



USAID IMPACT AND PERFORMANCE EVALUATION OF THE WAACHE WASOME PROJECT (TANZANIA)

Endline Report

Prepared under Contract No.: GC-I0F-0033M/AID-OAA-M-I3-00010 , Tasking N 015.03

USAID READING AND ACCESS

USAID IMPACT AND PERFORMANCE

EVALUATION OF THE WAACHE WASOME PROJECT (TANZANIA)

Endline Report

(MAY 2022)

Prepared under Contract No.: GC-I0F-0033M/AID-OAA-M-I3-00010, Tasking N 015.03

Submitted to:

Benjamin Silla, USAID COR

Submitted by:

Alejandro Ome, Alicia Menendez, Letitia Onyango, Jessica Wallach, and Marie-Celine Schulte

Contractor:

NORC at the University of Chicago
4350 East West Highway, 8th Floor
Bethesda, MD 20814
Attention: Varuni Dayaratna

DISCLAIMER

The authors' views expressed in this publication do not necessarily reflect the views of the United States Agency for International Development or the United States Government.

TABLE OF CONTENTS

LIST OF ACRONYMS.....	v
EXECUTIVE SUMMARY	vi
CONTEXT AND PROJECT BACKGROUND	vi
EVALUATION PURPOSE & EVALUATION QUESTIONS.....	vi
IMPACT EVALUATION	vii
PERFORMANCE EVALUATION	vii
EVALUATION METHODS AND DATA COLLECTION.....	vii
IMPACT EVALUATION	vii
PERFORMANCE EVALUATION	viii
A1. PROJECT BACKGROUND.....	viii
RESULTS AND CONCLUSIONS.....	x
A. INTRODUCTION	1
A2. CONTEXT.....	1
A3. EVIDENCE	1
B. EVALUATION PURPOSE AND QUESTIONS.....	2
IMPACT EVALUATION.....	2
PERFORMANCE EVALUATION.....	3
C. EVALUATION METHODS.....	4
C1. IMPACT EVALUATION.....	4
IMPACT ON GIRLS.....	5
PARENTS	6
TEACHERS.....	6
C2. PERFORMANCE EVALUATION	6
C3. DATA COLLECTION	7
QUANTITATIVE DATA	7
QUALITATIVE DATA.....	8
D. IMPACT EVALUATION RESULTS	9
D1. IMPACT ON ATTENDANCE, RETENTION, AND PROMOTION	9
PROGRAM TAKE-UP	14
D2. INTERMEDIATE OUTCOMES - GIRLS	16
PROTECTIVE ASSETS AND ATTITUDES TOWARDS GENDER NORMS, ROLES & EQUITY	16
ATTITUDES TOWARDS SRGBV	17

D3. INTERMEDIATE OUTCOMES - PARENTS.....	20
PROGRAM TAKE-UP	21
CHANGES BETWEEN BASELINE AND ENDLINE.....	22
D4. INTERMEDIATE OUTCOMES - TEACHERS.....	25
PROGRAM TAKE-UP	25
PROGRAM IMPACT	27
E. DISCUSSION	35
TAKE-UP AND SUBSTITUTION.....	35
IMPACT ON INTERMEDIATE OUTCOMES.....	36
POWER AND SAMPLE SIZE.....	37
F. PERFORMANCE EVALUATION RESULTS	37
F1. POY CLUBS.....	37
F2. LIMCA GROUPS.....	44
F3. SRGBV TRAINING.....	49
ALTERNATIVE DISCIPLINE METHODS	52
F4. SUSTAINABILITY.....	55
G. CONCLUSIONS AND PROGRAMMATIC RECOMMENDATIONS	57
REFERENCES.....	61
ANNEX I. RELEVANT LITERATURE.....	65
ANNEX II. SELECTING ARD GIRLS IN THE CONTROL GROUP.....	67
ANNEX III. INSTRUMENTS.....	69
ANNEX IV. DATA COLLECTION	70
ANNEX V. MONTHLY ATTENDANCE DATA	80
ANNEX VI. SURVEY ATTRITION	82
ANNEX VII. ADDITIONAL DESCRIPTIVE STATISTICS AND RESULTS	83
ANNEX VIII. GIRLS' ATTITUDES TOWARDS SRGBV.....	86
ANNEX IX. LIMCA AS A HEDGING INSTRUMENT.....	91
ANNEX X. INTRAHOUSEHOLD INVESTMENTS IN EDUCATION	94
ANNEX XI. DETERMINANTS OF SCHOOL RETENTION	97
ANNEX XII. STATEMENT OF WORK.....	101

LIST OF TABLES

Table 1: Girls' retention rates using administrative and survey data	xi
Table 2: Fieldwork results – Quantitative data	7
Table 3: Fieldwork results – Qualitative data	8
Table 4: Retention rates by gender and treatment status	9
Table 5: The impact of WW on attendance	10
Table 6: Number of girls surveyed at baseline and endline by treatment and POY club status.....	11
Table 7: Survey and school retention as fraction of the girls surveyed at baseline, by treatment status ..	12
Table 8: Impact of WW on ARD girls' school retention.....	13
Table 9: Impact of WW on school retention of girls randomly selected for POY clubs.....	14
Table 10: Club participation.....	15
Table 11: Meeting attendance and credit access (N=305).....	21
Table 12: Participation in savings clubs according to the girl survey (N=1792).....	21
Table 13: Teachers' knowledge about child protection regulations and types of SRGBV (n=589).....	29
Table 14: Program take-up and substitution	36
Table 15: Weighting the control group so it is comparable to ARD girls.....	68
Table 16: Characteristics of panel and non-panel girls at baseline	82
Table 17: Protective assets descriptive statistics at endline.....	84
Table 18: Sociodemographic characteristics at baseline of the panel of LIMCA parents.....	85
Table 19: Teachers' characteristics	85
Table 20: List experiments on SRGBV	89
Table 21: Experienced negative shocks and type of shocks (fraction).....	91
Table 22: Determinants of school retention – Linear probability models.....	98

LIST OF FIGURES

Figure 1: Theory of Change.....	x
Figure 2: Reasons for girls not being enrolled in school according to teachers (N=646 girls)	10
Figure 3: Self-reported reasons for having dropped out of school (N=189)	12
Figure 4: Satisfaction with POY clubs (N=713).....	15
Figure 5: Mean difference between treatment and control at endline on protective assets and gender attitudes with 95 percent confidence interval	17
Figure 6: Perceptions of SRGBV by Treatment Group	19
Figure 7: Fraction of parents answering agree or strongly agree.....	22
Figure 8: Changes in parenting style scores	24
Figure 9: Fraction of Parents saying statements on disciplinary strategies are true.....	24
Figure 10: Days of SRGBV training (N=280*).....	26
Figure 11: Attendance at Lunch and Learn sessions (N=278).....	27
Figure 12: Fraction of teachers that agree to gender-related questions (N=589).....	28
Figure 13: Fraction of teachers that think they have a role and responsibility to intervene when students experience violence from... (N=589).....	29
Figure 14: Teachers' perceptions of physical violence in schools (N=589).....	30
Figure 15: Teachers' perceptions of proportions of teachers and students who perpetrate sexual violence at school (N=589)	31

Figure 16: Frequency at which teachers engage in positive, non-violent versus harsh disciplinary methods	32
Figure 17: Teachers self-reported response on to whom they will report in the school	34
Figure 18: Teachers self-reported response on to whom they will report outside the school	35
Figure 19: LIMCA members with children (under age 19).....	47
Figure 20: Attendance rates by gender and treatment status - 2018	80
Figure 21: Attendance rates by gender and treatment status - 2019	80
Figure 22: Attendance rates by gender and treatment status – 2020.....	81
Figure 23: Attendance rates by gender and treatment status - 2021	81
Figure 24: Beliefs on peers’ opinions on SRGBV by treatment group – Emotional violence	86
Figure 25: Beliefs on peers’ opinions on SRGBV by treatment group – Physical violence	87
Figure 26: Beliefs on peers’ opinions on SRGBV by treatment group – Sexual violence.....	88
Figure 27: Fraction of households that received a loan from LIMCA by whether they experienced a shock in previous 12 months	92
Figure 28: Response to shocks (multiple responses allowed).....	93
Figure 29: LIMCA loan uses (multiple responses were allowed)	93
Figure 30: Number of children/youth living with the respondent by age and gender.....	94
Figure 31: School enrollment by age, gender, and wave of data collection	95
Figure 32: Average time spend studying at home and on chores by age and gender (hours per day)	96

LIST OF ACRONYMS

ARD	At Risk of Dropping Out
DBE	Department of Basic Education
COSTECH	Tanzania Commission for Science and Technology
EDR	Evaluation Design Report
EEV	Economic Empowerment Volunteers
FGD	Focus Group Discussion
GoT	Government of Tanzania
IE	Impact Evaluation
KII	Key Informant Interview
LIMCA	Livelihood Improvement for Mothers and Caregivers of Adolescents
LGA	Local Government Authority
LGL	Let Girls Learn
MDES	Minimum Detectable Effect Size
NORC	NORC at the University of Chicago
PE	Performance Evaluation
POY	Protect Our Youth
PMP	Performance Monitoring Plan
RCT	Randomized Controlled Trial
RTI	Research Triangle Institute
SRGBV	School-Related Gender-Based Violence
TDHS	Tanzania Demographic and Health Survey
USAID	United States Agency for International Development
WEI	World Education Inc.
WW	Waache Wasome

EXECUTIVE SUMMARY

CONTEXT AND PROJECT BACKGROUND

Waache Wasome ('Let Them Learn' in Kiswahili) is a five-year program aimed at increasing adolescent girls' enrollment and retention in secondary school in four districts of Tanzania. The program was designed and implemented by the Bantwana Initiative of World Education (WEI/B). Under the Reading and Access contract, NORC conducted an evaluation of *Waache Wasome*, adopting a distinctive impact and performance evaluation (IE+) methodology.

Waache Wasome (WW) followed a holistic approach to try and keep girls in school, working directly with them, as well as within their school and home environments. To achieve its goal of increasing school retention, WW sought to:

- 1. Build the agency, knowledge, and protective assets of girls in secondary school.** The program implemented Protect Our Youth (POY) clubs, providing safe spaces in schools where adolescent girls could learn to navigate the risks they face and build protective assets that reduce vulnerabilities, helping them to remain and succeed in school.
- 2. Increase family commitment and capacity to invest in girls' education.** Savings groups for caregivers of adolescents, called Livelihood Improvement for Mothers and Caregivers of Adolescents (LIMCA), were assembled as part of program activities. LIMCA was also used as a channel to deliver training to improve parental skills and gender norms.
- 3. Foster a girl friendly and supportive school environment.** To promote a supportive environment in schools, the program conducted two activities: (1) training on school-related gender-based violence (SRGBV) prevention and response for teachers and school administrators, and (2) peer study groups.

The underlying theory of change for WW is that the interventions described above will provide both a supportive environment and a comprehensive set of resources, so girls would be more likely to succeed and stay in school.

EVALUATION PURPOSE & EVALUATION QUESTIONS

The objectives of this evaluation are to:

- Evaluate the impact of the WW program as a whole (including all its components) on school retention and protective assets.
- Assess the performance of three WW components, namely POY clubs, LIMCA, and SRGBV training.
- Draw conclusions based on impact and performance evaluation findings and develop recommendations for fine-tuning the program to improve impact and implementation and ensure sustainability.

The impact and performance evaluation questions addressed in the study are as follows:

IMPACT EVALUATION

- IE1. What is the impact of the program on retention, attendance, and promotion?
- IE2. What is the impact of the POY club component on retention, attendance, and promotion for girls attending treatment schools?
- IE3. Does the program strengthen girls' protective assets?
- IE4. Is the WW package changing gender norms?
- IE5. Does the program change attitudes on SRGBV?
- IE6. Does the program improve attitudes towards gender equitable access to education within the household?
- IE7. Does the program change parenting skills?
- IE8. Does the program change gender attitudes and perceptions of teachers in school?
- IE9. Does the program change teachers' knowledge, attitudes, and practices related to SRGBV?

PERFORMANCE EVALUATION

- PE1. How are girls and boys benefitting from POY clubs?
- PE2. Is LIMCA benefitting its members, girls, and other stakeholders as intended by WW?
- PE3. What are the benefits of SRGBV prevention and response teacher training to teachers, students, and other stakeholders?
- PE4. Are the interventions addressed in PE1-3 sustainable?

EVALUATION METHODS AND DATA COLLECTION

WW was implemented in the districts of Karatu DC, Musoma DC, Musoma MC, and Tarime DC. Of 67 schools targeted to benefit from WW, the evaluation excluded the six Karatu DC schools that were part of the pilot, which started in 2017. Thus, a total of 61 schools are part of the evaluation.

IMPACT EVALUATION

To evaluate the impact of the program, NORC worked with WEI to conduct a randomized control trial. The evaluation followed a pipeline design where 30 schools were randomly selected to the treatment group (start treatment in February 2018) and 31 schools to the control group (start treatment in February 2020).

Two types of data were collected to estimate the impact of the program on girls' outcomes: survey data and school administrative data. Survey data was used to document program impacts on retention rates and intermediate outcomes, namely girls' protective assets, gender attitudes, and perceptions on SRGBV. Baseline data was collected in February 2018 and endline data was collected in February 2020.

School administrative data was used to evaluate the program on school attendance. Administrative data was collected for the years 2018-2021.

To document research questions related to parenting skills and parents' attitudes towards gender equitable access to education, we use data collected through parents' surveys. The main, and to some extent, only direct interaction between the program and parents will be LIMCA; therefore, it is natural to focus the evaluation efforts on the impact of this component when we are trying to evaluate the effect of WW on parents.

The evaluation focuses on a sample of parents that signed up for LIMCA. Specifically, we surveyed them after savings groups were constituted, so we have a baseline for LIMCA parents, and survey these same parents again at endline. No comparison group was surveyed, so the analysis can be described as a before-and-after evaluation. Parent baseline data was collected in February 2019 and endline data in February 2020.

Finally, teachers in each of the treatment and control schools were surveyed at endline only (February 2020). In this survey we recorded teachers' gender attitudes, teachers' knowledge, attitudes, and practices on SRGBV. Using these data, we compare teachers in treatment and control schools to estimate the impact of the program on teachers.

PERFORMANCE EVALUATION

The performance evaluation of WW assesses the performance of three interventions: POY clubs, LIMCA activities, and the SRGBV training for school teachers/administrators. The evaluation assesses implementation, intermediate outcomes, sustainability, and scalability of each intervention. For this, in June 2020, NORC collected data from KIs with teachers, principals and other stakeholders. In September 2020 we conducted FGDs with girls, boys, and parents. Finally in October-November 2021 we conducted FGDs with girls that had participated in POY clubs. All qualitative data was collected in treatment schools.

AI. PROJECT BACKGROUND I

Waache Wasome ('Let Them Learn' in Kiswahili) is a five-year program to increase the secondary school enrollment and retention of adolescent girls in four districts of Tanzania. The program was designed and implemented by the Bantwana Initiative of World Education Inc. (WEI) and funded by the United States Agency for International Development (USAID). Under the USAID Reading and Access Evaluations contract, NORC conducted an evaluation of *Waache Wasome*, adopting a distinctive impact and performance evaluation (IE+) methodology.

Waache Wasome (WW) employs a holistic approach to keep girls in school, working directly with them, as well as within their school and home environments.

The program's target areas were the district of Karatu District Council (DC), one of seven districts in the Arusha Region of northern Tanzania, and three districts in the Mara Region, Musoma DC, Musoma Municipal Council (MC), and Tarime DC. The program was implemented in 67 public schools, including

¹ A detailed description of the project is available in NORC (2018) and references cited there.

all 60 public schools in Karatu DC, Musoma DC, and Musoma MC, and seven schools in Tarime DC. WEI and USAID selected these districts because data on violence against children and other conditions made them good candidates for a program like WW to have substantial impact, and because of WEI's presence in Arusha.²

In 2017, the first year of WW implementation, WEI implemented activities in six schools of Karatu DC. These pilot activities served to test various components of the larger intervention program. In February 2018, the program began expanding to 30 additional schools. Two years later, in February 2020, the program expanded further into the remaining 31 schools. Program activities ended in December 2021. In August 2020 a project amendment extended activities into one additional district (Arusha City Council) to treat a total of 108 schools (38 in Arusha, and three more in the original four districts). The 41 additional schools were not included in this evaluation.

To achieve its goal of increasing school retention, WW seeks to:

1. Build agency, knowledge, and protective assets of girls in secondary school:

The program implemented Protect Our Youth (POY) clubs, providing safe spaces in schools where adolescent girls could learn life skills to navigate the risks they face, and where they could build protective assets that reduce vulnerabilities, helping them remain and succeed in school.

At the beginning of each school year, WEI established four POY clubs, two in Form 1 and two in Form 3, in each school. Each POY club had a maximum of 40 participants: Ten girls at risk of dropping out (ARD), classified by their teachers when POY clubs were assembled and encouraged to enroll in the clubs, 15 girls randomly selected, and 15 boys randomly selected. Club sessions were led by two peer educators supported by teacher matrons and patrons, Local Government Authorities (LGAs), Ward Education Officers, and project staff.

2. Increase family commitment and capacity to invest in girls' education:

WEI invited mothers and caregivers of adolescents to join Livelihood Improvement for Mothers and Caregivers of Adolescents (LIMCA) groups. LIMCA is a community-based savings and lending platform that supports socio-economic change at the household level.

The 20 to 30 member groups were established at the village level and supported by trained Economic Empowerment Volunteers (EEVs). In addition, the savings groups were used to deliver training to improve parental skills and gender norms.

3. Foster a girl-friendly and supportive school environment:

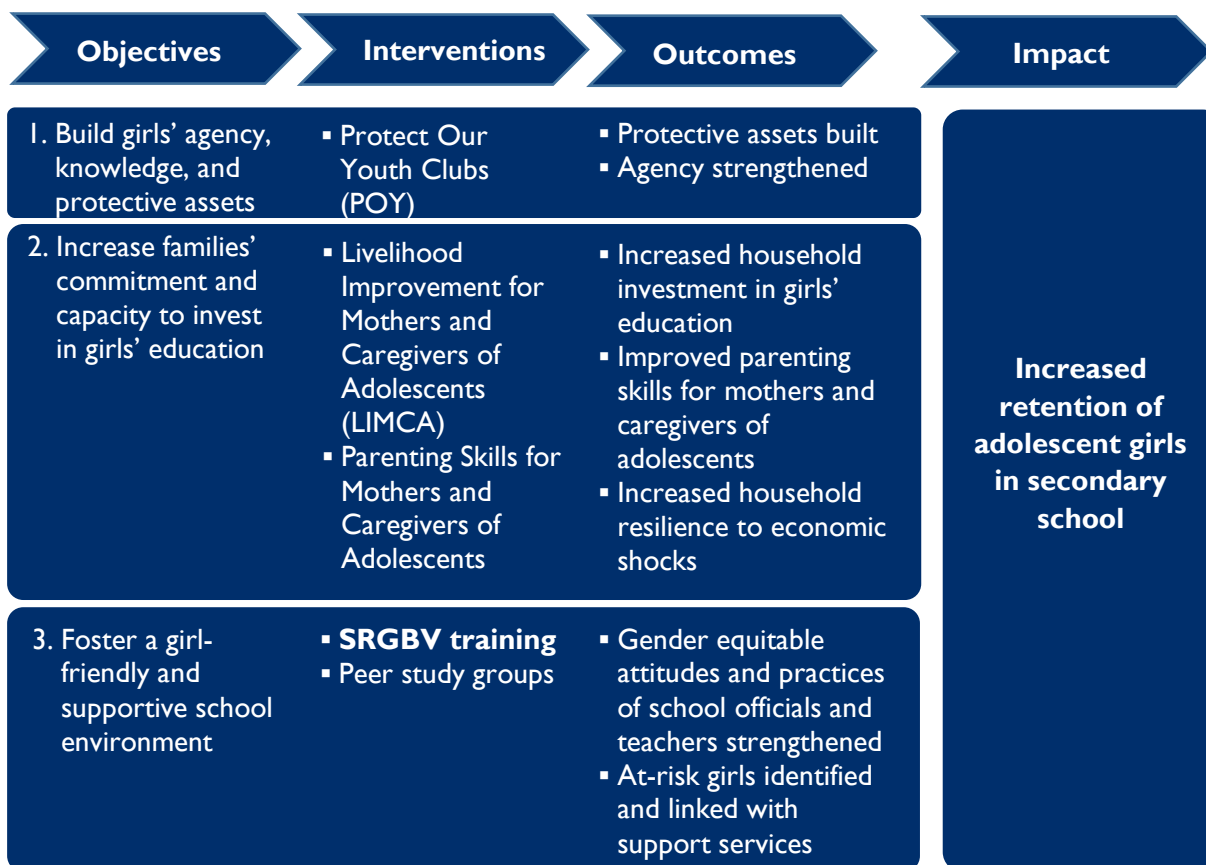
To promote a supportive environment in schools, the program conducted two activities: (1) training in school-related gender-based violence (SRGBV) prevention and response for teachers

² The Population Council created an index of nine indicators to categorize regions as hotspots for youth including: school attendance (ages 10-14); school attendance and parental residence; early childbearing; early marriage; partner age differences (greater than 10 year age difference); sexual violence; HIV prevention knowledge; comprehensive HIV knowledge; and HIV prevalence. The regions selected for the intervention are identified as hotspots for girls' vulnerability and HIV risk and low secondary enrollment rates (WEI, Program Concept note, Version June 20, 2016).

and school administrators, and (2) peer study groups, in which underperforming students meet with high performers to review academic content.³

The underlying theory of change was that the interventions described above would provide both a supportive environment and a comprehensive set of resources so girls would be more likely to succeed and stay in school. **Figure 1** shows how each program intervention maps onto outcomes, the aforementioned three objectives, and final impact.

Figure 1: Theory of Change



Source: Derived from WW's conceptual framework (WEI 2017)

RESULTS AND CONCLUSIONS

We do not find evidence that the program increased school retention. As **Table 5:** shows, administrative data indicates that retention was 2 percentage points higher for the treatment girls than for control girls between 2018 and 2019, but this positive effect is not observed in the administrative data between 2018 and 2020. Using survey data, we find that retention was 4 percentage points higher for the control girls than for treatment girls. In any case, none of these differences are statistically significant. These results indicate that the program did not have a major impact on school retention. The

³ A fourth objective originally considered was to "Increase access to youth-friendly adolescent reproductive health services." However, this component was dropped from the intervention after negotiations with the Government of Tanzania (GoT). A fifth objective included in the original Program Description, to "Provide alternative education pathways for girls who have dropped out of secondary school due to pregnancy and/or early marriage," is not discussed in this report because it is not part of the evaluation.

differences between treatment and control on protective assets are also small and most of them not statistically significant.

Table 1: Girls' retention rates using administrative and survey data

	Control	Treatment	Difference
One-year retention (2018-2019)			
Administrative data	0.85	0.87	0.02
Two-years retention rate (2018-2020)			
Administrative data	0.66	0.65	-0.01
Survey data	0.72	0.68	-0.04

Standard errors for the significance test are clustered at the school level. Survey data corresponds to retention rate for girls attending the same school at baseline and endline.

There are two results that could explain why we find no impact on retention. First, we show that there are other programs similar to WW's POY clubs in control schools. This affects the validity of the impact evaluation results, but more importantly questions whether this type of program should be expanded where similar interventions are already available. Second, administrative data indicate that there are no major differences between girls and boys retention rates in either the treatment or the control schools. If programs like WW are based on the assumption that girls are more likely to drop out of school than boys, these assumptions should be revised in future programming.

Both quantitative and qualitative data indicate that girls were satisfied with POY clubs. However, FGD show that it was unclear whether improvements in protective assets from POY club participation prevented dropping out of school. While for some girls the protective assets that POY clubs helped them build were a large factor in their decision to stay in school, girls also reported that the decision to drop out is driven by other factors, especially poverty, which POY club participation could not affect.

Regarding results for parents, we do not find significant changes in parenting skills between baseline and endline, and only small changes in attitudes towards gender equity. It is worth noting that there was only one year between baseline and endline, and parenting skills may need a long time to change.

Qualitative data show that LIMCA group members perceived group activities positively. Groups were successful in changing perspectives about girls' education, and motivating parents' support of girls within their communities.

We find that SRGBV teacher training had a positive impact on teachers' knowledge about SRGBV, as measured by their capacity to recognize SRGBV events and their understanding of key SRGBV regulations. However, more work should be done so teachers can properly navigate SRGBV response procedures, including how to report SRGBV events. More training sessions, regular refreshers, and supportive supervision could help with this objective.

One key success of the program was in raising awareness about SRGBV, especially corporal punishment. We found teachers in the treatment schools were less likely to cane students and assign chores at school as punishment than their counterparts in the control group. It is important to keep in mind that

WEI and other implementers in Tanzania operate under a regulatory framework that allows corporal punishment in schools, so efforts to reduce it are bounded by this reality.

Finally, we find that there are no substitutes for SRGBV training for teachers in these communities. This highlights the importance of continuing with SRGBV training for teachers, and during their preservice training at the teacher colleges.

A. INTRODUCTION

A2. CONTEXT

Public education in Tanzania consists of one year of preschool, seven years of primary education, four years of ordinary secondary education and two years of advanced secondary education. Gender gaps in ordinary secondary school completion are still observed in Tanzania, despite the fact that primary school completion rates are higher for girls than for boys. In 2015, the completion rate for ordinary secondary schools was 31.1 percent for males and 26.7 percent for females, while for primary school the completion rate was 75.1 for males and 83.7 percent for females.⁴

Violence against children is a pervasive problem in Tanzania. A national survey conducted by the United Nations International Children's Emergency Fund (UNICEF) in 2009 on females and males ages 13 to 24 found that roughly 30 percent of females and 13.4 percent of males reported having experienced sexual violence before the age of 18 (UNICEF 2011). In addition, 75 percent of both females and males reported experiencing physical violence during childhood. Schools do not necessarily constitute a safe space for children, as more than 50 percent of both males and females experienced physical violence by teachers. It is important to highlight that corporal punishment is legal in Tanzania, in particular caning, which poses a challenge for interventions like WW to have a strong impact on corporal punishment in schools. Moreover, 15 percent of females reporting sexual violence said at least one incident occurred on school grounds (UNICEF 2011).

Furthermore, during adolescence, girls in Tanzania face several challenges that discourage them from continuing their secondary school education. These factors include early marriage, teenage childbearing and gender disparities in cultural norms. The 2015-2016 Tanzania Demography and Health Survey found that 3.4 percent of adolescent girls aged 15 to 19 were married at age 15. Among women aged 20 to 24, 30.5 percent were married at age 18. Teenage childbearing is relatively common in Tanzania. Twenty-seven percent of adolescent girls aged 15 to 19 have begun childbearing (Ministry of Health, Community Development, Gender, Elderly and Children et al. 2016). In 2016, teenage pregnancy caused an estimated 3,700 girls in the country to drop out of school (UNICEF 2018).

Clearly, the relationship between these factors and education is not unidirectional, as lack of education can also precipitate early marriage and teenage childbearing. In effect, access to education can affect the lives of girls in multiple ways. Girls with more education tend to have fewer children and to delay marriage. They also find more success in the labor market. Observational analyses show that failing to educate girls in the developing world translates into lower future earnings for them and higher rates of stunting among their children (Wodon et al. 2018).

A3. EVIDENCE

A growing body of literature analyzes the impact of life skills training on girls. A number of recent studies have found small but positive impacts on school enrollment (Ashraf et al. 2018, Buchmann et al. 2018, Koroknay-Palicz & Montalvao 2019, Edmonds et al. 2019), although other studies found no significant impacts of life skills training on schooling (Austrian et al. 2018).

⁴ Adapted from UNESCO Institute for Statistics (UIS), <http://data.uis.unesco.org>, extracted September 20, 2018.

With respect to later outcomes, research has also shown that life skills training can have positive impacts on employment and income generation activities (Bandiera et al. 2020, Adoho et al. 2014), although there is also evidence that these effects can fade over time (Acevedo et al. 2017).

A review of these and other related studies is available in Annex I.

B. EVALUATION PURPOSE AND QUESTIONS

The objectives of this evaluation are to:

- Evaluate the impact of the WW program as a whole (including all its components) on school retention. To document the mechanisms through which the program may impact retention, the evaluation also estimates the effect of the program on girls' protective assets, gender attitudes, and perceptions on gender related violence; parents' attitudes towards investing in girls' education and parenting skills; and teachers' attitudes towards gender norms and SRGBV.
- Assess the performance of three WW components, namely POY clubs, LIMCA, and SRGBV training.
- Draw conclusions based on impact and performance evaluation findings and develop recommendations for fine-tuning the program to improve impact and implementation and ensure sustainability. To this end, we identify program components that worked/did not work, reasons for high/low impacts, and obstacles and drivers to improving impact on key outcomes of interest.

The impact and performance evaluation questions addressed in the study are as follows:

IMPACT EVALUATION

IE10. What is the impact of the program on retention, attendance, and promotion?

IE1.1. What is the impact of the program on the most vulnerable girls? (We define most vulnerable girls as ARD)

IE1.1. What is the impact of the POY club component on retention, attendance, and promotion for girls attending treatment schools?

INTERMEDIATE OUTCOMES-GIRLS

IE12. Does the program strengthen girls' protective assets?

IE13. Is the WW package changing gender norms?

IE14. Does the program change attitudes on SRGBV?⁵

⁵ Note we focus on attitudes towards SRGBV rather than actual incidence of SRGBV events. This is because recording SRGBV incidence would require the deployment of substantial resources to manage disclosed events (this includes referral processes, long-term psychological and legal counsel, and other follow-up activities), which surpassed the scope of the evaluation.

INTERMEDIATE OUTCOMES-PARENTS

- IEI5. Does the program improve attitudes towards gender equitable access to education within the household?
- IEI6. Does the program change parenting skills?

INTERMEDIATE OUTCOMES-TEACHERS

- IEI7. Does the program change gender attitudes and perceptions of teachers in school?
- IEI8. Does the program change teachers' knowledge, attitudes, and practices related to SRGBV?

PERFORMANCE EVALUATION

POY CLUBS

- PEI. How are girls and boys benefitting from POY clubs?
 - PEI.1. What are the main challenges POY clubs face? How is WW resolving these implementation challenges? Do challenges vary by district?
 - PEI.2. Are members participating actively in POY clubs? Are they satisfied with POY club activities? What benefits are perceived by participants?
 - PEI.3. Is the POY club curriculum user-friendly? What lessons/components of the POY club curriculum are most useful to adolescents?
 - PEI.4. How is WW engaging LGAs in POY clubs?
 - PEI.5. What kind of supportive supervision do patrons/matrons and peer educators of POY clubs receive? Do they perceive this supportive supervision as contributing to the results of POY clubs?
 - PEI.6. Are POY clubs having any unintended (positive or negative) consequences?
 - PEI.7. Do POY clubs influence the general school environment? If yes, how? If no, why not?
 - PEI.8. How can POY clubs be made more effective?

LIMCA

- PE2. Is LIMCA benefitting its members, girls, and other stakeholders as intended by WW?
 - PE2.1. Is LIMCA helping to reduce the socioeconomic vulnerability of participating households? Why or why not?
 - PE2.2. Are parents and caregivers investing in girls' education as a result of LIMCA participation? How many parents and caregivers are being supported by LIMCA groups?
 - PE2.3. What kind of supportive supervision do EEVs receive from local nongovernmental organization partners? Do EEVs perceive this supportive supervision as contributing to the results of LIMCA groups?
 - PE2.4. How is WW engaging LGAs in LIMCA?
 - PE2.5. What should be done to make this WW intervention more effective?

SRGBV TRAINING

PE3. What are the benefits of SRGBV prevention and response teacher training to teachers, students, and other stakeholders?

PE3.1. Are teachers satisfied with SRGBV training? Why or why not?

PE3.2. Have teachers' knowledge and attitudes towards SRGBV changed? Why or why not?

PE3.3. Which elements of the following training components are (i) easiest and (ii) most difficult for teachers to implement after receiving the training? Why?

— SRGBV prevention and response

— Gender responsive pedagogy

— Alternative forms of discipline

PE3.4. What kinds of support do teachers receive in implementing what they have learned? Are they satisfied with the support they are receiving?

PE3.5. How are LGAs (specifically Gender and Children Desk Police Officers) supporting SRGBV training for teachers?

PE3.6. How are stakeholders (parent-teacher associations, school boards, etc.) supporting SRGBV training for teachers?

PE3.7. What should be done to make this WW intervention more effective? Specifically, what can be done to improve SRGBV training, teachers' understanding of it, and communication/collaboration between LGAs and teachers?

SUSTAINABILITY

PE4. Are the interventions addressed in PEI-3 sustainable?

PE4.1. What actions are being taken to continue the interventions addressed in PEI-3 after WW ends?

PE4.2. Are these interventions likely to continue after WW ends? Why or why not?

C. EVALUATION METHODS

WW was implemented in the districts of Karatu DC, Musoma DC, Musoma MC, and Tarime DC. Of 67 schools targeted to benefit from WW, the evaluation excluded the six Karatu DC schools that were part of the pilot. Thus a total of 61 schools are part of the evaluation.⁶

CI. IMPACT EVALUATION

To evaluate the impact of the program, NORC worked with WEI to conduct a randomized control trial. Between December 2017 and January 2018, we randomly assigned 30 schools to the treatment group (start treatment in February 2018) and 31 schools to the control group (start treatment in February

⁶ Originally WEI had planned to treat Tarime Town Council (TC) instead of Tarime DC. Tarime TC was replaced by Tarime DC because WEI learned programs similar to WW were already operational in Tarime TC. However, while Tarime TC has only seven schools, Tarime DC has 30. To accommodate the program budget, WEI selected seven of the 30 schools in Tarime DC as replacements for the Tarime TC schools, based on their geographic proximity to Tarime TC.

2020). The fact that starting in 2020 all schools would be treated helped fostering school administrators' acceptance of the evaluation design. Below we describe how this evaluation assessed the impact on girls, parents, and teachers.

IMPACT ON GIRLS

We collected two types of data to estimate the impact of the program on girls' outcomes: survey data and school administrative data. Survey data is used to document the average impact on enrollment rates and intermediate outcomes, such as, girls' protective assets, gender attitudes, and perceptions of gender-based violence. Survey data also serves to document impacts on enrollment and intermediate outcomes for girls at risk of low school attendance and dropping out. School administrative data is used to evaluate the effect of the program on girls' attendance.

NORC collected survey data in February 2018 (baseline) and in February 2020 (endline). School administrative data was collected in November 2018 and September 2019. Additional waves of administrative data were collected in January 2021 and October 2021 to record attendance during 2020 and 2021, respectively. The objective of collecting attendance data for 2020 and 2021 was to document girls' school trajectories throughout the four years of secondary school.

SURVEY DATA

The sample frame for the girls' survey was the school roster of girls enrolled in Form I in January 2018 in the 61 study schools. The goal was to sample 38 randomly selected girls in each of these 61 schools, a total of 2,318 girls. This sample allows estimation of a minimum effect of seven percentage points of increased school retention.⁷

To evaluate the impact of the program, we compare retention rates and other outcomes of interest between treatment and control groups at endline. Using this method, we can estimate experimentally the impact of the program on average, but we cannot estimate experimentally the impact on ARD girls (IEI.I), since these girls were identified by teachers, and only in treatment schools. Hence, we use matching methods to select girls in control schools comparable to ARD girls in treatment schools. In Annex II we explain the implemented method in more detail.

The girls' survey includes modules focused on:

- Basic socio-demographic information (parental education, household size, and household asset ownership)
- Program participation
- Girls' protective assets (self-efficacy, self-esteem, communication skills, decision-making and social networks)
- Gender norms and equity
- Perceptions on SRGBV (bullying, sexual harassment, and corporal punishment)

⁷ For this estimate we used the observed Intra-cluster correlation of 0.03, a power of 80 percent and alpha of 5 percent.

SCHOOL ADMINISTRATIVE DATA

School administrative data is used to record enrollment and attendance. To collect this, enumerators visited each of the 61 schools in the study in November 2018, September 2019, January 2021 and October 2021. To evaluate the average impact on attendance in treatment schools, we compare enrollment rates between treatment and control schools.

PARENTS

To answer research questions related to parenting skills and parents' attitudes towards gender-equitable access to education, we use data collected through a parent survey. The primary, and to some extent, unique, direct interaction between the WW program and parents is through the savings LIMCA savings groups. Therefore, we focus on them to evaluate the impact of the program on parents.

During the design phase it was unclear whether it would be reasonable to replicate the methodology used for girls and survey random samples of parents in treatment and control schools. Because not all parents in treatment schools would necessarily join a LIMCA group, a random sample of parents might yield very few who were LIMCA members. Moreover, if the LIMCA group take up rate was very low, it was not clear if it would be helpful to conduct an impact evaluation for the parent component of WW. To address this, we decided during the design phase of the project that if the LIMCA group take-up rate was very low, we would conduct a before-and-after performance evaluation, surveying only parents who were LIMCA members (NORC 2018). On October 22, 2018, using LIMCA take-up data shared by WEI, NORC estimated a LIMCA take-up rate of 14 percent.⁸ This rate was deemed too low for an impact evaluation and, as such, we opted to conduct a before-and-after evaluation.

TEACHERS

For the teachers survey no baseline data was collected. At endline, roughly ten teachers in each of the 61 schools were interviewed. The teacher survey includes modules for gender attitudes and perceptions and practices on corporal punishment, sexual harassment, and bullying/emotional abuse. Using this data we compared teachers in treatment and control schools to estimate the impact of the program on teachers.

C2. PERFORMANCE EVALUATION

The performance evaluation assesses the implementation, intermediate outcomes, sustainability, and scalability of three WW interventions, POY clubs, LIMCA groups, and SRGBV training for school teachers and administrators, using qualitative data collected in July and October 2020. In July we conducted key information interviews (KIs) with teachers, head teachers, LGAs and EEVs. In October we conducted KIs with peer educators, focus group discussions (FGDs) with LIMCA parents, and FGDs with girls and boys. Additional qualitative data, specifically FGDs with girls who stayed in school and girls who stopped going to school, were collected from September to November 2021.

We describe the evaluation of each component below.

⁸ According to WEI records, in late 2018 there were 297 parents of girls in Form I that were part of a LIMCA group. In 2018 there were 2,164 Form I girls in the 30 treatment schools, for an approximate LIMCA take-up rate for parents with daughters in FI in treatment schools of 14 percent.

POY CLUBS

To explore the implementation and efficacy of POY clubs, we use data from FGDs with girls and boys in POY clubs, and KIs with peer educators, teachers, and LGAs. With the data we address question PE1 and its sub-questions.

LIMCA

To document implementation, benefits, and efficacy of LIMCA groups, we use data from FGDs with LIMCA parents and with girls, and KIs with EEVs and LGAs to address question PE2 and its corresponding sub-questions.

SRGBV PREVENTION AND RESPONSE TRAINING

To document perceived benefits, support, and efficacy of the SRGBV Prevention and Response Training, we use data from KIs with teachers and head teachers, FGDs with girls and boys, and KIs with LGAs. The data is used to answer question PE3 and its sub-questions.

C3. DATA COLLECTION

NORC contracted EDI Limited, a local Tanzanian firm, to conduct all data collection activities. Endline instruments are available in Annex III.

QUANTITATIVE DATA

Table 2 shows when each wave of data was collected, as well as fieldwork results. A detailed description of data collection activities is available in Annex IV.

Table 2: Fieldwork results – Quantitative data

Instrument	Fieldwork	Sample		
		Target	Completed	Completion rate %
Girl baseline	12th February – 12th March, 2018	2,318 (38 girls in 61 schools)	2,220	96
Attendance data 2018	26th November - 5th December, 2018	10,388 ^(a)	9,661	93
Attendance data 2019	23rd September - 11th October, 2019	9,661 (all students recorded in 2018)	8,215	85
Parent baseline	4th February to 15th February, 2019	450 (15 parents in 30 schools)	324	72 ^(b)
Girl endline	10th February - 10th March, 2020	2,220	1,815	82
Parent endline		324	305	94
Teacher		610 (10 teachers in 61 schools)	588	96

Instrument	Fieldwork	Sample		
		Target	Completed	Completion rate %
Attendance data 2020	18th January to 29th January, 2021	11,049 (all students recorded since 2018)	7,245	66
Attendance data 2021	9th October to 29th October, 2021	11,087 (all students recorded since 2018)	6,819	61

(a) In 2018 attendance was recorded in 187 classrooms, 93 percent of the 202 Form I classrooms in the evaluation schools. In the remaining 15 classrooms it was not possible to record attendance. In most cases this was because the teacher was not available. We estimate the target for this survey by assuming we surveyed 93 percent of it.

