



USAID READING AND ACCESS IMPACT AND PERFORMANCE EVALUATION OF THE WAACHE WASOME PROJECT (TANZANIA) BASELINE REPORT

Prepared under Contract No.: GC-I0F-0033M/AID-OAA-M-13-00010, Tasking N015.03

This publication was produced at the request of the United States Agency for International Development. It was prepared independently by NORC at the University of Chicago. 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.

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

Baseline Report

(September 2019)

Prepared under Contract No.: GC-I0F-0033M/AID-OAA-M-I3-00010, Tasking N015.03

Submitted to:

USAID

Submitted by:

Alejandro Ome, Alicia Menendez, Ilse Paniagua, and Varuni Dayaratna

Contractor:

NORC at the University of Chicago
4350 East West Highway, 8th Floor
Bethesda, MD 20814
Attention: Varuni Dayaratna
Tel: 301-634-9414; E-mail: Dayaratna-varuni@norc.org

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.....	iv
EXECUTIVE SUMMARY	v
CONTEXT AND PROJECT BACKGROUND	v
EVALUATION PURPOSE & EVALUATION QUESTIONS	vi
Impact evaluation.....	vi
Performance evaluation.....	vi
EVALUATION METHODS	vii
Impact evaluation.....	vii
Performance evaluation.....	viii
BASELINE DATA COLLECTION	viii
FINDINGS AND CONCLUSIONS.....	viii
A. INTRODUCTION	i
A1. CONTEXT.....	i
A2. PROJECT BACKGROUND.....	i
Evidence	4
B. EVALUATION PURPOSE & EVALUATION QUESTIONS.....	5
IMPACT EVALUATION.....	6
PERFORMANCE EVALUATION.....	6
C. EVALUATION METHODS.....	8
C1. IMPACT EVALUATION.....	8
Impact on girls.....	8
Parents	10
Teachers	10
C2. PERFORMANCE EVALUATION	10
C3. BASELINE DATA COLLECTION.....	11
Instrument design	11
Fieldwork and results	12
D. FINDINGS AND CONCLUSIONS	12
D1. Socio-demographic characteristics	12
D2. Outcome variables – Protective assets and gender norms	17
Descriptive analysis	17
Determinants of protective assets and gender norms	26

Balance of protective assets and gender norms.....	30
D3. Outcome Variables – SRGBV.....	30
D4. Conclusions	33
REFERENCES.....	36
ANNEX I. INSTRUMENTS	39
ANNEX II. DATA COLLECTION	68
ANNEX III. BALANCE BETWEEN RANDOMLY SELECTED GIRLS FOR POY CLUBS AND CONTROLS	72
ANNEX IV. DISTRIBUTIONS OF PROTECTIVE ASSET INDICES AND GENDER NORMS.....	73
ANNEX V. ANALYSIS ON GENDER NORMS ITEMS.....	75
ANNEX VI. BIVARIATE REGRESSIONS	77

LIST OF TABLES

Table 1: Descriptive statistics for Tanzania (mainland + Zanzibar), and the regions of Arusha and Mara (%)	2
Table 2: POY Club membership in 30 treatment schools.....	12
Table 3: Sociodemographic characteristics	13
Table 4: Education level of the parents (%).....	14
Table 5: School-related variables	16
Table 6: Primary School Leaving Exam grade (self-reported).....	16
Table 7: Expected maximum level of education.....	17
Table 8: Self-esteem index.....	19
Table 9: Self-efficacy index.....	21
Table 10: Decision making index	22
Table 11: Communication skills index.....	23
Table 12: Social networks index	24
Table 13: Attitudes towards gender norms, roles & equity index	26
Table 14: Bivariate regressions between protective assets and selected socio-demographic characteristics (N=2220†).....	27

Table 15: Bivariate regressions between selected sociodemographic characteristics (N=2220†)	28
Table 16: Bivariate regressions between protective assets and school-related variables (N=2220†).....	28
Table 17: Attitudes towards SRGBV.....	32
Table 18. Baseline data collection results	71
Table 19. Balance between POY Club-randomly selected girls and girls in the control group.....	72
Table 20. Factor loadings for the 11 items on gender norms.....	75
Table 21. Context for gender norms questions.....	75
Table 22: Bivariate regressions between protective assets.....	77

LIST OF FIGURES

Figure 1: Theory of Change.....	4
Figure 2: Disability prevalence - Self-reported (%).....	15
Figure 3: Self-esteem (%).....	19
Figure 4. Self-efficacy (%).....	20
Figure 5: Decision making (%)	21
Figure 6: Communication skills (%)	23
Figure 7. Social networks (%)	24
Figure 8. Attitudes towards gender norms, roles & equity (%).....	25
Figure 9. Correlations between self-esteem score and other protective assets scores	29
Figure 10. Mean indices for protective assets for girls in control schools, girls in treatment schools and ARD girls	30
Figure 11. Perceptions of SRGBV	31
Figure 12. Protective indices distributions.....	74

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 Demography 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 is designed and implemented by the Bantwana Initiative of World Education (WEI/B). Under the Reading and Access contract, NORC is conducting an evaluation of *Waache Wasome*, adopting a distinctive impact and performance evaluation (IE+) methodology.

