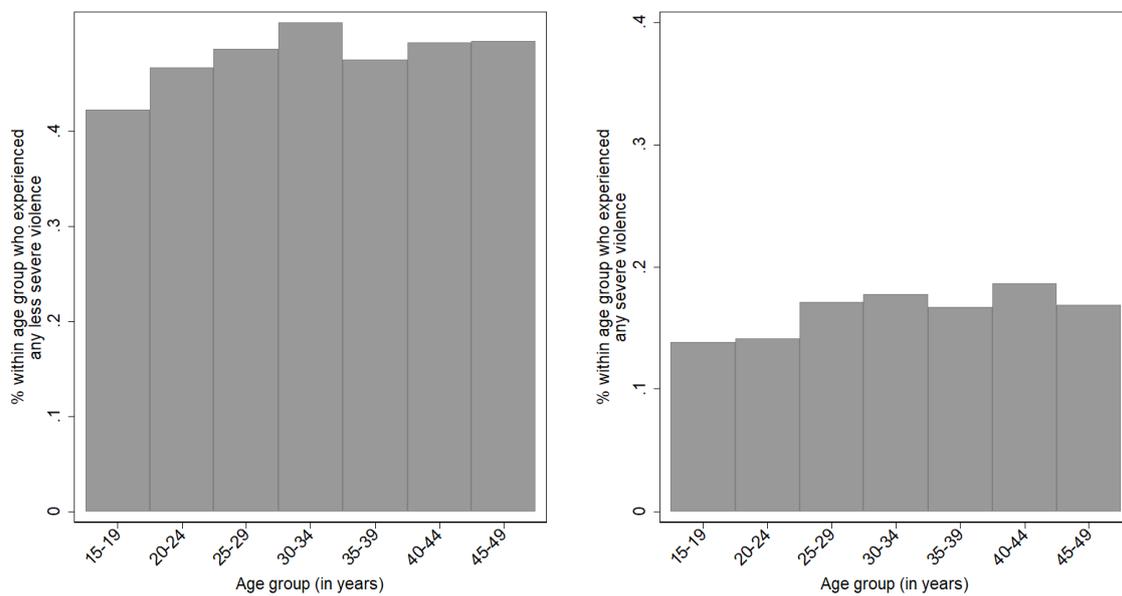
Source: Demographic and Health Surveys for Bangladesh

Figure A2: Average age at first co-habitation



Source: Demographic and Health Surveys for Bangladesh

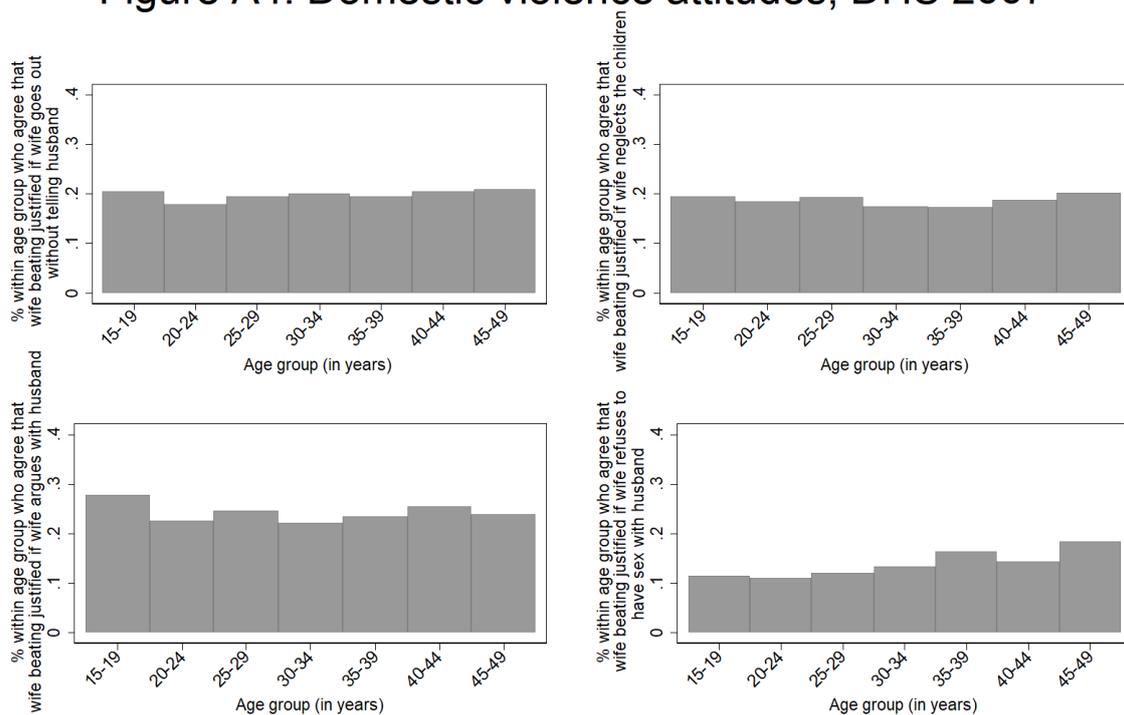
Figure A3: Domestic violence incidence, DHS 2007