(b) While the initial survey target was 450 parents (15 parents in 30 treatment schools), when data was collected, fewer than 15 parents were eligible for the survey in several schools. Eligibility was restricted to adults that were part of a LIMCA group in February 2019 and that were parents of girls 10-21 years old.

QUALITATIVE DATA

NORC had planned to conduct FGDs in May 2020 with youth and parents, and KIIs with peer educators, teachers, head teachers, economic empowerment volunteers, and local government agencies. Due to the COVID-19 pandemic, schools in Tanzania closed in March 2020. Given the risks associated with collecting data in person, we opted to collect qualitative data in two parts. In July we conducted KIIs by phone with teachers, head teachers, economic empowerment volunteers, and local government agencies. We postponed FGDs with girls, boys, and parents until October, when schools were open and in-person data collection was authorized. We also conducted KIIs with peer educators in October, to avoid difficulties associated with phone interviews of youth.

In addition, in September-November 2021, NORC conducted FGD with both girls still in school and girls that had dropped out.

Table 3: shows when each wave of qualitative data was collected, as well as fieldwork results. A detailed description of the data collection activities is available in Annex IV.

Table 3: Fieldwork results – Qualitative data

Instrument	Fieldwork	Sessions	Total Number of Participants ⁹
Teacher KIIs	3rd June – 30th June 2020	16	16
Head teacher KIIs		6	6
EEV KIIs		8	8
LGA Official KIIs		12	12
Girls FGD	21st September - 14th October, 2020	6	35
Boys FGD		6	32
Parents FGD		6	38
Peer leader KII		12	12
Girls FGD (Girls who stayed in school)	9th October 8th November, 2021	6	43

⁹ Focus groups had 5-8 participants per session

Instrument	Fieldwork	Sessions	Total Number of Participants ⁹
Girls FGD (Girls who did not complete secondary school)		6	30

D. IMPACT EVALUATION RESULTS

DI. IMPACT ON ATTENDANCE, RETENTION, AND PROMOTION

IEI. What is the impact of the program on retention, promotion, and attendance?

ADMINISTRATIVE DATA

To analyze the impact of the program on retention, promotion, and attendance, we first analyzed school administrative data collected in 2018, 2019 and 2020. In 2018, we collected data for 9,656 Form 1 students (52 percent male and 48 percent female). In 2019, out of the 9,656 students surveyed in 2018, 8,215 were enrolled for a retention rate of 85 percent, while in 2020 this figure was 6,359, for a retention rate of 66 percent. The Form 1-Form 2 promotion rate was also 85 percent, as the number of students surveyed in 2019 still in Form 1 was negligible.

Table 4 shows retention rates by gender and treatment status. Retention rates between 2018 and 2019 are higher for both males and females in the treatment group, although the differences are not statistically significant. Retention rates between 2018 and 2020 are higher for both males and females in the control group, but again, the differences are not statistically significant.

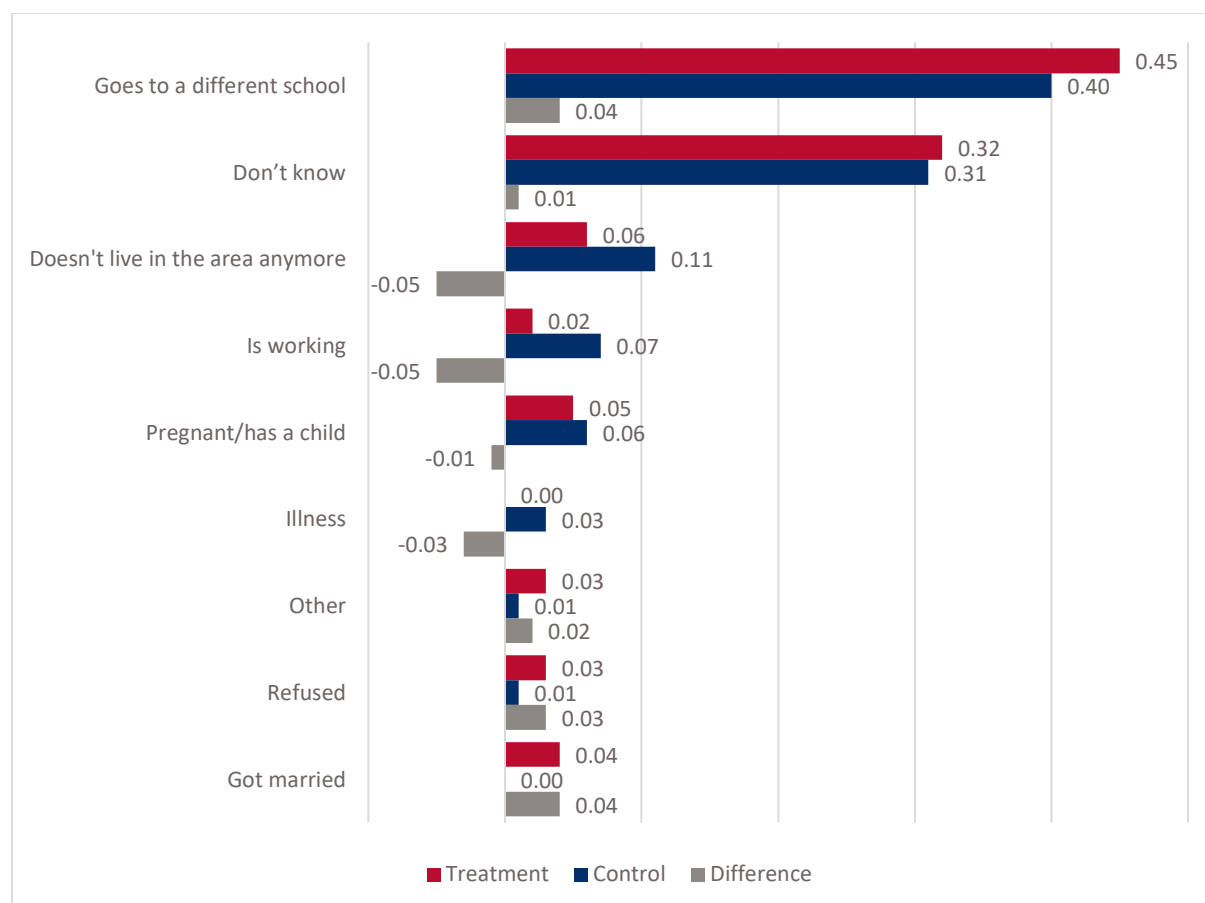
Table 4: Retention rates by gender and treatment status

	2018-2019			2018-2020		
	Control	Treatment	Difference	Control	Treatment	Difference
Females (N=4599)	0.85	0.87	0.02	0.66	0.65	-0.01
Males (N=5057)	0.84	0.85	0.01	0.68	0.65	-0.03

Standard errors for the significance test are clustered at the school level. Sample sizes for retention in 2019 are lower than for 2018-2020 (2 for females and 6 for females) due to item-specific missing data.

We asked teachers why each of the students enrolled in 2018 but not in 2019 were no longer in school. **Figure 2** shows teachers' responses for the 646 girls that were enrolled in 2018 but not in 2019. The primary explanation given was that girls had transferred to another school. Transfers are not extremely rare. Since the abolishment of secondary school fees enrollment soared, and due to capacity, many students are assigned to schools that are not the closest to their homes. They have no choice but to report to the school they are assigned but often change to a school closer home when they can. If some girls in fact transferred to a different school, then the real school retention rate is higher than our estimated rate. Notably, in about one third of cases, teachers did not know why girls were not enrolled in school.

Figure 2: Reasons for girls not being enrolled in school according to teachers (N=646 girls)



Note: Reasons in the “Other” category include inability of students to pay fees, natural disasters, and having to take care of a family member. The “Don’t know” option includes both cases where the teacher did not know the students and cases where they did but did not know the reason for not being enrolled.

Monthly attendance data was collected from school records. **Table 5:** shows the impact of the WW program on attendance by both total number of days attended in each year and yearly attendance rate.¹⁰ We focus on data for 2018 and 2019, as this was the period when the treatment group was treated, and the control was not. In the first and second columns we show the number of days students attended school in the treatment and control groups. The third column shows the difference between the two. Columns 4-6 show the same results for the attendance rate. The differences between treatment and control groups indicate a positive relation between treatment status and attendance for girls, although none of the coefficients are statistically significant. Panel A shows results for 2018 for boys and girls. Girls in the treatment group went to school on average six more days than did their counterparts in the control group. Differences in attendance rates are not statistically different from 0.

Table 5: The impact of WW on attendance

Number of days attended			Attendance rate		
Treatment	Control	Diff.	Treatment	Control	Diff.
(1)	(2)	(3)=(1)-(2)	(4)	(5)	(6)=(4)-(5)

A. Year 2018

¹⁰ Annex V shows similar results looking at the monthly-level data for years 2018-2021.

	Number of days attended			Attendance rate		
	Treatment (1)	Control (2)	Diff. (3)=(1)-(2)	Treatment (4)	Control (5)	Diff. (6)=(4)-(5)
Girls (N=4596)	128.6	123.0	5.60	0.85	0.83	0.02
Boys (N=5057)	124.2	120.6	3.57	0.82	0.82	0.00
<i>B. Year 2019</i>						
Girls (N=4681)	118.6	112.2	6.36	0.82	0.80	0.02
Boys (N=4940)	113.4	108.5	4.93	0.79	0.79	0.00

Note: Some students attended school more days than they were supposed to. This was pervasive in 2019, as schools prepared for the national exams and teachers asked students to attend on weekends. For the number of days, we added up all recorded days attended, so the figure includes days in excess of the number students were supposed to attend. For the attendance rate we truncated the yearly attendance rate to 1. Standard errors for the significance test are clustered at the school level.

Overall results using administrative data do not indicate that WW had an impact on retention or attendance.

SURVEY DATA

Of 2,220 girls surveyed at baseline, 1,815 were surveyed again at endline (82 percent). Of those, 1,554 (86 percent) were enrolled in the same school as at baseline, 71 (4 percent) were enrolled in a different school, and 190 (10 percent) were not enrolled in school. **Table 6** shows the number of girls surveyed at baseline and endline by treatment and POY club status.

Table 6: Number of girls surveyed at baseline and endline by treatment and POY club status

	Treatment			Control
	POY club			
	All	ARD	Randomly selected	
Baseline	1099	272	386	1121
Endline	891	217	311	924

For the purpose of analyzing the impact of the program on school retention, we assume that girls not surveyed at the endline are out of school, although it is possible some of them transferred to another school.

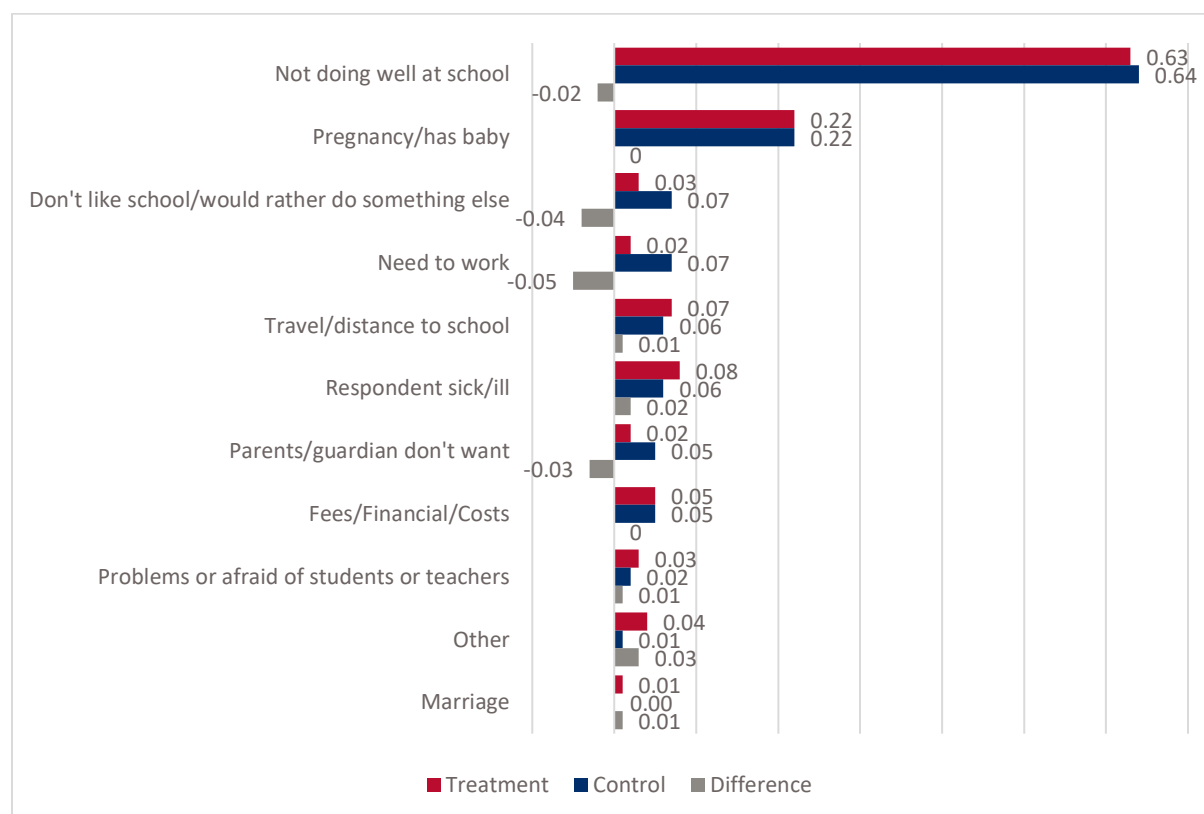
Table 7: shows the number of girls enrolled in school as a fraction of girls surveyed at baseline, by treatment group. The first row shows results for girls enrolled in (any) school and the second for girls enrolled in the same school as at baseline. Although the results indicate higher school retention rates for the control group than for the treatment group, the differences are not statistically significant. This means **there are no statistically significant differences in school retention rates between treatment and control groups.**

Table 7: Survey and school retention as fraction of the girls surveyed at baseline, by treatment status

	Treatment (1)	Control (2)	Diff. (3)=(1)-(2)
<i>A. All girls</i>			
Currently in school	0.72	0.75	-0.03
Currently in same school	0.68	0.72	-0.04
N	1099	1121	
<i>B. Gils surveyed at baseline and endline only</i>			
Currently in school	0.88	0.91	-0.02
Currently in same school	0.84	0.87	-0.03
N	891	924	

At endline, we were able to ask 190 girls who had dropped out why they left school. As shown in, about two thirds of respondents, both in the treatment and control groups, attributed the reason to their poor performance in school. The second most predominant reason was pregnancy or having a baby, mentioned by over a fifth of respondents in both groups. As was the case with the administrative data, there do not seem to be any meaningful differences in reasons for dropping out between treatment groups.

Figure 3: Self-reported reasons for having dropped out of school (N=189)



Notes: The sample of 189 is one less than the number of girls interviewed not enrolled in school because of one refusal. 'Need to work' includes working at the house/farm or for money.

The fact that not all girls surveyed at baseline were surveyed at endline could bias the impact evaluation results, especially if the endline samples of treatment and control groups are different in ways that can affect the estimation of the impact of the program. As we show in Annex VI, although there are some differences at baseline between girls that are observed both at baseline and endline and girls that are observed at baseline only, these differences do not affect the comparability between treatment and comparison groups, as the differences between these two groups are small and not significant.

IE2. What is the impact of the program on the most vulnerable girls?

Vulnerable girls are defined as ARD. At the beginning of 2018, teachers in treatment schools identified girls as ARD. Of the 1,099 girls surveyed at baseline in the treatment group, 272 were ARD.

The criteria teachers used to identify ARD girls included: frequent absences, academic struggles, and exposure to any type of violence, among other factors. Given this selection process, we expected ARD girls to come from more disadvantaged socioeconomic backgrounds than the average girl in the treatment or control groups. To construct a valid comparison group for ARD girls, we sought to identify girls in the control group who were also from disadvantaged backgrounds, using a matching method described in detail in Annex II.

We used this method to evaluate the impact of WW on school retention, promotion, and attendance. **Table 8** shows the results. The first row shows the fraction of ARD girls in school when endline data was collected. Sixty-nine percent of ARD girls were in school, while the figure for comparable girls in the control group is 72 percent, however the difference is not statistically significant. The second row shows the fraction of girls still in school who were in Form 3 (the remainder were in Form 2 or Form 1). The difference between the treatment and control groups is small and not statistically significant. The third row shows the number of days girls surveyed reported attending school the previous week. In this case as well, the difference is not statistically significant.

Table 8: Impact of WW on ARD girls' school retention

	Treatment (1)	Control (2)	Diff. (3)=(1)-(2)
Fraction currently in school (N=1391)	0.69	0.72	-0.03
Fraction in Form 3 (N=1021)	0.94	0.93	0.01
Previous week attendance (days) (N=1023)	4.57	4.64	-0.07

Notes: See Annex II for details on the matching method.

Standard errors for the significance test are clustered at the school level.

In sum, our analysis does not indicate WW had a positive impact on retention, promotion or attendance for ARD girls.

IE3. What is the impact of the POY club component on retention, attendance, and promotion for girls attending treatment schools?

Girls in POY clubs include ARD girls and girls randomly selected for participation in POY clubs. All ARD girls were selected for POY club, so Question IE1.1 already addressed the impact of the program, including POY club participation, on ARD girls. In this subsection we analyze the impact of POY club participation on girls randomly selected for participation in the clubs.

In principle, criteria affecting girls randomly selected for participation in the clubs do not differ from those of the average girl in treatment or control schools, precisely because they were chosen for POY club participation entirely by chance.¹¹ Along these lines, **Table 9** shows results of retention, promotion, and attendance for girls randomly selected for POY clubs and girls in the control group.

The results indicate that there are no statistically significant differences on retention-related outcomes between girls randomly selected for POY club participation and girls in the control group.

Table 9: Impact of WW on school retention of girls randomly selected for POY clubs

	Treatment (1)	Control (2)	Diff. (3)=(1)-(2)
Fraction currently in school (N=1235)	0.88	0.91	-0.02
Fraction in Form 3 (N=1111)	0.91	0.93	-0.02
Previous week attendance (days) (N=1112)	4.66	4.63	0.03

Standard errors for the significance test are clustered at the school level.

Multiple factors may explain the lack of impact of WW on school retention. In the remainder of this report we explore the reasons for these results, as we look at program take-up and the impact of the program on intermediate outcomes for girls, parents, and teachers. However, at this point it is useful to describe the main mechanisms that could explain the lack of results on school retention.

First, it is possible that the program **take-up** was low. Second, girls, teachers, and/or parents in the control group may have participated in similar activities to WW, diluting the impact we can estimate by comparing treatment and control groups. This evaluation problem is often called **substitution**. Third, the program probably **did not affect the intermediate outcomes** as intended, namely girls' protective assets and attitudes towards gender equity and SRGBV, parents' attitudes towards gender equity and parenting styles, and teachers' SRGBV knowledge, attitudes and practices. Fourth, **the intermediate outcomes were not the right “levers”** needed to increase school retention. And fifth, it is possible the **sample size** was too small to detect small but still meaningful impacts.

In the following sections we discuss the evidence in support of these different explanations, and how they may explain the results.

PROGRAM TAKE-UP

Table 10: shows the fraction of girls who participated in youth clubs in 2018, including POY clubs. **Eighty percent of girls in the treatment group participated in WW's POY clubs in 2018, compared to a negligible fraction in the control group.** These results show the high level of POY club participation observed in the treatment group. Girls in the treatment group who participated in POY clubs in 2018 attended 3.5 meetings per month. We asked students who had never participated in POY clubs if they had participated in other school-based youth clubs. Notably, **31 percent of students**

¹¹ It is possible that girls randomly selected for the POY club are also different from the average girl because they did have to show interest in participating. However, results (not shown here) indicate there are no major differences at baseline between girls randomly selected for participation in POY groups and control group girls, which suggests that comparing the endline outcomes between the two groups is a reasonable approximation of the program impact.

in the control group said they had participated in youth clubs, suggesting that treatment **substitution** may be a key factor explaining WW's absence of impact. Regarding peer study groups, these groups are popular across the country, so it is not surprising that a large fraction of girls have participated in them in both study groups.

Table 10: Club participation

	Treatment (1)	Control (2)	Diff. (3)=(1)-(2)
Participated in a POY club in 2018 (N=1815)	0.80	0.03	0.77***
<i>For those that participated in 2018:</i>			
Number of sessions attended per month (N=737)	3.52	2.58	0.94*
Participated in other youth clubs in last 2 years (N=1036)	0.21	0.31	-0.11
Participated in peer study groups (N=1814)	0.86	0.78	0.08**

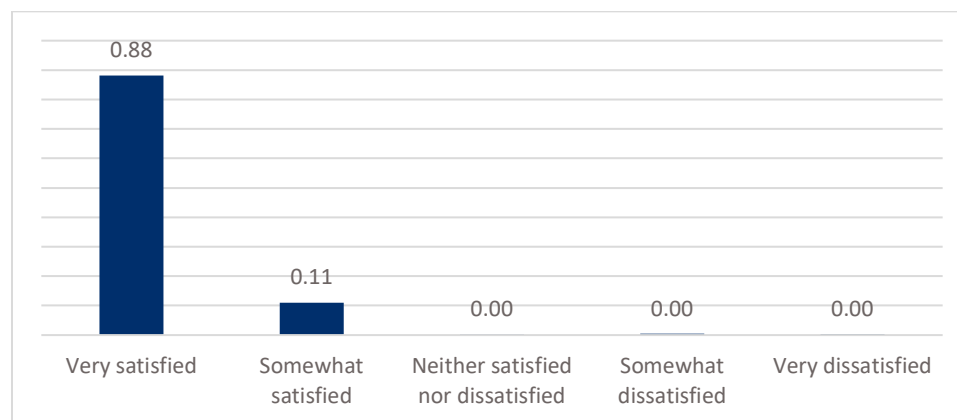
Notes: The sum of the second and third sample sizes is lower than the first sample size because we only asked girls who had never participated in POY clubs about other youth clubs, and some girls had participated in POY clubs, but in 2018.

Standard errors for the significance tests are clustered at the school level.

* p<0.05 ** p<0.01 *** p<0.001

The satisfaction level with the POY clubs was high. As **Figure 4** shows, 88 percent of girls in the treatment group who attended a POY club in 2018 were very satisfied, 11 percent were somewhat satisfied, and negligible fractions were dissatisfied.

Figure 4: Satisfaction with POY clubs (N=713)



Notes: Survey limited to girls in the treatment group who attended a POY club in 2018.

While most participant girls were very satisfied with the POY clubs, 25.5 percent in treatment schools who participated in POY clubs had suggestions for improving them. Suggestions included providing study materials and posters, extending the duration of training, and holding more frequent meetings.

A few important conclusions regarding club participation can be derived from the results. First, **participation in POY clubs was high**. Four out of 5 girls in treatment schools participated in them, implying that low take-up of this program component is not a concern. Second, other alternatives were available in control schools, so **program substitution**, that is, the fact that girls in the control group attended similar programs than girls in the treatment group, is a real concern for the impact evaluation. About one third of girls in control schools said they had participated in youth clubs. Of these, 60 percent said they were or had been part of a Fema club, supported by a nongovernmental organization

called Femina Hip. Fema clubs promote, among other skills and behaviors, youth leadership and life skills (Femina Hip 2020). If a substantial fraction of girls in control schools are receiving the same services as girls in treatment schools, it becomes more difficult to estimate a program impact experimentally. Third, peer study groups are widely available in schools in general, so we should probably not expect to observe an impact of these clubs on the outcomes of interest.

D2. INTERMEDIATE OUTCOMES - GIRLS

PROTECTIVE ASSETS AND ATTITUDES TOWARDS GENDER NORMS, ROLES & EQUITY

IE4. Does the program strengthen girls' protective assets?

IE5. Is the WW package changing gender norms?

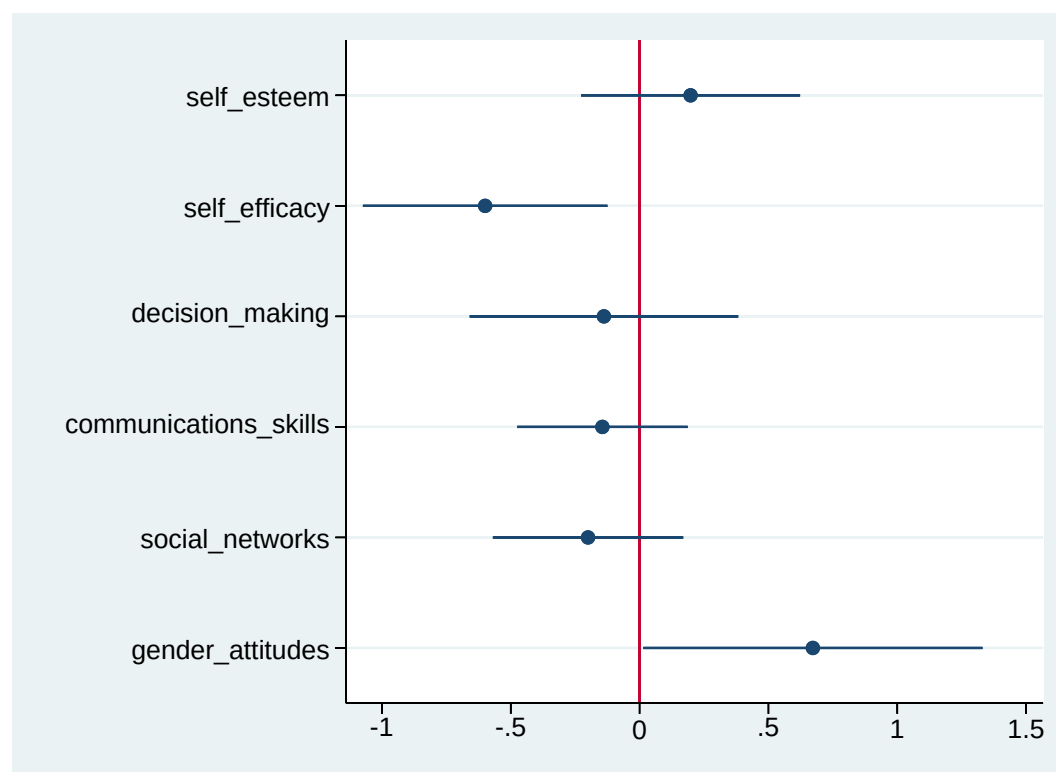
Protective assets are resources, knowledge, skills, or competencies girls can employ for themselves and others to overcome an obstacle or challenge. In this context, a protective asset is a store of value related to what a girl can do or be (Population Council 2015). Gender refers to the array of socially constructed roles and relationships, personality traits, attitudes, behaviors, values, and relative power and influence that society ascribes to the two sexes on a differential basis (USAID 2018). Gender equitable norms and attitudes can positively contribute to creating safe and supportive learning environments for girls. The WW program seeks to foster a friendly and supportive environment for girls in school.

WW's POY clubs offer a life skills curriculum to help girls develop protective assets such as decision-making, self-esteem, self-efficacy, communication skills, and relationship skills/social networks. The POY clubs' curriculum also includes sessions devoted to gender attitudes and gender equity.

To estimate WW's effect on girls' protective assets and gender attitudes, both the baseline and endline surveys included rubrics to measure self-esteem, self-efficacy, decision-making, communication skills, and social networks. Each of these rubrics of between six and eleven items employed a 4 point Likert scale (either strongly agree/agree/disagree/strongly disagree, or not true/slightly true/true/very true). We summarized these items by constructing indexes for each of the measured protective assets. For example, for self-esteem the endline score was 21.3 on average, out of a maximum possible value of 30. Descriptive statistics for all six scores are available in Annex VII.

To estimate the impact of the program, we calculated the difference at endline between treatment and control groups. **Figure 5** shows results, including the 95 percent confidence interval. The x-axis shows the difference between treatment and control for each index. A positive value for a given index means the program had a positive impact on the corresponding protective asset. The difference between the treatment and control groups for self-esteem is positive but not statistically significant (the confidence interval line crosses the zero line). The difference for self-efficacy is negative and significant, meaning the program *reduced* girls' self-efficacy. The results for decision-making, communication skills, and social networks are negative but not statistically significant. We found a positive and statistically significant effect only for gender attitudes. As a robustness check, we also estimated regressions using the baseline value of the corresponding protective score as a control, but the results (not shown) are consistent with what **Figure 5** shows.

Figure 5: Mean difference between treatment and control at endline on protective assets and gender attitudes with 95 percent confidence interval



Standard errors are clustered at the school level.

The results indicate the program did not have an impact on girls' protective assets. We do not find statistically significant differences for any of the protective assets, and for self-efficacy we find the program had a negative effect. The result for self-efficacy are counterintuitive, given that the project's aim was to increase self-efficacy. It is possible that girls in the control group were exposed to other programs that more effectively changed this protective asset. It is possible that the program affected what girls see themselves as capable of doing, made them more realistic about obstacles they face, and made them more willing to disclose efficacy-related challenges.

We found only positive impacts on gender attitudes, implying that, thanks to the program, girls have more equitable attitudes towards gender than their counterparts in the control group.

ATTITUDES TOWARDS SRGBV

IE6. Does the program change attitudes on SRGBV?

Incidents of SRGBV have been recorded in many settings, cultures, and schools around the world. Distinct acts of SRGBV may be psychological or socioemotional, physical, sexual, or economic, and often interlink in both their natures and consequences for child development. Although both boys and girls experience various forms of SRGBV, boys are often targeted for more severe forms of harsh physical punishment at school, while girls face higher rates of sexual violence in and around school. When schools are not safe for girls and boys, their ability to learn and develop suffers. Girls' access to education and learning is impacted when classroom and school climates are not safe and supportive.

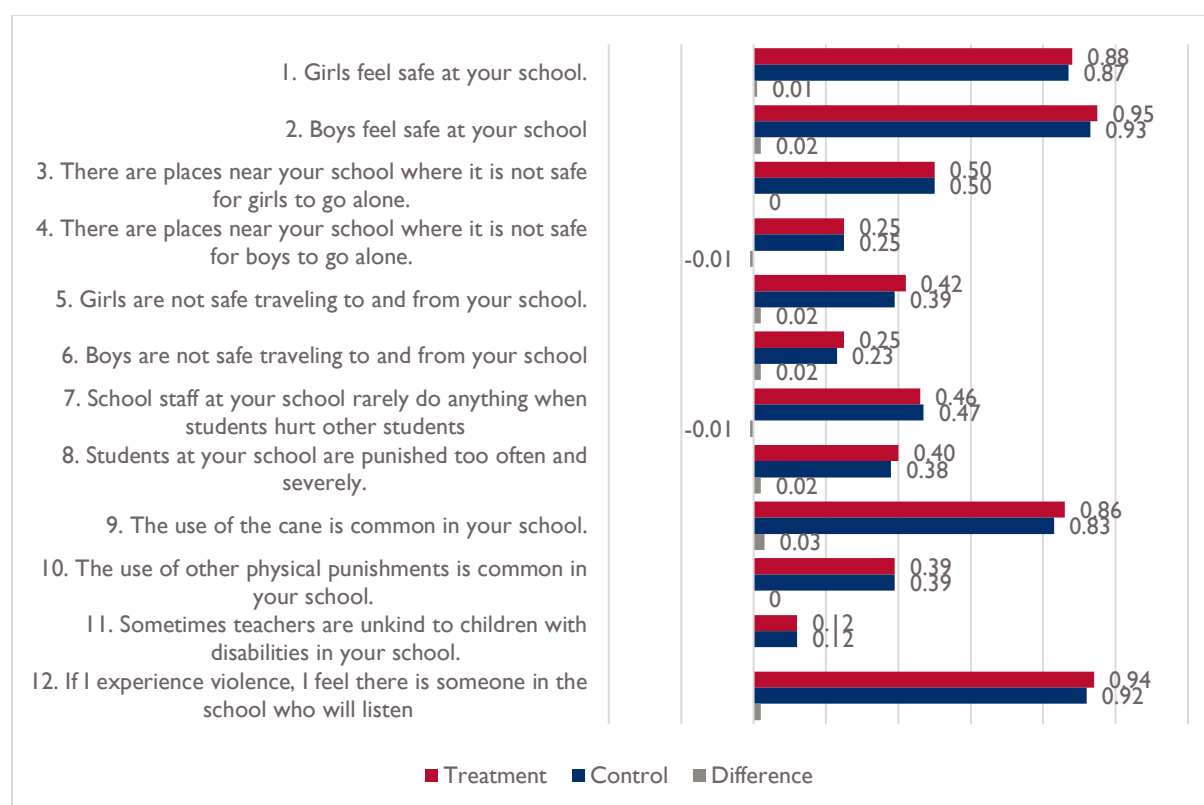
Perpetrators of gender-based violence can be peers, teachers, and other adults in a girl's life. The WW program aimed to give girls a safe and supportive learning environment by providing SRGBV training to students, parents/caregivers, and teachers.

Figure 6 shows results for a series of twelve statements related to perceived frequency and nature of SRGBV, as well as safety more broadly defined. To construct these statements, we reviewed instruments from NORC (2017) and RTI (2016). Responses followed a 4-point Likert scale, namely strongly agree, agree, disagree, or strongly disagree. **Figure 6** shows the fraction of respondents answering agree or strongly agree to each of the twelve items. It is important to keep in mind that these questions collected girls' perceptions of SRGBV, rather than actual incidence, so even if treatment had an impact on SRGBV incidence, it is possible that reporting bias may skew the responses. That would explain, in part, the lack of differences between treatment and control. For example, it is possible girls in the treatment group have grown more aware of SRGBV in their schools thanks to WW, so they could have developed greater knowledge and capacity than their counterparts in the control group to identify forms and acts of SRGBV.

There are no statistically significant differences in the frequency and nature of SRGBV perceived by girls in the two groups. In both groups most girls agreed or strongly agreed with the statement "Girls feel safe at your school" (item 1), and most girls also agreed or strongly agreed with the analogous statement for boys (item 2). The results are different for safety outside the school. Roughly half the girls agreed or strongly agreed with the statement "There are places near your school where it is not safe for girls to go alone." (item 3); however, when the statement is about boys (item 4), only 25 percent of girls agreed or strongly agreed with the statement. In other words, the girls perceive more safety risks for themselves near or outside school than for boys. A similar result can be seen for safety traveling to and from school (items 5 and 6).

We do not find statistically significant differences between treatment and control schools on the role of teachers in preventing violence in schools, as the results for item 7 indicate. About 39 percent of girls agreed or strongly agreed with the statement "Students at your school are punished too much for little things" (item 8), and 84 percent agreed or strongly agreed with the statement "The use of the cane or another form of physical discipline is common in your school" (item 9). A much lower fraction, 39 percent, agreed or strongly agreed with the statement "The use of other physical punishments is common in your school". Together, items 8-10 indicate teachers in surveyed schools use harsh discipline methods. Most girls do not think teachers are unkind to children with disabilities. Finally, most girls agreed or strongly agreed with the statement "If I experience violence, I feel there is someone in the school who will listen" (item 12), and again there are no differences between treatment and control schools.

Figure 6: Perceptions of SRGBV by Treatment Group



Although it is not possible methodologically to derive conclusions on the impact of the program on SRGBV incidence using these data, a few important results are worth highlighting:

First, for the most part, girls in both treatment and control schools seem comfortable with the safety/SRGBV environment in schools, with most reporting feeling safe in schools and having someone to talk to if they experience violence at school. However, girls' perceptions of their own safety at school are lower than girls' perceptions of boys' safety at school. Further, girls perceive that corporal punishment is used too severely in school. This suggests that SRGBV policy and programs should include a focus on preventing corporal punishment. Evidence from other studies suggests links between corporal punishment and the risk to students, particularly girls, of emotional, physical, and sexual violence in schools (NORC 2020). To date, 128 countries have banned corporal punishment in schools, acknowledging it entails multiple types of violence against children, all of which harm child development and academic achievement (see: endcorporalpunishment.org/schools/). A focus on corporal punishment prevention would be an important strategy for increasing overall school safety. But such a strategy would be complicated in Tanzania, where corporal punishment is not only pervasive, but legal.

And second, girls do not feel safe around school or traveling from home to school. This is challenging because as a primarily school-based program, WW has very few tools to address this safety concern, which can affect attendance and, ultimately, school retention.

To measure attitudes towards SRGBV at baseline we presented girls with different hypothetical scenarios where girls like them were exposed to different types of SRGBV. The baseline data showed that most girls rejected any form of SRGBV, so we concluded that there was little room for the

program to change the results on these items.¹² It is possible that social desirability bias could be driving these baseline results, if we assume that girls would like to present themselves as having more gender equitable attitudes towards SRGBV than they really have. To address this we incorporated a series of list experiments to measure both the impact of the program using a method in which girls feel their responses would be more confidential, and the extent of the social desirability bias, if any.¹³ The list experiment results, discussed in Annex VII, show that while social desirability bias does play a role in SRGBV responses, we do not find major differences on attitudes between treatment and control groups even when this elicitation technique is used.

Another possibility is that, while girls may have gender equitable attitudes towards SRGBV, they may perceive their peers as holding more inequitable views. If this is the case, it is possible that even if most girls have equitable views, they will not demonstrate them, because they perceive the social norm on the subject as less equitable. To measure estimates of attitudes and norms that socially condone and normalize SRGBV, in the endline survey we included items where we ask girls not about their own individual opinions, but what they think their peers' views are about different forms of SRGBV. The results of this exercise, described in detail in Annex VII, do not find statistically significant differences between treatment and control groups on girls' beliefs of the opinions of their peers on SRGBV.

The results in this section indicate the program did not have an impact on girls' intermediate outcomes. We did not find statistically significant differences for any of the protective assets, and for self-efficacy we found that the program had a negative effect. We found only positive impacts on gender attitudes.

Finally, we did not find major impacts on the attitudes of girls towards SRGBV using list experiments, or their beliefs about the opinions of their peers on SRGBV.

D3. INTERMEDIATE OUTCOMES - PARENTS

As explained in section A1, one objective of WW was to increase family commitment and capacity to invest in girls' education. To this end, WW helped assemble LIMCA local savings groups in target areas to deliver training on parenting skills and gender equity. LIMCA groups were also geared to provide a way for parents to absorb economic shocks that could affect their ability to invest in their children's education, and their consumption patterns.

To research parent attitudes and actions, we conducted a before-and-after evaluation. The initial target was to survey 450 parents (15 parents in each of 30 schools). But by the time data was collected, fewer than 15 parents in several schools were eligible for the survey. Only adults who were part of a LIMCA group in February 2019 and were parents of girls 10 to 21 years old were eligible. At baseline we surveyed 324 parents, distributed in 22 treatment schools. At endline we surveyed 305 of the 324 parents surveyed at baseline (94 percent). All the results in this section are restricted to the 305 parents

¹² These scenarios have been used successfully in previous studies (RTI (2016) and NORC (2017)).

¹³ In a list experiment respondents are randomly divided into two groups (A/B). Both groups are presented with the same set of three facts or opinions on innocuous topics (e.g. the government should invest more in hospitals). Group A is also presented with a statement on a potentially sensitive topic (e.g. caning is an effective discipline method), while group B is presented with only the common set of statements. Respondents from both groups are asked to indicate the *number* of statements they believe to be true. Because respondents do not respond to each item, but only to the total number of statements they believe to be true, they are provided an added level of confidentiality. Both groups should be expected, on average, to respond in agreement with the same total number of innocuous statements, so the mean difference in the total count between groups A and B approximates the fraction of people for which the sensitive item is true.

observed both at baseline and endline. Seventy-eight percent of parents in the LIMCA groups were women. Further descriptive statistics for this sample are available in Annex VII.

PROGRAM TAKE-UP

Baseline data was collected in February 2019, by which time parents had been accessing credit and participating in LIMCA meetings for a few months. As **Table 11**: shows, at baseline 82 percent of respondents said they were attending LIMCA meetings every week, and 83 percent said they had taken a loan with LIMCA. At endline the percentage of respondents saying they attended meetings every week was 93 percent, and the fraction accessing credit was 85 percent. For those accessing credits, the median loan went from 100,000 Tanzanian shillings at baseline (roughly US\$45), to 150,000 Tanzanian shillings at endline. Notably, LIMCA loans were predominantly used to pay for school-related costs, and to invest in economic activities, as we explain in more detail in Annex IX. It is important to keep in mind that **these figures correspond to parents already enrolled in LIMCA groups**. As explained in section C1, by the end of 2018 we estimated that only 14 percent of parents of girls in treatment schools were enrolled in LIMCA groups, indicating that LIMCA participation rate was rather low, at least when baseline data was collected.