Waache Wasome (WW) takes a holistic approach to try and keep girls in school, working directly with them, as well as within their school and home environments. Specifically, WW seeks to increase school retention by advancing three specific 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, and 3) Foster a girl friendly and supportive school environment.

To achieve each program objective, the following activities will be conducted:

- 1. Build the agency, knowledge, and protective assets of girls in secondary school.** The program implements Protect Our Youth (POY) clubs, providing safe spaces in schools where adolescent girls can 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), will be assembled as part of the program activities. LIMCA will also be 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 will conduct two activities: (1) training on school-related gender-based violence (SRGBV) prevention and response for teachers and school administrators, and (2) subject clubs.

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 are more likely to succeed and stay in school.

The program will be implemented in a total of 68 public schools, including all 60 of the public schools in Karatu DC, Musoma DC, and Musoma MC; and eight schools in Tarime DC. In 2017, the first year of WW implementation, WEI implemented activities in six schools in the district of Karatu DC. These six schools served as a pilot of the intervention to test and fine tune different program components. Starting in February 2018, the program expanded to 30 additional schools. Two years later, in February 2020, the remaining 32 schools will receive the WW interventions.

EVALUATION PURPOSE & EVALUATION QUESTIONS

The purpose of this evaluation is to estimate the impact of the program on girls' school retention, among other outcomes of interest. NORC will also conduct a performance evaluation of some of the program components. More specifically, the overarching objectives of this evaluation are to:

Objective 1: Evaluate the impact of the program on school retention and on multiple intermediate outcomes, including 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 school-related gender-based violence (SRGBV).

Objective 2: Assess the WW performance relating to three of the program components along the lines discussed in this report.

Objective 3: Draw upon impact evaluation (IE) and performance evaluation (PE) data meaningfully to present findings, conclusions and recommendations related to program components that what worked/did not work; reasons for high/low impacts; obstacles and drivers to improving impact on key outcomes of interest; etc. Develop recommendations for fine-tuning the program to improve impact and implementation and ensure sustainability.

The impact and performance evaluation questions that will be addressed are shown below.

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 LGL WW package changing gender norms (girls and boys)?
- IE5. Does the program change attitudes on SRGBV?
- IE6. Does the program change parenting skills?
- IE7. Does the program improve attitudes towards gender equitable access to education within the household?
- 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 the SRGBV prevention and response teacher training to teachers, students, and other stakeholders?
- PE4: Are the interventions addressed in PE1-3 sustainable?

EVALUATION METHODS

WW is being implemented in the districts of Karatu DC, Musoma DC, Musoma MC, and Tarime DC. Of the 68 schools targeted to benefit from WW, the evaluation excludes one control school in Tarime DC and the six Karatu DC schools that were part of the pilot.¹

IMPACT EVALUATION

To evaluate the impact of the program, NORC is conducting a randomized experiment. Between December 2017 and January 2018, 30 schools were randomly assigned to the ‘treatment’ group (start treatment in February 2018) and 31 schools were assigned to the control group (start treatment in February 2020). Below we describe how the program will evaluate the impact on girls, parents and teachers.

To evaluate the impact of the program on retention and other outcomes of interest, we will compare mean outcomes of interest between treatment and control schools at endline (February 2020).

Two types of data will be collected to estimate the impact of the program on girls’ outcomes: survey data and school administrative data. Survey data will be used to document program impacts on retention rates and intermediate outcomes, namely girls’ protective assets, gender attitudes, and perceptions on SRGBV.

Survey data will also be used to estimate the impact of the program for girls At Risk of Dropping out (ARD). ARD girls are classified as such by their teachers when POY clubs are assembled, and can choose to enroll in the clubs.

School administrative data will be used to evaluate the program on school attendance.

To document research questions related to parenting skills and parents’ attitudes towards gender equitable access to education, we will 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 will focus on a sample of parents that signed up for LIMCA. Specifically, we will survey them after savings groups are constituted, so we have a baseline for LIMCA parents, and survey these same parents again at endline. No comparison group will be surveyed, so the analysis can be described as a before-and-after evaluation.