Notes: Less severe violence includes pushing, slapping, punching or kicking while severe violence includes choking/burning, or threatening with a weapon.

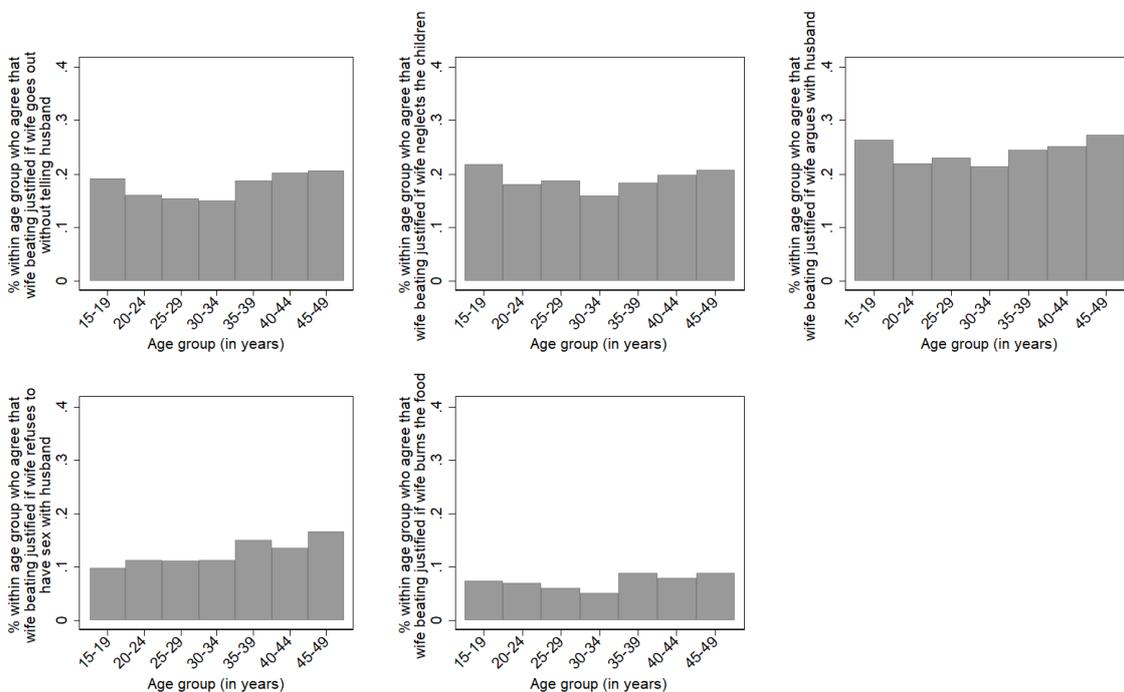
Source: Demographic and Health Survey, Bangladesh 2007

Figure A4: Domestic violence attitudes, DHS 2007



Source: Demographic and Health Survey, Bangladesh 2007

Figure A5: Domestic violence attitudes, DHS 2014



Source: Demographic and Health Survey, Bangladesh 2014

Appendix B: The ADP program (*for online publication only*)