Table 11: Meeting attendance and credit access (N=305)

	Baseline	Endline
Attended meetings every week (fraction)	0.82	0.93
Taken a loan (fraction)	0.83	0.85
Amount (Median TZS)	100,000	150,000

Notes: Sample sizes are lower for some outcomes due to item-specific missing data. The median loan figures include borrowers only.

The fact that parents were already attending meetings when baseline data was collected could appear to compromise the validity of the baseline. However, training on parenting skills and gender norms began in March 2019, a few weeks after baseline data was collected. Because the research questions below are about attitudes towards gender equity and parenting skills, the baseline was not contaminated by the training on these topics.

Although we did not survey parents in the control group, and in principle cannot document whether control parents had access to services similar to LIMCA groups, in the endline survey conducted with girls in the group, we did include a question on whether their parents were currently part of a savings club. The results, shown in **Table 12**:, indicate that 38 percent of parents, both in treatment and control groups, participate in savings clubs.

Table 12: Participation in savings clubs according to the girl survey (N=1792)

	Treatment	Control	Difference
Fraction of parents participating in savings clubs	0.38	0.38	-0.00

Assuming most parents in the treatment group who were in a savings club were in a LIMCA group, the take-up rate by early 2020 was higher than our late 2018 estimate (14 percent), but still relatively low. Most importantly, **parents were as likely to participate in savings clubs in control as in treatment groups**, suggesting that LIMCA substitution was probably substantial. Although we cannot tell if parents in control groups were exposed to similar content in terms of gender attitudes and

parenting training as parents in LIMCA groups, it is reasonable to assume there were no major differences in terms of access to savings/credit instruments.

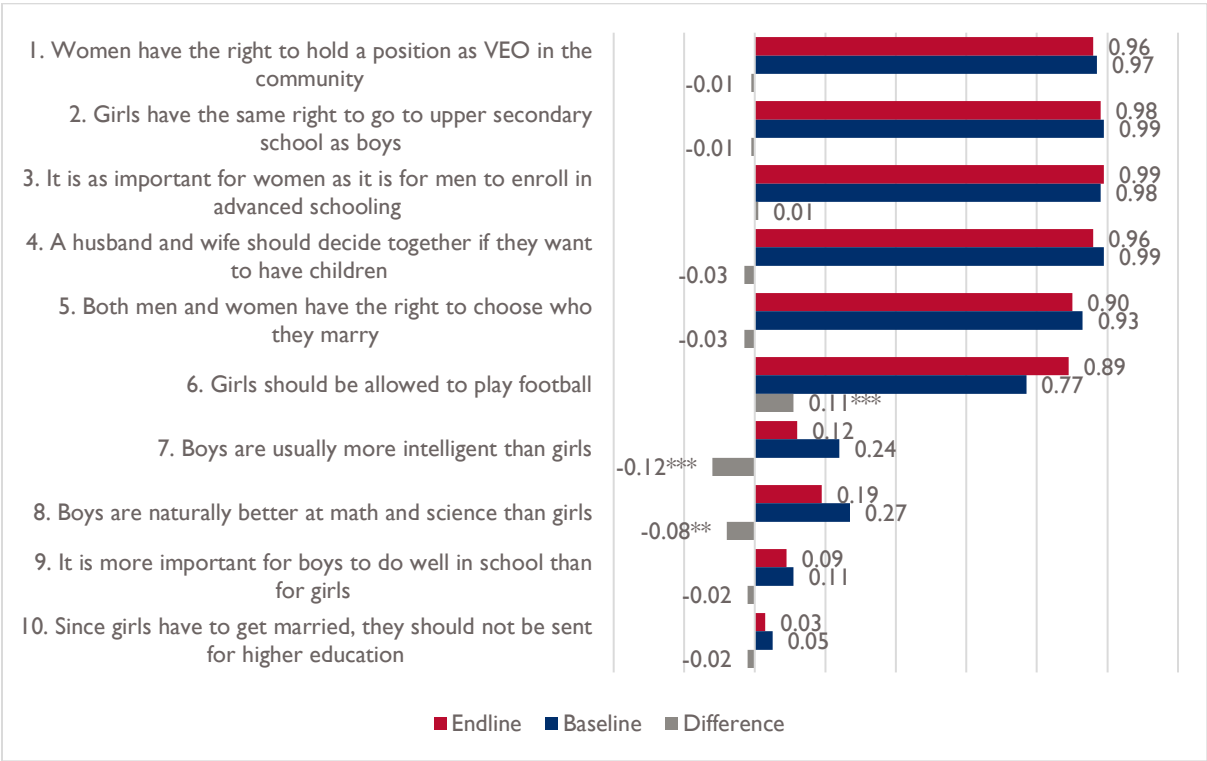
CHANGES BETWEEN BASELINE AND ENDLINE

IE6. Does the program improve attitudes towards gender equitable access to education within the household?

To analyze changes in parents attitudes towards gender equity we asked them about gender roles at school and at home and how much time their sons and daughters spend studying at home and on chores. The survey also included questions about domestic gender-based violence.

We presented parents with a list of 10 statements about gender and used a 4-point scale to record their responses, either strongly agree, agree, disagree, or strongly disagree.¹⁴ **Figure 7** shows the fraction of respondents answering agree or strongly agree to each of the 10 items. For the first six items the expected response from someone with progressive views towards gender equity are agree or strongly agree, while the opposite response would be expected for items seven to 10.

Figure 7: Fraction of parents answering agree or strongly agree



The results show that respondents have equitable gender views. More than 90 percent, both at baseline and endline, agreed or strongly agreed with statements 1-5, and less than 10 percent agreed or strongly agreed with statements 9 and 10. Notably, for the items where we saw fewer progressive views at baseline, namely items 6, 7 and 8, **by endline, there was an increase in the percent of parents**

¹⁴ The rubric was constructed using items from the Gender Equitable Index (GEI) (Care, 2014) and NORC (2017).

with more progressive/equitable views. For example, the percent of parents who agreed or strongly agreed that girls should be allowed to play football (item 6) was 77 percent at baseline and 89 percent at endline, and the percent of parents who agreed or strongly agreed that boys are naturally better at math and science than girls (item 8) declined from 27 percent at baseline to 19 percent at endline. Aside from the changes for items 6-8, the one other change between baseline and endline that is statically significant is for item 4. For this item, the results indicate that the percent of parents who agreed or strongly agreed that a husband and wife should decide together if they want to have children dropped from 99 to 96 percent. Although the results for item 4 indicate that the percent of parents with less gender equitable views increased after the program, the bottom line is that both at baseline and endline the percent who agreed with the statement was quite high, and the difference between the two moments, albeit statically significant, is small.

While attitudes towards gender norms are important inputs to characterize parents and explore how participation in LIMCA groups affected how much they value girls' education, a more direct way of addressing this is to study school-investment decisions made by households. To this end we asked respondents the school enrollment status of each child/youth living with them (10 to 21 years old), and the amount of time children/youth spent studying at home. As we explain in more detail in Annex X, we find that **girls are marginally more likely to be in school than boys both at baseline and endline, and for those in school we do not find major differences on time spent studying between boys and girls, both at baseline and endline.** This suggests that parents in LIMCA groups valued girls' education at baseline, so it is not clear there was much opportunity to improve their decisions in terms of school enrollment, or giving girls enough time to study at home, at least not relative to what parents were doing with boys. It is important to keep in mind that this sample was selected from parents who both had daughters in schools and decided to be part of a LIMCA group, a group of parents who may value girls' education more than the average parent in their communities.

IE7. Does the program change parenting skills?

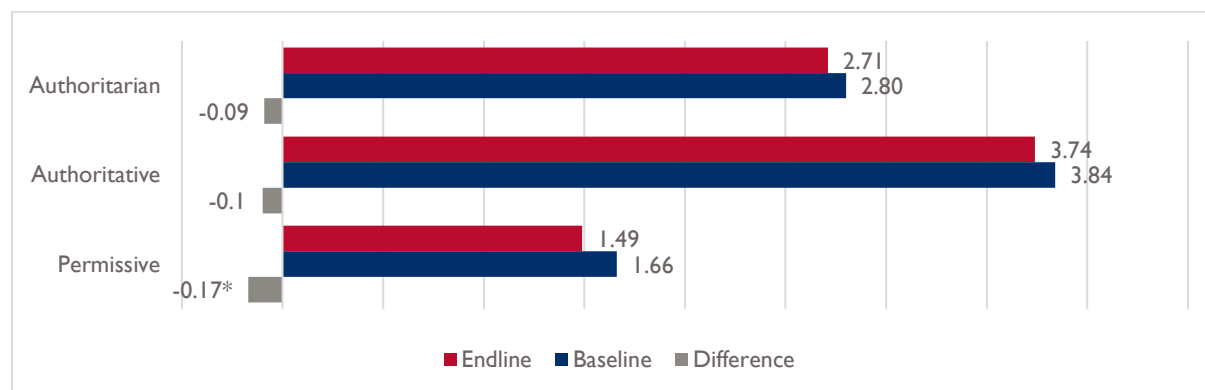
Parenting skills training introduces participants to different parenting styles, alternative discipline methods, and to techniques to communicate with children and adolescents. The training addresses risks faced by children and adolescents and how parents can help keep them safe. It focuses on adolescent development, including the role of parents in sex education and how parents can support adolescents during puberty.

To assess whether respondents modified their parenting styles before and after training, we used items from the short version of the Parenting Styles and Dimensions Questionnaire (Robinson, et. al 2001). The rubric includes 32 statements with a 5-point scale anchored by never (1) and always (5), and measures authoritative, authoritarian and permissive parenting.¹⁵ For brevity we randomly selected 15 of these items to approximate scores for the three parenting styles. LIMCA group training emphasized the importance of communication with children, including explaining the reasons for rules, so the expectation was that parents would tend to be more authoritative and less authoritarian. **Figure 8 shows that changes in the scores between baseline and endline are relatively small, and only**

¹⁵ Baumrind (1978) classifies parental disciplinary patterns in three main categories: authoritarian, permissive and authoritative. Authoritarian parents use forceful measures to exercise control over children, and do not foster discussions with children on the reason for the rules they put in place. Permissive parents are much more lenient than authoritative parents, and exercise little control on children's behavior. Authoritative parents provide children with some level of freedom and self-determination. They exercise control when children disobey but are willing to discuss reasons for rules.

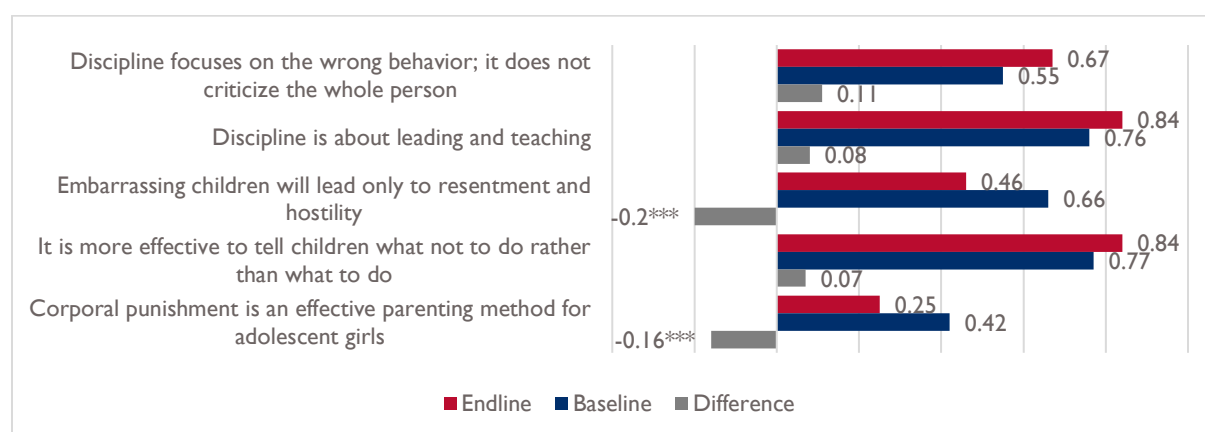
the one for permissive parenting is significant. It is possible this reflects the fact that parents became less permissive as children grew into adolescents.

Figure 8: Changes in parenting style scores



We also asked parents whether they believe five statements, mostly related to different discipline strategies, to be true or false. The responses are summarized in **Figure 9**. Given the content of the training, the expectation was that the fraction of parents saying statements 1 to 3 were true would increase between baseline and endline, and the fraction of parents saying statements 4 and 5 were true would decrease. The results show that there are significant differences only for items 3 and 5. The results for statement 3 indicate that the fraction of parents who think that embarrassing children will lead to resentment and hostility fell 20 percentage points between baseline and endline. As with any before-and-after results it is difficult to attribute this change to the program, as other things are changing at the same time in the lives of the respondents and their children. In this particular case, for example, it could be that as children grow, parents perceive they would be less sensitive to this type of disciplining. The results for statement 5 show that the fraction of parents who think that corporal punishment is an effective discipline strategy fell 16 percentage points between baseline and endline.

Figure 9: Fraction of Parents saying statements on disciplinary strategies are true.



A few important conclusions can be derived from the parent survey data. First, **parents seem to hold equitable views on gender, even at baseline**. This was reflected not only in parents' responses to different gender-related items, but also on their reports on gender gaps (or lack of them) on children's school enrollment or time spent studying at home. The fact that parents in this sample already held

gender equitable views at baseline could explain why we did not find major differences on these outcomes between baseline and endline, although for some gender-related items, we did find that the fraction of parents with gender equitable views grew between baseline and endline. **If parents already hold gender equitable views on access to education, then programs like WW should shift the focus from trying to affect parental attitudes around gender equity as an instrument to increase girls' education.**

Second, although we did not find major differences in parenting styles between baseline and endline, we did find evidence suggesting parents are more aware at endline than at baseline that corporal punishment is not an effective parenting method. It is possible that parenting styles are not very malleable, and a few sessions on parenting may not be enough to induce change. Focusing on a topic as specific and highly consequential as corporal punishment seems to be more productive. **These results suggest that programs trying to change parenting styles should focus more on raising awareness of the impact of corporal punishment, and less on parenting more holistically defined.**

Some of these responses may be plagued by social desirability bias. Although this is a valid concern, in terms of school enrollment and attendance the results from administrative data discussed in section D1 indicate that school retention and attendance gender gaps in these communities are small and, if anything, favor girls. This is consistent with responses from parents in the LIMCA group sample supporting gender-equitable access to education, suggesting that social desirability bias may be a minor concern in this context.

Finally, it is important to keep in mind that, when the LIMCA sample was selected, only 14 percent of eligible parents had signed up. Thus, while the attendance data reported in section D1 suggest equitable access to education across the board in the target districts, we cannot conclude that the views of that relatively small group on gender equity are representative of the majority of parents in the study districts.

D4. INTERMEDIATE OUTCOMES - TEACHERS

To assess whether the program changed the attitudes and practices of teachers related to SRGBV, we surveyed 292 teachers in treatment schools and 297 teachers in control schools (almost 10 per school).¹⁶ Teachers' descriptive statistics are available in Annex VII.

PROGRAM TAKE-UP

The main component of WW's Objective 3 (Foster a girl-friendly and supportive school environment) was to train teachers on SRGBV prevention and response. WEI implemented a cascade approach in which 33 district master trainers (henceforth master teachers) trained all teachers in the 30 treatment schools in 2018. The four-day training sessions were supplemented with a one-day refresher session in 2019, for a total of five days of training. Training addressed:¹⁷

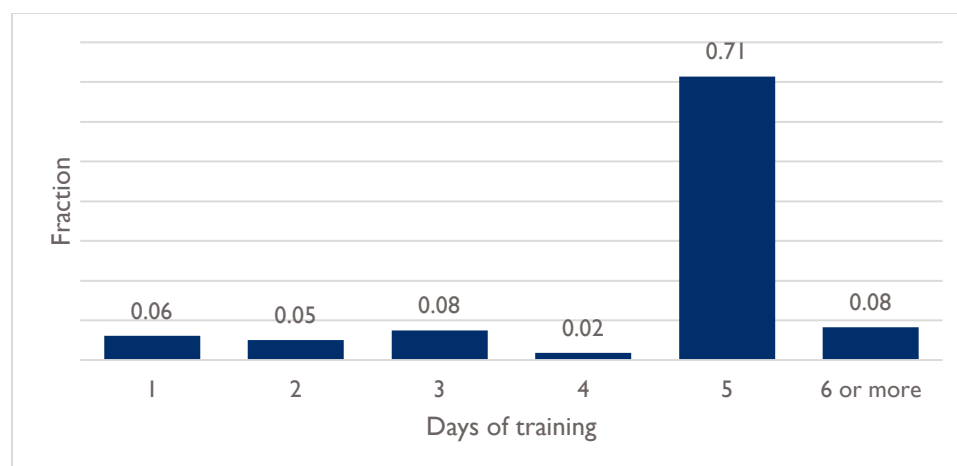
¹⁶ We planned to survey 10 teachers in each of the 61 study schools, for a total of 610 teachers. In each school we pre-selected up to 15 teachers (10 primary and 5 replacements), prioritizing teachers who taught Form 1 in 2018 or Form 2 in 2019. After screening out volunteer teachers and teachers that were not working at the school by the end of 2018, as well as three who declined to participate in the survey, we yielded a sample of 589 teachers (96 percent of the target), out of which 80 were replacements, for a replacement rate of 13.6 percent.

¹⁷ Source: WEI (2018)

- 1) Existing laws and policies that address violence against children
- 2) Child rights case studies
- 3) Assessments of knowledge on the existing national laws and policies
- 4) Defining gender versus sex
- 5) How gender affects school outcome
- 6) Supporting students with special needs
- 7) Positive discipline
- 8) Identification of harmful classroom practices
- 9) Strengthening school-based regulations to improve school environment
- 10) Reviewing and refining action plans for improving school environment
- 11) Case monitoring

Of the 292 teachers in the treatment group 283 participated in SRGBV training (97 percent), while only 12 teachers in the control group (4 percent) participated in SRGBV training (presumably provided by a different agency). **Figure 10** shows that of the teachers in the treatment group who participated in training, 71 percent attended all 5 days offered.

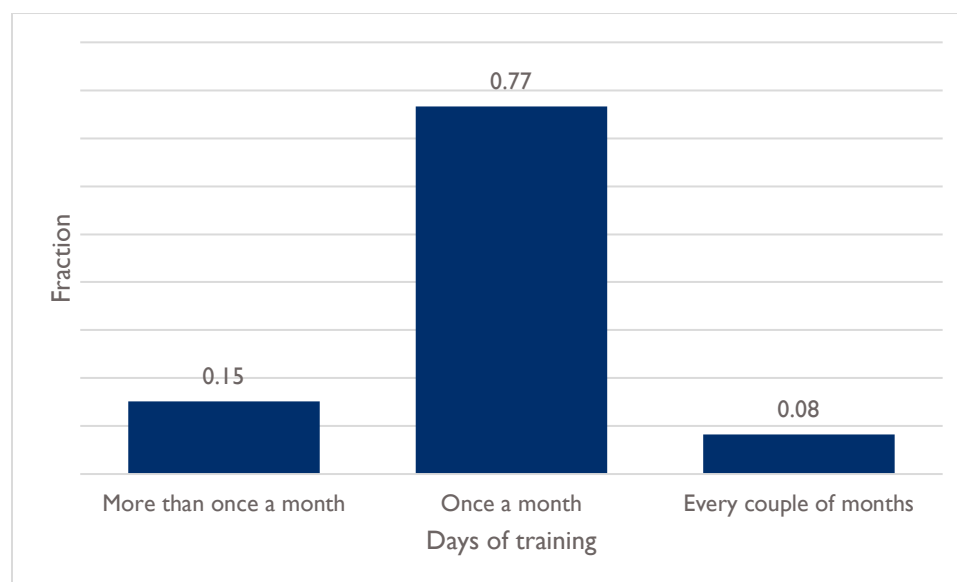
Figure 10: Days of SRGBV training (N=280*)



*The sample size for the figure is 3 teachers lower than the number of teachers that attended training, due to no response on the number days of training question.

As a mechanism for teachers to continue working on their SRGBV prevention and response-related skills, the program implemented Lunch and Learn sessions, held monthly to allow sharing of general or specific gender-based violence issues. Ninety-five percent of teachers in the treatment group said they had attended at least one Lunch and Learn session at school, compared to none in the control group. As expected, the majority of those teachers (77 percent) said they attended sessions once a month, as **Figure 11** shows.

Figure 11: Attendance at Lunch and Learn sessions (N=278)



The results of this section show that most teachers attended SRGBV training. More importantly, control teachers did not attend SRGBV training, so **substitution of this program component should not be a concern for the impact evaluation.**¹⁸

PROGRAM IMPACT

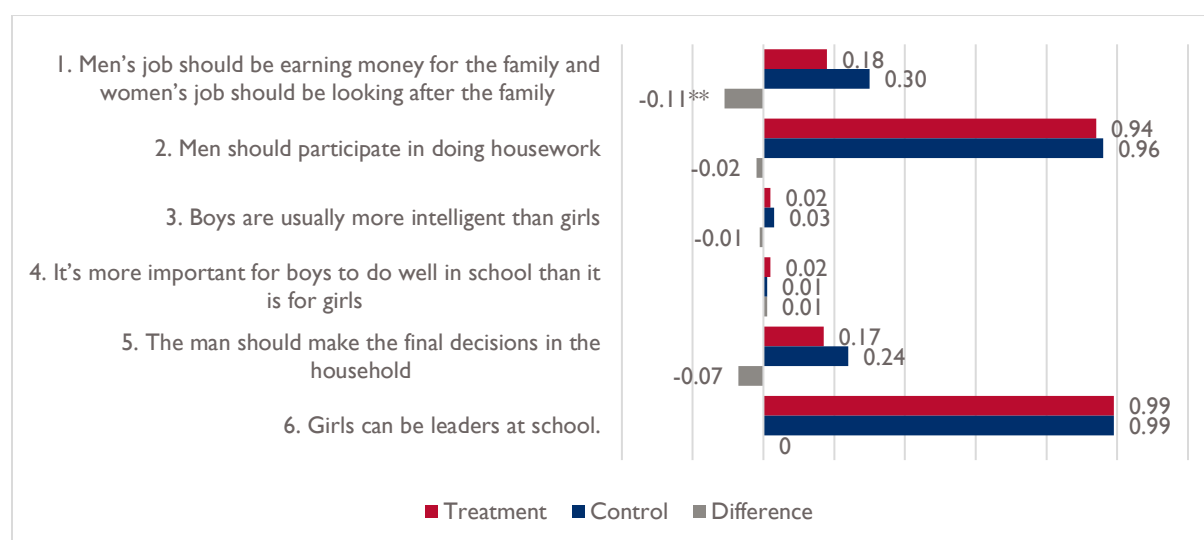
IE7. Does the program change gender attitudes and perceptions of teachers in school?

To assess whether the program had an impact on attitudes and perceptions towards gender equity, we presented teachers with a list of statements about gender and used a 4-point scale to record their responses, either strongly agree, agree, disagree, or strongly disagree.¹⁹ **Figure 12** shows the fraction of respondents answering agree or strongly agree to six selected items. Note that for items 2 and 6 the expected response from respondents with gender equitable views are agree or strongly agree, while the opposite response would be expected for items 1, 3, 4 and 5. There are statistically significant differences between treatment and control groups only for the first item (“Men’s job should be earning money for the family and women’s job should be looking after the family”), as 30 percent of the control group agrees or strongly agrees with the statement while only 18 percent of the treatment group do.

¹⁸ In this context program substitution refers to the possibility that teachers in the control group attend SRGBV training similar to the one offered by WW in treatment schools.

¹⁹ The rubric was constructed using items from the Gender Equitable Index (GEI) (Care, 2014) and Menendez, Haugan, and Grotophorst (2018).

Figure 12: Fraction of teachers that agree to gender-related questions (N=589)



Note: Sample size is one observation smaller for some items due to item-specific missing data. Standard errors for the significance test are clustered at the school level.

* p<0.05 ** p<0.01 *** p<0.001

Teachers in both treatment and control groups report largely gender equitable views. This makes it difficult to identify program impact. This is apparent when we look at all items except for 1 and 5.²⁰ The results for these two items suggest the program had the expected effect on attitudes towards gender equity, although the difference is statistically significant at standard levels of confidence for item 1 only, while the difference for item 5 is only marginally significant (p-value= 0.08).

Therefore, the results indicate that **WW did have a positive albeit small impact on teachers' attitudes towards gender equity.**

IE8. Does the program change teachers' knowledge, attitudes, and practices related to SRGBV?

We evaluated teachers' knowledge of SRGBV prevention and response utilizing two types of questions. First, we asked direct questions about Tanzanian regulations on child protection. This included questions on whether children in Tanzania become adults when they get married and the maximum number of times teachers are allowed to cane students (four times). Second, we asked teachers to determine if a series of situations constitute SRGBV.

Table 13: summarizes the results of these questions. Next to each score's label we show its range. In the first row we show the number of items answered correctly regarding child protection regulation. We asked three questions, the minimum score 0 and the maximum 3. Teachers in the treatment group answered correctly 0.2 items more than the control group. To assess teachers' ability to identify SRGBV, we asked them two different types of questions. Type I questions involved asking teachers if different scenarios involved a given form of SRGBV (e.g. If a teacher made threats to a student, should that be considered a form of psychological violence?). The rubric also included items that were not

²⁰ Furthermore, we are not showing results for another five items in gender attitudes that also show that teachers have largely gender equitable views in both treatment and control groups.

SRGBV to penalize respondents who would simply answer “yes” to everything.²¹ The results indicate that teachers in the treatment group performed better than their counterparts in the control group, although the difference is not statistically significant. Type II questions presented teachers with five scenarios and asked them to determine the type of violence identifiable in those scenarios, (i.e. psychological, physical, sexual, or economic), or if the scenario did not describe SRGBV at all. While in Type I questions the form of violence teachers had to identify was indicated (e.g. is X event economic violence?), Type II questions are more challenging because they involve selecting the type of violence taking place. Again, treatment teachers outperformed control teachers on this type of question, and in this case the difference was statistically significant.

Table 13: Teachers’ knowledge about child protection regulations and types of SRGBV (n=589)

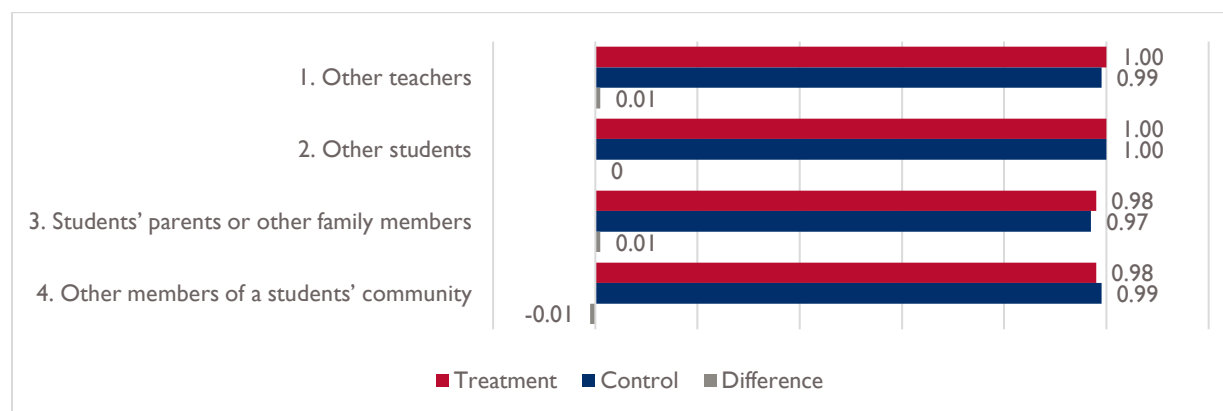
	Treatment (1)	Control (2)	Difference (3)=(1)-(2)
Knowledge of the law (0-3)	2.50	2.29	0.20**
<i>SRGBV identification</i>			
Question type I (0-4)	2.00	1.84	0.16
Question type II (0-5)	3.97	3.79	0.17*

Standard errors for the significance tests are clustered at the school level.

* p<0.05 ** p<0.01 *** p<0.001

To assess attitudes and perceptions of teachers towards SRGBV, we asked teachers whether they thought they had a role and responsibility to intervene when students experience violence from other teachers, other students, students’ parents, or other members of the community. As the results in **Figure 13** show, practically all teachers think they have a responsibility to intervene in these situations, regardless of their treatment status.

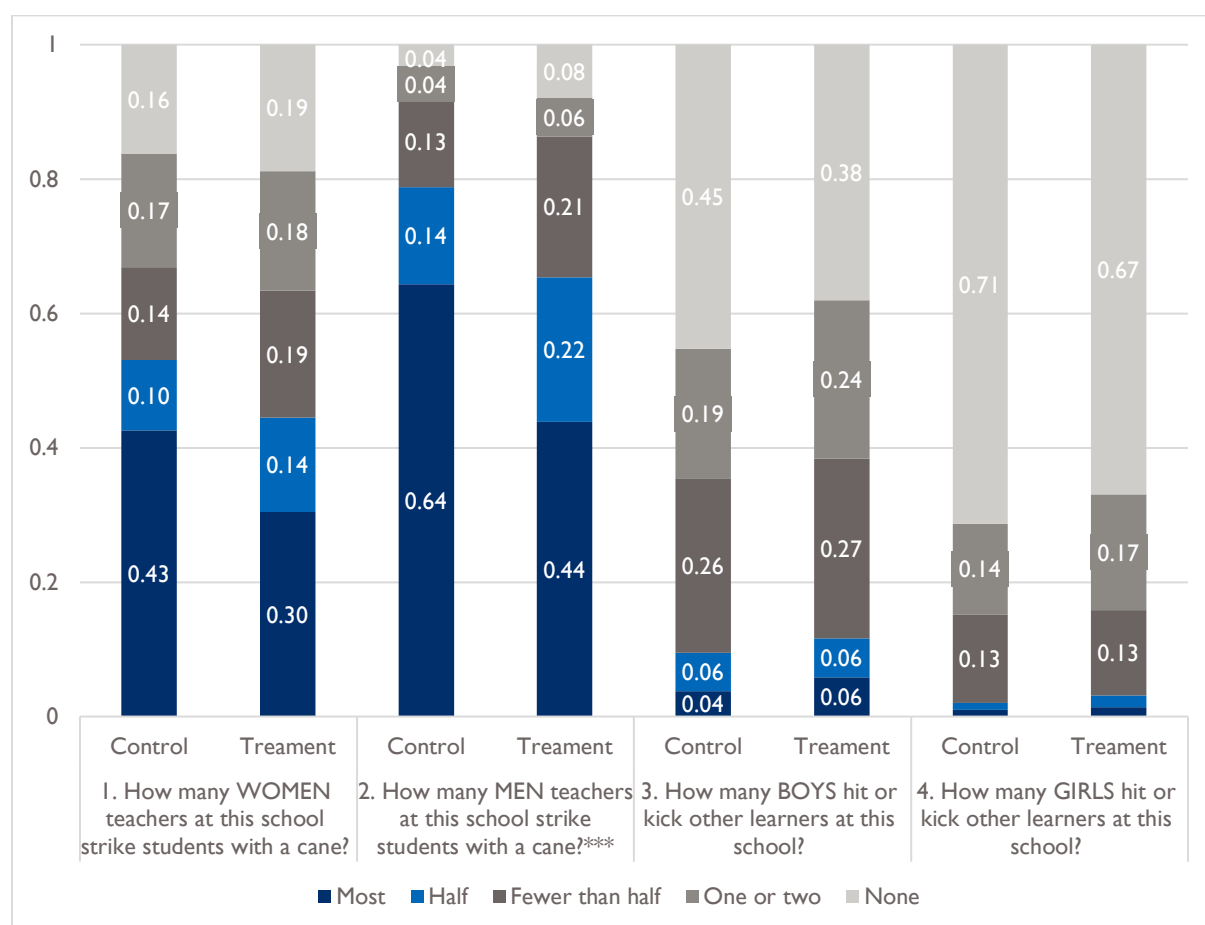
Figure 13: Fraction of teachers that think they have a role and responsibility to intervene when students experience violence from... (N=589)



²¹ To assess for example how well teachers can identify psychological SRGBV we asked them four questions, three of which described a form of psychological SRGBV, and a fourth question that did not involve SRGBV. If a teacher answers all four questions correctly, s/he gets 1 point, otherwise s/he gets 0. Similar batteries of questions were asked for physical, sexual and economic SRGBV. The score for type I questions shows the result of adding the results for the four batteries, so a teacher could have a minimum of score of 0 and a maximum of 4.

Teachers overwhelmingly report believing they need to intervene when there is violence in schools. To further explore teachers' beliefs and expectations about SRGBV in their schools, we asked them how widely teachers and students perpetrate physical and sexual violence at school. **Figure 14** summarizes the results for physical violence. Item 1 asks how many women teachers engage in striking students with a cane in their school. Forty-three percent of teachers in the control group of schools and 30 percent in the treatment group of schools said most women teachers cane students. The differences between the two groups are not statistically significant. The results indicate that male teachers in the control group are more likely to cane students than their counterparts in the treatment group of schools (item 3), and in this case the difference is statistically significant. However, it is important to consider the possibility that this may be driven by social desirability bias, as teachers exposed to SRGBV training may be more likely to report lower incidence, even if they do not change their behavior. The fact that girls' reports do not indicate differences between treatment and control groups on teacher caning (see **Figure 6**), suggests social desirability bias may be playing a role in teachers' responses. Items 3 and 4 ask how many boys and girls, respectively, hit or kick other students at school. The results indicate that boys engage much more in this type of behavior than girls, and there are no major differences by treatment status.

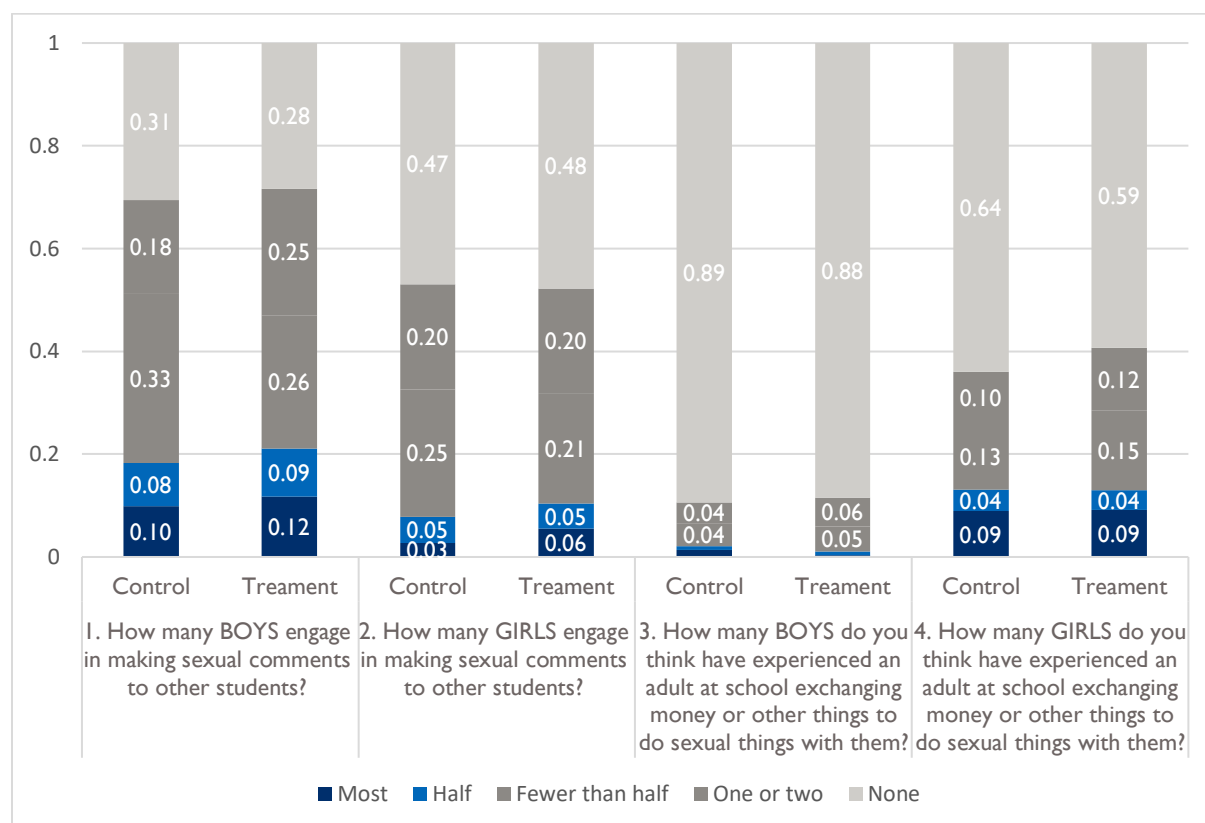
Figure 14: Teachers' perceptions of physical violence in schools (N=589)



Note: To estimate whether differences between treatment and control groups are statistically significant, we ran a type of regression model called ordered probit (Green, 2008, p. 831) of the survey responses on the treatment dummy. Stars at the end of each statement (* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$) indicate the level of significance of the treatment dummy. Standard errors for the significance tests are clustered at the school level.

Figure 15 shows respondents' perceptions of the proportions of teachers and students who perpetrate sexual violence at school. As the results for Items 1 and 2 show, respondents reported that both boys and girls make sexual comments to other students, although boys are more likely to than girls. Items 3 and 4 ask teachers how many children (girls and boys, respectively) they think have experienced an adult at school exchanging money, gifts or favors for sexual acts. Almost 90 percent of teachers said they think no boys have experienced this at their schools, but for girls the results suggest a much higher prevalence. We do not find statistically significant differences between treatment and control groups for any of the items on sexual violence.

Figure 15: Teachers' perceptions of proportions of teachers and students who perpetrate sexual violence at school (N=589)



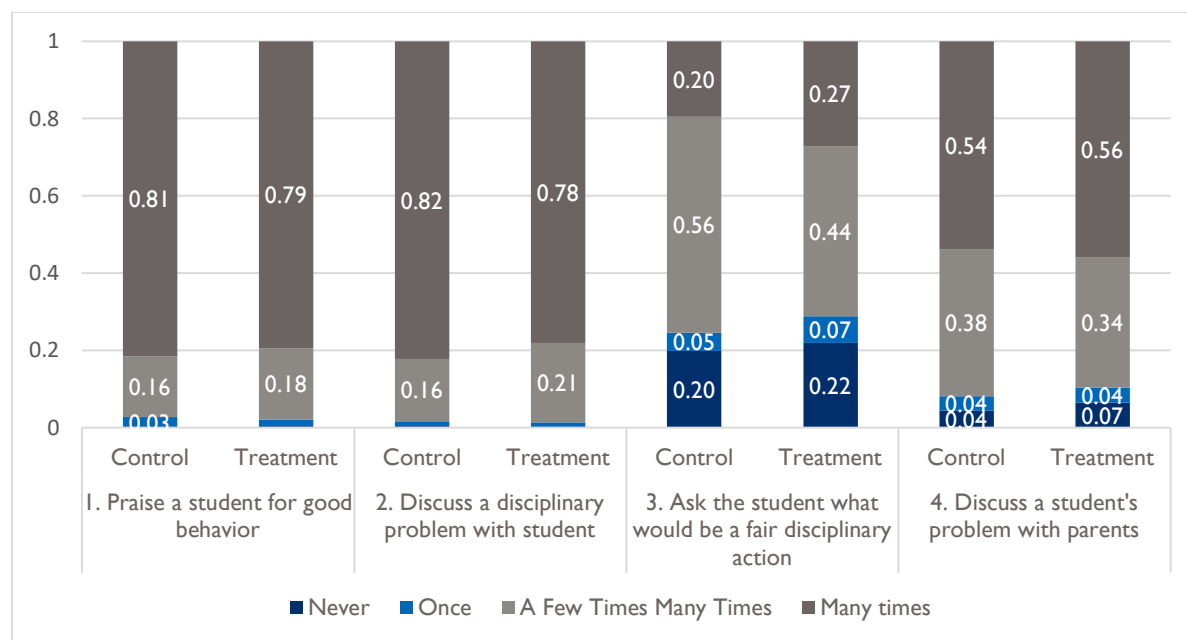
Note: To estimate whether differences between treatment and control groups are statistically significant, we ran an ordered probit model of the survey responses on the treatment dummy. Stars at the end of each statement (* p<0.05 ** p<0.01 *** p<0.001) indicate the level of significance of the treatment dummy. Standard errors for the significance tests are clustered at the school level.