Finally, teachers in each of the treatment and control schools will be surveyed at endline (February 2020). In this survey we will record teachers’ gender attitudes, teachers’ knowledge, attitudes, and

¹ Originally WEI had planned to treat Tarime TC instead of Tarime DC. Tarime TC was replaced by Tarime DC because WEI learned that there were already programs similar to WW in Tarime TC. However, while Tarime TC has only seven schools, Tarime DC has 30. To accommodate the program budget, WEI selected eight of the 30 schools in Tarime DC as replacements for the Tarime TC schools, based on their geographic proximity to Tarime TC. NORC had budgeted to survey all 7 schools in Tarime TC. When 8 schools were included from Tarime DC as a replacement for the 7 schools from Tarime TC, NORC had to drop (randomly) one of the extra control schools that had been added.

practices on SRGBV. Using this data we will compare teachers in treatment and control schools to estimate the impact of the program on teachers.

PERFORMANCE EVALUATION

The performance evaluation of WW will assess the performance of three interventions: POY clubs, LIMCA activities, and the SRGBV training for school teachers/administrators. The evaluation will assess implementation, intermediate outcomes, sustainability, and scalability of each intervention. For this, in February 2020, NORC will collect and analyze data from FGDs with girls, FGDs with boys, FGDs with parents, and KIs with teachers, principals and other stakeholders.

As a follow-up for the performance evaluation, in April 2021 we will randomly select six schools and conduct FGDs with girls that were in Form I in 2018 and who are still in school in April 2021, and FGDs with girls that dropped out.

BASELINE DATA COLLECTION

Baseline survey data was collected between February 12 and March 12, 2018. A total of 2,220 Form I girls were surveyed. Of these, 189 girls were replacements, for a replacement rate of 8.5 percent. The target was to survey 2,318 (61x38) girls, so the completion rate was 96 percent. The target was not reached because, in some schools, there were fewer than 38 girls.

Using data collected by WEI on POY Club membership, NORC identified 700 girls in the survey data as POY Club members.

FINDINGS AND CONCLUSIONS

The baseline survey included questions on girls' socio-demographic characteristics, items used to measure girls' protective assets and gender norms, and items used to measure girls' perceptions and attitudes towards SRGBV.

In terms of sociodemographic characteristics, we found that girls come from relatively disadvantaged backgrounds. Only a third of the girls had had breakfast on the day we conducted the interview, and only 28 percent of girls live in households connected to the electric grid.

We calculate protective asset indices to measure latent traits that we expect to be correlated with school retention. The main conclusions of the data collected on protective assets are as follows:

- Responses reflect an average or above average level of protective assets, in particular for self-esteem and self-efficacy. For other protective asset indices it is more difficult to establish comparisons with the existing literature because we adapted scales for brevity and our specific project context. However, for the most part we found that girls are thinking about the consequences of their decisions, they have listening and communication competencies, and positive social networks with their friends and other individuals in their lives.
- For gender norms, we found that girls believe that there should be gender equity between males and females, especially when we asked items contextualized in school.

- Most girls reject any form of SRGBV, as recorded by their attitudes to different scenarios we proposed in the survey.

In terms of balance between treatment and control schools, for the majority of outcomes we did not find significant differences between these two groups. Not surprisingly, we did find that ARD girls come from more disadvantaged backgrounds than average girls in the treatment or the control group. As explained in the Evaluation Design Report (EDR), for the impact evaluation we will use matching methods to construct a valid comparison group for ARD girls.

A. INTRODUCTION

AI. CONTEXT

Education in Tanzania follows a 1-7-4-2 structure, with 1 year of preschool, 7 years of primary education, 4 years of lower secondary education and 2 years of higher secondary education. Gender gaps in lower secondary school completion are still observed in Tanzania, despite the fact that primary school completion rates are actually higher for girls than for boys. In 2015, lower secondary completion rate for males was 31.1 percent, but for females was 26.7 percent; while for primary school the figures are 75.1 for males and 83.7 percent for females.²

As in other developing countries, violence against children is a pervasive problem in Tanzania. A national survey conducted in 2009 on females and males 13 to 24 years of age, showed that roughly 30 percent of females reported having experienced sexual violence before the age of 18, while the figure for males was 13.4 percent (UNICEF 2011). In addition, experiencing physical violence during childhood was reported by three fourths of both females and males. 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. Moreover, 15 percent of the females that reported sexual violence said at least one incident happened in school (UNICEF 2011).