BRAC has innovated a range of club-based adolescent development programmes which expose adolescents to a variety of activities such as (i) livelihood training courses, (ii) special network for (female) adolescent photographers, (iii) communication, awareness and advocacy through dialogues among adolescents, their parents and influential persons in the community²⁵, and (iv) the Adolescent Peer Organised Network (APON). All educational activities are organized in adolescent clubs (aka “Kishori clubs”) whereby lessons are delivered in structured courses. These clubs offer a safe space where adolescent girls can read, socialise, play games, take part in cultural activities and have an open discussion on personal and social issues with their peers. These clubs are set up at the village level using a former BRAC school building as the venue. In 2016, there were around 8,100 adolescent clubs all over Bangladesh. There are also different versions of the ADP in terms of the combination of activities conducted at the club. Our study involves a simpler version of the ADP intervention where the only activity component (or intervention) is APON. This focuses on life-skill based education on different social and health-related issues (such as reproductive health, sexual abuse, children's rights, gender, sexually transmitted infections, sexual harassment, child trafficking, substance abuse, violence, family planning, child marriage, dowry, and acid attacks) facilitated by adolescents’ peers. Club-based lessons on life-skills aside, APON activities also include book exchange, reading, playing indoor and outdoor games, performing cultural programmes and observing different international and national days.

In terms of pedagogic structure there are two hours session in a week which take place every Thursday at the club. In total, four sessions take place in a month and forty-eight sessions altogether in one year. The APON/Life skill-based education offers, in total, twelve subjects on different social and health-related issues in which thirty-one learning stories are articulated. Clubs are managed by an adolescent leader who is responsible for implementing all club activities. The leader is chosen based on leadership abilities.

Each club consists of 25-35 adolescent members of age 10-19 years, with 75% girls and the rest boys. Participation in the club is conditioned by socioeconomic status. Adolescents who dropped out from school and come from a poor socio-economic background are given priority. While individual adolescents from the eligible groups self-selected in an ADP club (i.e. participation is non-random), all eligible adolescents in ADP program village were equally exposed (i.e. intervention exposure is random) -- the study design randomly assigned treatment (i.e. the ADP Program placement) at the village level. Moreover, both program and non-program villages have benefited from BRAC’s non-formal education in the past.

Comparison with similar interventions in developing countries

There are also a number of other developing country studies that have evaluated related programs and their impact on gender attitudes and outcomes. These include the “empowerment and livelihoods for adolescents” (ELA) training scheme in Uganda (Bandiera et al., 2018), BALIKA (Bangladeshi Association for Life Skills, Income, and Knowledge for Adolescents) in Bangladesh (Amin et al. 2018) and Kishori Kendra (KK) scheme of training-based gender empowerment and financial incentives to delay marriage in Bangladesh (Buchmann et al. 2017). Both ELA and KK include safe space components where, in clubs, adolescents receive life skill lessons about gender rights and sexual education. However, they differ in other aspects. For instance, ELA simultaneously provides a vocational training component for income generating activities while KK includes a financial incentives component to delay marriage (in addition to a six month

²⁵ Those activities involves various initiative such as interactive popular theatre, adolescent fairs, cultural competition, sports for development and so on.

empowerment program) as well as an additional treatment arm offering empowerment plus incentive.

The existing developing country interventions differ considerably in terms of intensity of the treatment, target population, design and geographic coverage. For example, the ADP scheme's dosage was 96 hours total for one year. In contrast, girls in the safe space groups in Bangladesh received about 200 hours of training in over six months (Buchmann et al. (2017), 144 hours total in BALIKA scheme in Bangladesh (Amin et al. 2018) and over 500 hours in five sessions per week for two years in the ELA project in Uganda (Bandiera et al. (2018)).²⁶

In this context, ELA and KK are both variants of the scheme to which our sample respondents are exposed. However, in contrast to Bangladesh ADP scheme, ELA and KK are multifaceted programs and their evaluations have a longer window (4 years post-intervention). Moreover, these interventions did not have a conclusive impact on attitudes towards child marriage and do not report an impact on domestic violence.²⁷ Buchmann et al. (2017) contains data on a rich set of outcomes on age at marriage as well as indices for gender attitudes but the estimated impact on the standard empowerment component is insignificant. While ELA is reported to be effective in improving girls' expectations for ages at first marriage for women, the most suitable age to start childbearing and delaying pregnancy, it's not known what the impact would have been in the absence of livelihood training. Compared to KK, BALIKA and ELA, BRAC's ADP scheme studied in this paper only focuses on the standard empowerment component. While these differences can undermine the size of the program impact, they don't necessarily explain the greater support for attitude towards domestic violence among ADP participants which remains a puzzle.