To assess teachers' practices related to SRGBV, we presented teachers with a series of statements describing non-violent and violent discipline methods and asked them how often they had used each method during the previous term. **Figure 16** summarizes the responses. Panel A shows results for positive, non-violent, discipline methods. We found no statistically significant differences between treatment and control groups for any of the items in panel A. Panel B shows results for other, more psychologically or physically violent discipline methods. In this case we found statistically significant differences between treatment and control groups for Item 2, implying that teachers in the treatment group are less likely to assign students chores as punishment than teachers in the control group. We also found somewhat larger differences for Items 1 and 5, indicating that teachers in the treatment

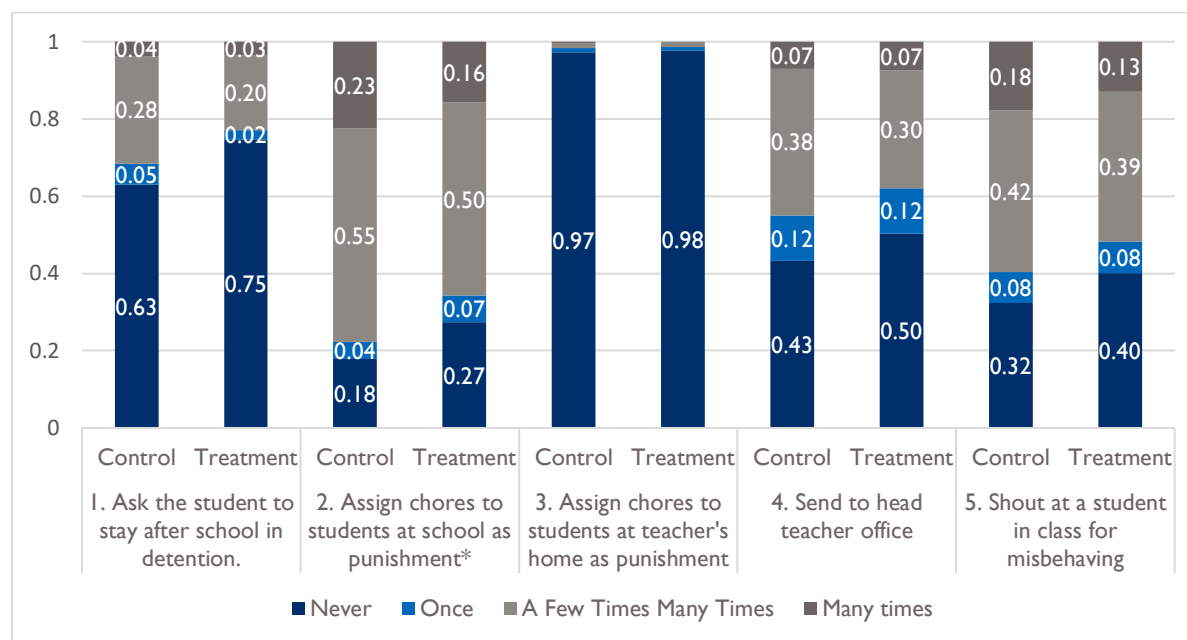
group are less likely to engage in harsh discipline methods. Even so, the differences are only marginally statistically significant (p-values of 0.06 and 0.07, respectively).

Figure 16: Frequency at which teachers engage in positive, non-violent versus harsh disciplinary methods

Panel A. Positive, non-violent disciplinary methods



Panel B. Psychologically or physically harsh disciplinary methods



Note: To estimate whether differences between treatment and control groups are statistically significant, we ran an ordered probit model of the survey responses on the treatment dummy. Stars at the end of each statement (* p<0.05 ** p<0.01 *** p<0.001) indicate the level of significance of the treatment dummy. Standard errors for the significance tests are clustered at the school level.

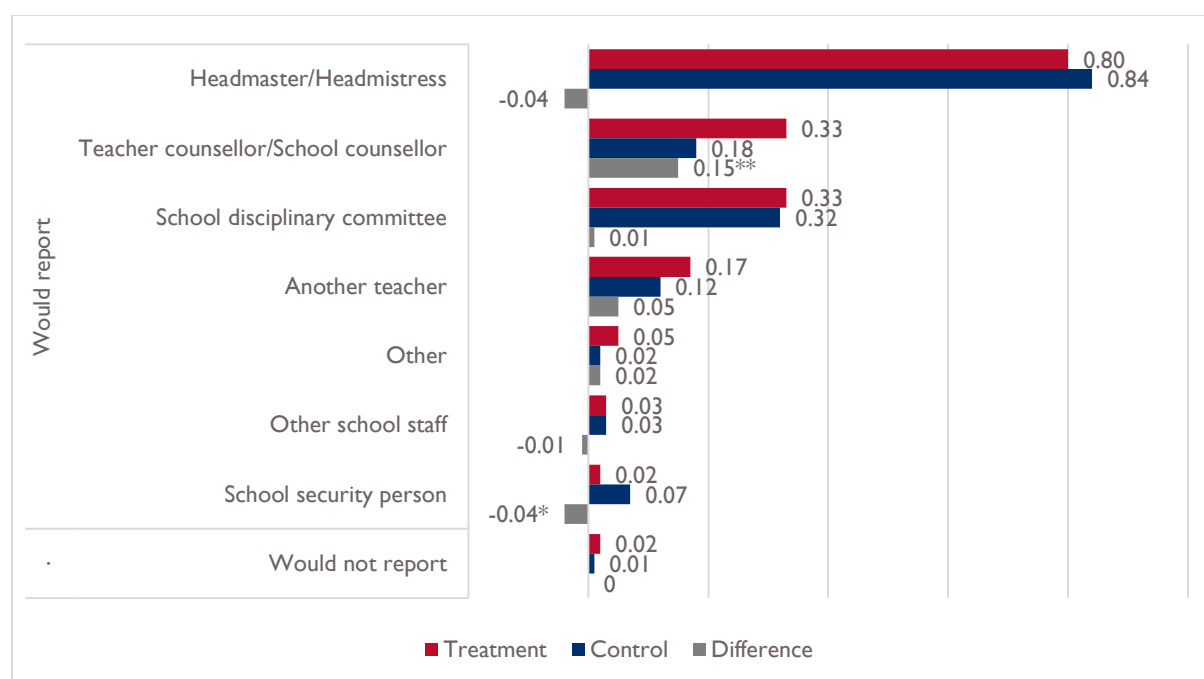
These results suggest that, **while the program did not have a major impact on the likelihood of teachers engaging in positive discipline methods** such as praising students and fostering greater communication, **teachers in the treatment group seem to be less likely to engage in psychologically or physically harsh forms of discipline** such as assigning chores as punishment.

Results regarding teachers' practices and perceptions suggest WW had an impact on some relevant SRGBV outcomes. Most notably, **teachers, especially male teachers, are less likely to cane students in treatment than in control schools. Teachers in treatment schools, meanwhile, are less likely to assign chores at school as punishment than their counterparts in the control group.** However, as mentioned before, some of these results may be driven by social desirability bias rather than actual behavior change.

Finally, we asked teachers if they would report within the school an incident in which another teacher or school staff person gave good grades, money, or other factors in exchange for sex with a student, and to whom they would make such report. **Figure 17** shows the results. While it would be compliant with Tanzanian policy for a teacher to report SRGBV to the head teacher, internationally recommended SRGBV reporting practices indicate that it would be better for teachers to whom a child discloses violence to report to a school counselor, who should have special training to handle SRGBV reports within a mechanism that does not require reporting through school hierarchy. In this case, the school counselor would then report to a designated officer at a local government level with a formal mandate for child protection, such as a Social Welfare Officer, for investigation and making referrals for support services for the child survivor. Another option would be that the teacher would report directly to a formally mandated child protection officer, again for example, a Social Welfare Officer. This option could help maintain the child's privacy and confidentiality in order to support their protection, and should result in referrals for support services for direct assistance to the child (e.g., skilled psychosocial counseling, medical care, safe house, legal aid).

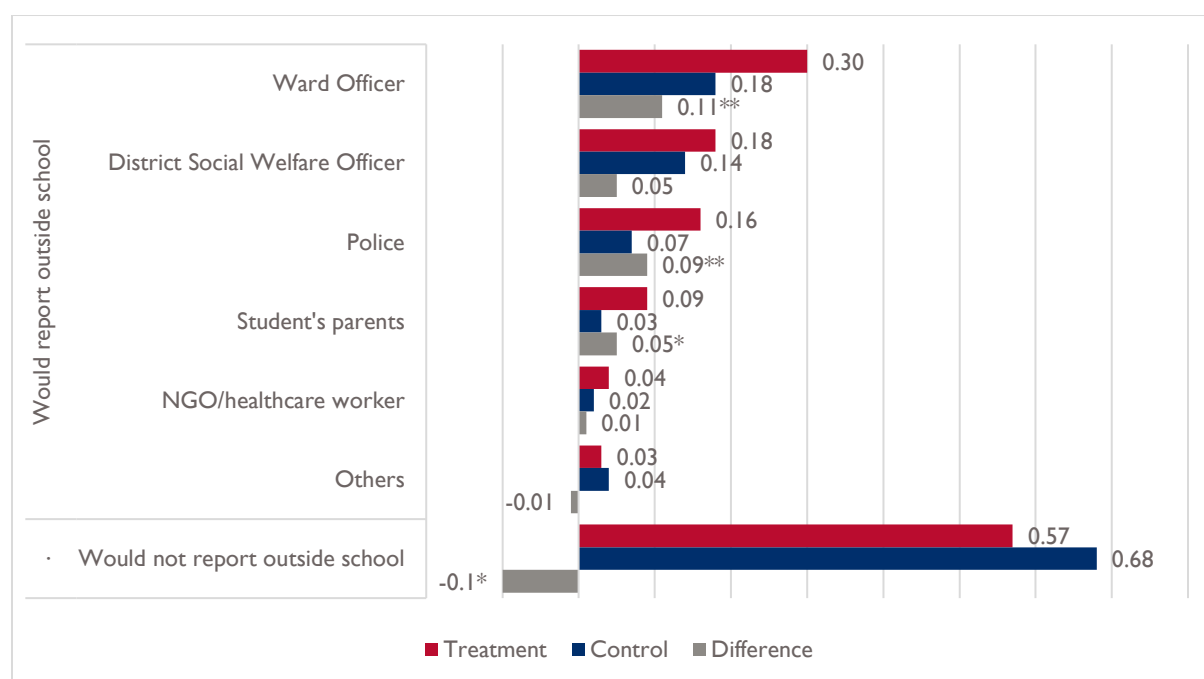
Reporting to a mandated child protection officer should also help protect the integrity of a potential follow-up law enforcement investigation if all involved uphold the privacy and confidentiality of the child's case, limiting information only to formal investigators, doctor or nurse, a psychosocial counselor, social worker. Reporting to a head teacher or up the school hierarchy, unfortunately raises risks of potential gatekeeping, stigma and even retaliatory attacks against the child. Head teachers certainly have a duty of care toward learners, yet they do not have a legal mandate as a child protection officer or professional skills for formal investigation of sexual or severe physical violence against a child or for making referrals to support services for the child's recovery. Still, although teachers in the treatment group are almost twice as likely to report to the counselor than teachers in the control group, **the vast majority of teachers in both treatment and control groups reported they would tell the headmaster or mistress of the school**, a potential gatekeeper who may interfere with confidential reporting to outside child protection authorities who could assist the student with referrals to support services. Notably, very few teachers said they would not report the incident.

Figure 17: Teachers self-reported response on to whom they will report in the school



We also asked teachers whether they would report to authorities outside the school, and if so, to whom. Teachers in the treatment group were ten percentage points more likely to do so than teachers in the control group, as **Figure 18** shows. However, among teachers who report outside school, those in the treatment group are more likely to report directly to police or the student's parents than are teachers in the control group, which while may seem like the right thing to do in the context, may not result protection and direct support to or justice for the child survivor. While contacting a Gender and Children's Desk Police Officer should result in appropriate referrals to survivor support services, the preferred course of action would be to contact the Social Welfare Officer provided such an officer is available and resourced to connect the child with needed child survivor-specific referral services for their recovery and justice. Further, a teacher reporting directly to the police a child's disclosure of sexual or severe physical violence could be accused of defamation of the alleged perpetrator. If a student reports to a teacher having experienced sexual violence or physical violence with injuries, it would be preferable the teacher report directly not to the head teacher, parents, or police, but to the Social Welfare Officer, a doctor or nurse, or a nongovernmental organization officer with a defined role in gender-based violence response or child protection.

Figure 18: Teachers self-reported response on to whom they will report outside the school



The results of this section indicate that: **The program had a positive impact on teachers' knowledge.** We find that teachers in the treatment group were more likely to correctly answer questions on children's protection regulation and were better at identifying different forms of SRGBV than teachers in the control group. Most teachers in both study groups believe they need to intervene when there is violence in schools. And while we did not find many significant differences in terms of prevalence of physical or sexual violence as perceived by teachers, we found evidence that teachers in the treatment schools **are less likely to cane students and assign chores at school as punishment than their counterparts in the control group.** Lastly, while most teachers said they would report sexual violence incidents within school, the majority of both treatment and control teachers said they would report to the head teacher.

E. DISCUSSION

A number of factors may explain why the study found no impact of WW on school retention. In this section we review them.

TAKE-UP AND SUBSTITUTION

Table 14 summarizes the results of program take-up and substitution for girls, parents, and teachers, and their corresponding program components (POY clubs, LIMCA and SRGBV training). A checkmark (✓) indicates there were no major challenges, and an X that there were. For the Girls/POY clubs we indicated that while participation in POY clubs was high, other alternatives were available in control schools, so program substitution seems to be a problem. For parents we showed that take-up was relatively low, and that other alternatives, similar to LIMCA, seem to be available in the control areas (at least according to the survey of girls). However, it is worth highlighting that while other savings groups may be available, it is not necessarily the case that they would have the same education-oriented aspects

LIMCA groups have; therefore, while as a smoothing consumption instrument LIMCA groups may have substitutes, that is probably not the case for other LIMCA components, like parenting training. Finally, for teachers we showed that take-up was high and substitution was low. In summation, it seems that high program substitution (in the case of girls and parents) and to a lesser extent low program take-up (parents) are significant limitations for the evaluation and could explain the lack of results on school retention.

Table 14: Program take-up and substitution

	Take-up	Substitution
Girls/POY clubs	√	X
Parents/LIMCA	X	X
Teachers/SRGBV training	√	√

IMPACT ON INTERMEDIATE OUTCOMES

A second potential explanation is that the program **did not affect the intermediate outcomes** that it intended. This could be because of low take-up and/or high substitution, and/or because the program itself was not an effective approach to impact the intermediate outcomes. Although we cannot document which of these mechanisms affected the results, the bottom line is that, in terms of girls' intermediate outcomes, namely protective assets, attitudes towards gender norms, and attitudes towards SRGBV, the program affected only attitudes towards gender equity.

Regarding parents' intermediate outcomes (attitudes towards gender equity and parenting styles), we cannot estimate the impact for these outcomes without a control group. But given the take-up and substitution problems described below, and the lack of a finding of significant change between baseline and endline regarding either parents' gender attitudes or parenting styles, it is reasonable to assume the impact for parents' intermediate outcomes were small or negligible.

Finally, the program had some positive but weak impacts on teachers' intermediate outcomes, including attitudes towards gender equity, teachers' knowledge of SRGBV, and use of corporal punishment as a disciplinary method.

Overall, the program did not appear to have a large impact on intermediate outcomes, which could explain the lack of results on school retention.

What explains school retention?

The third factor involves determining if the intermediate outcomes the project sought to impact are the right levers needed to increase school retention. For example, will improvements in girls' attitudes towards gender equity lead to greater school retention? This is important because it affects whether future programs should focus on impacting these constructs (i.e. protective assets, SRGBV), regardless of whether WW specifically succeeded in achieving such an impact.

One way to study this is to analyze whether outcomes the program sought to affect, like self-esteem or SRGBV are positively correlated with retention. If, for example, girls with high self-esteem at baseline are more likely to stay in school at endline than girls with low self-esteem, then a case can be made that

increasing self-esteem will lead to increased retention. In Annex XI we analyze the determinants of school retention and derive the following conclusions:

- Protective assets, including self-esteem, are weakly but positively correlated with school retention.
- SRGBV (as perceived by girls) is (negatively) weakly correlated with school retention.
- Girls' age and school attendance are strong predictors of school retention.
- Household variables (e.g. parental education, household wealth) do not to explain school retention in this sample.

In sum, these results suggest that, while protective assets and SRGBV are only weakly correlated with school retention, they do affect it.

POWER AND SAMPLE SIZE

Finally, it is possible that the program has positive impacts on retention or other outcomes, but these impacts are too small for the sample to estimate them with acceptable precision. We estimate that in accordance with the sample design, we should be able to find a minimum effect of seven percentage points of increased school retention.²² Therefore, it is possible the program increased retention by say four or five percentage points, but this sample cannot estimate it with acceptable precision.

While we did not find positive impacts on retention using survey data, administrative data showed that the 2018-2019 retention rate for girls was only two percentage points higher for the treatment group than for the control group. These results do not refer to a sample of students in the 61 schools, but to all Form I students 2018 students in the 61 schools. Therefore, the difference in retention rates between the treatment and control groups is not an estimate of the true difference, but the actual difference at the population level.

Significantly, the 2018-2019 retention rate of the treatment group exceeded that of the control group by 2 percentage points. However, these gains were not observed in the 2018-2020 retention rate.

F. PERFORMANCE EVALUATION RESULTS

FI. POY CLUBS

To assess the impact of POY clubs, our analysis drew from focus groups with POY club members conducted in 2020 and 2021, along with KIIs with peer educators, matron, and patron teachers, and LGAs, conducted in 2020. During the FGD conducted in 2020, POY club members were asked to reflect on their experiences with POY clubs, perceived benefits, and levels of participation. During the FGD conducted in 2021, girls were asked about their plans and goals, how they remembered and what they learned from their POY club experiences, how safe they felt in their schools and linkages between POY club participation and staying in school. In 2021 half of the focus groups were conducted with girls who had dropped out of secondary school, and half with girls who stayed in school. Peer educators provided insights on POY club implementation, successes, and challenges, while teachers provided

²² For this estimate we used the observed Intra-cluster correlation of 0.03, a power of 80 percent and alpha of 5 percent.

anecdotes on POY club attendance and school environment. Findings on the perceived benefits of POY clubs were mixed, and differences were noted between boys' and girls' perceptions of POY clubs.

PEI. How are girls and boys benefitting from POY clubs?

Across focus groups, boys and girls cited confidence building activities as the primary benefit of POY clubs. The activities included skill-building for problem solving, communication, and training on bodily autonomy. Other key benefits included the provision of school materials and study activities. Additional benefits and implementation insights are below.

PEI.1. What are the main challenges POY clubs face? How is WW resolving these implementation challenges? Do challenges vary by district?

Peer educators were asked to reflect on POY club startup, including challenges encountered in administering the POY club curriculum, coordinating with others, and motivating attendance. According to peer educators, POY club members were enthusiastic at startup, but encountered challenges in implementation. In Karatu, peer educators reported challenges in maintaining the attention of POY club members due to the timing of group meetings. Given that group meetings were held after school, peer educators and group members said it was difficult to stay focused at the end of the day, and without food.

Another reason why is that you may tell a member to respond to a question and they say we cannot respond because we are hungry. (Karatu Peer Educator, F, 2020)

During focus groups discussions, POY club members mentioned similar challenges related to the timing of club meetings and other logistics. In Musoma, POY club members and peer educators reported challenges in attending POY clubs due to the long distances they have to travel after school.

Peer educators also reported challenges in communicating POY club objectives and content. The educators said that girls experiencing low levels of self-esteem were reluctant to engage in these conversations.

Data indicate that implementation challenges were more affected by school level differences, such as level of matron/patron teacher involvement and meeting times. No major differences in implementation challenges between regions were documented.

PEI.2. Are members participating actively in POY clubs? Are they satisfied with POY club activities? What benefits are perceived by participants?

Across districts, POY club members and peer educators reported that POY Club attendance generally tapered off over time. Over time, peer educators had difficulty maintaining enthusiasm for POY Club activities. Similarly, POY Club members felt that once they attained a certain level of knowledge, POY Club meetings were not as illuminating as they were before.

When we were in the first year, they were attending in large numbers and they were very fond of it but as the time went on, the more educated they became the less in attendance. I mean when we got into second year, those girls we started with in the first year, started declining in attendance but those who

were in form one who also formed their clubs, were the ones who were attending to a large extent because they were eager to know and they loved it than those who already knew about the club.
(Musoma POY Club Member, F, 2020)

As previously noted, confidence building activities were perceived as beneficial, especially for girls. Such activities helped girls assert themselves in peer-to-peer interactions and understand their rights. Girls also said the activities helped them feel confident asking teachers for help with topics they did not understand, and in communicating with teachers in general.

We had POY leaders and when we attended POY clubs, they used to teach us confidence. For example, you might find that a teacher has taught in class and a student did not understand and has no confidence to go and ask a teacher for more clarification and this leads him to failing exams. (Karatu POY club Member, F, 2020)

Confidence building activities also helped girls manage pressure from older men in their communities to engage in sexual activity. **Girls reported that such relationships, and pregnancies that were often the result, were a significant barrier to completing their schooling. POY clubs have given them the confidence to manage such advances.** Specifically, girls said they learned in POY clubs how to assert autonomy over their own bodies.

Yes. I have experienced change. First of all, I have become self-aware of myself, I am confident because a boy might come telling me that he wants to fall in love with me. I am confident to tell him that I don't want that. This means that I am confident to refuse anything I don't want. (Musoma POY club Member, F, 2020)

Among boys, confidence building activities were also well-received. **However, boys perceived more benefits from study groups, the provision of study materials, and modules on time management.** Girls also found these modules to be helpful but cited them less often than boys.

To assess what benefits former members saw in POY club participation, during the 2021 FGD we asked about the links between POY club participation and the decision to stay or drop out of school. It was unclear whether improvements in protective assets from POY club participation prevented dropping out of school. The decision to drop out is driven by many other factors. In situations of household poverty in particular, girls faced limited prospects and for some, romantic relationships (leading to early pregnancy) were their most effective strategy for ensuring that their needs could be met. Therefore, improving protective assets alone would not be enough to keep these girls in school.

To illustrate this, a former POY club member in Musoma, when asked about challenges girls face in pursuing secondary education, responded,

"In the deepest poverty, a girl faces many obstacles, hence school becomes hard for her [...] If someone sees that there is no money at home, she decides to find a man who can provide money, so she can buy clothes. That's how it is. So, they end up engaging in romantic relationships." (Former POY club member, Musoma, F, 2021)

Pregnancy was the main reason why girls dropped out of school, according to the responses from the 2021 FGDs. Pregnancy was the most frequently mentioned reason both in terms of the participants'

own experiences and their responses when asked why girls drop out in general. There were also mentions of “getting married while they are still in school” (Musoma, F, 2021) and girls being married off so that their families could receive bride price. Other reasons for drop out that were mentioned were distance from school, illness, poverty (not being able to afford school uniforms or fees) and failing classes.

Furthermore, many girls had competing obligations for their time related to farming, earning money, or doing unpaid home labor that prevented them from staying in school despite improvements in their protective assets. Parents who depend on girls to do things like fetch water or cook might be reticent to re-register them in school.

However, several girls in the focus groups with POY club members who stayed in school said that their participation in POY clubs helped them to stay:

“I was meeting people in the street advising me, maybe just stop studying, we saw many who study but they didn't succeed, but when I joined the POY club, I got rid of those peer groups. That's why I've fulfilled my dreams and now I'm going to finish Form Four.” (Former POY club member, Musoma, F, 2021)

Girls who stayed in school did so because they were determined to achieve their goals which often required higher education. A POY club member in Karatu said that she wants to “be among prominent scientists in Africa.” (Karatu, F, 2021) Many girls aspired to be doctors, lawyers, or teachers. Almost all girls who stayed in school wanted to continue with their education after secondary school, but many also spoke of income generating activities (e.g. selling vegetables, hair dressing or tour guiding). These girls did not see income generating activities and schooling as mutually exclusive.

Girls who dropped out of school planned to pursue or continue some kind of paid work or income generating activity, often with little preference for what kind. Specific types of work that girls planned to do included being shop keepers or engaging in commerce (usually for clothes or cosmetics), entrepreneurship, hairdressing, tailoring/sewing, hotel management, and farming. In the long term, many of these girls did want to go back to school. A former POY club member in Karatu reasoned, “I think if you are educated employment will last longer than temporary jobs.” (Karatu, F, 2021) When asked what they would like to study when they go back to school, most girls specified vocational skills such as needlework or hotel management, as opposed to finishing secondary school and going to university.

A big difference between the plans of girls who dropped out of school versus those who stayed in school was their timing of marriage and children. Girls who dropped out brought up plans of marriage unprompted, before being asked directly about plans related to marriage or children. They specified that these plans were for the short term, as near as in one year. In contrast, across the focus groups for girls in school, they spoke of delaying marriage:

“When you have finished everything concerning education, etc., everyone will need to have a child. But for now, no.” (Former POY club member, Karatu, F, 2021)

Girls in school reported that COVID-19 was disruptive to their learning, but did not significantly influence their plans, and girls who dropped out said that COVID-19 did not affect their decision making.

It was clear that for a significant number of girls who participated in POY clubs and decided to stay in school, the protective assets that POY helped them build were a large factor in their decision to stay. This was reflected not only in direct statements indicating as much, but also in the fact that girls who stayed clearly took the POY content seriously, retained the lessons, and applied them to their lives.

PEI.3. Is the POY club curriculum user-friendly? What lessons/components of the POY club curriculum are most useful to adolescents?

Peer educators reported the POY club curriculum was generally user-friendly, especially regarding time management, confidence building, and study habits. However, peer educators found modules around gender norms and gender-based violence to be challenging, especially when discussing gender norms with boys.

The topics that I was taught and felt apprehensive about were sex and gender, and I didn't know how to present them. I was uncomfortable talking to boys about cooking and sharing chores at home. There were a lot of questions and some of them were challenging to the extent that I could not provide answers.
(Musoma Peer Educator, F, 2020)

Discussions about responsibilities at home suggest that changes in gender norms and attitudes are hampered by the reinforcement of gender norms at home. Boys remained resistant to the idea of doing chores that are traditionally “feminine.” When asked about the types of responsibilities students have at home, boys and girls expressed very different perspectives about the types of chores they were involved in.

Boys perceive their home responsibilities to be fair, and reflective of appropriate social norms. Boys felt that their responsibilities at home should reflect their physical strength, and girls should continue to support maintenance of the home.

Participant 5: May be to add on, in our locality the work that needs energy and strength are always for men, and other tasks like cooking and cleaning, are for girls. (Karatu POY Club members, M, 2020)

Boys also expressed gender inequitable views during discussions about education. Boys had mixed perceptions regarding the extent to which girls' education should be prioritized. Some boys said that government programs targeted to girls gave girls an unfair advantage.

Despite conversations with peer educators about gender-based violence, boys continued to express gender-inequitable attitudes regarding factors contributing to girls dropping out of school. For example, during a discussion on challenges girls may face related to sexual assault on their way to school, some boys expressed victim-blaming attitudes that girls invited sexual advances and could not blame their perpetrators:

Participant 4: *A girl face many challenges in her daily life. For example, when a girl walks to school, she can encounter street boys that can rape her.*

Participant 5: *If girls want to achieve their goals, they can do it. About the street boys, I think girls bring that on themselves.*

Mod: *So, you mean girls themselves create room for those practices?*

Participant 5: *Yes, because when they are dismissed at school, instead of going home they can be loitering on the roads. (Karatu POY club Members, M, 2020)*

These findings suggest persisting challenges in facilitating conversations around gender norms among POY club members. Female peer educators were uncomfortable discussing issues of gender roles and gender-based violence with boys, and unable to change the gender-inequitable attitudes of boys. Although girls found discussions about gender norms to be affirming, boys were more resistant to the gender equity-focused portions of the curriculum.

There was a clear difference between girls who dropped out and girls in school in terms of how much they remembered or retained from POY clubs in 2021. Girls who dropped out retained much less. They were not able to remember specific activities or modules. One former POY club member in Musoma, when asked about what POY clubs had taught her, responded, “It [POY Club Content] doesn’t affect me currently. It doesn’t touch my life at all.” (Musoma, F, 2021) For many of the specific questions about POY activities, most dropped out girls simply stayed silent.

Girls who stayed in school clearly retained much more from their participation in POY clubs than their peers who dropped out. They were able to identify specific activities and skills, and they both remembered modules on GBV without being prompted and were able to define it correctly and in detail, such as differentiating between psychological and physical GBV. These girls also remembered what they had been taught about how to report cases of GBV. They remembered that they could report to counselors, teachers, or parents, depending on who they were most comfortable with.

Girls in the 2021 FGDs identified self-esteem, self-awareness, assertive communication, and time management as the most important skills that POY clubs helped them to build. When asked if the clubs changed what they thought girls are capable of, a POY club member in Karatu responded, “I never thought that there was a possibility of me being one among scientists, but from POY club lessons I am motivated, and I have realized that there is a possibility of being one of them.” (Karatu, F, 2021)

Girls who had participated in different forms said that they liked the content of form 3 the best, because they were older and more able to understand the importance of what they were learning. In form one in particular, they said that they did not take the content seriously and made jokes about it, but by form three, they were better critical thinkers, and it was easier to process the lessons.

PE 1.4: How is WW engaging LGAs in POY clubs?

Findings on LGA involvement in POY clubs were limited. A few LGAs reported meeting briefly with POY clubs within their communities and encouraging members to continue their participation. While LGAs spoke positively about POY clubs, the majority of LGAs interviewed did not indicate extensive engagement with them. As such, they offered limited perspectives on their impact. Similarly, peer educators said their interactions with LGAs were limited, citing a dearth of LGA availability and time. Peer educators also reported feeling intimidated interacting with LGAs and would have liked to have additional support from WW in facilitating those relationships.

My request is for local government authority officers to visit POY clubs in schools, identify the challenges which we are facing as students. Because, as students, we cannot get time to start looking for the chairperson or social welfare officers due to their congested schedule. (Ichungu Secondary School Peer Educator, M, 2020)

PE 1.5. What kind of supportive supervision do patrons/matrons and peer educators of POY clubs receive? Do they perceive this as contributing to the results of POY clubs?

Teachers and peer educators of POY clubs reported mixed experiences regarding supportive supervision. Some peer educators reported not receiving visits at all, while others reported attending an additional training in Arusha administered by WW. Peer educators who attended this training found it useful and expressed a desire for additional support. In Karatu, peer educators reported receiving supportive supervision for POY club elections. Teachers and peer educators did not comment on the relationship between supportive supervision and POY clubs results.

PE 1.6. Are POY clubs having any unintended (positive or negative) consequences?

There were no observed direct unintended consequences of POY clubs in discussions. Indirectly, however, peer educators faced certain challenges in facilitating gender norms discussions effectively. The opportunity to discuss these topics without effective guidance may have inadvertently reinforced inequitable and harmful gender norms.

PE 1.7. Do POY clubs influence the general school environment? If yes, how? If no, why not?

Findings from teachers and POY club members suggest clubs have some impact on the school environment. In schools where groups continue to be active, members attribute changes in student-teacher relationships to POY clubs. As outlined earlier, **confidence building activities have encouraged girls to be more proactive about seeking academic support from teachers.** Additionally, some teachers suggest the formation of study groups has impacted the overall school environment and encouraged non-members to form study groups.

PE 1.8. How can POY clubs be made more effective?

To promote increased attendance and group cohesion, POY club members suggested:

- Providing peer educator training materials in the local language to all POY club members so they can identify and prepare to lead new activities and discussions.
- Offering training opportunities to enable group members to lead groups when peer educators are unavailable.
- Improving training of and supportive supervision for peer educators, especially those less comfortable speaking in front of a group about gender norms.
- Providing official POY club T-shirts to easily identify group members and facilitate feelings of group belonging.

- Ensuring gender-matching in both peer educator trainers and peer educators, and between peer educators and POY groups—which should be single-sex—to facilitate sensitive discussions around SRGBV and gender norms.
- Strengthening Teen Awareness Group feedback meetings and facilitate introductions, training, and dialogues between peer educators and Social Welfare Officers on child protection procedures, including referral and response to gender-based violence reports from children.
- Designating private spaces for meetings. Group members and peer educators report challenges identifying consistent meeting places. In some cases, groups change locations often, sometimes resulting in members missing meetings.

In addition to these recommendations, findings suggest that POY clubs could be bolstered by additional and consistent supportive supervision from WW, including more coordination between peer educators and teachers.

F2. LIMCA GROUPS

Findings about LIMCA group efficacy and implementation drew from focus groups with LIMCA group members, interviews with EEVs, and LGAs. During focus group discussions, LIMCA group members were asked to reflect on the recruitment process and how they came to learn about the LIMCA program. Similarly, EEVs were asked to reflect on the challenges they experienced while recruiting prospective members and strategies they used to overcome them and maximize participation. Findings from LIMCA group members and EEVs suggest a few key challenges at start-up may have impacted participation rates in the groups and their overall efficacy.

EEVs reported conducting recruitment through public community meetings arranged by ward offices. During these meetings, EEVs outlined key objectives of LIMCA groups, including group savings and support for girls' education. EEVs reported a number of challenges in motivating prospective members to join. EEVs report that some prospective group members saw no value in joining LIMCA groups and expressed fears that their money would be stolen. **LIMCA group members initially hesitant to join confirmed this finding, citing prior negative experiences with savings groups.**

LIMCA recruitment is challenging; prospective members would say 'aah, these people can be tricky, so why follow up, there are so many groups, there are many organizations that came here to gather the groups and then they left us in the air, they don't come to visit us anymore.' (Tarime EEV, F, 2020)

EEVs also reported challenges motivating prospective group members who had access to traditional bank loans, or high levels of comfort with other savings groups in the area. **In some farming communities, there were several savings groups related to farmer development and livestock keeping.**

The existence of other savings groups presented some gender-related challenges in enrolling members in LIMCA groups. As men are traditionally the ones participating in savings group activities, women reported challenges convincing their spouses to allow them to participate in LIMCA groups.

Lastly, EEVs cited differences in political ideologies as an obstacle to recruitment. This was especially pronounced in Mara, where the political party affiliation of EEVs or other group members may have

impacted the extent to which community members were interested in joining. Despite these initial challenges, EEVs said they felt equipped to manage hesitations to join and highlighted the educational aspects of LIMCA to distinguish it from other savings groups.

PE2: Is LIMCA benefitting its members, girls, and other stakeholders as intended by WW?

Findings suggest LIMCA group membership has had some positive impact on economic outcomes, investment in girls' education, and improved parenting skills. Overall, LIMCA group members reported benefits from improved loan access, money management, and group investment in girls' education funds.

PE2.1: Is LIMCA helping reduce the socioeconomic vulnerability of participating households? Why or why not?

We found that LIMCA groups helped reduce the socioeconomic vulnerability of participating households through improved loan access and better money management.

During focus group discussions, respondents were asked how they earned income. The majority respondents reported engagement in some combination of small business, trade, agriculture, livestock rearing, and fishing.

LIMCA group members said their involvement in the group has improved their management of emergency expenses related to illnesses, deaths, and loss of income. Parents cited their access to a new social network (the LIMCA group) was a valuable aspect of the activity.

I was living a vulnerable and threatening life before joining LIMCA. But now, my life and economic status have improved so much. My group can lend me either 50 or 100 thousand Tsh, I can repay them easily and they trust me. (LIMCA group member, Mara, 2020)

Group members reported involvement in LIMCA groups helped them manage emergency expenses more easily. Group members reported experiencing economic shocks related to their harvest and illnesses. In these cases, group members said LIMCA loans served as the primary strategy for managing sudden financial expenses:

My husband was very sick, he was admitted for operation so he couldn't provide for our family anymore. I was the only one remaining to carry my family burden and during that time, LIMCA was my only solution. They lent me 50 thousand Tsh that assisted me so much (LIMCA group Member, Mara, 2020)

LIMCA group members also experienced challenges in improving their financial outcomes due to loan repayment. Across districts, group members had trouble repaying the loans provided through the group, leading to strained group relations. Group members report that in some situations, they take loans from the group, and have to use their repayment money to address another emergency expense:

We face a lot of challenges. For instance you can take money and you plan to use it for a specific thing, but unfortunately you get ill or get into an accident that needs money, so you need to take that money planned for a certain task and use it to cover the accident/illness etc. (LIMCA group member, Karatu, 2020)

Findings suggest the majority of groups are able to recover and maintain membership levels even when members have difficulty repaying. However in Mara, groups report several members have dropped off due to difficulty repaying loans. It is worth highlighting that survey data show that only between 5-6 percent of parents said they spent the LIMCA loan repaying other debts (see Annex IX).

To improve longer term outcomes, group members said they would benefit from additional business and money management training, as well as sufficient seed capital to launch their businesses. Notably, EEVs shared this sentiment as well, revealing group members regularly asked about additional training opportunities for business development.

PE2.2: Are parents and caregivers investing in girls' education as a result of LIMCA participation? How many parents and caregivers are being supported by LIMCA groups?

Across focus groups, group members reported the opportunity to support children's education as a motivator for joining LIMCA groups. While some report joining to support their own child's education, many came to appreciate the importance of supporting the education of other girls in the community.

Group members acknowledged that cultural norms created expectations that girls would get married immediately after or before finishing secondary school. As such, they dismissed the importance of investing time and resources in the education of girls. However, caregivers report that their involvement in LIMCA shifted their perspectives:

For sure LIMCA has given us good lessons. In the past we invested much to the boys' education and we tended to abandon girls. LIMCA however, has introduced a positive mentality to me towards girls. This has surely helped me. (LIMCA group Member, Mara, 2020)

Notably, group members said participating in LIMCA also shifted their perspectives on extending their support to girls beyond finishing secondary school. As one group member reports:

For me it was very important to learn about the value of girls' education. Last year my daughter almost quit university because of the fees, but thanks to LIMCA me and my family now understand the importance of education, so we helped her to stay in school. (LIMCA group Member, Karatu, 2020)

Group members in Karatu said parental skills training helped them improve communication with their daughters and better understand different factors affecting their school attendance. That gave them new insights into the importance of providing material support for girls, including for menstrual care, to increase the likelihood that their daughters would stay in school:

It has changed a lot. Nowadays we know well and understand their needs and conditions, because when I fail to fulfill her needs, this may cause her to feel uncomfortable at school and when this happens, she can miss classes. (LIMCA group Member, Karatu, 2020)

Group members cited similar factors, and in fact used group funds to address infrastructural barriers to menstrual management at school. One group in Mara reported dedicating funds to building a separate room where girls could change and provided sanitary napkins for girls at school.

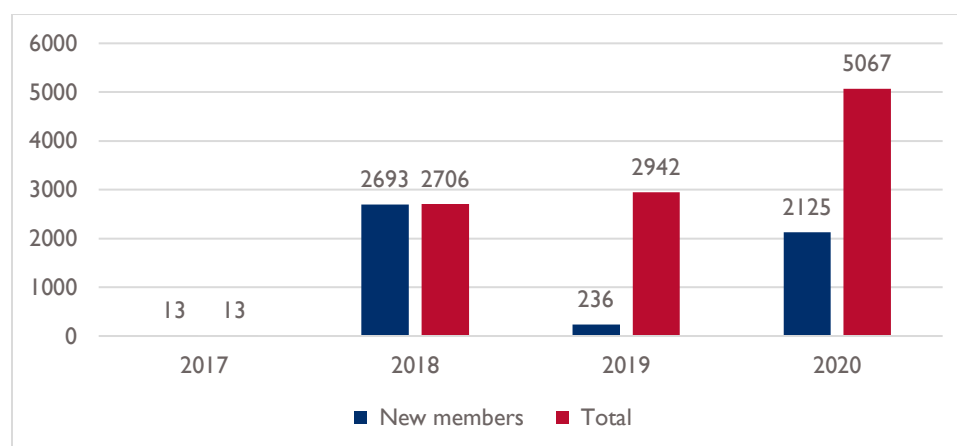
It has changed to a great extent. In the past girls would not attend schools if they were menstruating. This made girls feel like if they go to school, they will be disturbed. After LIMCA however, we do take items at school and a special room for girls has been built. This has contributed to increased child's school attendance. (LIMCA group Member, Mara, 2020)

Findings indicate greater challenges in Musoma in changing perspectives on investment in girls' education. Group members there said they were frustrated with the amount of energy dedicated to girls' education, without similar efforts devoted to that of boys. For other group members in Musoma, the focus on a community fund was a deterrent for participation, as they said that their daughters, specifically, were not benefitting from the fund.