Furthermore, girls in Tanzania face several challenges during adolescence that prevent them to stay in school. These factors include early marriage, teenage childbearing and gender disparities in cultural norms. Data from the Tanzania Demography and Health Survey (TDHS) of 2015-2016 show that 3.4 percent of women age 15-19 were married at age 15, and for women age 20-24, 30.5 percent were married at age 18. Teenage childbearing is relatively common in Tanzania. Twenty-seven percent of women age 15-19 have begun childbearing (Ministry of Health, Community Development, Gender, Elderly and Children et al. 2016). Teenage pregnancy caused close to 3,700 girls to drop out of school in 2016 in Tanzania (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 girls' lives in multiple ways. Girls with more education tend to have fewer children and delay marriage, they also observe better outcomes in the labor market, and know how to take better care of themselves and their children. Observational analyses show that the costs of not educating girls in the developing world include lower future earnings for these girls and higher stunting rates for their children (Wodon et al. 2018).

A2. PROJECT BACKGROUND³

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 is designed and implemented by the Bantwana Initiative of World Education (WEI/B). Under the Reading and Access contract, NORC is conducting an evaluation of *Waache Wasome*, adopting a distinctive impact and

² Adapted from UNESCO Institute for Statistics (UIS), <http://data.uis.unesco.org>, extracted September 20, 2018.

³ A detailed description of the project is available in the EDR and references cited there.

performance evaluation (IE+) methodologies. EDI Limited is the local firm contracted by NORC to conduct data collection.

Waache Wasome (WW) takes a holistic approach to try and keep girls in school, working directly with them, as well as within their school and home environments. Specifically, WW seeks to increase school retention by advancing three specific objectives:⁴

1. Build the agency, knowledge, and protective assets of girls in secondary school;
2. Increase family commitment and capacity to invest in girls' education; and
3. Foster a girl friendly and supportive school environment.

The target areas of the program are the district of Karatu DC, one of the seven districts in the region of Arusha, and the districts of Musoma DC, Musoma MC, and Tarime DC, located in the Mara region. The program will be implemented in a total of 68 public schools, including all 60 of the public schools in Karatu DC, Musoma DC, and Musoma MC; and eight schools in Tarime DC. World Education, Inc. (WEI) and USAID selected these districts based on data on violence against children and other conditions that make them good candidates for a program like WW to have a substantial impact, and because of WEI's presence in Arusha.⁵

Data from the TDHS 2015-2016 show that, in effect, the target districts are located in areas of high risk for women, in particular in the case of the districts in the Mara region. For example, Table 1 shows that women in Mara are less likely to attend secondary school, more likely to have begun childbearing as teenagers, and more likely to have experienced physical and sexual violence than women in Tanzania (mainland + Zanzibar).

Table 1: Descriptive statistics for Tanzania (mainland + Zanzibar), and the regions of Arusha and Mara (%)

	Tanzania	Arusha	Mara
Attending secondary school (women age 14-17)	23.8	30.1	22.1
Have begun childbearing (women age 15-19)	26.7	14.9	37.4
Have experienced physical violence since age 15 (women age 15-49)	39.5	34.1	61.2
Have experienced sexual violence in past 12 months (women age 15-49)	8.7	4.4	14.2

Source: Ministry of Health, Community Development, Gender, Elderly and Children et al. 2016.

In 2017, the first year of WW implementation, WEI implemented activities in six schools in the district of Karatu DC. These six schools served as a pilot of the intervention to test and fine tune different

⁴ 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.

⁵ The Population Council created an index of 9 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).