²⁶ Another recent study is Dhar et al (2018) which examines a school based randomized intervention in a north Indian state (Haryana) where gender discrimination is entrenched. In contrast to club-based safe space interventions in Bangladesh and Uganda, the intervention in Dhar et al (2018) is integrated within regular classrooms/schools and conditional on school attendance and government school enrolment. The sample includes both rural and urban locations. While dosage was only a total of 20 hours in the secondary school-based program in Haryana (India), it's a non-community-level multi-year school-based intervention. This study finds a positive effect of the intervention on adolescent's support for gender equality. The evaluation study (i.e. Dhar et al 2018) relies on aggregate indices of gender attitudes and don't report treatment effect for attitude questions specific to the appropriate age of marriage for girls.

²⁷ This is also true for Dhar et al (2018) who report positive impacts on gender attitudes. However, they don't report the results separately on attitude towards domestic violence and child marriage.

APPENDIX TABLES

Table A1: Comparison of selected characteristics: baseline vs. end-line samples of girls

VARIABLES	Full sample			End-line sample		
	N	Mean	Std. Dev.	N	Mean	Std. Dev.
Adolescent girl's characteristics at baseline						
Age	3240	13.46	1.53	2020	13.27	1.53
Non-Muslim	3240	0.15	0.36	2020	0.15	0.36
Never in school	3240	0.01	0.11	2020	0.01	0.10
Currently studying	3240	0.80	0.40	2020	0.84	0.37
Years of schooling completed	3240	5.37	2.14	2020	5.29	2.09
School attended type (madrasa)	3197	0.08	0.27	1999	0.07	0.25
Ever attended BRAC non-formal school	3240	0.13	0.34	2020	0.13	0.33
Adolescent girl's characteristics at end-line						
ADP exposed (residing in ADP program area)				2020	0.50	0.50
Currently studying				2020	0.56	0.50
Years of schooling completed				2004	8.47	2.57
Married				2020	0.28	0.45
Child marriage				556	0.83	0.38
Mother's characteristics at baseline						
Age	3206	39.61	7.72	2008	39.58	7.66
Years of schooling completed	3206	1.82	2.90	2008	1.83	2.90
Mother beaten by her husband	3206	0.16	0.36	2008	0.15	0.36
Mother married early	3206	0.72	0.45	2008	0.71	0.45
Mother pregnant early	3206	0.47	0.50	2008	0.46	0.50

Table A2: Balance tests ADP intervention

Variables	Non-program group		Program group		Diff
	Mean	SD	Mean	SD	P value
≤ Primary education	0.422	0.494	0.427	0.495	0.813
Age	17.467	1.717	17.532	1.740	0.396
Married	0.276	0.447	0.276	0.447	0.995
Attitude DV (1=a wife can be hit, slapped, kicked or physically hurt by the husband under any circumstances)	0.050	0.218	0.059	0.235	0.390
Attitude CM (1=a girl should be married off before 18)	0.021	0.142	0.021	0.143	0.964
LE DV answer	2.333	0.637	2.333	0.639	0.991
LE CM answer	2.614	0.612	2.580	0.596	0.216
Mother beaten by her husband at least once in the last 12 months	0.151	0.358	0.154	0.361	0.865
Mother was married before age 18	0.706	0.456	0.712	0.453	0.778
Mother was pregnant before age 18	0.451	0.498	0.460	0.499	0.680
Mother practises purdah when going out	0.882	0.323	0.888	0.315	0.628
<i>N</i>	1017		1003		

Table A3. Average marginal effects after probit regressions for responses to the direct attitude questions