For example, in the beginning we were 25 group members but after we received our first shares, five members dropped out. The reason that made them drop out was a claim that their children are not considered for the contributions that we bring to school. They claimed that their daughters in school are not given sanitary napkins. That is why they dropped out. (LIMCA group Member, Musoma, 2020)

In terms of parent participation, LIMCA groups increased membership especially in 2018 and 2020, reaching more than 5 thousand members by 2020, as **Figure 19** shows.

Figure 19: LIMCA members with children (under age 19).



Source: WEI administrative records.

PE2.3: What kind of supportive supervision do EEVs receive from local NGO partners? Do EEVs perceive this supportive supervision as contributing to the results of LIMCA groups?

EEVs receive supportive supervision through WEI's community-based implementing partner organizations on a regular basis, with some additional periodic support from WW staff. EEVs report typically receiving a check-in phone call (a practice increased during 2020, taking the place of in-person visits, due to COVID-19 safety protocols and travel restrictions) but said this support could be more time-intensive. EEVs indicate that supportive supervision visits would be a good opportunity to build on the training they received. Nonetheless, EEVs report that supportive supervision has helped increase the efficacy of their groups:

The supportive supervision sessions are helpful, especially when it comes to helping group members who do barn farming. (EEV, Musoma, 2020)

EEVs interviewed in Karatu did not report receiving supportive supervision. They said that while WW has visited to observe groups, it has not provided other support. It must be noted that some interview respondents appeared to conflate financial support with supportive supervision visits.

PE2.4: How is WW engaging LGAs in LIMCA?

EEVs report extensive engagement with LGAs, including ward and street executive officers. These collaborations have been beneficial in moving community projects forward. While LGAs are supportive of LIMCA groups in general, EEVs report they are sometimes limited in their ability or desire to attend meetings or respond to requests for their time or presence.

When you ask a LGA to come to meeting they would say ‘Maybe another time’. This is because officials are few, and groups are many, and all groups their support. They keep promising that they will come... The other challenge, sometimes the government people are difficult - they don’t want to volunteer as we volunteer. You can call them, and they will ask you if there is an allowance. (EEV, Mara, 2020)

In Musoma, LGA involvement was critical to gaining buy-in from group members. LGAs had extensive involvement in group formation, helping facilitate group constitutions and resolve conflicts.

Since I started the group, I have called ward and village administrators and development officials. And I can’t start a group without a constitution, and the ward and village administrator and development officials are the ones that endorse those constitutions. When a problem arises in the group, we called them to solve it, especially when the problem is about loan returns. So yes, the cooperation with them helps me a lot. (EEV, Musoma, 2020)

As previously noted, there were unique challenges in retaining group members in Musoma. LGA engagement helped address some of these challenges and motivate remaining group members to participate. Findings suggest that consistent LGA engagement facilitates LIMCA group success due to their ability to facilitate group participation and manage conflict resolution. Findings also indicate LGAs have varying levels of engagement with LIMCA, and in some cases are unwilling or unable to engage with group members without financial compensation.

PE2.5: What should be done to make this WW intervention more effective?

Group members and EEVs described challenges in implementing parenting skills training and improving economic outcomes. These challenges demonstrate opportunities to improve the efficacy of LIMCA activities.

First, EEVs reported **difficulty effectively communicating the objectives of parenting skills training**. Even though EEVs generally included discussions on parental skills in every meeting, they said the discussions met significant resistance. EEVs in Mara reported that the majority of parents adopted an authoritarian style and were resistant to training.

Second, EEVs report **more pronounced challenges for women EEVs working to communicate parenting styles to male-headed households**. Although they managed to gain group trust over time, some EEVs recommend having mixed-gender co-facilitators help support training on parental

supervision skills. Parents recommended administering more intensive parenting skills training so that incoming members were up to date:

Parenting skills training should be repeated every year, because there are some new members every year in need of that content, so we would like to ask WEI to include that activity on the year budget so new members can be trained. (EEV, Karatu, 2020)

Third, EEVs and group members said LIMCA groups should offer additional training in business management to introduce new subjects, and to strengthen their capacity in business development training skills. In Karatu, group members asked to receive training directly from WEI. They said EEVs were not always equipped to answer questions that arose during training.

EEVs and group members across districts agreed that providing additional training would significantly improve the efficacy of LIMCA groups. Group members also expressed a strong desire to continue LIMCA groups but did not feel they could do so without the provision of additional training.

Overall, LIMCA group members reported positive perceptions of LIMCA group activities, especially the creation of the education fund. Groups were successful in changing perspectives about girls' education and motivating parents to support girls within their communities. Across districts, LIMCA groups were strengthened when LGAs were closely engaged with group members, attended meetings, and participated in decision-making processes. LGA involvement was also an important component of gaining initial buy-in from prospective group members. In addition to these successes, data revealed some opportunities for improvement and future sustainability. Parents are motivated to continue with LIMCA groups but request additional training and more direct engagement with WEI. EEVs also said they could benefit from additional training, as it would help them lead discussions about parenting skills and address technical questions related to business management.

F3. SRGBV TRAINING

PE3: What are the benefits of the SRGBV prevention and response teacher training to teachers, students, and other stakeholders?

To assess the impact of SRGBV training, analysis primarily focused on findings from interviews with teachers to understand the logistics of the training and reporting structures within schools. Focus groups with boys and girls provided additional perspectives on school environment, disciplinary norms, and the extent to which these have changed since the introduction of SRGBV training. Overall, the training improved teacher knowledge about SRGBV, but had limited impact on attitudes and practices concerning corporal punishment.

PE3.1: Are teachers satisfied with the SRGBV training? Why or why not?

Generally, teachers were satisfied with topics covered during SRGBV training and said that these topics were relevant to their work. Teachers interviewed said the training equipped them with tools to manage students with behavioral issues and understand how to improve communication with students. Some teachers said they were initially suspicious of training's focus on alternative teaching methods. But they said with continued engagement, they came to find value in training.

The training helped me understand the value of the training content. At first, I thought that with this training they wanted to destroy the education system, but soon after I got the training, I came to realize the training was very needed and very useful. And I am sure the training changed almost everyone, and I see the gap in terms of how to look at some violent topics and how different organizations dealing with the cases versus how it was in the past. (Karatu Teacher, F, 2020)

While some teachers became increasingly open to the SRGBV subject matter presented in the training, others maintained their stance that the training did not align with their values, especially concerning corporal punishment.

Teachers expressed some dissatisfaction with the duration of the training, noting that given the newness of some of the content, they would have benefitted from repeated sessions throughout the year. Teachers across districts were also demotivated by the removal of Lunch and Learn financial incentives. Teachers report these benefits were critical for maintaining momentum and gaining buy-in from teachers who were more resistant to the changes in teachers' practices proposed by SRGBV training. Many saw these Lunch & Learns as an opportunity to share strategies for implementing lessons learned from SRGBV training.

PE3.2: Have teachers' knowledge and attitudes towards SRGBV changed? Why or why not?

As discussed in section D4, quantitative findings demonstrate significant improvement in teachers' knowledge about SRGBV. During KIs teachers discussed their changed knowledge in SRGBV and the impacts of gender norms in the school environment. When asked whether their attitudes towards SRGBV had changed, the majority of teachers reported their attitudes had changed significantly, and that they were sharing that knowledge with other adults:

Yeah, all of that is because of the training, I remember for the last two years, 2018 and 2019, whenever we have a parents' meeting, we use that time to talk about violence, so we found ourselves transferring knowledge to school staff and to the parents. (Mara Teacher, M, 2020)

However, when asked their perspectives on appropriate discipline methods or perspectives on the hypothetical prohibition of corporal punishment, it was clear that their attitudes had not changed much. Teachers felt that the prohibiting corporal punishment would make it impossible to keep children under control. Teachers also said that without addressing disciplinary practices at home, prohibiting corporal punishment in school would be challenging given that parents are still using various forms of physical punishment at home.

Okay, this comes from my heart. If we completely stop caning students, the discipline will disappear. Like seriously, at one time I talked to even WEI people, if we stop caning students completely, then we need to change the parents' nature, since these students are raised at their home. (Karatu Teacher, F, 2020)

These contradictions were consistent across districts, suggesting teachers had yet to recognize the effectiveness of positive, non-violent disciplinary methods in place of caning.

Findings from focus groups with students also suggest persistent challenges in altering teacher attitudes on sexual violence. While the majority of students expressed high levels of comfort in seeking support

from teachers to report sexual violence, a few girls had lower levels of confidence, citing fear that their teachers would not believe them, or would blame them:

The reasons are fear and not to have confidence, for example, when a girl is raped, she will be afraid to go and report to the teacher that she was raped because she will be asked in what place was she raped, so when she says in the shrubs, they might tell her “You brought it upon yourself”, [laughter]. That’s why she is scared. (Karatu Student, non POY, F, 2020)

Overall, SRGBV training has been impactful in improving teachers’ knowledge about SRGBV, but challenges in changing attitudes remain. This is most prevalent with regard to attitudes about corporal punishment, and to a lesser extent victim-blaming attitudes.

In the 2021 FGDs girls were asked about their feelings of safety in their schools. While all girls expressed strong negative feelings around physical discipline from teachers, those who dropped out were also adamant that this was not the deciding factor in them doing so.

Respondents across the 2021 focus groups reported that discipline was usually the same between boys and girls, so teachers would cane both boys and girls the same. One exception to this was that girls reported that in the past, male teachers would cane both girls and boys on the buttocks, but in the present, male teachers only cane girls on the palms, not the buttocks. There were also mentions of male teachers going into latrines to find girls who skip class. When asked about feelings of safety at school, girls in multiple FGDs (both girls who dropped out and girls who stayed) brought up menstruation difficulties, especially having no water in latrines and no space for them to deal with menstruation as well as having to do strenuous punishments while menstruating. One former POY club member in Musoma said:

“Girls here did not feel safe because, you get punished if you arrive late at school, sometimes when you are punished to do squat-jumping and you are undergoing menstruation cycle. So things get worse, you become scared. When you refuse to do squat-jumping, they start caning your back thinking that you are rude without knowing that you are going through a menstrual cycle and that’s why you cannot do that.” (Former POY club member, Musoma, F, 2021)

Both the rigidity of physical discipline being enforced on menstruating girls and the state of girls’ latrines that made it difficult to deal with menstruation caused girls to skip school.

PE3.3: Which elements of the following training components are easiest and most difficult for teachers to implement after receiving the training? Why?

Findings indicate SRGBV training was easiest to implement, while alternative disciplinary methods were hardest.

SRGBV prevention and response training was well-received. However, this was partly due to the tendency of teachers to think of SRGBV as overt harassment and other forms of sexual violence between peers, and not corporal punishment. In this regard, teachers focused on their response to sexual violence reports within school, and how to support students who wish to report SRGBV. Notably, students also perceived SRGBV in this way, and the majority said they felt confident in seeking teacher support after exposure to sexual violence.

ALTERNATIVE DISCIPLINE METHODS

Throughout interviews, teachers were asked to reflect on the types of disciplinary methods used in school before training, and any differences between pre-training and post-training methods. The majority said that prior to the training, corporal punishment was the primary disciplinary method, often accompanied with verbal reprimand.

We used to cane students, this was the best because it will make the student and a teacher to end their issues instantly, no need to follow up...Before the training and WW program, we were using that method, since it was the easiest way to punish a student, because student by nature are scared of strokes. (Karatu Teacher, F, 2020)

Teachers across schools demonstrated understanding of the value of alternative discipline methods in fostering a more positive relationship with students, and the incorporation of alternative discipline methods into current methods.

But instead you befriend the students, and also when you go to teach them they will appreciate you, they will see you as their parents or their guardian, they will trust you. (Mara Teacher, F, 2020)

Teachers also saw value in using educative alternative discipline methods to build skills in subjects such as writing, or biology. For example, in lieu of physical punishment, some teachers described assigning a student a research activity, such as investigating malaria transmission, and to submit their findings in essay form. However, teachers did not favor sole use of alternative, positive, and non-violent discipline methods. In their view, this would compromise their authority within school, and undermine the expectations of parents and students, as students expect to receive physical punishment. Discussions with students also reflect this, revealing that boys and girls perceive corporal punishment to be normal, and typical of disciplinary methods used in school.

During group discussions, students were asked to reflect on the types of disciplinary methods most often used, and the extent to which they felt these were appropriate. Boys reported the use of a variety of disciplinary methods, depending on the perceived infraction. Generally, teachers use caning as a primary disciplinary method, but may substitute chores such as cleaning, fetching water, and other physical labor. There were gendered differences in the types of physical tasks assigned as punishment:

When we have class disruption, boys are told to go and dig a pit, and girls are told to go to collect garbage and bring it to the toilet. (Karatu Student, M, 2020)

Boys and girls demonstrated similarities and differences in perspectives on whether caning was an appropriate method of discipline. The majority of respondents said caning was appropriate, provided it was limited, and only administered when the teacher was no longer angry or behaving harshly. Others said assigning chores was more effective. However, both boys and girls said that without caning, students would lose respect for teachers.

What I see is that students won't respect teachers because they are not punished and will not follow school instructions. They will fail in the exams because they are not punished and they are getting to school late. (Karatu Girls, non POY, 2020)

Girls were more resistant to prohibiting corporal punishment, saying it was needed to keep students in line. Girls also said prohibiting corporal punishment would make girls more likely to engage in sexual behaviors, and risk becoming pregnant. This finding is especially important as it suggests girls in the areas studied believe discipline facilitates good “morals.” Across discussions, both girls and boys said girls who get pregnant lack morals or discipline. Therefore, many girls perceive the use of disciplinary methods, especially corporal punishment, as a means of ensuring “proper” behavior among girls and minimizing sexual promiscuity. This view was more prominent among girls in Musoma.

Overall, findings suggest alternative discipline methods were most difficult to implement. This is largely due to norms and expectations concerning corporal punishment and limited application of alternative discipline methods. Although teachers said alternative discipline methods were valuable, they said they would not be effective as the primary or only form of discipline. Similarly, students were opposed to harsh corporal punishment, but did not feel that it should be omitted altogether.

PE3.4: What kinds of support do teachers receive in implementing what they have learned? Are they satisfied with the support they are receiving?

The differences between the reported support given to teachers and master teachers was stark. Teachers said that support from WW to implement lessons from SRGBV training was limited. The discontinued financial support of Lunch and Learns was the key factor for dissatisfaction with the support received. Without these incentives, teachers were no longer attending sessions, and consequently, no longer discussing their strategies for and challenges in implementing SRGBV training. Teachers said the sessions had been critical platforms for discussion of SRGBV implementation strategies, and for collectively identifying solutions to challenges to implementation.

In contrast, master teachers said that WW staff were always available, and they could call upon them as needed to clarify training elements or provide solutions to emerging challenges in implementing SRGBV training.

The support was there, because the leaders of WEI were close to us after the training, and I was happy doing their job. Even my fellow teachers were happy with me and they took me as the expert since I was selected to attend the training so that when I come back I train other colleagues. (Musoma Master Teacher, M, 2020)

PE3.5: How are LGAs (specifically the Gender and Children Desk Police Officers) supporting SRGBV training for teachers?

According to the surveyed teachers, the availability and depth of LGA support for SRGBV training was dependent on a range of factors, including proximity to schools, levels of coordination between LGAs and schools, and administrative challenges. In some areas, LGAs were closely connected with schools, and not only attended WW-administered training, but also school staff meetings, facilitated training with students, and actively participated in meetings with parents to discuss issues related to student wellbeing, including SRGBV.

For example we have the ward police come to train the students about reporting [SRGBV] cases. Education officers are always with us. Social welfare officers are always with us to make sure the school environment is clean, and no violence can occur in our locality. (Musoma Master Teacher, M, 2020)

However, active LGAs were in the minority. The majority of LGAs reported intermittent involvement, and often in response to reported sexual violence cases, or other issues arising at school. Additionally, only master teachers reported direct engagement with LGAs responding to SRGBV reports. Across districts, teachers reported that hierarchical structures affected their levels of engagement with and support from LGAs. Typically, teachers said they thought that SRGBV cases must first go to head teachers before rising to the LGA level. Therefore, any other LGA interaction would only occur if or when LGAs conduct visits to schools.

I have never interacted with those authorities, I can say zero relationship, because any report must go to head teachers. (Musoma Teacher, M, 2020)

Teachers and LGAs reported administrative and logistical challenges in SRGBV reporting. First, long distances to LGA offices were an important challenge for teachers and LGAs to collaborate on SRGBV training and reporting. When LGA offices were too far away, teachers were more likely to say that submitting a report to them was not worthwhile, and to discourage teachers from submitting reports.

The challenge we have when the teacher wants to report the case, the distance, for example from our school to their office if it is a public transport is TZS 1000/=, and if it is by using a boda boda it is TZS 3000/=. (Master Teacher, Mara, M, 2020)

LGAs were also cognizant of this and expressed similar challenges in conducting school visits with greater frequency. Women LGAs also expressed that conducting school visits was especially challenging for them, as traveling long distances to school or community sites presented additional safety concerns.

Teachers and LGAs also reported administrative challenges related to LGA workloads, with limited capacity resulting in delayed follow-up and resolution of reported cases. LGAs said their current workloads could not accommodate the volume of cases received, leaving them to prioritize the most severe cases, while others went unaddressed.

PE3.6: How are stakeholders (parent-teacher associations, school boards, etc.) supporting SRGBV training for teachers?

Parent-teacher associations' support for SRGBV training varied widely. In areas where parent interest for supporting SRGBV training was high, schools had strong existing mechanisms for involving parents, and began engaging parents early on. Additionally, these schools had strong linkages to LGAs, who were also deeply involved in training activities and facilitated parent buy-in. In these cases, master teachers, school board chairpersons, and parent association representatives took part in SRGBV training, and were actively involved in dialogue around SRGBV during other school meetings.

In instances where parent associations with low interest in supporting SRGBV training, teachers said convincing parents to prioritize support for the training was challenging. While parents were not opposed to SRGBV training, they often said addressing more immediate needs would improve outcomes for girls. In their view, addressing issues of SRGBV would be futile without first addressing more pressing food insecurity, infrastructure, and safety issues. Consequently, their efforts were focused on creating mechanisms for school feeding, improving latrines, and identifying funding opportunities for building dormitories so that girls would not be exposed to risks of gender-based violence on their way to and from school. These priorities were also reflected in parent discussions, with parents saying these

needs were most pressing and leveraged education funds to promote food contributions and infrastructural development.

F4. SUSTAINABILITY

PE4: Are the interventions addressed in PEI-3 sustainable?

To assess sustainability, analysis drew on data about respondent desire and capacity to continue WW-funded activities without external material assistance. Analysis also drew from successes and challenges identified in focus group discussions and interviews. Findings indicate varying levels of sustainability of WW intervention activities. The primary factors affecting the sustainability of interventions were social norms, perceived benefits, level of sense of ownership of intervention activities, coordination with existing community mechanisms, and support from LGAs.

LIMCA

Across interventions, LIMCA groups hold the most potential for sustainability, as group members in some communities felt a strong sense of ownership, and thus expressed a strong desire to continue with LIMCA. In areas with strong LIMCA groups, parents used community funds to address barriers to school attendance. In addition to the education fund, LIMCA group members in some communities mobilized resources to create food contributions, with some crops intended for sale donated to schools instead. COVID-19-related closures presented challenges to LIMCA group members across districts. However, many groups pivoted during school closures, donating their food contribution directly to needy student families. Parents in some communities raised funds to build changing areas for girls to improve menstrual management at school and made plans to construct dormitories to address safety concerns. In most cases, these initiatives were directed by group members themselves, demonstrating a strong sense of ownership.

These benefits were facilitated through strong support from LGAs. LGA involvement in LIMCA groups helped to resolve conflict and led to changing perspectives around the importance of investing in girls' education. Groups with weaker LGA support were not as fast to recover from tension arising from unpaid loans or disagreements on how to prioritize education funds.

WW facilitated the sustainability of these interventions through the creation of management committees and EEV training. Findings showed that in instances where EEVs were unable to attend meetings, LIMCA group members were well-positioned to step up and move group activities forward.

POY CLUBS

POY clubs showed lower levels of sustainability, as group members perceived diminishing benefits over time. Group members were initially enthusiastic about participation, demonstrating higher levels of participation at start-up. Group members also found value in discussion topics, study groups, and the provision of study materials. However, these activities soon lost their novelty, and were not enough to maintain similar levels of participation. POY club members said that topics covered helped them navigate issues within secondary school but did not focus sufficiently on strategies for post-secondary success. Group members expressed a strong desire to discuss topics around tertiary education and training options, and business development.

In addition, the roll-out of POY clubs at different times created some tensions. Teachers revealed that some students became aware of POY club formation in other schools, and perceived that their schools were delayed. This may have created some resentment and affected their initial enthusiasm for participating in the clubs.

WW's efforts to promote sustainability of POY clubs included training peer educators and close involvement of matron/patron teachers and LGAs. However, these efforts were not enough either to promote a strong sense of ownership among group members and stakeholders or to maintain interest in POY clubs.

SRGBV TRAINING

SRGBV trainings showed the lowest levels of sustainability relative to LIMCA and POY clubs, due to challenges related to entrenched existing social norms, limited feelings of ownership and coordination with communities, varying levels of LGA support and capacity, and varying levels of skill among facilitators in managing group discussions about sensitive topics around SRGBV. As findings demonstrated, SRGBV trainings were challenged by prevailing beliefs and social expectations that perpetuate inequitable gender norms and corporal punishment.

Findings demonstrated contradicting perspectives around changes in attitudes towards SRGBV. Teachers did not recognize corporal punishment as a form of SRGBV. Teacher definitions of SRGBV focused mainly on sexual violence, rather than encompassing multiple forms of psychological, physical, and sexual violence in and around school. While teachers found value in learning alternative discipline methods, they were not willing to adopt them alone. Further, teachers demonstrated a lack of clarity in what constitutes alternative disciplinary methods to include those that are non-violent, positive, and educative, not simply any method that avoids physical punishment, such as manual labor at school or suspension. Prevailing social norms around teacher-student power dynamics and gender also remain a significant barrier to the sustainability of SRGBV activities in the absence of consistently skilled facilitation to dialogue through those challenges. Findings indicated that beliefs about how to regulate girls' behavior are deeply internalized. Female teachers and students were more strongly opposed to the prohibition of corporal punishment.

Teachers also failed to form a strong sense of ownership of SRGBV training activities. Some male and female master teachers said they were empowered to carry out SRGBV training elements within schools. But teachers trained by master teachers said this approach affected their ability to champion SRGBV issues within their schools, motivate fellow teachers to adopt lessons learned from SRGBV training, and do collective exploration to address challenges in implementing SRGBV training in school. For many teachers, financial benefits were a strong motivation for participation in Lunch and Learn activities. In the absence of these benefits, teachers did not feel compelled to participate as often.

Lastly, limited coordination with LGAs and community members affected the sustainability of SRGBV activities. As previously noted, LGA involvement and parent committees helped facilitate the prioritization of SRGBV training and critical dialogue between parents, LGAs, and teachers on issues of SRGBV. Teachers and LGAs expressed a strong desire to better integrate SRGBV training into existing community leadership mechanisms. In their view, this would address key barriers to SRGBV training and reporting related to limited LGA staff capacity and distances to LGA offices. LGAs also noted that extending training into existing community-level leadership structure would help address the volume of

cases that often limits their capacity, as the majority of violence against girls takes place within school, home, or community. Though WW worked to involve LGAs in the implementation of SRGBV training, LGAs said they felt this was applied inconsistently, affecting the overall sustainability of intervention activities.

If the intention is to eliminate violence it must extend the service to all places- they should go through all the councils and through the community. Also the collaboration, you can't implement something and jump to a certain level and run to another level so if they want their project to be sustainable, that is, all levels can be involved in one way or another. Because at the end of the day they as stakeholders are leaving but the government will remain, so when you do something half way up and skip another level, when you leave we will never know what you did, what is going on or what you have left, you will go away with your project as you are. (LGA, Karatu, F, 2020)

G. CONCLUSIONS AND PROGRAMMATIC RECOMMENDATIONS

Our main conclusions and associated recommendations are as follows:

SCHOOL RETENTION AND POY CLUBS

We did not find conclusive evidence that the program increased school retention. Using administrative data we found that retention was 2 percentage points higher for the treatment girls than for control girls between 2018 and 2019, but this positive effect was not observed in the administrative data between 2018 and 2020. Using survey data, we found that retention was between 2 and 3 percentage points higher for the control girls than for treatment girls. In any case, none of these differences are statistically significant.

These results suggest that the program did not have a major impact on school retention. Furthermore, there are two additional results that could explain why we found no impact on retention.

First, we found that there are other programs similar to WW's POY clubs in these districts. This affects the validity of the impact evaluation results, but more importantly questions whether this type of program should be expanded where similar interventions are already available. Second, although our qualitative findings indicate that boys and girls drop out due to different reasons, administrative data indicated that there are no major differences between girls and boys retention rates in either the treatment or the control schools. If programs like WW are based on the assumption that girls are more likely to drop out of school than boys, these assumptions should be revised in future programming.

In terms of POY club implementation, qualitative data showed the following areas for potential improvement:

- Some peer educators reported difficulty dealing with certain topics and answering certain questions during POY sessions. Peer educators should receive more support through supervision visits and refresher training. Also, where some topics are harder for girls to discuss regardless of their training, other educator models, such as matrons or older peers (Form 3 students as educators of Form 1 students), could be explored.
- POY club members' enthusiasm for participating in the clubs diminishes over time. While partly due to logistical challenges, some group members noted difficulties connecting with the content

after moving through modules on confidence building, studying, and personal autonomy. Future POY club activities may review the need for POY clubs at Form 2 and 4, and consider identifying priorities for members in Form 3, including post-secondary success.

- POY clubs happen after the school day so girls have to stay late in the afternoon. It may be necessary to schedule POY clubs so they conclude early enough for participants to get home before dark.
- It was unclear whether improvements in protective assets from POY club participation prevented dropping out of school. While for a significant number of girls the protective assets that POY clubs helped them build were a large factor in their decision to stay in school, girls also reported that the decision to drop out is driven by factors like poverty, which POY club participation could not affect.

It is worth noting that both quantitative and qualitative data indicate that girls were satisfied with POY clubs. This does not necessarily mean that the program should have positive impacts but is important to keep in mind that the program could be having effects on outcomes we do not observe or cannot observe yet. As discussed in section A3, and in more detail in Annex I, research has shown that life skills training can have positive impacts on employment and income generation activities (Bandiera et al. 2020, Adoho et al. 2014). Although this is out of the scope of this evaluation, further research could assess whether WW impacts later outcomes on treated girls.

LIMCA AND PARENTING

We did not find significant changes in parents' parenting skills between baseline and endline, and only small changes in attitudes towards gender equity. It is worth noting that there was only one year between baseline and endline, and parenting skills may need a long time to change, even if continued training eventually has an impact. Qualitative data indicate some challenges regarding the delivery of LIMCA training, especially regarding parenting skills. It is possible that some of the most sensitive sections of the LIMCA training, such as parenting skills, should be delivered by well-known community leaders and not by EEVs in order to gain buy-in.

However, qualitative data also show that LIMCA group members perceived group activities positively, especially the creation of the education fund. Groups were successful in changing perspectives about girls' education, and motivating parents' support of girls within their communities. Across districts, LIMCA groups were strengthened when LGAs were closely engaged with group members, attended meetings, and participated in decision-making processes. LGA involvement was also an important component of gaining initial buy-in from prospective group members.

In sum, while quantitative data does not indicate that LIMCA was successful in achieving the intended outcomes, qualitative data suggest there were positive impacts on attitudes towards girls education and awareness about the importance of girls' education. Along these lines, if parenting skills training delivery is improved, LIMCA groups should be considered as a promising instrument to improve girls' education.

SRGBV TRAINING

One key success of the program was in raising awareness about SRGBV, especially corporal punishment. Our findings suggest parents were more aware at endline than at baseline that corporal punishment is not effective. More importantly, we found teachers in the treatment schools were less likely to cane

students and assign chores at school as punishment than their counterparts in the control group, although some of these results are probably driven by social desirability bias. Furthermore, it is important to keep in mind that WEI and other implementers in Tanzania operate under a regulatory framework that allows corporal punishment in schools, so efforts to reduce it are obviously bounded by this reality. Therefore, even if programs like WW do reduce corporal punishment, parallel work should be conducted at the policy level. In the future, to reduce SRGBV it will ideal to engage with education policy makers in a broader sense.²³

SRGBV training had a positive impact on teachers' knowledge, but more work should be done so teachers can properly navigate SRGBV response procedures, including how to report SRGBV events, in ways that uphold confidentiality and support child protection follow-up. More training sessions, regular refreshers, and supportive supervision could help with this objective.

The qualitative work suggested several avenues for improving program implementation:

- Although we found evidence that the SRGBV training did help improve teachers' knowledge, qualitative work shows some confusion on basic concepts that could be clarified through further training. For example, it appears SRGBV is understood as sexual violence only, not as corporal punishment or emotional/physical violence. Though alternative discipline methods were deemed potentially effective, teachers were hesitant about using alternative discipline methods exclusively, and abolishing corporal punishment altogether.
- Qualitative data also indicate that girls face major challenges to deal with menstruation in school, and that teachers are not necessarily supportive in these scenarios.
- Findings suggest that in many cases, teachers report any incidences of sexual violence to head teachers, who then report them to local authorities. This reporting model leaves head teachers and master teachers with the power to determine which cases are worth raising to local authorities, which goes against recommended SRGBV reporting practices and raises risks of retaliation and revictimization to a child violence survivor. Further training should address this problem and teach teachers how to properly report SRGBV to a designated, skilled and mandated school counselor or formal child protection officer. This counselor should be trained and supervised in confidential, "do no harm," child protection and survivor-centered practices, and reporting to a legally mandated child protection officer, such as a Social Welfare Officer.

Unlike POY clubs, there are no substitutes for SRGBV training for teachers in these communities. This highlights the importance of continuing with SRGBV training for teachers in these and similar communities, and also during their preservice training at the teacher colleges.

²³ Insights from the INSPIRE model Seven Strategies for Preventing Violence Against Children include recommendations for implementing a combination of direct and indirect strategies at multiple levels (World Health Organization 2018).

SUSTAINABILITY

Findings indicate the following opportunities to improve the overall sustainability of WW intervention activities:

- Integrating community-level leadership, such as councils and parents groups in SRGBV training with clear explanation with roles and responsibilities of a child GBV-survivor centered reporting mechanism. This would uphold confidentiality of the child's case while connecting the child survivor with referrals to needed external psychosocial, medical, legal, safe refuge, or other support services. Referrals should be overseen by a Social Welfare Officer or other legally mandated child protection officer rather than by LGA officers who lack child protection mandate or training.
- Matching peer educators by sex with single-sex POY clubs and providing them mentorship and supportive supervision.
- Findings reveal that across WEI activities, LGA engagement was not always consistent. However, in areas where LGAs were active, teachers and EEVs had positive perceptions of the impact of LGA involvement on various activities. Though LGAs said they wanted to be more involved, they said they encountered challenges in securing transportation to local communities and in balancing their work duties with supporting WEI. Future activities should facilitate more opportunities for LGAs to engage with WEI activity components, including EEVs, peer educators, and teachers. Additionally, it may be beneficial to explore methods of providing transport fare for LGAs traveling to support WEI activities.

REFERENCES

- Acevedo, Paloma, Guillermo Cruces, Paul Gertler, Sebastian Martinez. 2017. Working Paper No. 23264. National Bureau of Economic Research. Cambridge, Massachusetts.
- Adoho, Franck, Shubha Chakravarty, Dala T. Korkoyah, Jr., Mattias Lundberg, and Afia Tasneem. 2014. “The Impact of an Adolescent Girls Employment Program: The EPAG Project in Liberia”. Policy Research Working Paper No. 6832. World Bank, Washington, DC.
- Alfassi. 2003. “Promoting the Will and Skill of Students at Academic Risk: An Evaluation of an Instructional Design Geared to Foster Achievement, Self-Efficacy and Motivation.” *Journal of Instructional Psychology* Volume: X-Y.
- Ashraf, Nava, Natalie Bau, Corinne Low, and Kathleen McGinn. 2018. “Negotiating a Better Future: How Interpersonal Skills Facilitate Inter-generational Investment.” *Population Center Working Papers* 4-28-2018.
- Austrian, K., Soler-Hampejsek, E., Mumah, J., Kangwana, B., Wado, Y., Abuya, B., Shah, V., and Maluccio, J. 2018. *Adolescent Girls Initiative-Kenya: Midline Results Report*. Population Council, Nairobi, Kenya.
- Bandiera, Oriana, Niklas Buehren, Markus Goldstein, Imran Rasul, and Andrea Smurra 2019. “The Economic Lives of Young Women in the Time of Ebola Lessons from an Empowerment Program.” *Policy Research Working Paper* No. 8760. World Bank, Washington, D.C.
- Bandiera, Oriana, Niklas Buehren, Robin Burgess, Markus Goldstein, Selim Gulesci, Imran Rasul, and Munshi Sulaiman. 2020. “Women’s Empowerment in Action: Evidence from a Randomized Control Trial in Africa.” *American Economic Journal: Applied Economics* 12(1): 210–259
- Bandura, A. 1997. *Self-efficacy: The exercise of control*. W H Freeman/Times Books/ Henry Holt & Co, New York, New York.
- Baumrind, D. 1978. “Parental disciplinary patterns and social competence in children.” *Journal of Youth & Society* 9: 239–276.
- Benard, B. 2004. *Resiliency, What We Have Learned*, WestEd, San Francisco, CA.
- Ozler, Berk, Kelly Hallman, Marie-France Guimond, Elizabeth A. Kelvin, Marian Rogers, and Esther Karnley. 2020. “Girl Empower – A gender transformative mentoring and cash transfer intervention to promote adolescent wellbeing: Impact findings from a cluster-randomized controlled trial in Liberia” *SSM*. Volume 10 April 2020, 100527.
- Buchmann, Nina, Erica Field, Rachel Glennerster, Shahana Nazneen, Svetlana Pimkina, and Iman Sen. 2018. “Power vs Money: Alternative Approaches to Reducing Child Marriage in Bangladesh, a Randomized Control Trial.” *Mimeo*. Duke University.
- CARE USA .2014. Gender Equitable Index (GEI) Toolkit. CARE USA: Atlanta.

- Cohen, S., and Wills, T. A. 1985. "Stress, Social Support, and the Buffering Hypothesis." *Psychological Bulletin*, 98: 310–357.
- Dhar, Diva, Tarun Jain, and Seema Jayachandran. 2018. "Reshaping Adolescents' Gender Attitudes: Evidence from a School-Based Experiment in India." August, 2018.
- Drago, Francesco. 2011. "Self-esteem and Earnings." *Journal of Economic Psychology* no. 32, 2011: 480–488.
- Edmonds, Eric, Ben Feigenberg, and Jessica Leight. 2019. Advancing the Agency of Adolescent Girls. Unpublished manuscript.
- Femina Hip. 2020. *Fema Clubs & Networks*. Viewed July 2, 2020.
- <http://www.feminahip.or.tz/how-we-do-it/community-mobilization/fema-clubs/#:~:text=We%20now%20have%20active%20Fema,Dar%20es%20Salaam%20and%20Mtwara.http://www.feminahip.or.tz/how-we-do-it/community-mobilization/fema-clubs/#:~:text=We%20now%20have%20active%20Fema,Dar%20es%20Salaam%20and%20Mtwara.>
- Graczyk, P.A. 2000. "Adolescent Peer Relationships and Their Association with Emotional and Physical Wellness." PhD diss., Northern Illinois University. *Dissertation Abstracts International* 6:, 12-B.
- Green, William H. 2008. *Econometric Analysis*. 6th Edition. New Jersey: Pearson Prentice Hall.
- Guerrero, G., Sugimaru C., Cussianovich A., Fraine, B., and Cueto, S. .2016. "Education Aspirations among Young People in Peru and their Perceptions of Barriers to Higher Education." *Young Lives Working Paper* 148. Department of International Development at the University of Oxford.
- Hallman, Kelly. 2016. "Girl Empower Intervention Baseline Survey. Innovations for Poverty Action." Population Council. International Rescue Committee.
- Jensen, Robert and Emily Oster. 2009. "The Power of TV: Cable Television and Women's Status in India." *The Quarterly Journal of Economics*, Vol. 124, No. 3, Aug., 2009: 1057-1094
- Jones, Shelley. 2015. Gender Issues in Mathematics Education in Tanzania. A Literature Review. Capacity Development for Mathematics Teaching in Rural and Remote Communities in Tanzania project.
- Koroknay-Palicz, Tricia and Joao Montalvao. 2019. The Impact of Social and Emotional Learning: Experimental Evidence from Adolescent Girls in Liberia. 3rd IZA Workshop on Gender and Family Economics. April 12-3, Viña del Mar, Chile.
- Meeks-Gardner, J., Guerra, N. G., Williams, K. R., and Walker, I. 2011. "The Jamaica Youth Survey: Assessing core competencies and risk for aggression among Jamaican youth." *Caribbean Child Development Quarterly*, no. 57: 35-53.
- Menendez, Alicia, Gregory Haugan, and Zoe Grotophorst. 2018. Apateni Mwayi Atsikana Aphunzire Amaa Evaluation and Research: Baseline Report for School Construction Activity in Machinga and Balaka Districts.

- Ministry of Health, Community Development, Gender, Elderly and Children [Tanzania Mainland], Ministry of Health [Zanzibar], National Bureau of Statistics, Office of the Chief Government Statistician, and ICF .2016. *Tanzania Demographic and Health Survey and Malaria Indicator Survey (TDHS-MIS) 2015-16*. Dar es Salaam, Tanzania, and Rockville, Maryland.
- NORC at the University of Chicago. 2017. *USAID/Uganda Performance and Impact Evaluation for Literacy Achievement and Retention Activity (LARA)*. Bethesda, MD: NORC.
- https://pdf.usaid.gov/pdf_docs/PA00TGQQ.pdf
- NORC at the University of Chicago. 2018. *USAID/Tanzania Waache Wasome Evaluation Design Report*. Bethesda, MD: NORC.
- NORC at the University of Chicago. 2019. *USAID/Tanzania Waache Wasome Baseline Report*. Bethesda, MD: NORC.
- NORC at the University of Chicago. 2020. “Corporal Punishment in Schools in Low-income Countries: An Evidence Review.” *Select Gender-based Violence Literature Reviews*. Washington, DC: USAID/NORC.
- Population Council. 2015. *Building Assets toolkit: Developing positive benchmarks for adolescent girls*. Place needed: Population Council.
- Robinson, C. C., Mandleco, B., Olsen, S. F., and Hart, C. H. 2001. “The Parenting Styles and Dimensions Questionnaire (PSDQ)”. *Handbook of family measurement techniques: Vol. 3. Instruments & index*: 319 - 321. B. F. Perlmutter, J. Touliatos, & G. W. Holden (Eds.), Thousand Oaks: Sage.
- Rosenberg, M. 1965. *Society and the Adolescent Self-image*. Princeton, NJ: Princeton University Press.
- RTI International. 2016. *Conceptual Framework for Measuring School-Related Gender-Based Violence*. U.S. Agency for International Development. Washington, DC
- RTI International 2016. Literature Review on School-Related Gender-Based Violence: How it is Defined and Studied. Washington, DC: U.S. Agency for International Development.
- Schmitt, David P. and Juri Allik. 2005 :Simultaneous Administration of the Rosenberg Self-Esteem Scale in 53 Nations: Exploring the Universal and Culture-Specific Features of Global Self-Esteem.” *American Psychological Association Journal of Personality and Social Psychology*. Vol. 89, No. 4: 623–642.
- Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy scale. In J. Weinman, S. Wright, & M. Johnston, *Measures in health psychology: A user’s portfolio. Causal and control beliefs* (pp. 35-37). Windsor, UK: NFER-NELSON.
- Schwarzer, Ralf & Urte Scholz (2000). Cross-cultural assessment of coping resources: the general perceived self-efficacy scale. Paper presented at the Asian Congress of Health Psychology 2000: Health Psychology and Culture, Tokyo, Japan, August 28-29.