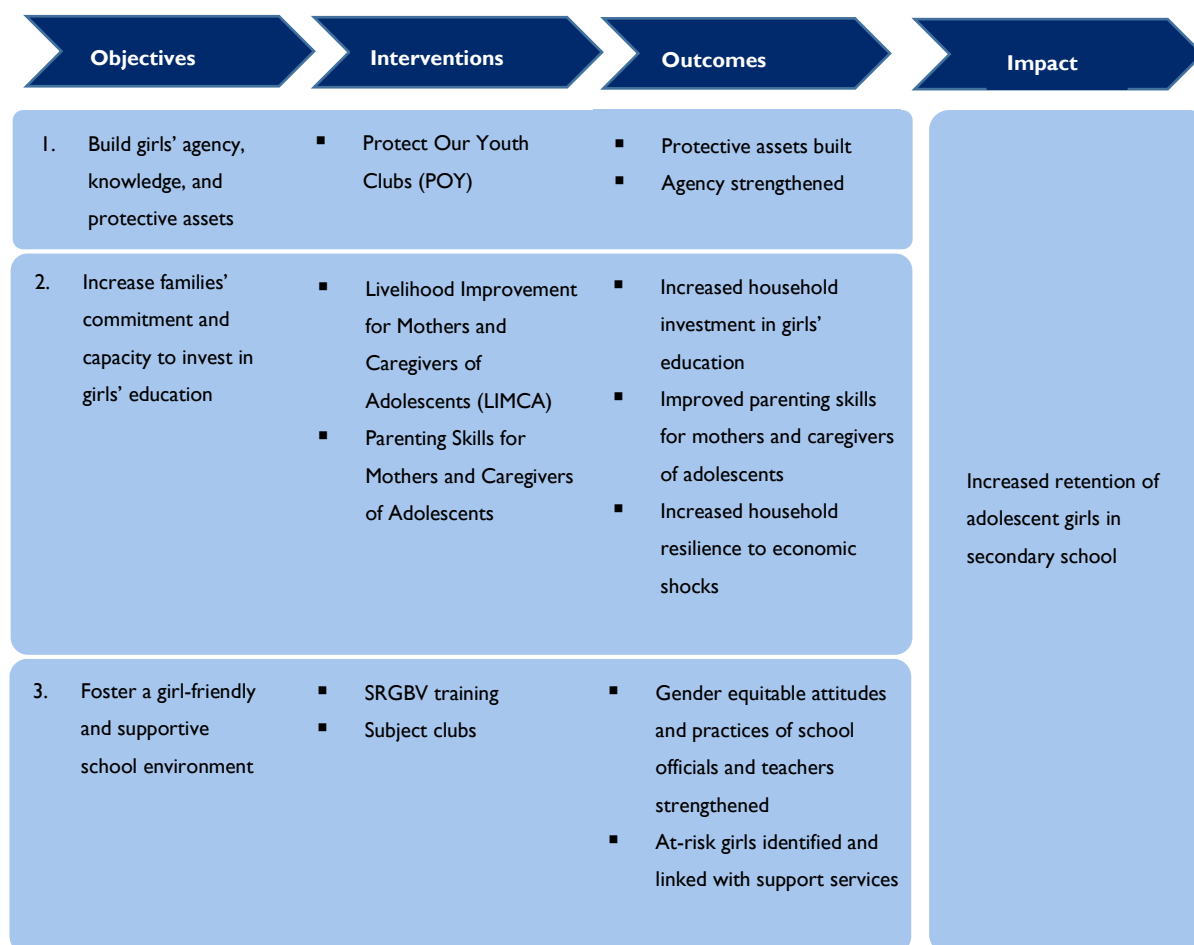
program components. Starting in February 2018, the program expanded to 30 additional schools. Two years later, in February 2020, the remaining 32 schools will receive the WW interventions.

To achieve each program objective, the following activities will be conducted:

1. **Build the agency, knowledge, and protective assets of girls in secondary school.** The program implements Protect Our Youth (POY) clubs, providing safe spaces in schools where adolescent girls can learn to navigate the risks they face and build protective assets that reduce vulnerabilities, helping them to remain and succeed in school.
At the beginning of each school year between 2018 and 2021, WEI will establish four POY Clubs - two in Form 1 and two in Form 3 - in each school. Each POY Club will have 40 participants: 10 ARD girls selected by teachers, 15 girls randomly selected, and 15 boys randomly selected. Club sessions will be 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 will invite mothers and caregivers of adolescents to join local savings groups called Livelihood Improvement for Mothers and Caregivers of Adolescents (LIMCA). LIMCA groups of 20-30 members will be established at the village level and be supported by trained Economic Empowerment Volunteers (EEVs). In addition, LIMCA will be 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 will conduct two activities: (1) training on school-related gender-based violence (SRGBV) prevention and response for teachers and school administrators, and (2) subject clubs.

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 are 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)

EVIDENCE

A growing body of literature analyzes the impact of interventions aimed at developing protective assets on future outcomes, including educational attainment, employment and fertility. Ashraf et al. (2018) analyzed the impact of providing negotiation and interpersonal communication training to adolescent girls in Zambia. Using administrative data, these authors found that the program reduced dropout rates and increased school attendance.

Buchmann et al. (2017) 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 conditional on delaying marriage. The study found that 4.5 years after program implementation, the financial incentives reduced child marriage and teenage childbearing, and increased schooling, while the empowerment program had a positive impact on schooling only.

With respect to labor outcomes, Acevedo et al. (2017) analyzed the impact on earnings and personal skills of training youth in the Dominican Republic. More specifically, the evaluation design randomized participants in two treatment groups and a control group. Treatment 1 included soft-skills training and internship, 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 do 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.

Regarding attitudes towards gender roles and norms, a few studies have documented the impact of modifying gender norms and roles on school-related and other sociodemographic 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 that 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 that 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.

In sum, the little evidence available on programs similar to WW indicates that this type of intervention can have an impact on participants' gender attitudes and self-esteem, and that this can translate into positive impacts on school attendance.