Variables	(1) DQ DV	(2) DQ DV	(3) DQ DV	(4) DQ DV	(5) DQ CM	(6) DQ CM	(7) DQ CM	(8) DQ CM
ADP exposure		-0.003 (0.015) [0.018]	-0.003 (0.015) [0.018]	-0.003 (0.015) [0.018]		-0.002 (0.009) [0.010]	-0.003 (0.009) [0.011]	-0.002 (0.009) [0.012]
≤ Primary education			0.033 (0.015) [0.016]	0.028 (0.015) [0.018]			0.014 (0.009) [0.009]	0.013 (0.010) [0.011]
Age			-0.001 (0.004) [0.006]	-0.002 (0.005) [0.006]			-0.005 (0.003) [0.002]	-0.005 (0.003) [0.002]
Married				0.015 (0.017) [0.020]				0.005 (0.011) [0.014]
Constant	0.058 (0.007) [0.010]				0.019 (0.004) [0.004]			
Pseudo R2	.	0.0001	0.011	0.013	.	0.0004	0.035	0.036
Observations	1,026	1,026	1,018	1,018	994	994	986	986

Notes: The dependent variables are dummies equal to one if the respondent is in favour of Domestic Violence (DV) or Child Marriage (CM). DQ refers to Direct Question. The sample used for the DQ DV (DQ CM) corresponds to the list experiment DV control group (list experiment CM control group). Data on education is missing for 16 observations. Robust standard errors in parentheses, and bootstrap standard errors clustered at the NGO branch level in brackets.

Table A4. Social desirability bias (based on probit regressions)

	LE		DQ		Differences		Test	P-value
	Est.	SE	Est.	SE	Est.	SE		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: DV								
No covariates	0.297	0.028	0.058	0.007	0.239	0.029	8.384	0.000
Covariates	0.296	0.028	0.058	0.007	0.238	0.029	8.296	0.000
ADP non-exposed	0.240	0.040	0.059	0.01	0.180	0.041	4.388	0.000
ADP exposed	0.354	0.039	0.056	0.01	0.297	0.040	7.349	0.000
Secondary/tertiary education	0.268	0.037	0.046	0.009	0.222	0.038	5.931	0.000
Primary education (or below)	0.335	0.045	0.074	0.012	0.261	0.047	5.550	0.000
Not married	0.320	0.034	0.053	0.009	0.266	0.035	7.602	0.000
Married	0.235	0.055	0.069	0.016	0.166	0.058	2.880	0.004
Panel B: CM								
No covariates	0.239	0.026	0.019	0.004	0.219	0.027	8.203	0.000
Covariates	0.235	0.027	0.019	0.004	0.215	0.027	8.018	0.000
ADP non-exposed	0.232	0.038	0.021	0.006	0.212	0.038	5.505	0.000
ADP exposed	0.237	0.037	0.018	0.006	0.219	0.038	5.830	0.000
Secondary/tertiary education	0.166	0.035	0.013	0.005	0.152	0.035	4.291	0.000
Primary education (or below)	0.328	0.043	0.026	0.008	0.302	0.044	6.906	0.000
Not married	0.246	0.032	0.018	0.005	0.228	0.033	6.994	0.000
Married	0.205	0.054	0.023	0.01	0.182	0.055	3.336	0.001

Notes: LE refers to List Experiment, DQ to Direct Question, DV to Domestic Violence and CM to Child Marriage. Predictions are based on the probit model for the direct question (columns (4) and (8) of Table A3), and on the linear model for the indirect question (columns (4) and (8) of Table 3). The results are averaged over the sample distribution of covariates. Standard errors are robust.

Table A5: Design effects. Estimated respondent types for the list experiments.

	LE on DV		LE on CM	
	Est	SE	Est	SE
$\pi(y = 0, t = 1)$	0.000	0.000	0.000	0.000
$\pi(y = 1, t = 1)$	0.064	0.012	0.026	0.008
$\pi(y = 2, t = 1)$	0.196	0.021	0.160	0.022
$\pi(y = 3, t = 1)$	0.037	0.006	0.050	0.007
$\pi(y = 0, t = 0)$	0.000	0.000	0.000	0.000
$\pi(y = 1, t = 0)$	0.043	0.006	0.024	0.005
$\pi(y = 2, t = 0)$	0.406	0.019	0.262	0.016
$\pi(y = 3, t = 0)$	0.255	0.015	0.478	0.017

Notes: The table shows the estimated proportion (and standard error) of respondent types, π_{yt} , characterised by the total number of affirmative answers to the control questions, y , and the truthful answer for the sensitive item. The null hypothesis of no design effects implies that $\pi_{yt} \geq 0$ for all y and t . Since all proportions are positive, we cannot reject the null. The SE of these tests are robust.