Tafere, Y., (2014) “Education Aspirations and Barriers to Achievement for Young People in Ethiopia, Young Lives” Working Paper 120, Young Lives, Department of International Development at the University of Oxford.

UNICEF (2011). Summary Report on the Prevalence of Sexual, Physical and Emotional Violence, Context of Sexual Violence, and Health and Behavioural Consequences of Violence Experienced in Childhood. Violence against Children in Tanzania: Findings from a National Survey, 2009. Dar es Salaam, Tanzania: UNICEF Tanzania, Division of Violence Prevention, National Center for Injury Prevention and Control, Centers for Disease Control and Prevention, and Muhimbili University of Health and Allied Sciences.

UNICEF (2018). Tanzania/Education. Retrieved from:

<https://www.unicef.org/tanzania/education.html>, September 21, 2018.

USAID (2007). Gender Terminology. November.

USAID (2018). How-To Note Gender Integration in Education Programming. June.

WEI (2018). FY18 Annual Progress Report October 1, 2017 – September 30, 2018.

Wodon, Q., C. Montenegro, H. Nguyen, and A. Onagoruwa (2018). Missed Opportunities: The High Cost of Not Educating Girls. The Cost of Not Educating Girls Notes Series. Washington, DC: The World Bank.

World Health Organization (2018). INSPIRE Handbook: Action for implementing the Seven Strategies for Ending Violence against Children. Geneva: Licence: CC BY-NC-SA 3.0 IGO.

ANNEX I. RELEVANT LITERATURE

A growing body of literature analyzes the impact of interventions aimed at developing protective assets on future outcomes, including educational attainment, employment and fertility.

In terms of education, Ashraf et al. (2018) analyzed the impact of providing negotiation and interpersonal communication training to adolescent girls in Zambia. Using administrative data, the authors found that the program reduced dropout rates and increased school enrollment four years after the program by 4.4 percentage points (10 percent).

Buchmann et al. (2018) studied the impact of two different interventions aimed at reducing child marriage and teenage pregnancy and increasing girls' schooling in Bangladesh. One intervention was an empowerment program that included an educational component (i.e., basic literacy, numeracy) and a social competency component (i.e., life skills, and nutritional and reproductive health knowledge). The second intervention provided financial incentives conditioned on delaying marriage. The study found that 4.5 years after program implementation, financial incentives reduced child marriage and teenage childbearing, and increased schooling, while the empowerment program had a positive impact on schooling only. Specifically, girls in the empowerment program were 2.9 percentage points (10 percent) more likely to be in school than the control group. Notably, there were positive and relatively large effects on income-generating activities (e.g. employment status, hours worked and wages) for the empowerment arm, which were not replicated in the financial incentives arm.

Koroknay-Palicz & Montalvao (2019) analyzed the impact of a mentorship program in Monrovia, Liberia, aimed at increasing girls' socioemotional skills, including trust building, communication skills, managing emotions, and problem solving. The study found a 3.6 percentage points gain in primary school completion (6.2 percent), and a 3.4 percentage points gain in transition to secondary school, (9.6 percent).

Edmonds et al. (2019) evaluated a life skills curriculum and mentoring program for adolescent girls in India, delivered by trained local women. The objective was to increase girls' agency to stay in school. They found a decline of 3 percentage points in the dropout rate (25 percent).

Bandiera et al. (2019) analyzed a program in Sierra Leone providing young girls and young women vocational training and information on health and reproductive issues. The timeline of the evaluation coincided with the Ebola outbreak of 2014. Among other government responses to the crises, schools were closed through the 2014-15 academic year, which had major effects on human capital accumulation for school-age children. The authors show that while in villages highly affected by the Ebola crisis girls were less likely to return to school once schools reopened, being in the treatment group reduced these adverse effects substantially.

Austrian et al. (2018) reviewed the midline results of a multi-arm intervention in Kenya that included cash transfers conditioned on school attendance, weekly girls' group meetings led by local women covering health and life-skills topics, and training on financial education and savings accounts. Although the study found positive impacts on schooling for some treatment arms, the results do not suggest that the life skills component has a differential impact on school attendance.

With respect to labor outcomes, Acevedo et al. (2017) analyzed the impact on earnings and personal skills of training youth in the Dominican Republic. The evaluation design randomized participants in two treatment groups and a control group. Treatment 1 included soft-skills training and internships and treatment 2 included all treatment 1 components plus job training. Soft skills training focused on promoting self-esteem, communication skills, and conflict resolution, among other traits. Twelve months after the intervention, the evaluation documented a positive impact on earnings for females in both treatment groups. The fact that there were no major differences between the two treatment groups suggests that job training was of low marginal value. Although the authors did not find impacts on labor outcomes for either treatment group 36 months after the intervention, females in the treatment groups were found to have higher self-esteem than their counterparts in the control group.

Bandiera et al. (2020) analyzed a program in Uganda providing adolescent girls (both in and out of school) training on both vocational and life skills. Four years post-intervention, the study found that girls in the treatment group were more likely to engage in income-generating activities than in the control group, and less likely to have a child.

Adoho et al. (2014) analyzed the impact of technical and life skills training for young women in Liberia. They found positive impacts on employment and earnings.

Regarding attitudes towards gender roles and norms, several studies have documented the impact of modifying gender norms and roles on school-related and other socio demographic outcomes. For example, Jensen and Oster (2009) analyzed the impact of access to cable television on women's status in India. They used a three-year panel dataset to show that access to cable television reduced the likelihood women report that it is acceptable for husbands to beat their wives. Access to cable television was also found to reduce fertility rates, and increased school enrollment for young children.

More recently, Dhar et al. (2018) analyzed the short-term impacts of a school-based intervention aimed at reducing gender norms disparities. The intervention is based on classroom discussions about gender equality, girls' education, and harassment, among other topics. The authors used a randomized experimental design to measure the impact of the program on students' gender attitudes, aspirations, and behaviors. They found the program positively affects gender attitudes, measured by an index that summarizes a number of gender norms-related items. The intervention also reduced gender stereotyped behavior.

Ozler et al. (2020) studied Girl Empower, a life skills program aimed at reducing sexual violence for adolescent girls fielded in rural Liberia. Over a period of 10-months, 84 villages were randomly assigned to two treatment groups and one control group. One of the treatment groups received the basic Girl Empower intervention, which included a life skills curriculum facilitated by local female mentors, caregiver discussion groups, and seed money for girls to start savings accounts. The other treatment groups receive the same services plus financial incentive to attend the life skills curriculum sessions. The study found that while the program did not have an impact on sexual violence or schooling, it had positive and significant impacts on gender attitudes, life skills and sexual and reproductive health.

ANNEX II. SELECTING ARD GIRLS IN THE CONTROL GROUP

To evaluate the impact of the program, we compare retention rates and other outcomes of interest between treatment and control groups at the endline. While we can estimate experimentally the impact of the program on average, we cannot estimate experimentally the impact on the most vulnerable girls (IEI.I). This is because teachers selected ARD girls for POY clubs participation based on a series of criteria that made these girls more likely to drop out (e.g. frequent absenteeism, struggling academically, etc.). Given this selection process we expect ARD girls to come from more disadvantaged socioeconomic backgrounds than the average girl in the treatment or control groups.

If ARD girls come from more disadvantaged backgrounds than the average girl in the sample, we need to search for girls in the control group who also come from disadvantaged backgrounds, so we can construct a valid comparison group for ARD girls.

Therefore, to evaluate the impact of the program on ARD girls, we use quasi-experimental methods. Specifically, we use propensity score matching (PSM) to weight the comparison group so it is observationally equivalent to the ARD girls.

More specifically, to estimate the impact of the program on ARD girls, we calculate the probability of being selected as ARD girl in treatment schools estimating:

$$[\Pr(D)_i = 1] = \Lambda[\mathbf{x}_i' \boldsymbol{\delta} + u_i]$$

where $\Lambda[\cdot]$ is the logistic function, D_i is a dummy variable for being selected as ARD girl, $\mathbf{x}_i$ is a vector of girl characteristics that include age, parental education, household size and possible protective scores at baseline, u_i is an error term, and $\boldsymbol{\delta}$ is a vector of parameters to be estimated. Note that this regression is estimated using only girls in treatment schools. We use the estimated parameters in this model $\hat{\boldsymbol{\delta}}$ to calculate the propensity score $\hat{p}$ (that is, the probability of having been selected as ARD girl) for girls in the control group. Finally, we use $\hat{p}$ to construct weights so girls in the control group who were more likely to be classified as ARD are upweighted, and girls who were less likely are down-weighted. This weighted control group constitutes the matched control group that we will use to estimate the impact of the program on ARD girls.

The results of this balancing exercise are shown in **Table 15**. The first and second columns show descriptive statistics for ARD girls and the “raw” (unweighted) control group, respectively, while the third column shows the differences between these two groups. ARD girls own fewer household assets, spend more time traveling to school, and their parents have less education, than girls in the control group, on average. The fourth column shows the weighted control group. In this case the differences with respect to ARD girls, shown in the fifth and final column, are much smaller, and none of them are significant, showing how the PSM weighting process balances the two samples.

Table 15: Weighting the control group so it is comparable to ARD girls

	Unweighted			Weighted	
	ARD girls (1)	Control (2)	Diff. (3)=(1)-(2)	Control (4)	Diff. (5)=(1)-(4)
Age (years)	14.61	14.63	-0.02	14.74	-0.13
Number of household members	6.55	6.64	-0.10	6.75	-0.20
Did you eat anything today? (fraction)	0.28	0.36	-0.08	0.27	0.01
Electric grid (fraction)	0.21	0.29	-0.08	0.20	0.00
Cell phone ownership (fraction)	0.93	0.96	-0.03	0.94	-0.01
Asset index	3.80	4.45	-0.65**	3.96	-0.17
Time to get to school (minutes)	52.43	43.03	9.40**	46.39	6.05
Parental education (fraction)					
No schooling	0.06	0.04	0.02	0.05	0.01
Some primary	0.11	0.07	0.04*	0.09	0.02
Complete primary	0.62	0.58	0.04	0.66	-0.04
Some secondary	0.07	0.08	-0.01	0.07	0.00
Complete secondary	0.11	0.16	-0.06*	0.09	0.02
Higher than secondary	0.03	0.07	-0.04*	0.04	-0.01
N	271	1120		1120	

Notes: One of the ARD girls is dropped due to lack of common support. Weights used were kernel Epanechnikov.

Standard errors for the significance test are clustered at the school level.

* p<0.05 ** p<0.01 *** p<0.001

This weighted control group is the one used in the estimation of the impact of the program on ARD girls, shown in **Table 8** of the main body of the report.

ANNEX III. INSTRUMENTS

ANNEX IV. DATA COLLECTION

Data collection for this project was conducted by Economic Development Initiatives Limited (EDI) Ltd, with close supervision from NORC at the University of Chicago. This annex provides an overview of the fieldwork set-up, sampling, and data collection protocols.

QUANTITATIVE DATA COLLECTION²⁴

The girls' surveys include modules to capture information about girls' protective assets, attitudes towards gender norms, and perceptions on SRGBV, as well as basic socio demographic information, including parental education, household size and household asset ownership. To construct this instrument, NORC's team reviewed several widely used and validated instruments, including Tanzania's Demographic Health Survey, the Self-Efficacy Questionnaire for Children (SEQ-C); the Rosenberg Self-Esteem Scale; and USAID's SRGBV toolkit.

To design the parent and teacher surveys, we reviewed and drew on the content of the corresponding training manuals used by WW. The parent survey also included items to measure parenting styles from the Parenting Styles and Dimensions Questionnaire (Robinson, et. al 2001) and attitudes towards gender equity from the Gender Equity Index (Care, 2014). The teacher survey also included items from the Gender Equity Index. All endline surveys included a module to measure participation in WW activities.

I. Administrative data collection 2018 (October – November 2018)

Survey Development and Pretesting

The survey was subjected to thorough desk testing throughout the preparation phase by EDI and NORC. The questionnaire structure and length were tested again during interviewer training

Training

Training took place October 29 to November 2, 2018, with additional refresher training on November 22. The training sessions took place in Bukoba, Kagera. The interviewer training was attended by a total of ten interviewers. All interviewers were introduced to the project and given an initial overview of the survey questionnaire and protocols. Of the ten interviewers trained, eight were selected to be a part of the field teams, based on their performance in training. During the training phases, trainees conducted mock interviews to further test the questionnaires and the interviewers' familiarity with the tools. During this phase the interviewers also contributed to further updating the questionnaires, manuals, and translations.

Team Setup

The field work was carried out by eight interviewers, overseen by a Field Coordinator who remained in the field for the duration of the data collected activities. The interviewers were selected on the basis of having worked previously with NORC, or on other projects involving visiting schools. All teams were overseen by a Project Coordinator, a Data Processing Officer and a Team Leader.

²⁴ The data collection of the girl baseline survey, conducted in early 2018, is described in NORC (2019).

Fieldwork Summary

EDI attempted to start fieldwork on November 5, 2018. However, on that first day it became clear schools would not be permitting our activities in coming weeks, when national and local exams would be administered in all secondary schools. The decision was therefore made to halt data collection until after exams had been completed. From November 6 to 9, EDI contacted each school in the sample to ascertain the most convenient dates for us to visit the schools and complete the activities. Based on these calls EDI and NORC decided to go back to the field and visit all 61 schools from November 26 to December 5. All details that follow refer to this second, main, fieldwork activity.

In total, 61 surveys were completed as part of the fieldwork, with 9,661 students recorded.

2. Parent Baseline (January – February, 2019)

Questionnaire Development and Pretesting

During the preparation phase from January 7 to 22, 2019 a number of changes were made to the questionnaire. These questions were requested by NORC and were configured into SurveyBe by the EDI Team Leader. Fieldwork protocols were reworked, questions were added and removed, and response lists were also updated as necessary. The survey was subjected to thorough desk testing throughout the preparation phase, both by EDI staff in the UK and Tanzania.

Training

Training for this round of the survey took place from January 29 to February 1, 2019. The training sessions took place in Bukoba, Kagera. The interviewer training was attended by a total of 14 interviewers as well as the EDI coordination team, and a representative of NORC. All interviewers were introduced to the project and given an initial overview of the survey questionnaire and protocols. Of the 14 interviewers trained, 12 were selected to be a part of the field teams, based on their performance in training. All 12 of the selected interviewers had worked on the girl survey fielded in 2018. During the training phases, trainees conducted mock interviews to further test the questionnaires, and the interviewers' familiarity with the tools. During this phase the interviewers also contributed to further updating the questionnaires, manuals, and translations.

Contacting parents

Sampled parents were contacted through WW's EEVs, and asked to go to their children's schools for an interview. Only parents from schools where there were at least five eligible parents (parents that were in a LIMCA group and were parents of a school-age girl) were interviewed. The target was to survey 333 parents in 22 treatment schools.

Data Collection Summary:

Data collection took place February 2 to 17, 2019 in four districts of Tanzania. In total, 324 surveys in 22 schools were completed as part of the fieldwork. The field work was carried out by 12 interviewers, split into three teams, each led by a senior interviewer.

3. Administrative data collection 2019 (September – October 2019)

Survey Development and Pretesting

Only minor changes were introduced with respect to the instrument used in 2018. The survey was subjected to thorough desk testing throughout the preparation phase. The questionnaire structure and length were tested again during interviewer training.

Training

Training for this round of the survey took place September 18 to 20. The training sessions took place in Bukoba, Kagera. The interviewer training was attended by a total of ten interviewers. All interviewers were introduced to the project and given an initial overview of the survey questionnaire and protocols. Of the ten interviewers trained, nine were selected to be a part of the field teams, based on their performance in training. During the training phases, trainees conducted mock interviews to further test the questionnaires, and the interviewers' familiarity with the tools. During this phase the interviewers also contributed to further updating the questionnaires, manuals, and translations.

Fieldwork Summary:

Data main collection activity took place September 23 to October 11 across four districts of Tanzania. In total, 61 surveys were completed as part of the fieldwork.

4. Endline Girl, Parent and Teacher Surveys (January – March 2020)

Questionnaire Development and Pre-testing

Testing was split into two activities. First, EDI Global Project Coordinator and Data Manager desk-tested the survey tools from January 15 to 17, 2020. Second, a field pre-test of the girl, parent and teacher quantitative surveys were conducted in four schools in Karatu District January 21 to 23, 2020. The four pretest schools had received the WW interventions but were not part of the sample, making them ideal for survey testing. The pre-testing was overseen by EDI and NORC staff.

Pre-visits

School pre-visits took place January 2031, 2020. A team composed of EDI Global pre-visit interviewers visited the 61 schools to inform the schools of the upcoming survey activities and compile lists of the girls surveyed at baseline to see who was still in school and who had dropped out. A full list of teachers from the school with accompanying eligibility questions was also compiled during the pre-visit. Finally, the pre-visit teams also delivered parent consent forms to the school for the sampled girls and invitation letters for parents to be interviewed.

Two days before the visits, the supervisors of each team contacted the school Headmasters to remind them and to check the number of parent consent forms returned to the school. During the two days in the schools, the teams carried out interviews with girls after collecting consent forms signed by parents and teacher interviews.

Girl survey

During the school pre-visit for the endline survey, information on girls interviewed at baseline were collected to determine if the girls were still attending the same school, had transferred to a different school, moved to another community, dropped out of school or their status was unknown to school authorities.

Each supervisor was provided lists of up to 38 girls per school to be interviewed as well as information on the girls no longer in school. This list was given to the Headmaster, who was then responsible for finding the selected students. Each interviewer was assigned a girl to be interviewed by the supervisor. Interviewer teams had two days to complete the girl interviews at each school. One tracking interviewer with the teacher team moved independently from the main team and was responsible for tracking girls who dropped out of school and who were still living in the community.

A total of 1,815 girls, 82 percent of the 2,220 surveyed at baseline, were interviewed. The survey lasted approximately 45 minutes on average.

Teacher Survey

During school pre-visits conducted prior the main data collection a full list of the names of every teacher in the school was compiled. From this list, 15 (ten sample and five replacement) teachers were randomly sampled in SurveyBe according to the following priority ranking:

- First priority - Teachers who taught Form 1 in 2018 or Form 2 in 2019.
- Second priority – If additional teachers needed to reach a sample of 15 teachers (ten for the main sample and five replacement teachers), SurveyBe randomly selected any other teachers listed in the roster until reaching a sample of 15.

Teachers who complied with the first category were the first priority to be interviewed because of the likelihood they had taught the respondents interviewed in the baseline Girl Survey. Each supervisor was provided a list of teachers per school with their IDs to be interviewed and the replacement to be used if needed. The supervisor assigned two interviewers from the team, five teachers each to be interviewed during the two-day school visit.

A total of 588 teachers were interviewed. The survey lasted approximately 35 minutes on average.

Parent Survey

School pre-visit interviewers were tasked to list parents interviewed at baseline in 2019. After completing the parent pre-visit survey, the pre-visit interviewer provided the parent interview letters to the Headmaster and these were passed onto students if parents still have a child (either the sampled girl or other child) in the school.

For parents with no child currently enrolled in this school, EDI Global contacted them later using contact information details collected in 2019 and using the contact of the EEV that was in charge of the LIMCA group of each particular parent. Only parents interviewed at baseline were eligible for the endline survey (no replacements).

EEVs were used to contact parents for interviews. When contacted, parents were asked to come to their child's school to conduct the interview.

Sampled parents were provided a date and time to arrive at the school for an interview on the invitation letters. They were asked to inform the interviewer if they were unable to attend the school and the interviewer in partnership with the Project Coordinator decided the best course of action as to the location of the interview.

The field teams visited 22 treatment schools February 10 to 22, 2020. Due to a variety of issues that prevented parents from attending the school to conduct the interview, field teams were asked to visit households of parents who were unable to come to school. From February 24 to March 6, supervisors responsible for the school survey teams also conducted parent interviews of some respondents who were unavailable on the designated interview day.

A total of 305 parents, 94 percent of the 324 surveyed at baseline, were interviewed. The survey lasted approximately 40 minutes on average.

5. Administrative data collection 2020 (January 2021)

Survey Development and Pretesting

During the preparation phase a small number of changes were made to the questionnaire. The questionnaire reference data was updated to include all the students recorded in 2019. The survey was subjected to thorough desk testing throughout the preparation phase. The questionnaire structure and length were tested again during interviewer training.

Training

Training for this round of the survey took place on the 14th and 15th of January in Bukoba, Kagera. The interviewer training was attended by a total of 9 interviewers. All interviewers were introduced to the project and given an initial overview of the survey questionnaire and protocols. During the training phases, trainees conducted mock interviews to further test the questionnaire, and the interviewers' familiarity with the tools. During this phase the interviewers also contributed to further updating the questionnaire, manual, and translations.

Due to Covid-19 pandemic on arrival staff underwent a health questionnaire and contactless temperature check to ensure no interviewers showed symptoms of infection. Following a clear health check trainee data collection staff were invited to sit socially distanced for the length of the training session.

Fieldwork Summary

The main data collection activity took place between January 18th and January 29th across four districts of Tanzania. In total, 61 surveys (one per school) were conducted.

6. Administrative data collection 2021 (October 2021)

Survey Development and Pretesting

Only minor changes were introduced with respect to the instrument used in early 2021. The survey was subjected to thorough desk testing throughout the preparation phase. The questionnaire structure and length were tested again during interviewer training. Reference data was updated to include all the students recorded in 2020/21.

Training

Training for this round of the survey took place on the 7th and 8th of October in Bukoba, Kagera. The interviewer training was attended by a total of six interviewers. All interviewers were introduced to the project and given an initial overview of the survey questionnaire and protocols. During the training phases, trainees entered historic data to further test the questionnaire, and grow the interviewers' familiarity with the tools. During this phase the interviewers also contributed to further updating the questionnaire, manual, and translations. Continuing with precautions in response to the COVID-19 pandemic, on arrival for the first day of training staff underwent a health questionnaire, and each day received a contactless temperature check to ensure no interviewer showed symptoms of infection. Following a clear health check trainee data collection staff were invited to sit socially distanced for the length of the training session.

Fieldwork Summary:

The main data collection activity took place between October 9th and October 29th across four districts of Tanzania. In total, 61 schools (one per school) were surveyed as part of the fieldwork.

QUALITATIVE DATA COLLECTION

Qualitative instruments sought to measure respondent perceptions of program efficacy, implementation challenges, and social norms related to gender. Similar questions on gender norms and program efficacy were asked across different respondents to allow for triangulation. All instruments were informed by review of WW program documents, including training manuals, quarterly reports, and annual reports.

Both quantitative and qualitative instruments were shared with USAID (E3/GenDev and the Tanzania Mission) and WEI for feedback.

NORC planned to conduct FGD with youth and parents, as well as KIIs with peer educators, teachers, head teachers, economic empowerment volunteers, and local government agencies in May 2020. Due to the COVID-19 pandemic, schools in Tanzania closed in March 2020. Given the risks associated with conducting in-person data collection in schools, we decided to split the qualitative data collection in two. In July we conducted KIIs by phone with teachers, head teachers, economic empowerment volunteers, and local government agencies. We postponed FGDs with girls, boys and parents until October, when schools were open and in-person data collection was authorized. We also conducted KIIs with peer educators in October, to avoid difficulties associated with interviewing youth on the phone.

I. Phone KIs (May - July 2020)

Data Collection Team

The data collection team was composed of one team leader, one project coordinator, three enumerators, and two quality control officers. The project coordinator served as an additional enumerator in this stage.

Training

Enumerator training was conducted over Skype, with enumerators, EDI, and NORC staff participating from home. The training was conducted May 19 to 29. The training schedule included two sessions per day, each a maximum of two hours. This ensured session content was delivered efficiently and afforded sufficient time for independent learning. The morning sessions were internal EDI Global session led by the Team Leader. Internal session topics included a review of the protocols and enumerator manual. The afternoon sessions were led by NORC, and sessions consisted of reviews of the survey tools, interviewing techniques, and protocols for reporting cases of violence.

Sampling

Teachers were selected from among those surveyed in February 2020. Six head teachers from six randomly sampled schools were interviewed. LGAs and EEVs were sampled from lists provided by WEI.

Pre-Calls

June 1 and June 2, 2020 were allocated as days to make pre-call arrangements. Pre-calls were introductory calls made by the interviewers to sampled respondents. After three failed call attempts, the respondent was replaced with another respondent by the team leader.

Data Collection and Results

Data collection took place June 3 to July 20, 2020. Sample respondents were spread over the four districts of Karatu DC in Arusha region and Musoma DC, Musoma MC, and Tarime DC in Mara region. KKIs were conducted with 16 teachers, 6 head teachers, 8 EEVs, and 12 LGA, for a total of 42 interviews. While the plan in the Evaluation Design Report was to conduct 12 KIs with teachers and 12 KIs with EEVs, per suggestions of WEI, it was decided to do 4 more KIs with teachers and 4 fewer with EEVs. The average duration of the interviews was 63 minutes.

The samples were provided by NORC and distributed by the Team Leader to the enumerator team. The sample was split to ensure the key informant groups and locations were distributed evenly across the enumerator team.

If a replacement was required, the enumerators contacted the Team Leader who then provided a replacement who ideally worked in the same district as the original respondent. The replacement rate was 26 percent.

Quality Control

All transcripts were double reviewed, firstly by an EDI Global staff member based in Uganda and secondly by the EDI Global Team Leader whose first language is English. This review process ensured every transcript could be refined in terms of grammar, formatting, and small changes to the English language to enhance meaning and clarity.

2. In-person KIIs and FGDS (August - October 2020)

Data Collection Team

The data collection teams were composed of one team leader, one project coordinator, three enumerators (1 FGD Facilitator, 1 FGD Notetaker, and 1 KII Interviewer), and two quality control officers. The project coordinator served as one of the three enumerators.

Training

Interviewer training was conducted from EDI Global's office in Bukoba, with session leads calling in via Skype. The training was conducted August 31 to September 4. The training schedule was built to have three full-day training sessions with an additional two days scheduled for practice FGDs and transcripts. Morning sessions focused on the review of the COVID-19 Return to Field Protocols, interviewer manual, and project details. The afternoon sessions were led by NORC, and sessions consisted of reviews of the survey tools, interviewing techniques, and protocols for reporting cases of violence.

On the first training day, trainees were required to fill out a health screening survey on arrival attesting they did not have any symptoms associated with COVID-19. The trainees also received COVID-19 protocols and successfully adhered to them throughout the training period.

Data Collection and Results

As this data collection activity was the first EDI Global project to return to the field after the COVID-19 closures, the project team provided additional field protocols approved by NORC and EDI Global to ensure fieldwork could be conducted safely.

Data collection took place September 21 to October 14, 2020. The sample respondents were attached to six schools spread over the four districts of Karatu DC in Arusha region and Musoma DC, Musoma MC and Tarime DC in Mara region. In total, 12 KIIs and 18 FGDS were completed.

Participants for the Girl and Parent FGDS were sampled from the girl and parent survey data collected in February 2020. The sample for the Boys FGDS (POY & Non-POY) was randomly sampled by the Team Leader, using administrative data from 2019 Admin Data Collection. Peer educators were also sampled by the Team Leader during fieldwork. Although the plan was to sample only girl-peer educators, in the selected schools, there were fewer than 12, so three boy-peer educators were interviewed as replacements. Six FGD were completed with girls, boys and parents, for a total of 18 FGD, as well as 12 KIIs with peer leaders. Most FGDS were approximately two hours in length (average duration of 125 minutes) and the KIIs lasted an average of 52 minutes.

Quality Control

All transcripts were double reviewed, first by an EDI Global staff member based in Uganda and second by the EDI Global Team Leader native language is English. This process ensured every transcript could be refined in terms of grammar, formatting, and word choice to enhance meaning and clarity. All quality control reviewers followed the transcriber guidelines provided by NORC to ensure consistency across the transcripts.

3. In-person FGDS (October-November 2021)

Data collection team:

The data collection team was composed of one Team Leader, one Project Coordinator, one FGD Facilitator, one FGD Notetaker and one Quality Control Consultant with expert information and guidance provided through EDI's Technical Oversight. The Project Coordinator and QC Consultant were also two of the three translation staff during this activity.

Training:

Training for this data collection was conducted from EDI Global's office in Bukoba, between the 6th and 8th of October 2021, with session leads calling in via Microsoft Teams. The Team Leader and session facilitators participated from their own homes in the UK, Germany and USA, with the FGD team in attendance at the Tanzania office in Bukoba, Kagera. The training schedule was built to have three full-day training sessions, including time scheduled for practice FGDs and transcription. Client led sessions which took place in the afternoons, while internal sessions were scheduled for the mornings and led by the Project coordinator. Morning sessions focused on reviewing the COVID-19 Protocols, interviewer manual and project details. The afternoon sessions were led by NORC and sessions consisted of reviews of the survey tools, interviewing techniques and protocols for reporting cases of violence. On the first training day, all trainees were required to fill out a health screening survey on arrival to ensure that they did not have any symptoms associated with COVID-19 followed by a contactless temperature check, that was then undertaken for each day of the training. The trainees were also explained the COVID-19 protocols and successfully adhered to them throughout the training period.

Data collection and results:

Data collection took place between 9th October and 8th November 2021. Participants were students or former students from six schools spread over the four target districts. In total 12 FGDs were completed, six with female POY club members that still attend school and six with former POY club members who have dropped out of school. Before undertaking data collection in schools, each school was visited by a field team to make them aware of the upcoming school visit and track prospective participants. All prospective participants were provided with a parental consent form and information sheet.

The initial sample, containing those known to have been POY club members, were provided by NORC and distributed by the Team Leader to the interviewer team to approach those listed upon visiting schools. If the original participants were unable or unwilling to participate in the FGD a replacement was selected.

The mean length of FGDs for in-school groups was 88 minutes, whereas dropout groups averaged 74 minutes.

Quality Control:

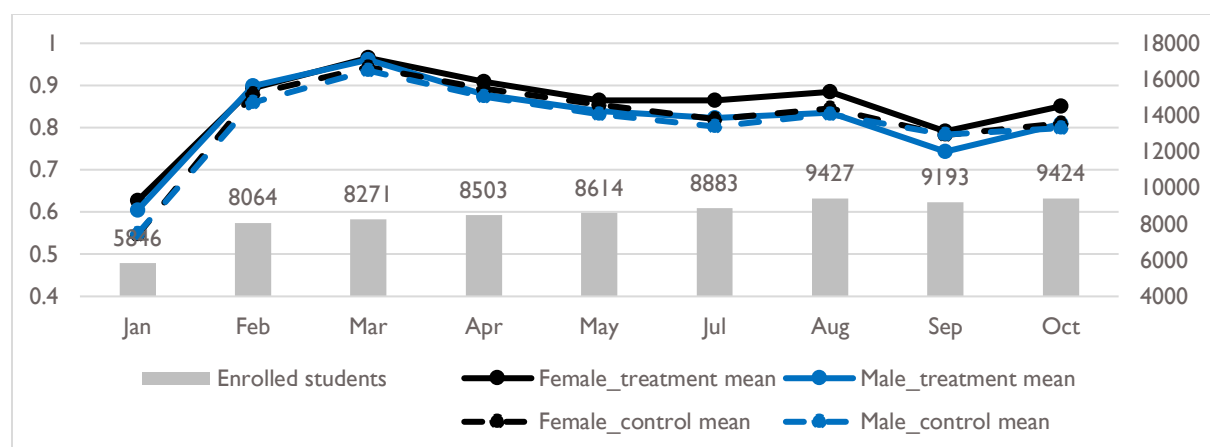
Three transcripts completed in the first week of translation were selected and reviewed by an experienced translation consultant based in Tanzania who possesses an advanced level of English language. All transcripts were double reviewed, firstly by an EDI Global staff member based in Tanzania and secondly by the EDI Global Team Leader whose mother tongue is English. All quality control reviewers followed the transcriber guidelines provided by NORC to complete the review, ensuring consistency across the transcripts..

ANNEX V. MONTHLY ATTENDANCE DATA

In this Annex we analyze the monthly attendance data collected administratively in 2018 and 2019.

Figure 20 shows attendance rates for 2018 by gender and treatment status. Bars located in the lower part of the figure show the total number of students in each month.²⁵ For both males and females attendance is a little higher at the beginning and throughout the year for treatment than for control schools.

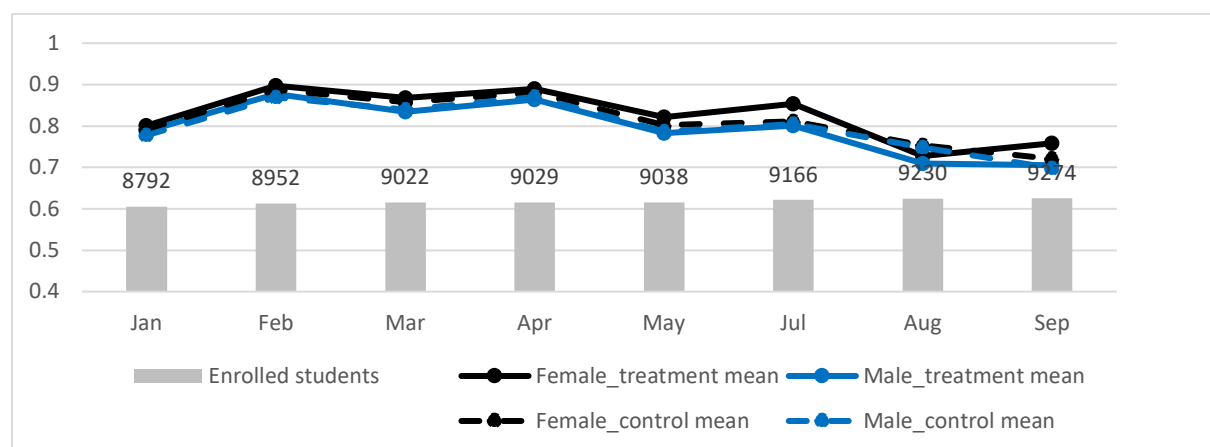
Figure 20: Attendance rates by gender and treatment status - 2018



Notes: We calculate attendance rate as the number of days a student attended school over the number of days s/he was supposed to attend. All attendance rates are truncated at 1 (in some cases attendance rates are higher than one because teachers sometimes have students come during the weekends, this happened rarely during 2018). There are no data points for June because students were on vacation most of the time.

Figure 21 shows attendance rates for 2019. In this year there is also a difference in favor of girls in the treatment group with respect to boys or girls in the control group, but the difference is very small.

Figure 21: Attendance rates by gender and treatment status - 2019

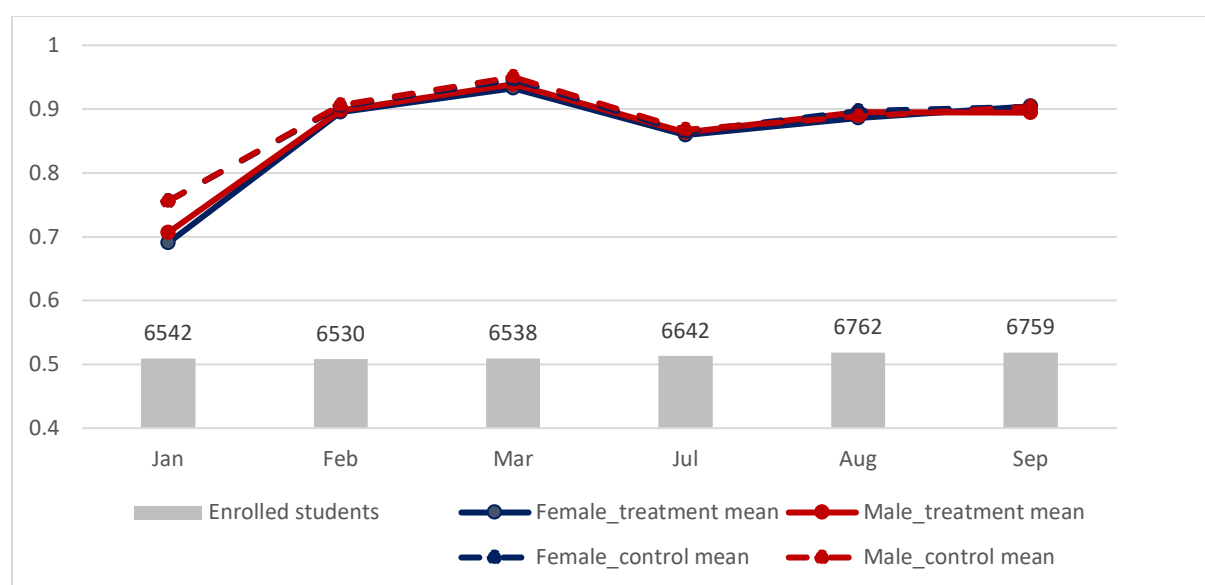


Notes: We calculate attendance as the number of days a student went to school over the number of days s/he was supposed to attend. All attendance rates were truncated at 1 (in some cases attendance rates are higher than one, this happened often during 2019 as schools prepared for the national exams and teachers asked students to come on the weekends). There are no data points for June because students were on vacation most of the time.

²⁵ Note that the number of students increases over time, especially between January and February. This is partly because more students enroll during the year, but also because, especially during the first half of the year, for some classes there were no attendance registers available when data was collected.

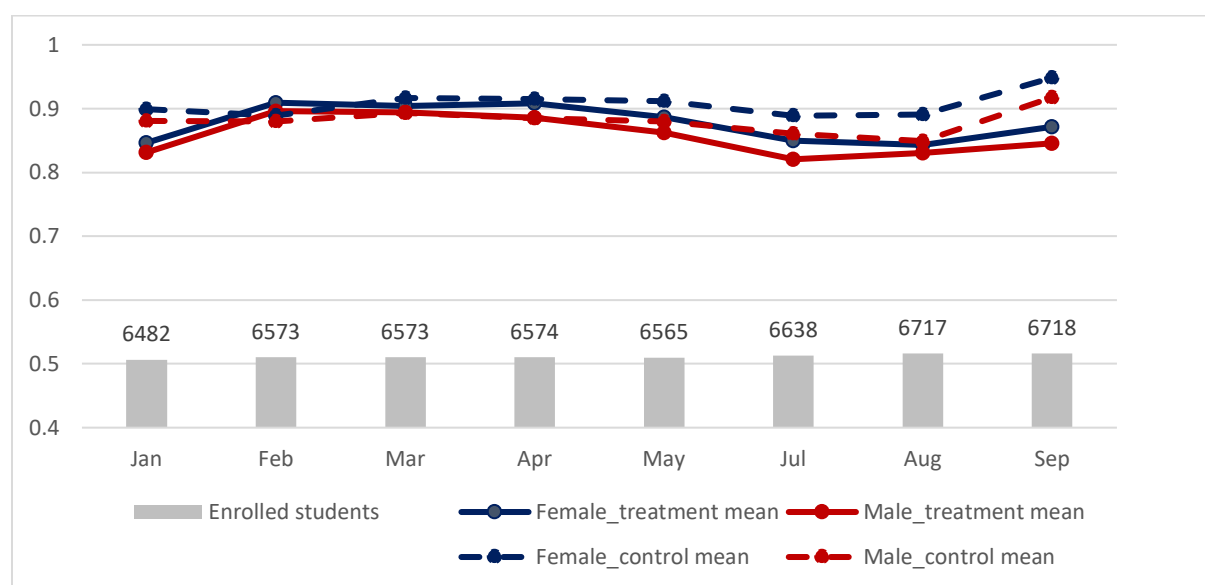
Finally, **Figure 22** and **Figure 23** show monthly attendance rates for 2020 and 2021, respectively. Note that during these years both treatment and control groups received WW interventions, so we do not expect to observe any major differences even if the program increased attendance. Again, there does not seem to be major differences between the two groups. Schools in Tanzania were closed between March 13th and June 29th due to Covid-19, which is why there are no data points for April-May in **Figure 22**. Notably, once schools reopened attendance went quickly back to pre-pandemic levels.

Figure 22: Attendance rates by gender and treatment status – 2020



Notes: We calculate attendance as the number of days a student went to school over the number of days s/he was supposed to attend. All attendance rates were truncated at 1. There are no data points for April-May due to Covid-related closures.

Figure 23: Attendance rates by gender and treatment status - 2021



Notes: We calculate attendance as the number of days a student went to school over the number of days s/he was supposed to attend. All attendance rates were truncated at 1. There are no data points for June because students were on vacation most of the time.