B. EVALUATION PURPOSE & EVALUATION QUESTIONS

The purpose of this evaluation is to estimate the impact of the program on girls' school retention, among other outcomes of interest. NORC will also conduct a performance evaluation of some of the program components. More specifically, the overarching objectives of this evaluation are to:

Objective 1: Evaluate the impact of the program on school retention. To document the mechanisms through which the program may impact retention, the evaluation will also estimate the effect of the program on multiple intermediate outcomes, including 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 school-related gender-based violence (SRGBV).

Objective 2: Assess the WW performance relating to three of the program components along the lines discussed in this report.

Objective 3: Draw upon impact evaluation (IE) and performance evaluation (PE) data meaningfully to present findings, conclusions and recommendations related to program components that what worked/did not work; reasons for high/low impacts; obstacles and drivers to improving impact on key outcomes of interest; etc. Develop recommendations for fine-tuning the program to improve impact and implementation and ensure sustainability.

The impact and performance evaluation questions that will be addressed are shown below.⁶

IMPACT EVALUATION

- IEI0. What is the impact of the program on retention, attendance, and promotion?
IEI.1: What is the impact of the program on the most vulnerable girls?
- IEI1. What is the impact of the POY Club component on retention, attendance, and promotion for girls attending treatment schools?

Intermediate outcomes-girls

- IEI2. Does the program strengthen girls' protective assets?
- IEI3. Is the LGL WW package changing gender norms (girls and boys)?
- IEI4. Does the program change attitudes on SRGBV?⁷

Intermediate outcomes-parents

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

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?⁸

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?

⁶ In March 2017, representatives from USAID (E3/GenDev and the Tanzania Mission), WEI, and NORC met in Arusha to start planning the evaluation of WW. Participants agreed on a list of research questions for the impact evaluation. Subsequent discussions between USAID and WEI also yielded a set of performance evaluation questions for the evaluation.

⁷ Note that 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.

⁸ As the instruments are designed, it will be decided how to record teachers' practices.

- PE1.2: Are members participating actively in POY Clubs? Are they satisfied with POY Club activities? What benefits are perceived by participants?
- PE 1.3: Is the POY Club curriculum user-friendly? What lessons/components of the POY Club curriculum are most useful to adolescents?
- PE 1.4: How is WW engaging LGAs in POY Clubs?
- PE 1.5: 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?
- PE 1.6: Are POY Clubs having any unintended (positive or negative) consequences?
- PE 1.7: Do POY Clubs influence the general school environment? If yes, how? If no, why not?
- PE 1.8: 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 NGO 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 the SRGBV prevention and response teacher training to teachers, students, and other stakeholders?
- PE3.1: Are teachers satisfied with the 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
 - The gender responsive pedagogy
 - The 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 the SRGBV training, teachers' understanding of it, and communication/collaboration between LGAs and teachers?

Sustainability

PE4: Are the interventions addressed in PE1-3 sustainable?

PE4.1: What actions are being taken to continue the interventions addressed in PE1-3 after WW ends?

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

C. EVALUATION METHODS

WW is being implemented in the districts of Karatu DC, Musoma DC, Musoma MC, and Tarime DC. These districts were purposefully selected (rather than randomly) based on data on violence against children and other conditions that make them good candidates for a program like WW to have a substantial impact,⁹ and because of the presence of WEI in Arusha. Of the 68 schools targeted to benefit from WW, the evaluation excludes one control school in Tarime DC and the six Karatu DC schools that were part of the pilot.¹⁰

CI. IMPACT EVALUATION

To evaluate the impact of the program, NORC is conducting a randomized experiment. Between December 2017 and January 2018, 30 schools were randomly assigned to the 'treatment' group (start treatment in February 2018) and 31 schools were assigned to the control group (start treatment in February 2020). Below we describe how the program will evaluate the impact on girls, parents and teachers.

IMPACT ON GIRLS

We collect 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 impact for an average girl on enrollment rates and intermediate outcomes, girls' protective assets, gender attitudes, and perceptions on gender related violence. Survey data also serves to document impacts on enrollment and intermediate

⁹ The Population Council created an index of 9 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, 2016).

¹⁰ Originally WEI had planned to treat Tarime TC instead of Tarime DC. Tarime TC was replaced by Tarime DC because WEI learned that there were already programs similar to WW in Tarime TC. However, while Tarime TC has only seven schools, Tarime DC has 30. To accommodate the program budget, WEI selected eight of the 30 schools in Tarime DC as replacements for the Tarime TC schools, based on their geographic proximity to Tarime TC. NORC had budgeted to survey all 7 schools in Tarime TC. When 8 schools were included from Tarime DC as a replacement for the 7 schools from Tarime TC, NORC had to drop (randomly) one of the extra control schools that had been added.

outcomes for girls at risk. School administrative data is used to evaluate the effect of the program on girls' attendance.