Table A6: Linear regression estimates for responses to the list experiments, controlling for different measures of maternal empowerment

VARIABLES	(5)	(6)	(7)	(8)	(5)	(6)	(7)	(8)
	LE DV	LE DV	LE DV	LE DV	LE CM	LE CM	LE CM	LE CM
LE	0.480 (0.286) <i>0.155</i>	0.414 (0.289) <i>0.206</i>	0.447 (0.286) <i>0.187</i>	0.535 (0.296) <i>0.130</i>	0.438 (0.279) <i>0.236</i>	0.436 (0.284) <i>0.235</i>	0.435 (0.280) <i>0.244</i>	0.452 (0.293) <i>0.210</i>
ADP exposure	-0.067 (0.038) <i>0.0724</i>	-0.067 (0.038) <i>0.0708</i>	-0.067 (0.038) <i>0.0705</i>	-0.069 (0.038) <i>0.0575</i>	-0.040 (0.038) <i>0.458</i>	-0.040 (0.038) <i>0.454</i>	-0.039 (0.038) <i>0.456</i>	-0.039 (0.038) <i>0.465</i>
ADP exposure x LE	0.115 (0.056) <i>0.00490</i>	0.114 (0.056) <i>0.00490</i>	0.115 (0.056) <i>0.00490</i>	0.116 (0.056) <i>0.00490</i>	0.007 (0.053) <i>0.893</i>	0.007 (0.053) <i>0.891</i>	0.007 (0.053) <i>0.896</i>	0.005 (0.053) <i>0.932</i>
≤ Primary education	-0.032 (0.038) <i>0.549</i>	-0.026 (0.038) <i>0.630</i>	-0.026 (0.038) <i>0.629</i>	-0.032 (0.038) <i>0.555</i>	-0.110 (0.039) <i>0.0138</i>	-0.112 (0.039) <i>0.0104</i>	-0.110 (0.038) <i>0.0116</i>	-0.116 (0.039) <i>0.0109</i>
Age	0.019 (0.011) <i>0.111</i>	0.019 (0.011) <i>0.125</i>	0.019 (0.011) <i>0.119</i>	0.018 (0.011) <i>0.122</i>	0.006 (0.011) <i>0.587</i>	0.007 (0.011) <i>0.561</i>	0.007 (0.011) <i>0.578</i>	0.007 (0.011) <i>0.558</i>
Mother beaten by her husband	0.065 (0.051) <i>0.273</i>				-0.019 (0.051) <i>0.731</i>			
Mother married early		-0.030 (0.042) <i>0.492</i>				0.026 (0.042) <i>0.588</i>		
Mother pregnant early			-0.037 (0.038) <i>0.250</i>				-0.010 (0.038) <i>0.802</i>	
Mother practices purdah				0.114 (0.056) <i>0.168</i>				0.108 (0.067) <i>0.155</i>
≤ Primary education x LE	0.044 (0.057) <i>0.420</i>	0.035 (0.057) <i>0.540</i>	0.036 (0.057) <i>0.522</i>	0.041 (0.057) <i>0.440</i>	0.141 (0.054) <i>0.0239</i>	0.141 (0.054) <i>0.0268</i>	0.141 (0.054) <i>0.0216</i>	0.144 (0.054) <i>0.0188</i>
Age x LE	-0.014 (0.016) <i>0.436</i>	-0.013 (0.016) <i>0.472</i>	-0.013 (0.016) <i>0.454</i>	-0.012 (0.016) <i>0.483</i>	-0.015 (0.016) <i>0.422</i>	-0.016 (0.016) <i>0.420</i>	-0.016 (0.016) <i>0.413</i>	-0.017 (0.016) <i>0.387</i>
Mother beaten by her husband x LE	-0.120 (0.079) <i>0.137</i>				0.029 (0.069) <i>0.710</i>			
Mother married early x LE		0.043 (0.061) <i>0.488</i>				0.013 (0.060) <i>0.813</i>		
Mother pregnant early x LE			0.021 (0.056) <i>0.711</i>				0.028 (0.053) <i>0.542</i>	