ANNEX VI. SURVEY ATTRITION

The fact that not all girls surveyed at baseline were surveyed at endline could bias the impact evaluation results. To analyze the differences between “panel” girls (those surveyed at baseline and endline) and “non-panel” girls (those surveyed at baseline only), **Table 16:** compares baseline sociodemographic characteristics between non-panel and panel girls. The first three columns show results for girls in the treatment group, specifically non-panel girls in column (1), panel girls in column (2) and the difference in column (3). Columns (4) to (6) do the same for the control group. Panel girls are a bit younger than non-panel girls in both the treatment and control groups. This could be because older girls are more likely to drop out of school, and therefore harder to locate for interviews. It is interesting to note that for most variables the differences between panel and non-panel girls are the same for the treatment and the control groups. Across both experimental groups panel girls have larger families with fewer assets, are more likely to have eaten something the day of the interview, and are less likely to live in homes that are connected to the electric grid than non-panel girls. These differences are small and in most cases not significant. Because most differences between panel and non-panel girls are similar for treatment and control girls, it is not surprising that the differences between treatment and control groups of the panel girls, shown in column (7), are all small and not significant.

Table 16: Characteristics of panel and non-panel girls at baseline

	Treatment			Control			Balance at endline
	Non-panel (1)	Panel (2)	Diff. (3)= (1)-(2)	Non-panel (4)	Panel (5)	Diff. (6)= (4)-(5)	Diff. (7)= (2)-(5)
Age	14.73	14.46	0.27***	14.84	14.58	0.26**	-0.12
Number of household members	6.39	6.52	-0.13	6.30	6.72	-0.42*	-0.20
Did you eat anything today?	0.32	0.36	-0.04	0.29	0.37	-0.07	-0.01
Electric grid	0.32	0.26	0.06	0.35	0.28	0.07	-0.02
Cell phone ownership	0.96	0.95	0.00	0.92	0.97	-0.05	-0.02
Asset index	4.25	4.18	0.07	4.63	4.41	0.23	-0.23
N	208	891		197	924		

Standard errors clustered at the school level in parentheses.

* p<0.05 ** p<0.01 *** p<0.001

We can derive a few important conclusions from these results. First, there are some differences between panel and non-panel girls, implying that attrition was not random. This could affect the external validity of the results. However, it seems that this did not affect the comparability between treatment and comparison groups of the panel sample, as the differences between these two groups, shown in column (7) of **Table 16:**, are small and not significant. This indicates that balance was not compromised by attrition, suggesting that the impact evaluation results are internally valid.

ANNEX VII. ADDITIONAL DESCRIPTIVE STATISTICS AND RESULTS

GIRLS' PROTECTIVE ASSETS

Protective assets can be defined as resources, knowledge, skills, or competencies that girls can employ to overcome an obstacle or challenge for themselves or others. In this context, a protective asset is a store of value related to what a girl can do or be (Population Council, 2015).

WW's POY Clubs offered a life skills curriculum to help girls develop protective assets such as decision-making, self-esteem, self-efficacy, communication skills, and relationship skills/social networks. The POY Clubs' curriculum also included sessions devoted to gender norms and equity issues. Below we discuss the reasons for measuring protective assets and how we measured them. Self-esteem: Positive self-esteem is important for adolescent girls because it helps them build confidence and agency during the challenging developmental stage that is adolescence. During adolescence, girls' self-esteem typically decreases, impacting their decision-making. To measure self-esteem we used the popular Rosenberg Self-Esteem Scale (RSES) (Rosenberg, 1965), which measures positive and negative self-perceptions using ten items. This rubric has been used in several studies; among other features, RSES has been shown to be a predictor of future earnings (Drago 2011).

Self-efficacy: Girls' self-efficacy contributes to girl's empowerment (Bandura, 1997). Self-efficacy is also an important social and emotional learning skill. Self-efficacy is more than just having the confidence but rather the belief that you can achieve something because of that confidence. To measure girls' self-efficacy, we used items from the Generalized Self-efficacy Scale (GSE) (Schwarzer & Jerusalem, 1995).

Decision making: Having decision-making knowledge and competencies helps girls make smart and healthy decisions about school, peer-relationships, and their future. Additionally, when girls are presented with a challenge, they can employ these decision-making skills. To measure decision-making skills, we used items from the Jamaica Youth Survey (Meeks-Gardner, et al, 2011).

Communication skills: Communication and listening skills are critical for girls' social and emotional development. Being able to communicate feelings with peers and adults help girls build social networks, develop healthy relationships and communicate with empathy and assertiveness. Communication skills are an important protective asset because girls are able to employ these skills when they need to express their feelings and show empathy to others. To measure girls' communication skills, we developed items that focus on girls' behaviors and communication skills for interacting with their friends and adults in their lives.

Social networks: Interpersonal relationships and social networks are indispensable in helping girls cope with stressors. These networks can act as sources of social support that protect girls from psychological distress. Learning from others' experiences may elucidate strategies to be employed to support girls during this vulnerable life stage. Girls were asked about their social networks with peers and adults using the Girl Empower (GE) Baseline Survey (Hallman, 2016).

Gender norms, roles & equity: Gender refers to the socially constructed roles and relationships, personality traits, attitudes, behaviors, values, relative power, and influence that society ascribes to sexes on a differential basis (USAID, 2018). Communities and societies create social norms of behavior, values,

and attitudes that are deemed appropriate for boys and girls and the relations between them. Often these gender norms and roles evolve into gender stereotypes and discrimination where girls and women are disproportionately affected because of their gender. Gender equity addresses fundamental social transformation, working with men and boys and women and girls, to bring about changes in attitudes, behaviors, roles and responsibility, and promote gender equity at home, school, in the workplace, and in the community (USAID, 2018). Gender equitable norms and attitudes can positively contribute to creating safe and supportive learning environments for girls. The WW program seeks to foster a friendly and supportive environment for girls in school. We asked girls about their beliefs on gender norms, roles, and equity using the Gender Equitable Index (GEI) (Care, 2014).

To estimate WW's effect on girls' protective assets and gender attitudes, both the baseline and endline surveys included rubrics to measure self-esteem, self-efficacy, decision-making, communication skills, and social networks. These rubrics included between six and eleven items each, using a 4-point Likert scale (either strongly agree/agree/disagree/strongly disagree, or not true/slightly true/true/very true). We summarized these by constructing indexes for each of the measured protective assets. More specifically, items were coded on a 4-point scale ranging from 0 (strongly disagree) to 3 (strongly agree).²⁶

Table 17: shows, for each calculated index, the mean, standard deviation, the observed maximum, and the intra-cluster correlation. We also calculated the Cronbach's alpha, which measures the extent to which items in a specific scale are correlated with each other; therefore, it is considered a measure for a scale's reliability. In general, alphas between 65 and 80 are considered acceptable.

Table 17: Protective assets descriptive statistics at endline

	Mean	Standard deviation	Observed maximum	ICC	Cronbach alpha	N
Self-esteem	21.3	3.0	28	0.04	0.66	1800
Self-efficacy	21.3	4.3	30	0.02	0.72	1812
Communication skills	14.7	3.1	21	0.01	0.64	1807
Decision making skills	19.0	3.6	27	0.05	0.67	1808
Social networks	13.0	2.9	18	0.03	0.67	1811
Gender attitudes	24.9	3.6	33	0.10	0.69	1810

PARENTS' DESCRIPTIVE STATISTICS

Table 18: shows sociodemographic characteristics at baseline of the panel of surveyed parents. Seventy-eight percent of parents are female. On average they are 44 years old. Sixty-nine percent have a primary education and only 12.2 percent have more than primary. On average parents have 2.8 children each, 1.1 male and 1.7 female. The gender imbalance for children is not surprising, given that only parents of girls 10-21 years old were screened for the survey.

²⁶ Some of the self-esteem and gender norms items were reversed coded because they had negative valence. Details on how each score was constructed are available in NORC (2019).

Table 18: Sociodemographic characteristics at baseline of the panel of LIMCA parents

Female (%)	78.4
Age (years)	44.0
Education (%)	
No schooling	4.6
Some primary	14.1
Primary	69.1
More than primary	12.2
Household assets (count)	3.6
Household size (count)	7.1
10-21 years old living in household (count)	
Girls	1.7
Boys	1.1
N	305

TEACHERS' DESCRIPTIVE STATISTICS

Table 19: shows sociodemographic characteristics of the teachers. Between a fourth and a third of teachers are female. Teachers are 34 years old on average, 41 percent have a teaching diploma, 56 percent a bachelor's degree, and only 3 percent an advanced degree. On average, teachers have eight years of experience. Note that, as expected, the differences between treatment and control groups are small and not statistically significant.

Table 19: Teachers' characteristics

	Treatment (1)	Control (2)	Difference (3)=(1)-(2)
Female (fraction)	0.29	0.25	0.05
Age (years)	34.26	34.80	-0.54
Education (fraction)			
Teaching diploma	0.41	0.41	0.00
Bachelor's degree	0.56	0.56	0.00
Advanced degree	0.03	0.03	0.00
Years of experience	7.85	8.33	-0.47
N	292	297	

Standard errors for the significance tests are clustered at the school level.

* p<0.05 ** p<0.01 *** p<0.001

ANNEX VIII. GIRLS' ATTITUDES TOWARDS SRGBV

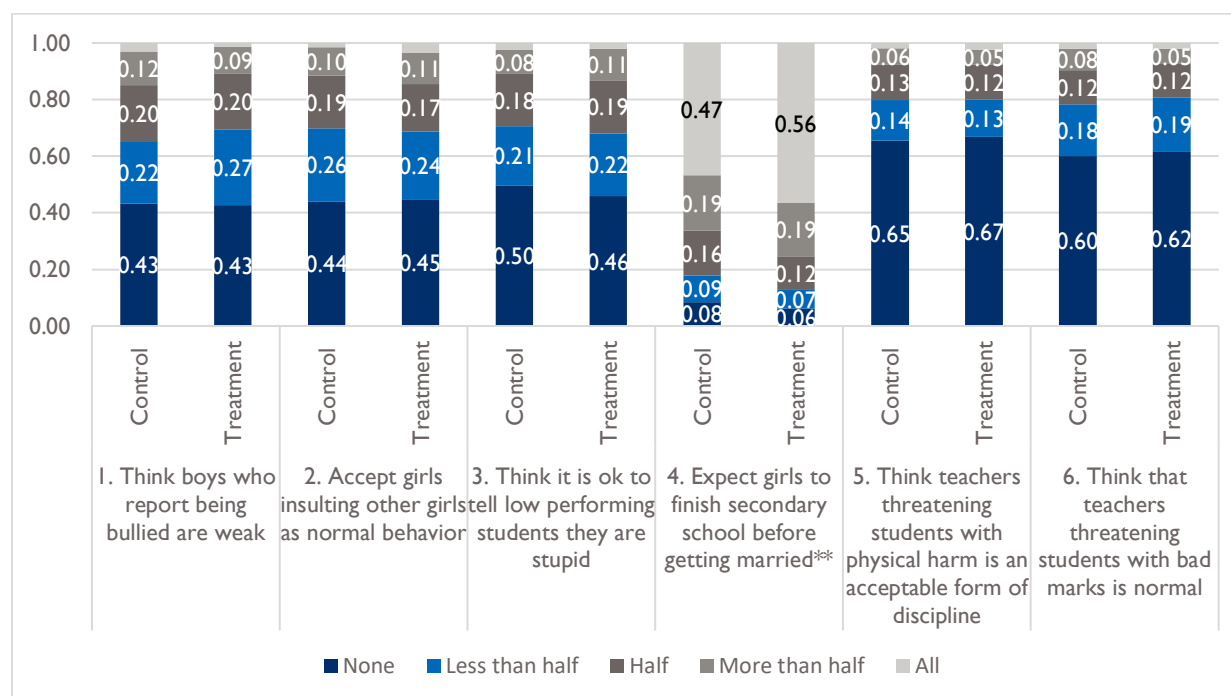
Beliefs on peers' opinions on SRGBV

To record how girls perceive their peers' views on SRGBV, we included questions to girls about their peers' opinions. For example, we ask girls to what extent peers "Think it is okay to tell low performing students they are stupid", and provided girls with a 5-point scale to record their responses, namely none [of their peers think that], less than half, half, more than half, and all.

Figure 24 shows results for items related to emotional violence. The differences between treatment and control are quite small and all but one are not statistically significant. The sole item where the difference between study groups is statistically significant (statistical significance is denoted by stars in the item transcription) is item 4 (Expect girls to finish secondary school before getting married). For this item 47 percent of the girls said all their peers think the statement is true in the control group, while for the treatment group the figure was 56 percent, indicating that the norm that girls finish secondary school before getting married is more pervasively held in treatment than in control schools.

For items on how students treat each other (items 1, 2 and 3) we find that more than half the girls indicate that some or all of their peers think those that are bullied are weak. This suggests there may be room for improvement on gender inequitable norms concerning emotional abuse in the school environment. Notably, girls' normative disapproval of emotional violence seems more pervasive when perpetrated by teachers, as indicated by the results for items 5 and 6. In these cases, about two-thirds of surveyed girls said none of their peers think the statement was true, suggesting that girls think their peers see teacher-perpetrated emotional violence as unacceptable.

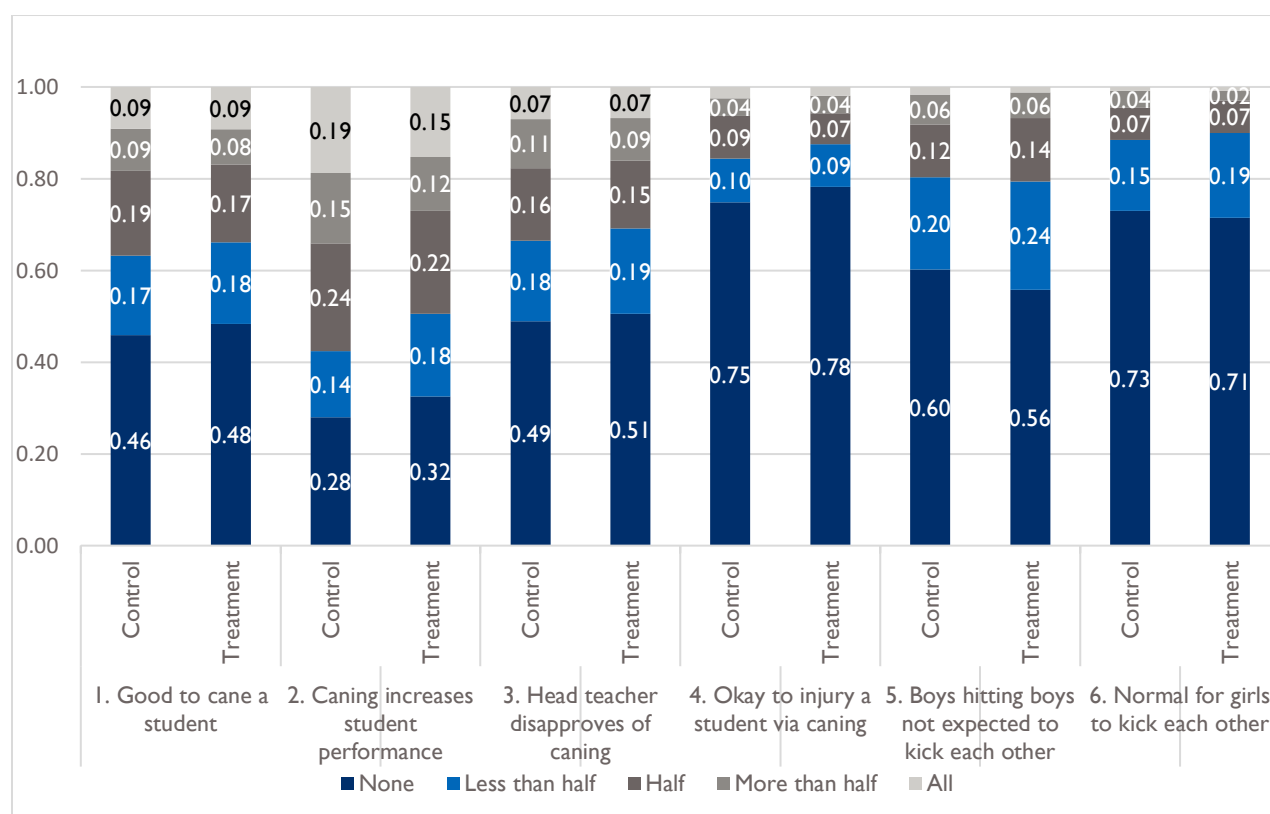
Figure 24: Beliefs on peers' opinions on SRGBV by treatment group – Emotional violence



Note: To estimate whether differences between treatment and control groups are statistically significant, we ran an ordered probit model of the survey responses on the treatment dummy. Stars at the end of each statement (* p<0.05 ** p<0.01 *** p<0.001) indicate the level of significance of the treatment dummy. Standard errors for the significance tests are clustered at the school level.

Figure 25 shows results for items on physical violence. We do not find any statistically significant differences between treatment and control for these items. Girls in both treatment and control schools said their peers, in general, reject corporal punishment at school, as reflected by the results on items 1 and 4. However, the results for the statement “Caning increases student performance” (item 2) indicates a certain level of acceptance that caning increases student performance as only 28 percent of the control group and 32 percent of the treatment group said none of their peers think the statement is true. Furthermore, although the differences between treatment and control groups are small, they are marginally statistically significant (p-value=6.1 percent), suggesting that girls in treatment schools think their peers have a level of rejection to caning as a discipline method greater than their counterparts in the control schools. Regarding intra-student physical violence, the results suggest there is a gender gap in terms of the expected use of violence at schools, while according to item 5 between-boys violence seems to be the norm, intra-girl physical violence is much less expected.

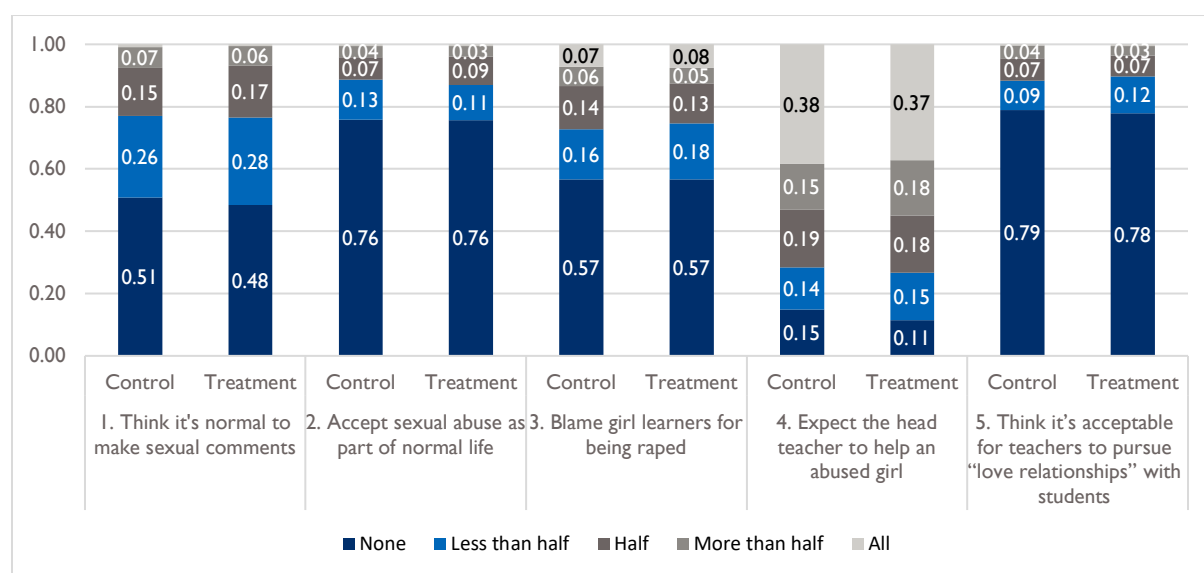
Figure 25: Beliefs on peers’ opinions on SRGBV by treatment group – Physical violence



Note: To estimate whether differences between treatment and control groups are statistically significant, we ran an ordered probit model of the survey responses on the treatment dummy. Stars at the end of each statement (* p<0.05 ** p<0.01 *** p<0.001) indicate the level of significance of the treatment dummy.

Figure 26 shows results for normative beliefs about peer views on sexual violence. We do not find statistically significant differences between treatment and control for any of these items. It is important to highlight that, while the majority of girls think most of their peers reject any form of sexual violence, there are non-negligible fractions who said that less than half or more of their peers think is normal to make sexual comments (item 1), blame girls for being raped (item 3), or that is acceptable for teachers to pursue love relationships with students (item 5). This suggests there is still work to be done regarding gender norms that perpetuate SRGBV in schools.

Figure 26: Beliefs on peers' opinions on SRGBV by treatment group – Sexual violence



LIST EXPERIMENTS

List experiments is a survey technique designed to elicit truthful responses to sensitive issues. In a list experiment respondents are randomly divided into two groups (A/B). Both groups are presented with the same set of three facts or opinions on innocuous topics (e.g. the government should invest more in hospitals). To group A is added a statement on a potentially sensitive topic (e.g. caning is an effective discipline method), while group B is given only the common set of statements. Respondents from both groups are asked to indicate the number of statements they believe to be true. Because respondents do not answer each item but indicate the total number of statements they believe to be true, they are provided an extra level of confidentiality. Both groups should, on average, say they agree with the same total number of innocuous statements, so the mean difference in the total count between groups A and B approximates the fraction of people for which the sensitive item is true.

In the endline surveys we included a few list experiments. In this section we present the main results.

Table 20: shows results for six list experiments on SRGBV included in the girl survey. Panel A shows results for two LE on bullying and emotional violence at school. The first item corresponds to the fraction of girls who think that the statement “Being mistreated by someone stronger or more powerful at school is the learner's own fault” is true. The first row shows the results when the item is asked directly. Twelve percent of girls in the control group said this statement is true, while only seven percent did in the treatment group. This implies a reduction of 5 percentage points due to the program. The results for the LE, in the second row, indicate that 30 percent in both treatment and control group think the statement is true; therefore, when girls felt their responses were more confidential, they were more likely to say the statement was true. This shows the importance of using LE to approximate a more accurate disclosure rate at which girls indicate they think this statement is true. If we asked the question directly only, we would likely underestimate the fraction of girls who think being mistreated at

school is the learner's fault. It also shows that there is not a statistically significant difference between treatment and control groups, regardless of how the question is asked.

The results for most of the other outcomes in

Table 20: are relatively similar. When the question is asked as a list experiment girls disclose less opposition to any form of SRGBV than when the question is asked directly, suggesting the prevalence of social desirability bias when questions are asked directly. However, we do not find statistically significant differences between treatment and control groups, regardless of how the question is asked.

One notable exception is the item 2 in Panel B: “Parents expect teachers to cane children at school if they misbehave”. For this item the results are a little surprising because, for both treatment and control groups, a higher fraction of girls said this was true when asked directly than as part of a list. It is possible that girls are aware parents are expected under community norms to expect teachers to cane students, but privately some girls are also aware that their parents may not agree with this normative expectation. Notably, when asked as part of a list, a higher fraction of girls in the treatment group think the statement is true than in the control group, indicating the program did have an impact on this parental expectation (at least as perceived by the girls).

Table 20: List experiments on SRGBV

	Control (1)	Treatment (2)	Difference (3)=(2)-(1)	Effect size (4)=(3)/(1)
A. Bullying and emotional violence at school				
<i>Being mistreated by someone stronger or more powerful at school is the learner's own fault</i>				
Direct	0.12	0.07	-0.05*	-0.39
LE	0.30	0.30	0.01	0.02
<i>Teachers shouting at a student in front of other students is part of good discipline practice</i>				
Direct	0.13	0.09	-0.04	-0.30
LE	0.36	0.38	0.02	0.06
B. Corporal/harsh punishment and physical violence at school				
<i>Caning is an effective way to discipline students.</i>				
Direct	0.35	0.35	-0.01	-0.02
LE	0.50	0.42	-0.08	-0.16
<i>Parents expect teachers to cane children at school if they misbehave.</i>				
Direct	0.79	0.75	-0.04	-0.05
LE	0.48	0.65	0.17*	0.37
C. Sexual violence at school				
<i>A man teacher has the right to expect sex from a girl if he gives her gifts or favors.</i>				
Direct	0.06	0.05	-0.01	-0.16
LE	0.18	0.15	-0.03	-0.16
<i>Girl students should report sexual abuse at school only if they have serious physical injuries.</i>				
Direct	0.56	0.51	-0.05	-0.09
LE	0.54	0.54	-0.00	0.00

Standard errors for the significance tests are clustered at the school level.

* $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

These results indicate that social desirability affects girl's responses when sensitive items are asked directly. Also, the LE results show there are no major differences on attitudes towards SRGBV between treatment and control groups.

ANNEX IX. LIMCA AS A HEDGING INSTRUMENT

LIMCA savings clubs offered a way for parents to protect themselves from economic shocks through saving and borrowing. This is important for keeping girls (and boys) in school if we consider that parents may offset financial downturns by reducing investments in education (e.g. pulling children out of school).

Table 21: shows the fraction of parents who experienced an economic shock in the last 12 months as well as the type of shock. At baseline 41 percent of parents (134 in total) said they experienced a negative economic shock, while at endline the fraction fell to 27 percent (a total of 83 cases). The table also shows the type of shock suffered. Both at baseline and endline the most predominant were shocks linked to weather irregularities, namely droughts and floods, followed by illness/accident of a household member.

Table 21: Experienced negative shocks and type of shocks (fraction)

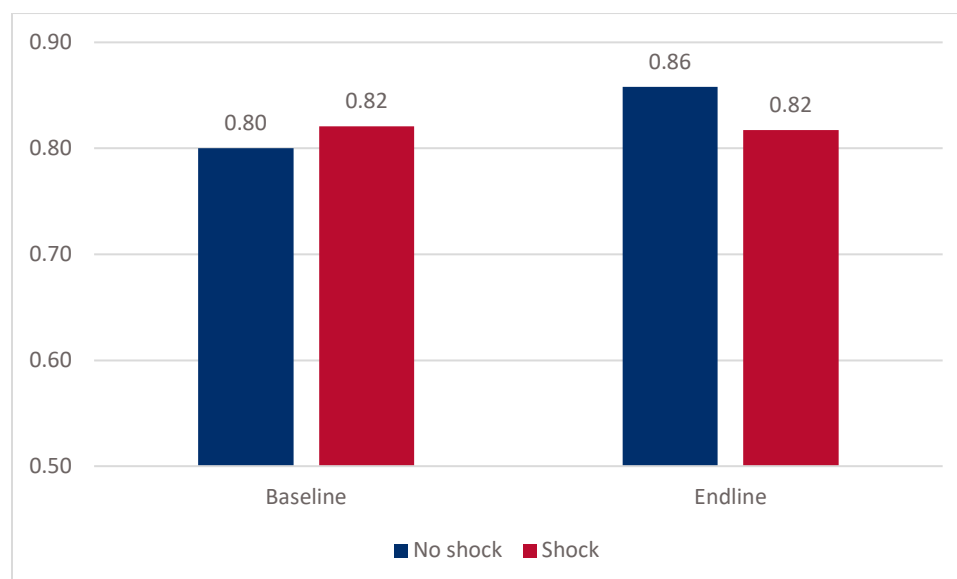
	Baseline	Endline
Experienced at least one shock	0.41	0.27
<i>Type of shock for those that experienced one</i>		
Drought/flood	0.37	0.39
Illness/accident of a household member	0.22	0.24
Reduced income for a household member	0.14	0.02
Pests/livestock diseases	0.13	0.05
High prices/costs	0.02	0.05
Death of a household member	0.02	0.08
Other natural phenomena	0.01	0.13
Other	0.07	0.04

Note: Some households suffered more than one shock over the last 12 months, at baseline 12 households suffered two shocks and 1 suffered 3, while at endline 3 suffered 2 shocks and 1 suffered 3. To avoid double-counting households that experience more than one shock we used the most important shock (according to the respondent).

It is unlikely that the sharp decline in the fraction of parents who experienced an economic shock was due to LIMCA. The program may have allowed participants to engage in economics activities that made them less vulnerable to economic shocks, but it is more likely that economic conditions improved in the region in general between baseline and endline.

For households that experienced a shock, we would expect that they would be more likely to borrow money from LIMCA. Indeed, as **Figure 27** shows, roughly 4 out of 5 participants who experienced a shock received a loan from LIMCA both at baseline and endline. However, among households that did not experience shocks the LIMCA borrowing rate was approximately also 4 out of 5. This indicates that, while having experienced a shock does not predict borrowing money from LIMCA, LIMCA helped households protect themselves from economic shocks.

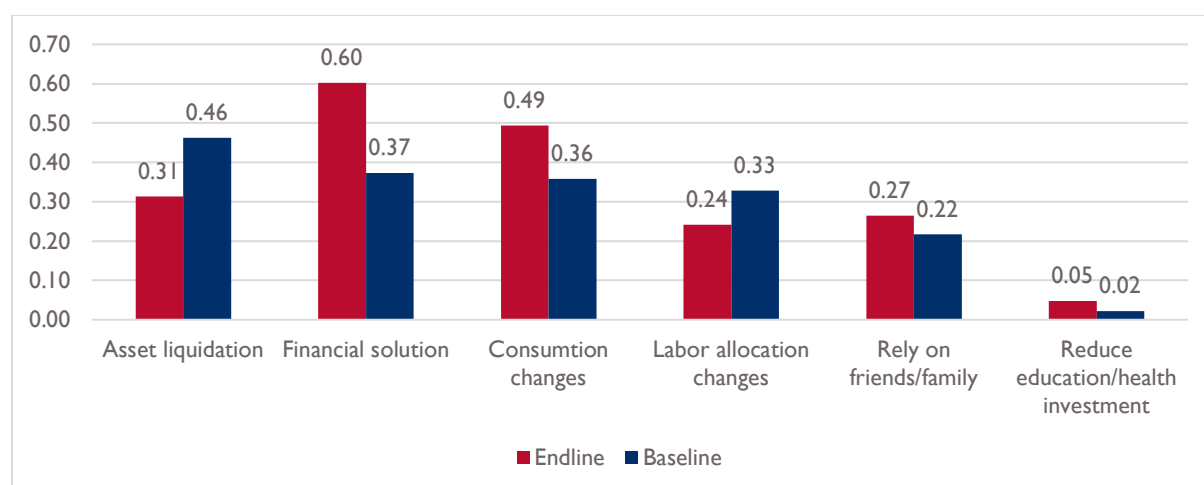
Figure 27: Fraction of households that received a loan from LIMCA by whether they experienced a shock in previous 12 months



For households that did experience a shock it is important to determine whether this impacted their investments in their children’s education, and more broadly, how this shock affected the household well-being. **Figure 28** shows how respondents said they faced shocks experienced the previous year. We classify responses in six main categories, namely asset liquidation (e.g. selling household and productive assets), financial solutions (spending savings or getting a loan), changing consumption patterns (e.g. reducing number of meals or switch to cheaper meals), modifying labor allocations (e.g. increase casual labor), receiving help from relatives, and reducing investments in health or education, including pulling children out of school. The results show that the fraction of households liquidating assets fell between baseline and endline, which is compensated by an increase in the fraction using financial solutions and, to a lesser extent, changes in consumption patterns. Liquidating assets is arguably the most disruptive way to offset a shock, as it affects the long term wellbeing of the household. At the same time, assuming interest rates are not predatory, as was probably the case, since most of these loans were likely with LIMCA, getting a loan is probably the least disruptive way to face a shock.²⁷ Although we found no correlation between receiving a loan and experiencing a shock, these results indicate that households experiencing a shock were more likely to borrow money and/or spend savings (financial solution) than to liquidate assets at the endline, while the opposite pattern was observed at baseline. Notably, the fraction of households reducing investments in education and health were negligible both at baseline and endline.

²⁷ Although we did not record the complete and detailed financial activity of the respondents, 87.5 percent of the respondents that used a loan to face a shock at baseline receive a loan from LIMCA, and this fraction at endline was 94 percent.

Figure 28: Response to shocks (multiple responses allowed)

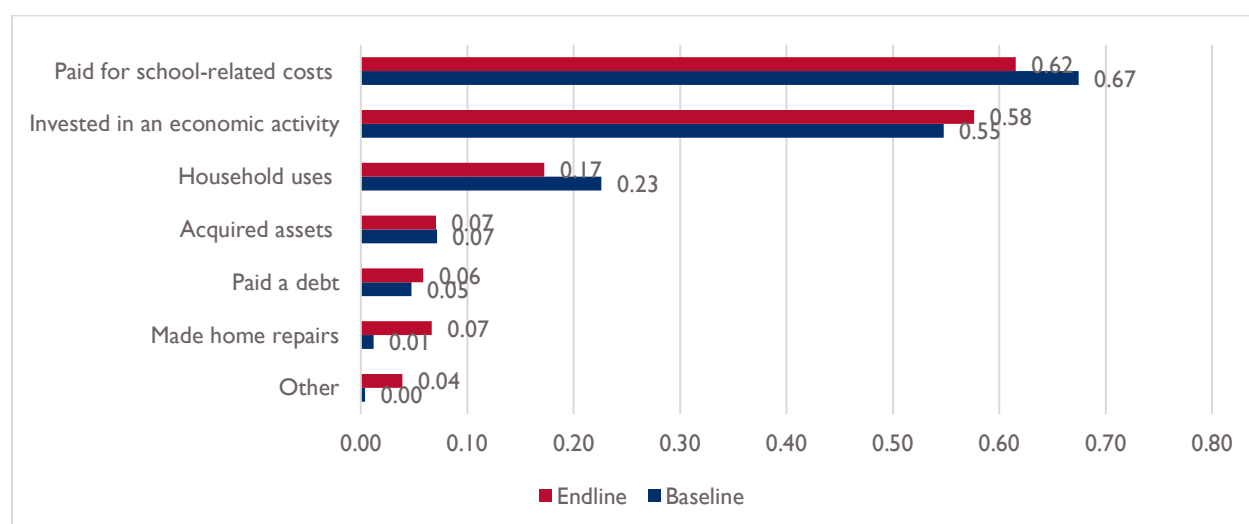


These results suggest that LIMCA loans were not used exclusively to offset economic shocks, but were an alternative to selling household and productive assets for those households that did experience shocks, especially at endline.

If LIMCA was not used only (or even predominantly) as a buffer to economic shocks, it is worth asking how people spend the money from these loans.

Figure 29 shows that roughly two thirds of the respondents said they used the LIMCA loan to pay for school-related costs (i.e., uniforms, books), followed by roughly half saying they invested the money in an economic activity, none of which sound like expenses associated to (unpredictable) shocks. Although the fungibility of money makes it difficult to determine how specific amounts were spent, these results corroborate that LIMCA was not used chiefly to face economic shocks.

Figure 29: LIMCA loan uses (multiple responses were allowed)



The results indicate that households experiencing shocks were not more likely to get a LIMCA loan, and while there is some evidence that LIMCA loans were used to offset financial shocks, LIMCA loans were predominantly used to pay for school-related costs, and to invest in economic activities.

ANNEX X. INTRAHOUSEHOLD INVESTMENTS IN EDUCATION

In this annex we document whether there are differences in enrollment and time children/youth spend studying at home between baseline and endline. For this, we asked parents the school enrollment status of each of the children/youth living with them (10-21 years old), as well as the amount of time children/youth spend studying at home. We collected data on 839 children/youth, 498 female and 341 male.²⁸ The gender imbalance comes primarily from girls 14-16 being overrepresented relative to boys, as **Figure 30** shows.

Figure 30: Number of children/youth living with the respondent by age and gender

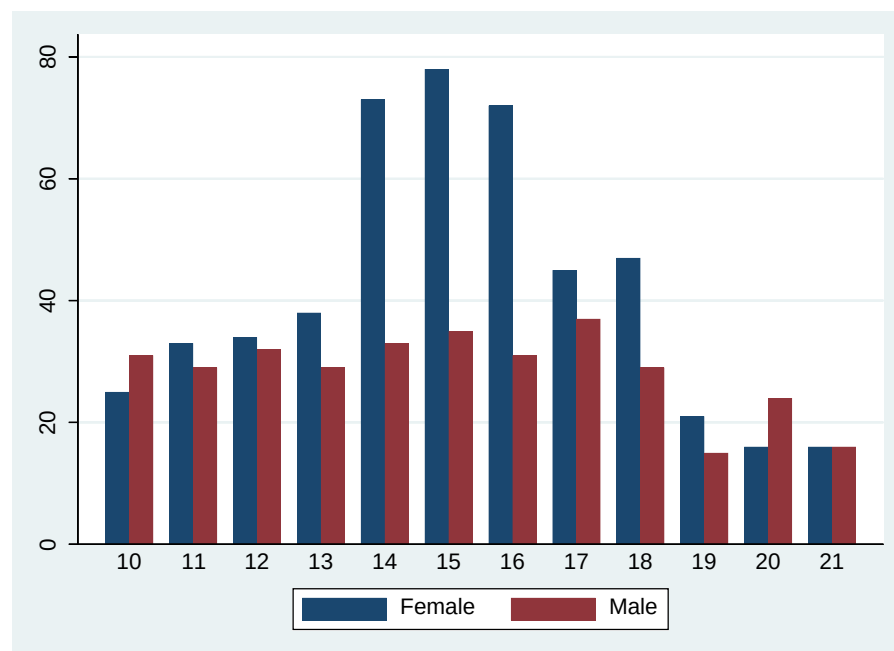
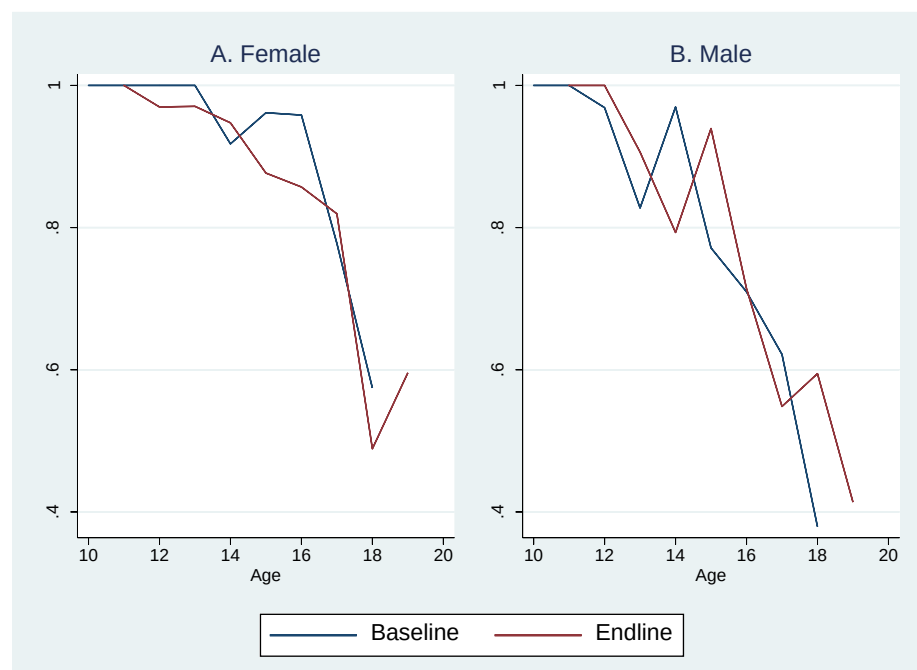


Figure 31 shows enrollment rates by age and wave of data collection for the children/youth living in the households of the surveyed parents. Panel A shows results for females. For girls ages 10-13 enrollment rates are quite high both at baseline and endline, but enrollment rates start to fall for girls 14 and older, and by age 18 only about half of girls are still in school. Note that after 17 there is a pronounced fall in enrollment rates, reflecting the fact that youth start Form I at age 14, on average, and after they finish secondary schools four years later, many stop attending school altogether. **Although enrollment rates are higher at baseline than endline for some ages, the differences are small.** Panel B shows results for boys. In general enrollment is lower for boys than for girls; for example, among boys as young as 16 enrollment rates are already below 80 percent both at baseline and endline. It is important to keep in mind that only parents with at least one girl in schools were eligible to be surveyed, so it was expected that girls are more likely to be enrolled in schools than boys in this sample.