NORC will collect survey data in February 2018 (for the baseline) and in February 2020 (for the endline). School administrative data will be collected in November 2018, November 2019, November 2020, and April 2021.

SURVEY DATA

The sample frame for the girls' survey is the school roster of girls enrolled in Form I in January 2018 in the 61 study schools. The goal is to sample 38 girls in each of these 61 schools, for a total of 2,318 girls. Assuming 20 percent survey attrition at endline, with this design we will be able to estimate a Minimum Detectable Effect Size (MDES) of 0.27.

To evaluate the impact of the program, we compare retention rates and other outcomes of interest between treatment and control groups at endline. While we can estimate experimentally the impact of the program on an average girl, we cannot estimate experimentally the impact on ARD girls or girls randomly selected for POY Clubs. In effect, two different selection processes were at play when POY Clubs were assembled. First, girls expressed their interest (or not) in participating in the club. Second, teachers selected ARD girls. These two selection processes make it difficult to identify girls in control schools that are comparable to any of these two 'types' of girls in treatment schools.

Therefore, to evaluate the impact of the program on ARD girls and girls randomly selected for POY Clubs, we will use quasi-experimental methods. Specifically, we will use matching techniques to construct comparison groups from the control schools that are comparable to the ARD girls and the girls randomly selected for POY Clubs.

The girls' baseline survey (see Annex I for the instrument), which will also be used at endline, includes the modules focused on:

- 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)
- Basic socio-demographic information (parental education, household size, and household asset ownership)
- Program participation

SCHOOL ADMINISTRATIVE DATA

School administrative data will be used to record enrollment and attendance. To collect this data, enumerators will visit each of the 61 schools in the study in November 2018, and again in November 2019, November 2020, and April 2021. To evaluate the impact of attendance for the average girl in a treatment school, we will compare enrollment rates between treatment and control schools. Because information on all students in these 61 schools will be recorded, we will be able to compare attendance rates between treatment and control exact precision.

PARENTS

To answer research questions related to parenting skills and parents' attitudes towards gender-equitable access to education, we will use data collected through a survey of parents. The main, and to some extent, unique direct interaction between the program and parents will be LIMCA; therefore, we focus on LIMCA to evaluate the effect of WW on parents.

The evaluation will focus on a sample of parents that signed up for LIMCA. Specifically, we will survey a sample of parents that signed up for LIMCA after LIMCA groups are constituted, so we have a baseline for LIMCA parents, and survey these same parents again at endline. No comparison group will be surveyed, so the analysis could be described as a before-and-after evaluation.¹¹ We will interview 15 LIMCA parents in each of the 30 treatment schools, for a total of 450 interviews.

TEACHERS

At endline, ten teachers in each of the 61 schools will be interviewed. The teacher survey will include modules for gender attitudes and perceptions and reported frequency of bullying/emotional abuse, sexual harassment, and corporal punishment. The survey will record teachers' opinions on the program's effectiveness in improving the school environment, effects on girls and boys in terms of their behavior at school, and on the likelihood that students drop out of school. Using this data we will compare teachers in treatment and control schools to estimate the impact of the program on teachers.

C2. PERFORMANCE EVALUATION

The performance evaluation will assess the implementation, intermediate outcomes, sustainability, and scalability of three WW interventions – POY Clubs, LIMCA activities, and the SRGBV training for school teachers/administrators – using qualitative data collected in February 2020 and April 2021. In February 2020, we will randomly select six schools for qualitative data collection, three from rural districts and three from peri-urban districts. In each school, three FGDs will be conducted, one with girls, one with boys, and one with parents enrolled in LIMCA. FGDs with girls and boys will include POY Club participants and non-participants. In addition, in each of these six schools, KIs will be conducted with two peer educators, two teachers, and the head teacher. Additional KIs will be conducted with LGAs and EEVs.

In April 2021, we will randomly select six schools and conduct FGDs with girls that were in Form I in 2018 and who are still in school in April 2021, and FGDs with girls that dropped out.

We describe the evaluation of each component below.

¹¹ In the Evaluation Design Report for this program (NORC, 2018) we explained that depending on the parents' take-up rate of LIMCA, we will conduct either an impact evaluation or a performance evaluation. If the LIMCA take-up rate was high, we were going to survey at endline a subsample of the parents of the girls surveyed in February 2018, both in treatment and control schools. If the LIMCA take-up rate was low, the evaluation was going to focus on a sample of parents that actually signed up for LIMCA, and survey them before and after they received parenting training, in order to conduct a performance evaluation. As a preliminary estimate, we considered 33 percent as a take-up rate threshold to decide whether a performance or impact evaluation would be conducted. In October 22 2018, WEI shared LIMCA take-up data and using this data NORC estimated a LIMCA take-up rate of 14 percent. This rate was deemed too low for an impact evaluation so it was decided that a before-and-after evaluation will be conducted.