Mother practices purdah x LE				-0.113 (0.088) <i>0.275</i>				0.013 (0.087) <i>0.856</i>
Constant	1.886 (0.197) <i>0.00</i>	1.926 (0.199) <i>0.00</i>	1.913 (0.197) <i>0.00</i>	1.815 (0.203) <i>0.00</i>	2.434 (0.193) <i>0.00</i>	2.410 (0.197) <i>0.00</i>	2.435 (0.193) <i>0.00</i>	2.328 (0.204) <i>0.00</i>
Mean of outcome	2.331	2.331	2.331	2.331	2.598	2.598	2.598	2.598
N	1,992	1,992	1,992	1,992	1,992	1,992	1,992	1,992
R-squared	0.058	0.057	0.058	0.059	0.044	0.045	0.044	0.048

Notes: LE refers to List Experiment, DV to Domestic Violence and CM to Child Marriage. The dependent variable is the response to the list experiment questions. It is either 0, 1, 2, 3 for the respondents in the control group or 0, 1, 2, 3, 4 in the treatment group. Estimated coefficients from the item count technique linear regression model. Robust standard errors in parentheses, and wild bootstrap p-values clustered at the NGO branch level in italics.

Table A7. Linear regression estimates for responses to the direct attitude questions, controlling for different measures of maternal empowerment.

VARIABLES	(1) DQ DV	(2) DQ DV	(3) DQ DV	(4) DQ DV	(5) DQ CM	(6) DQ CM	(7) DQ CM	(8) DQ CM
ADP exposure	-0.002 (0.015) <i>0.888</i>	-0.003 (0.015) <i>0.885</i>	-0.002 (0.015) <i>0.889</i>	-0.003 (0.015) <i>0.874</i>	-0.003 (0.009) <i>0.757</i>	-0.003 (0.009) <i>0.760</i>	-0.003 (0.009) <i>0.791</i>	-0.003 (0.009) <i>0.771</i>
≤ Primary education	0.033 (0.016) <i>0.0472</i>	0.034 (0.016) <i>0.0550</i>	0.033 (0.016) <i>0.0642</i>	0.033 (0.015) <i>0.0480</i>	0.015 (0.009) <i>0.101</i>	0.015 (0.009) <i>0.105</i>	0.016 (0.009) <i>0.0890</i>	0.014 (0.009) <i>0.123</i>
Age	-0.001 (0.004) <i>0.906</i>	-0.001 (0.004) <i>0.892</i>	-0.001 (0.004) <i>0.869</i>	-0.001 (0.004) <i>0.880</i>	-0.005 (0.002) <i>0.0222</i>	-0.005 (0.002) <i>0.0176</i>	-0.005 (0.002) <i>0.0173</i>	-0.005 (0.002) <i>0.0202</i>
Mother beaten by father	0.015 (0.022) <i>0.525</i>				-0.000 (0.014) <i>0.985</i>			
Mother married early		0.005 (0.016) <i>0.817</i>				-0.002 (0.010) <i>0.834</i>		
Mother pregnant early			0.019 (0.015) <i>0.397</i>				-0.015 (0.008) <i>0.0834</i>	
Mother practices purdah				0.019 (0.019) <i>0.428</i>				0.019 (0.005) <i>0.000300</i>
Constant	0.055 (0.078) <i>0.637</i>	0.056 (0.076) <i>0.603</i>	0.054 (0.077) <i>0.627</i>	0.045 (0.078) <i>0.687</i>	0.095 (0.046) <i>0.0143</i>	0.097 (0.046) <i>0.0100</i>	0.104 (0.045) <i>0.00910</i>	0.077 (0.043) <i>0.0288</i>
Mean of outcome	0.0582	0.0582	0.0582	0.0582	0.0194	0.0194	0.0194	0.0194
N	1,014	1,014	1,014	1,014	978	978	978	978
R-squared	0.006	0.005	0.007	0.006	0.007	0.007	0.010	0.009

Notes: DQ refers to Direct Question, DV to Domestic Violence and CM to Child Marriage. The dependent variables are dummies equal to one if the respondent is in favour of DV or CM. Robust standard errors in parentheses, and wild bootstrap p-values clustered at the NGO branch level in italics.