²⁸ Out of the 324 parents surveyed at baseline, 316 (97 percent) were living with at least one girl under their care (usually their daughter) between 10 and 21 years old. The reason why the remaining eight parents were not living with their daughters but were surveyed anyway is that these girls were in boarding school (7) or living elsewhere (1). For these 316 parents we collected data on their children/youth (10-21 years old). At baseline we collected data on a total of 891 children/youth. At the endline we collected data on the same children we asked about at baseline. At the endline we have data for a total of 839 children/youth, 52 fewer than at baseline, which correspond to the children of the 19 parents that were only surveyed at baseline.

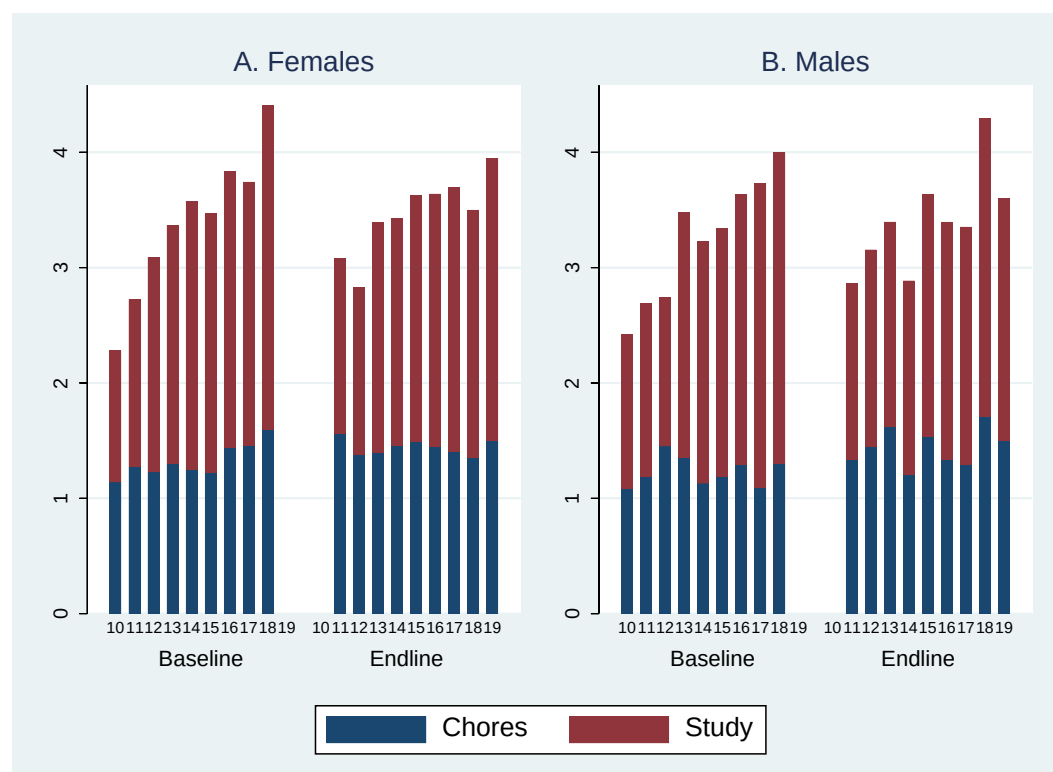
In other words, these results do not necessarily imply that girls are more likely to attend school than boys in the target districts.

Figure 31: School enrollment by age, gender, and wave of data collection



Aside from school enrollment, differences in the opportunities parents grant their children to succeed in school could manifest in how much time they spend studying at home. If girls spend more time on chores than boys, this could translate into lower academic performance and fewer opportunities to succeed in school (Jones 2015). To address this, respondents were asked how many hours a day children/youth spend studying at home, and working on chores, on average. These questions were only asked about children/youth enrolled in school. **Figure 32** summarizes the responses. Panel A shows results for females. At baseline females spend a little over an hour on chores irrespective of their age, while the number of hours they spend studying increases with age, going from just over an hour at age 10, to almost three hours at age 18. Note that the older children/youth are, the more selected the sample is in the sense that the data corresponds only to children/youth still in school. The endline results show the same basic pattern; females spend on average about 1.5 hours on chores regardless of their age, and the older they are the more time they spend studying at home. Panel B shows results for males. Overall the results are very similar to the females'. Males spend between 1 and 1.5 hours on chores regardless of their age, and the older they are they spend more time studying at home.

Figure 32: Average time spend studying at home and on chores by age and gender (hours per day)



The results indicate that parents in this sample already valued girls' education at baseline at least as much as they valued boys' education. Indeed, baseline enrollment rates for females are often higher than for males, and enrolled females seem to have as much time to study at home as enrolled boys, conditioning on age. As was discussed in subsection DI, there are no major differences in retention rates between males and females (either in the treatment or the control group). This could reflect gender equitable views on education among parents in these communities, even before participating in LIMCA. It may also reflect the characteristics of the sample itself. The surveyed parents, all of whom have daughters in schools and all of whom opted to join a LIMCA group, are as a group probably more likely than the average parent in these communities to value girls' education.

In sum, for this sample we find that girls are marginally more likely to be in school than boys both at baseline and endline, and for those in school we do not find major differences in time spent studying, both between boys and girls or between baseline and endline. This suggests that parents treated by LIMCA already valued girls' education, so it is not clear there was much space to improve their decisions in terms of school enrollment, or in giving girls enough time to study at home.

ANNEX XI. DETERMINANTS OF SCHOOL RETENTION

In this annex we use the baseline data to model school retention at endline. The objective is to explore whether the intermediate outcomes WW was trying to affect (e.g. self-esteem or SRGBV environment in school) are predictive of school retention, or if other factors, less malleable to interventions like WW, such as parental education or school characteristics, are in effect the main drivers of school retention.

To conduct this analysis we classify potential predictors of school retention as household-specific variables, girl or personal-specific variables, and school-specific variables. We estimate regression models of school retention and include iteratively household, personal and school variables. Mathematically, we estimate:

$$\Pr(R = 1) = \alpha + \mathbf{h}'\boldsymbol{\beta} + \mathbf{p}'\boldsymbol{\gamma} + \mathbf{s}'\boldsymbol{\delta} + e \quad (1)$$

where $\Lambda[\cdot]$ is the logistic function, R is a binary variable for school retention, $\mathbf{h}$, $\mathbf{p}$ and $\mathbf{s}$ are vectors of household, personal and school characteristics, e is an error term and α , $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\delta}$ are parameters to be estimated.

While it is perhaps natural to assume that most girls not surveyed at endline dropped out of school, it is possible that some of them transferred to other schools. For this reason, we present results for school retention including all girls surveyed at baseline, and assuming all girls not surveyed at endline dropped out of school (so retention $R=0$ if the girls was not surveyed), and also restricting the sample to only those girls surveyed both at baseline and endline, that is, the panel sample. In Error! Reference source not found. we present the results of these regressions and label the models using all girls surveyed at baseline “All”, and the models using panel girls only “Panel”.

In the first two columns we include only household variables as regressors, namely parental education (an indicator variable for whether the parent with more education completed primary school), the household asset index as a proxy for household wealth, and binary variables for whether the household is connected to the electric grid and whether the girl ate something before the interview. Parental education is statistically significant only for the Panel specification and eating anything before the interview is significant only for the All specification. Household assets and being connected to the electric grid do not seem to explain school retention.

In the third and fourth columns we include personal factors. Both age and the number of days girls missed school the previous week (our proxy for school attendance) are negatively correlated with school retention, regardless of whether the regressions use all or panel girls only. To approximate their cognitive skills we included an indicator variable that is 1 if girls scored A or B in the Primary School Leaving Exam (PSLE),²⁹ and 0 if they scored C, the minimum required to go to secondary school. The results show that the PSLE is positively correlated with school retention, but only when we restrict the analysis to panel girls. To approximate girls’ non-cognitive skills we used exploratory factor analysis to extract the first factor of the items in the self-esteem, self-efficacy, and communication and decision making skills rubrics. The factor is constructed in such a way that a higher value implies a higher level of

²⁹ Students in Tanzania are required to pass the PSLE at the end of primary school in order to continue to secondary school. The exam is conducted by the National Examination Council of Tanzania and covers mathematics, English, science, social studies, and Swahili.

these protective assets. As in the case for the PSLE, protective assets are positively correlated with school retention but only when the regression is restricted to panel girls.

In the fifth and sixth columns we include school factors. Time spent traveling to school is negatively correlated with school retention across both specifications. To approximate the gender-based violence school environment we used factor analysis and extract the first factor of the 11 SRGBV items included in the baseline survey,³⁰ the variable is constructed in such a way that higher levels of the variable imply a worse gender-based violence environment in school; the results indicate a negative relationship between SRGBV and school retention, but the parameter is significant only when all girls are included in the regression. Finally, results for schools size indicators show that, while the number of girls is positively correlated with school retention, the number of boys has a negative correlation.

In the seventh and eighth columns we include all the variables as predictors of retention. None of the household variables are significant regardless of whether the regressions include all or the panel sample only. Including household and school variables does not change results for the girl variables in a meaningful way; age and attendance are still significant in both columns (7) and (8), and the PSLE binary variable and the protective assets factor are significant for the panel sample only. Regarding school variables the results are quite different depending on the sample that is used; when all the girls are included in the regression only the school size variables are statistically significant; when only panel girls are considered, only the time to get to school is.

Table 22: Determinants of school retention – Linear probability models

	All (1)	Panel (2)	All (3)	Panel (4)	All (5)	Panel (6)	All (7)	Panel (8)
Household factors								
Parental education (more than primary)	0.013 (0.022)	0.036* (0.017)					0.0063 (0.021)	0.027 (0.015)
Asset ownership index	-0.0026 (0.0072)	0.0002 (0.0055)					-0.0075 (0.0073)	-0.0032 (0.0059)
Connected to the electric grid	-0.0035 (0.023)	0.024 (0.017)					-0.015 (0.024)	0.012 (0.016)
Ate something today	0.052* (0.026)	0.016 (0.019)					0.049 (0.026)	0.024 (0.020)
Personal factors								

³⁰ For the most part these are the same items as those shown in figure 6 where the item-level SRGBV perceptions results for the endline are shown.

	All (1)	Panel (2)	All (3)	Panel (4)	All (5)	Panel (6)	All (7)	Panel (8)
Age			- 0.060** * (0.0096)	- 0.036** * (0.0087)			- 0.063** * (0.0099)	- 0.033** * (0.0093)
Number of days missed school last week			- 0.057** * (0.014)	- 0.032** (0.011)			- 0.055** * (0.014)	-0.028* (0.011)
A/B PSLE score (161-200/121-160)			0.027 (0.018)	0.042** (0.013)			0.032 (0.018)	0.041** (0.014)
Protective assets factor			0.013 (0.0094)	0.025** (0.0082)			0.011 (0.011)	0.025** (0.0095)
School factors								
Time to get to school (log)					-0.030* (0.012)	- 0.037** * (0.0086)	-0.016 (0.013)	- 0.033** * (0.0092)
SRGBV factor					-0.026* (0.012)	-0.014 (0.0082)	-0.016 (0.013)	-0.0052 (0.0090)
School size								
Number of males in Form I (log-2018)					-0.11** (0.036)	-0.065 (0.033)	-0.11** (0.041)	-0.065 (0.037)
Number of females in Form I (log-2018)					0.11*** (0.029)	0.060 (0.032)	0.092** (0.034)	0.050 (0.034)
Treatment	-0.018 (0.024)	-0.023 (0.018)	-0.043 (0.024)	-0.033 (0.017)	-0.022 (0.023)	-0.013 (0.015)	-0.032 (0.024)	-0.028 (0.018)
N	2056	1684	2104	1723	2200	1800	1932	1583
Adjusted R2	0.0094	0.032	0.038	0.053	0.017	0.036	0.043	0.060
F-tests								
Household covariates	0.37	0.15					0.31	0.34
Girl covariates			0.000	0.000			0.000	0.000

	All (1)	Panel (2)	All (3)	Panel (4)	All (5)	Panel (6)	All (7)	Panel (8)
School covariates					0.001	0.001	0.047	0.0034

Note: All models include district fixed effects. Standard errors clustered at the school level in parentheses.

* p<0.05 ** p<0.01 *** p<0.001

This analysis has several key limitations. First, most if not all of the variables included in it are in jointly determined by the household, the girl, and the school. For example, the amount of time girls spend from home to school is as much a school variable as it is a household one. Moreover, a case could be made that protective assets are jointly determined by the household, the girls, and the school. However, the proposed classification provides a way to organize retention determinants that allows us to analyze what segment of the girls' ecology affords greater opportunities for potential channels to impact school retention. Second, we are restricted to the variables recorded in the survey, so other household, girl or school level variables we did not observe may explain retention. Third, and most importantly, we cannot claim that these relationships are causal, as other unobserved variables could be explaining both the covariates we are including and school retention.

Despite these limitations, a few implications of this analysis are worth highlighting. First, household variables do not seem to explain school retention in this sample. Parental education and whether girls ate something the day of the interview are significant contingent on the specification, and no household variable is significant once girl and school variables are included. This does not necessarily mean that household characteristics do not matter for retention. Again, we are using the variables that we have, and it could well be that other household variables not included in the survey have explanatory power on retention. However, a perhaps more plausible explanation for these null results is that households in the sample are already a selected subsample of households in Tanzania. Therefore, variation in parental education or household wealth observed in the sample is perhaps irrelevant for school retention, as households are homogenous in other unobservable factors that explain how these girls were able to attend secondary school at all.

Second, age is a strong predictor of school retention, meaning that older girls are more likely to dropout than younger girls; and school attendance is also a strong predictor of attendance. This is information readily available to school administrators, so it should be considered to predict dropping out in the context of the Dropout Early Warning Systems being developed by schools. Also, it is interesting that the coefficients for cognitive skills (PSLE score) and non-cognitive skills (protective assets factor) are significant for the panel sample (columns (4) and (8)), but not when the regression uses all the girls. A possible explanation for this is that cognitive and non-cognitive skills do explain school retention, but girls with higher levels of these skills transferred to other schools, so in the retention model where all girls are included the correlation between cognitive and non-cognitive skills and retention is diluted because the girls with high levels of cognitive and non-cognitive skills that we do not observe transferred to other schools. In any case, these results do not allow us to discard the possibility that protective assets are correlated with school retention. This suggests it is unlikely that the reason WW does not impact retention is because protective assets do not correlate with retention.

Third, although the SRGBV factor is significant for only one specification, it is interesting that it has explanatory power on school retention, and suggests that, just as protective assets, reducing gender-based violence in schools could increase school retention.

ANNEX XII. STATEMENT OF WORK



STATEMENT OF WORK OF The *Waache Wasome* Evaluation in Tanzania

July 18, 2017

I. PURPOSE OF THE EVALUATION

Waache Wasome is supported by USAID's Let Girls Learn Challenge Funds awarded by the Office of Gender Equality and Women's Empowerment (E3/GenDev) to World Education, Inc. *Waache Wasome* is implemented by World Education Inc. through Cooperative Agreement AID-OAA-A-17-00001 and managed by USAID/Tanzania. An Impact Evaluation with a Performance Evaluation component (IE+) will be funded by GenDev and carried out by NORC, through the Reading and Access Evaluations contract (Time and Material):GS-10F-0033M/AID-OAA-M-1300010 managed by the Office of Education (E3/ED). A GenDev team member will serve as the Activity Manager of this evaluation.

The *Waache Wasome* evaluation objectives are:

- Evaluate and report on the impact of the package of interventions on school attendance, promotion, retention, and wellbeing of girls in secondary schools benefiting from activities; and
- Assess and report on activity performance and the intermediate results of individual interventions in the context of girls' needs addressed by *Waache Wasome* and the *Waache Wasome* implementation experience. The interventions to be assessed, performance criteria relating to effectiveness and sustainability and the intermediate outcomes will be determined in consultation with USAID and WEI/B.

The evaluation's findings, conclusions and recommendations will enable key audiences, such as, USAID, World Education Inc., the Government of Tanzania (GoT), and the donor community to understand the impact of the *Waache Wasome* program model and package of interventions in the context of the needs of adolescent girls in Tanzania and the *Waache Wasome* implementation experience. The data generated by this evaluation will contribute toward building the growing body of evidence on programming for adolescent girls. Information generated by this evaluation can be applied to adapt, replicate and/or scale-up the *Waache Wasome* model or its individual interventions.

2. SUMMARY INFORMATION

Strategy, Project or Activity Name	Waache Wasome in Tanzania
Implementer	World Education Inc. Bantwana Initiative (WEI/B)
Cooperative Agreement/Contract #	AID-OAA-A-17-00001
Total Estimated Ceiling of the Evaluated Project/Activity (TEC)	\$ 10.5 million
Life of Strategy, Project, or Activity	December 15, 2016 to September 30, 2021
Active Geographic Regions	Waache Wasome will be implemented in two rural districts (Karatu DC and Musoma DC) and two periurban districts (Tarime TC and Musoma MC).
Development Objective(s) (DOs)	USAID/Tanzania CDCS DO 1: Development Objective 1: Tanzanian Women and Youth Empowered
USAID Office	USAID/Tanzania in partnership with E3/Office of Gender Equality and Women's Empowerment (GenDev)

3. BACKGROUND

3A. DESCRIPTION OF THE PROBLEM, DEVELOPMENT HYPOTHESIS, AND THEORY OF CHANGE

In Tanzania, gendered beliefs and practices in the home, school, and community environments combine to negatively affect girls' school enrollment and retention, as well as to limit girls' aspirations, attitudes, agency, and potential to learn. *Waache Wasome* is designed as a comprehensive model to address the full range of gendered beliefs and practices at the home, school, and community that negatively affect girls' enrollment and retention. *Waache Wasome* is rooted in the principle of putting girls at the center of creating their own agency while simultaneously addressing the intersecting social norms and economic barriers that exist within families, schools, and communities.

Waache Wasome Logic Model¹

Purpose: Increase enrolment retention of adolescent girls in secondary school and influence negative perceptions about the value of educating girls	
Cross Cutting Areas: Media Human centered design Gender transformative approaches Teen Advisory Group	Critical assumptions: Commitment of the GoT to implement relevant laws and national policies pertaining to girls' education (National Plan of Action to End Violence Against Women and Children; Circular No 5 relating to the Education and Training Policy of 2014; Law of the Child Act of 2009) does not diminish. GoT and LGAs continue to collaborate with WEI/Bantwana in program implementation Tanzania economy does not undergo any unforeseen shocks Line Ministries remain committed to enhancing and scaling up school based adolescent clubs

Sub-Purpose	Program Models	Outputs	Outcomes	Impact ²
1: Build the agency, knowledge, and protective assets of girls in secondary school	Protect our Youth (POY) Clubs	#POY club tools adapted (POY facilitator guide, participant guides, training modules, etc...); # POY peer educators trained; # school officials and teachers oriented to support POY clubs; # POY clubs established; # students enrolled in POY clubs; # students participating regularly in POY clubs;	• Protective assets of adolescent girls enhanced; • Aspirational attitudes of girls increased; • Household resilience to economic shocks enforced; • Commitment of mothers and caregivers to invest in girls' education strengthened; • Parenting skills of mothers and caregivers of adolescents improved; • Knowledge and attitudes on GBV improved; • Gender equitable attitudes and practices of school officials, teachers, students and parents strengthened; • School related GBV (bullying, corporal punishment, and sexual abuse) reduced; • Adolescents' knowledge of reproductive health improved; • Adolescent access to reproductive health services increased; • Unwanted pregnancies among secondary school girls decreased; • Improved parenting skills of pregnant adolescents and teen mothers; • More girls engaged in STEM	Attendance of adolescent girls improved; Retention of adolescent girls in secondary school improved; Enrollment and retention in alternative education pathways of adolescent girls who dropped out of secondary school due to unwanted pregnancy increased; A shift in perceptions about the value of educating adolescent girls
	Community Mentor Program (for in school girls)	# school officials and teachers oriented to support community mentor program; # community mentors trained; # adolescent girls participating in community mentor program		
2: Increase family commitment and capacity to invest in girls' education	Livelihood Improvement for Mothers and Caregivers of Adolescents (LIMCA)	# economic empowerment volunteers (EEVs) recruited and trained; # LIMCA groups formed; # mothers and caregivers of adolescents joined LIMCA groups; \$\$ cumulative savings and loans issued to LIMCA members;		
	Parenting Skills for Mothers and Caregivers of Adolescents	# caregivers of adolescent girls completing a parenting skills curriculum		
3: Foster a girl friendly and supportive school environment	GBV Training Peer Study Groups/ Subject Clubs	# teachers and school officials trained on in-school mechanisms for GBV prevention and response; # peer tutors trained to facilitate peer study groups; # adolescents participating in peer study groups/ Subject Clubs;		
	Design Squad/ Invent it Build it ICT Coding Camps for Girls	# adolescents participating in Design Squad/ Invent It Build It; # adolescents participating in ICT Coding Camps		

Sub-Purpose	Program Models	Outputs	Outcomes	Impact ²
4: Increase access to youth friendly adolescent reproductive health services	Adolescent Reproductive Health Education and Services	# adolescent girls participating in sports and recreation activities that aim to increase access to youth friendly reproductive health services; # adolescent girls accessing reproductive health services through sports and recreation activities; media coverage for sports and recreation activities		
5: Provide alternative education pathways for girls who have dropped out of secondary school due to pregnancy and/or early marriage	Alternative Education Pathway	# out of school adolescent girls and teen mothers enrolled and retained in Out of School Study Groups (OSSG) program; # adolescent girls and teen mothers accessing wraparound services through OSSG;		
	Early Childhood Stimulation	# volunteers trained to facilitate early childhood stimulation program; # pregnant girls, young mothers and caregivers participating in early childhood stimulation		

¹ This is the most recently updated Logic Model presented in the Waache Wasome Draft Monitoring, Evaluation and Learning Plan, June 8, 2017 prepared by WEI/B.

² Progression has not been included in the logic model because of new government regulations stipulating that secondary students must automatically progress, regardless of performance, from Form 1 to Form 2, and from Form 3 to Form 4. Learning outcomes are not been included in this logic model because this program does not address quality of education issues. WEI/Bantwana will track student learning outcomes; however, because the project was not designed to address quality of education issues, this is not included in the logic model.

3B. SUMMARY STRATEGY/PROJECT/ACTIVITY/INTERVENTION TO BE EVALUATED

3B.1. THE WHOLE-OF-GIRL APPROACH:

In March 2015, the U.S. Government launched Let Girls Learn (LGL), a global initiative which builds upon existing efforts and invests in new programs to expand education for adolescent girls. LGL seeks to ensure that adolescent girls around the world are able to access quality education enabling them to reach their full potential. LGL activities adopt a Whole-of-Girl approach by recognizing that multiple solutions are needed to address the complex challenges faced by girls. Barriers to girls' education pose formidable challenges.

The Whole-of-Girl approach is based on the recognition that discriminatory gender norms motivate some parents to invest in educating their sons rather than their daughters. Girls attending school risk harassment and violence on their way to school and in school. Girls drop out of school owing to pregnancy and early marriage. In some contexts, parents and teachers tend to support boys rather than

girls to make headway in their studies and careers. Whole-of-Girl interventions foster an enabling environment for adolescent girls' education, safety and well-being, wherein girls increase self-efficacy and are valued by their families, communities and other institutions.

3B.2. WAACHE WASOME:

Waache Wasome ('Let them Learn' in Kiswahili) is designed and implemented by the Bantwana Initiative of World Education (WEI/B). Its intent is to improve enrollment and retention of adolescent girls. *Waache Wasome* (WW) is adopting a holistic approach shaped by evidence and designed for replication and scale up. WW will serve adolescent girls by creating links between education, protection, health, and social welfare sectors. WEI/B's partners include regional, local and technical government agencies.

The overarching goal of WW is to increase retention of adolescent girls in school and reduce negative perceptions about the value of educating girls. WW has five objectives:

1. Build the agency, knowledge, and protective assets of girls in secondary school;
2. Increase family commitment and capacity to invest in girls' education;
3. Foster a girl friendly and supportive school environment;
4. Increase access to youth friendly adolescent reproductive health services; and,
5. Provide alternative education pathways for girls who have dropped out of secondary school due to pregnancy and/or early marriage.

WW targets girls aged 13-19 who: 1) have completed primary school and are at risk of or have failed to transition to secondary school; 2) have enrolled in lower secondary school and are at risk of dropping out; and 3) who have dropped out of secondary school because of pregnancy or early marriage.

WW is being implemented in two rural districts (Karatu DC and Musoma DC) and two peri-urban districts (Tarime TC and Musoma MC). These districts are likely to afford high program impact, due to the following factors: 1) situated in regions identified as hotspots for girls' vulnerability and HIV risk; 2) low secondary enrollment rates; 3) presence of other USG programs; and, 4) presence of trained village ParaSocial Workers or National Child Protection System operations.

4. EVALUATION QUESTIONS

Two sets of evaluation questions have been drafted to facilitate the IE+ approach adopted by the WW evaluation covering the IE and the PE component. The evaluation team will finalize the sets of evaluation questions in consultation with USAID and WEI/B before the evaluation begins based on the WW implementation situation. The IE will evaluate the impact of the full WW package of interventions on girl beneficiaries in secondary schools, beneficiary teachers and parents. The PE component will enable assessment of activity performance and intermediate outcomes relating to three distinctive WW interventions.

The evaluation will answer the following IE and PE topline questions. The evaluation is required to address – but not be limited to – the sub-questions/topics listed in bullets under each topline question.

4A. ILLUSTRATIVE IMPACT EVALUATION QUESTIONS

IE.Q.1: What impact does WW have on secondary school retention, attendance and promotion of girls?

- What impact has WW on school retention, attendance and promotion of the most vulnerable girls?
- What is the impact of the Protect Our Youth (POY) clubs on school retention, attendance and promotion of girls?

IE.Q.2: What impact does WW have on reducing pregnancy and early marriage among girl beneficiaries?

IE.Q.3: Has WW strengthened girls' protective assets?

- Has WW increased girls' self-efficacy?
- Has WW changed students' (girls and boys) attitudes regarding harmful gender practices?
- Has WW changed attitudes on SRGBV?
- Has WW improved (a) girls' knowledge of reproductive health; and (b) girls' access to reproductive health services; and
- Has WW reduced the likelihood of girls engaging in risky sexual behavior?

IE.Q.4: Has WW changed the gender norms of beneficiary parents?

- Has WW changed parenting skills affecting girls?
- Has WW improved attitudes toward gender equitable access to education within the household?

IE.Q.5: Has WW changed the gender norms of beneficiary teachers?

- Has WW changed the perceptions and attitudes of beneficiary school teachers toward girls?
- Has WW changed beneficiary school teachers' knowledge, attitudes and practices (to be determined as feasible) relating to SRGBV?

4B. ILLUSTRATIVE PERFORMANCE EVALUATION QUESTIONS:

PE.Q.1 Are girls and other stakeholders benefiting from POY clubs?

- What are the main POY (Protect Our Youth) Club challenges? Do POY Club implementation challenges vary by district? Is WW resolving POY Club implementation challenges or not?
- Are members participating actively in POY Clubs? Are members and teachers satisfied with POY Club activities? What benefits are perceived by participants?
- Is the POY Club curriculum user-friendly? What lessons/components of the POY curriculum are most useful to adolescents?
- How is WW engaging Local Government Authorities (LGAs) in POY clubs?
- What kind of supportive supervision do patron/matrons and peer educators of the POY clubs receive? Do they perceive this supportive supervision contributing to the results of POY clubs?

- Are POY Clubs having any unintended (positive or negative) consequences?
- Do POY clubs influence the general school environment? If yes, what types of influence are exerted by the POY Clubs? If no, why do POY Clubs lack broader influence?
- What should be done to make this WW intervention more effective?

PE.Q.2: Is LIMCA benefiting LIMCA members, girls and other stakeholders as intended by WW?

- What are the benefits of participating in LIMCA (Livelihoods Improvement for Mothers and Caregivers of Adolescents)? Is LIMCA helping to reduce the socio-economic vulnerability of participating households? Why or why not?
- Are parents and caregivers investing in girls' education as a result of LIMCA participation? How many parents of adolescent girls are being supported by LIMCA Groups?
- What kind of supportive supervision do EEVs (Economic Empowerment Volunteer) receive from local NGO partners? Do EEVs perceive this supportive supervision as contributing to the results of the LIMCA groups?
- How is WW engaging LGAs in LIMCA?
- What should be done to make this WW intervention more effective?

PE.Q.3: What are the benefits of the Gender Based Violence (GBV) prevention and response teacher training to teachers, students and other stakeholders?

- Are teachers satisfied with the GBV training? Why or why not?
- Have teachers' knowledge and attitudes toward GBV changed? Are there corresponding changes in their behavior and classroom practices?
- Which elements of the following training components are (i) easiest, and (ii) most difficult for teachers to implement after receiving the training? Why?
 - a) The GBV prevention and response training component;
 - b) The gender responsive pedagogy training component; and (c) The alternative forms of discipline training component.
- What kinds of support do teachers get to implement what they have learned?
- Are teachers satisfied with the support they get?
- How are LGAs (specifically G&C Desk Police Officers) supporting GBV training for teachers?
- How are stakeholders (PTAs, school boards, etc.) supporting GBV training for teachers?
- What should be done to make this WW intervention more effective?

PE.Q.4: Are the interventions addressed in PE.Q1-3 sustainable?

- What actions are being taken to continue the interventions addressed in PE.Q1-3 after WW ends?
- Are these interventions likely to continue after WW ends? Why or why not?

5. EVALUATION DESIGN AND METHODOLOGY

The evaluation team is tasked with designing and implementing an impact evaluation

(IE) to evaluate the impact of the full WW package on key outcomes described in the IE evaluation questions (see SOW section 4A). The team will use either an experimental or a quasi-experimental design for the IE. In addition, the WW evaluation in Tanzania should also design a performance evaluation component to assess three individual WW interventions by answering the PE questions (see SOW section 4B). The three interventions are:

- Protect Our Youth (POY) Clubs in secondary schools;
- LIMCA activities for parents and caregivers; and
- GBV training for school teachers.

Overall, this IE+ design should be developed in consultation with USAID and WEI/B. In particular, the evaluation team must work closely with WEI/B to cause minimum disruption to WW activity implementation when selecting and administering treatment group(s) and control/comparison group(s). Control and treatment groups should be selected in advance from the list supplied by WEI/B in consultation LGAs.

To answer the IE+ evaluation questions, the evaluation team will adopt a mixed methods approach to collecting, analyzing and reporting quantitative and qualitative data, including document analysis, surveys, focus group discussions and key informant interviews as appropriate and feasible. When finalizing the evaluation methodology, the evaluation team will consider adopting if appropriate, resources on data collection and analysis recommended by USAID, such as, the Conceptual Framework for Measuring School-Related Gender-Based Violence (SRGBV) produced by USAID's Africa Bureau.

In alignment with the requirements and standards in Section C: Scope of Work of the Reading and Access contract, the evaluation team will undertake and complete the following tasks, including:

- Finalize the IE and PE evaluation questions drafted to-date in consultation with USAID and WEI/B;
- Prepare the Evaluation Design Report and the corresponding evaluation work-plan and detailed evaluation budget using the templates recommended by the E3/Education R&A Contracting Officer's Representative (COR). The Evaluation Design Report should:
 - o Describe the evaluation methodology to collect and analyze quantitative and qualitative data, and disaggregate data by sex, age and other relevant variables;
 - o Provide information on the sampling frame, power calculations, control and treatment groups;
 - o Discuss internal and external validity and how generalizable the results will be;
 - o Identify the strengths and limitations of the proposed methodology as well as evaluation implementation challenges and proposed solutions.
 - o Discuss the approach to optimizing cost-efficiency;
 - o Describe the plan to leverage local knowledge to benefit the evaluation team;
 - o Describe the data collection tools for the surveys, focus group discussions and key informant interviews;

- Use the evaluation report template(s) recommended by the R&A COR. □ Prepare the data collection tools;
- Recruit and deploy an evaluation implementation team with the recommended qualifications, including Tanzanian nationals;
- Implement the IE+ evaluation in consultation with USAID and WEI/B;
- Report the findings, conclusions and findings of the IE+ evaluation in the context of the needs of adolescent girls in Tanzania and the WW implementation experience; and
- Implement the activities to learn from the IE+ evaluation's findings, conclusions and recommendations during the baseline, midline and end-line stages.

6. DELIVERABLES AND REPORTING REQUIREMENTS

6A. DELIVERABLES

With reference to the requirements for deliverables in the Reading and Access Evaluations contract, the evaluation team will deliver the following products:

- An Evaluation Design Report with a work-plan and detailed budget for the LGL WW IE+ evaluation with an in-house presentation on the evaluation design to USAID and WEI/B using audio-visual media to connect audiences;
- The set of data collection tools;
- Briefings for USAID/Tanzania and WEI/B during data collection visits in Tanzania;
- In 2018, a draft baseline report, the final baseline report and a presentation to USAID and WEI/B, using audio-visual media to connect audiences;
- In 2020, a draft mid-line evaluation report (covering IE and PE components) and the final mid-line evaluation report with an in-house presentation to USAID and WEI/B, using audio-visual media to connect audiences;
- In 2021, a draft final evaluation report, the final evaluation report and a 1-2 page summary Briefing Note (covering IE and PE components) with an inhouse presentation to USAID and WEI/B, using audio-visual media to connect audiences;
- Hold a webinar on the evaluation's findings, conclusions and recommendation for Agency-wide stakeholder audiences and post the webinar recording on Agency web sites recommended by USAID.

6B. REPORTING

The evaluation reports (mid-line and end-line) will discuss the IE+ findings, conclusions and recommendations in the context of the needs of girls in Tanzania and the WW implementation experience to-date. Each evaluation report (baseline, midline and end-line) should contain the following components:

- Executive Summary;
- Introduction (including evaluation purpose, audience, anticipated uses, and questions);

- Program Background (overview of the needs of adolescent girls in Tanzania, and the WW programming approach and activities);
- Methodology (overview of data collection and analysis methods, including limitations, anticipated challenges and solutions);
- Part 1: Findings, conclusions and recommendations for each IE topline question in the context of relevant needs of adolescent girls in Tanzania and the overall implementation experience across the full package of interventions and in Part 2 (in the mid-line and end-line evaluation reports): the findings, conclusions and recommendations for each PE topline question in the context of the needs of girls in Tanzania and the implementation experience of each individual WW intervention addressed; and Annexes (including evaluation SOW, references, list of respondents, and data collection instruments).

The WW evaluation reports will meet the evaluation quality criteria described in the USAID Evaluation Policy (see text box below):

USAID EVALUATION POLICY, APPENDIX I

CRITERIA TO ENSURE THE QUALITY OF THE EVALUATION REPORT

- The evaluation report should represent a thoughtful, well-researched and well organized effort to objectively evaluate what worked in the project, what did not and why.
- Evaluation reports shall address all evaluation questions included in the scope of work.
- The evaluation report should include the scope of work as an annex. All modifications to the scope of work, whether in technical requirements, evaluation questions, evaluation team composition, methodology or timeline need to be agreed upon in writing by the technical officer.
- Evaluation methodology shall be explained in detail and all tools used in conducting the evaluation such as questionnaires, checklists, and discussion guides will be included in an Annex in the final report.
- Evaluation findings will assess outcomes and impact on males and females.
- Limitations to the evaluation shall be disclosed in the report, with particular attention to the limitations associated with the evaluation methodology (selection bias, recall bias, unobservable differences between comparator groups, etc.).
- Evaluation findings should be presented as analyzed facts, evidence and data and not based on anecdotes, hearsay or the compilation of people's opinions. Findings should be specific, concise and supported by strong quantitative or qualitative evidence.
- Sources of information need to be properly identified and listed in an annex.
- Recommendations need to be supported by a specific set of findings.
- Recommendations should be action-oriented, practical, and specific, with defined responsibility for the action.

7. EVALUATION TEAM COMPOSITION

Within the Labor Categories described in the Reading and Access Evaluations contract, NORC will select an evaluation team that includes a Principal Investigator, statistician(s), gender in education specialist(s), administrative associate(s), and incountry data collectors. In alignment with the requirements and standards in Section C: Scope of Work of the Reading and Access Evaluations contract, the LGL WW evaluation team is tasked with designing and implementing an impact evaluation with a performance evaluation component developed in consultation with USAID and WEI/B.

The WW evaluation team will collectively reflect the following qualifications scaled appropriately to senior, mid-level and junior levels:

- Graduate degree(s) in the social sciences;
- Expertise in applying a variety of methods to gather and analyze quantitative and qualitative data, as well as designing and implementing impact evaluations and performance evaluations;
- Knowledge of USAID’s performance monitoring and evaluation guidance; □ International development programming experience;
- Demonstrated experience in gender equality and women’s empowerment research and/or programming;
- Specialist knowledge on the needs of adolescent girls and education systems in Tanzania;
- Experience in managing projects/activities funded by USAID;
- Specialist expertise in working with databases, and software to collect and analyze data and to disseminate findings;
- Data collection, data analysis, report writing and communication skills, including languages spoken in the WW implementation areas; and
- Excellent ability to work collaboratively with evaluation team members, USAID, WEI/B and its partners.

The Evaluation Design Report will propose the staffing plan for this evaluation, including, specific positions and CVs for proposed individuals to serve in the evaluation team positions. Each evaluation team member will sign USAID’s conflict of interest statement before conducting any documentary or field research. To leverage local knowledge, the contractor is encouraged to use host-country evaluation specialists with the requisite qualifications.

8. EVALUATION SCHEDULE

Tasks and Deliverables	Timeline	Illustrative Dates
1. Submit draft Evaluation Design Report to Activity Manager (AM)	20 business days following receipt of USAID approval of the evaluation Statement of Work	July 31, 2017
2. Presentation on Evaluation Design Report to USAID, and WEI/B	Within 10 business days following AM’s feedback on the draft Evaluation Design Report	August 24, 2017

Tasks and Deliverables	Timeline	Illustrative Dates
3. Submit Final Evaluation Design Report, incorporating USAID's feedback on the draft	10 business days after receiving written feedback from all USAID stakeholders on the draft Evaluation Design Report	September 14, 2017
4. Submit draft data collection tools for review and incorporate stakeholder feedback	October-November 2017	October 16 - November 15, 2017
5. Brief USAID/Tanzania and WEI/B, collect baseline data and draft baseline report	January-February, 2018	January 8 – March 15, 2018
6. Submit Draft Baseline Report to AM	40 business days following completion of data collection activities	March 31, 2018
7. Present baseline evaluation findings and recommendations to USAID audiences and WEI/B stakeholders	Within 10 business days after submission of Draft Evaluation Report	April 10, 2018
8. Submit Final Baseline Report to AM	Within 15 business days following receipt of final feedback from all reviewers	April 30, 2018
9. Brief USAID/Tanzania, and WEI/B, collect midline data and start drafting midline report	November – December 2019	November 4 2019 – December 31, 2020
10. Submit Draft Midline Evaluation Report to AM	45 business days following completion of data collection activities	February 25, 2020
11. Present midline evaluation findings and recommendations to USAID audiences and WEI/B stakeholders	Within 10 business days after submission of Draft Evaluation Report	February, 28, 2020
12. Submit Final Midline Report to AM	Within 15 business days following receipt of final feedback from all reviewers	March 31, 2020
13. Brief USAID/Tanzania, and WEI/B, collect end-line data and start drafting end-line report	May, 2021	May 31, 2021
14. Submit Draft End-Line Report to AM	45 business days following completion of data collection activities	August 10, 2021
15. Present end-line evaluation findings and recommendations to USAID audiences and WEI/B stakeholders	Within 10 business days after submission of Draft Evaluation Report	August 24, 2021
16. Submit Final End-Line Report and Briefing Note to AM	Within 15 business days following receipt of final feedback from all reviewers	September 15, 2021

Tasks and Deliverables	Timeline	Illustrative Dates
17. Provide evaluation debriefing to AM and E3/Education Office Evidence Team	Within 10 business days following the submission of the Final Evaluation Report	September 22, 2021
18 Hold and record webinar on evaluation	Within 5 business days following the submission of the Final Evaluation Report	September 24, 2021

9. OTHER REQUIREMENTS

All quantitative data collected by the evaluation team must be provided in machine-readable, non-proprietary formats as required by USAID's Open Data policy (see ADS 579). The data should be organized and fully documented for use by those not fully familiar with the project or the evaluation. USAID will retain ownership of the survey and all datasets developed.

All modifications to the required elements of the SOW of the contract/agreement, technical requirements, evaluation questions, evaluation team composition, methodology, or timeline, need to be agreed upon in writing by the COR. Any revisions should be updated in the SOW that is included as an annex to the Evaluation Report.

U.S. Agency for International Development

1300 Pennsylvania Avenue, NW

Washington, DC 20523