POY CLUBS

To explore the implementation and efficacy of POY Clubs, we will use data from FGDs with girls and boys in POY Clubs, and KIs with peer educators and LGAs. These interviews and discussions will generate information to address question PE1 and its sub-questions.

LIMCA

To document implementation, benefits, and efficacy of LIMCA, we will use data from FGDs with parents participating in LIMCA, FGDs 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 will use data from KIs with teachers and head teachers, FGDs with girls and boys, and KIs with LGAs. These data will be used to answer question PE3 and its sub-questions.

Lastly, we will review all of WEI's training documents, including the curriculum, SRGBV training materials, and LIMCA materials. This review will provide contextual information on specific training objectives, and allow us to assess the extent to which training objectives have been achieved.

C3. BASELINE DATA COLLECTION

INSTRUMENT DESIGN

The girls' baseline survey includes modules to capture information about girls' protective assets, attitudes towards gender norms, and perceptions on SRGBV. The questionnaire also includes basic sociodemographic 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; the CARE Gender Equity Index; and USAID's SRGBV toolkit. In October 2017, a draft survey was shared with USAID (E3/GenDev and the Tanzania Mission) and WEI for feedback. After several iterations, a final version of the survey was translated and pre-tested in Tanzania in January 2018.

Pretesting was split into two activities; the first was desk testing of the tools which took place January 17-18, 2018. Interviews were carried out in EDI Bukoba office, with EDI personnel not involved in the project participating as interviewees. The second consisted of a field pre-test of the girls' quantitative survey on January 29-31, 2018, in three schools in Kagera region, namely in Bukoba Municipal (MC) and District (DC). This activity was led by the NORC's gender expert, as well as the project PI. The schools used for the pre-testing were selected based on their similarity to the schools in the sampled districts – that is, schools in relatively more disadvantaged areas than the average school near Bukoba. The goal of the pre-test was to ensure that the instruments are capable of capturing the data required by the evaluation. This includes checking that the questions wording and translation are appropriate, that respondents understand the questions, and that response choices are appropriate. In addition, the pre-test was used to identify the need for any additional consistency checks, and to test the routing, length, and general flow of the questionnaire.

After this process, a final version of the survey was created and programmed for fieldwork.

FIELDWORK AND RESULTS¹²

Baseline survey data was collected between February 12 and March 12, 2018. Advance teams visited the schools one week before survey data collection began to randomly sample 46 girls in each school (38 to be interviewed and 8 for replacement). A total of 2,220 girls were surveyed. Of these, 189 girls were replacements, for a replacement rate of 8.5 percent. Our target was to survey 2,318 (61x38) girls so the completion rate was 96 percent. The target was not reached because, in some schools, there were fewer than 38 girls.

POY CLUB DATA

Between January and February 2018, WEI visited all 30 treatment schools to assemble the POY Clubs. In these visits, WEI, in collaboration with the school teachers, selected a total of 1,386 Form 1 girls for POY Clubs in the 30 treatment schools, approximately 46 Form 1 girls per school. Table 2 shows the distribution of these girls between ARD girls and girls selected randomly. In addition, three schools were small enough so all girls were able to enroll.

Table 2: POY Club membership in 30 treatment schools

POY Club status	Girls
At risk of dropping out	519
Randomly selected	801
All girls selected (3 schools)	66
Total	1,386

Source: WEI POY Club data.

POY Club data collected by the implementer was used to identify which of the surveyed girls had signed up for POY Clubs. NORC identified 700 girls in the survey data as POY Club members, of which 272 are ARD girls, 42 are girls from the three schools where all girls were automatically enrolled, and the rest are girls selected randomly for the POY Club.

D. FINDINGS AND CONCLUSIONS

The baseline survey included questions on girls' socio-demographic characteristics, items used to measure girls' protective assets and gender norms, and items used to measure girls' perceptions and attitudes towards SRGBV. In this section, we discuss the baseline results for these three sets of results.

DI. SOCIO-DEMOGRAPHIC CHARACTERISTICS

In this sub-section, we discuss girls' socio-demographic characteristics for treatment and control groups. We also show results for the subgroup of ARD girls. Because these girls were classified by teachers as

¹² A more detailed description of the data collection activities is available in Annex II.

girls at risk of dropping out, we expect them to come from more disadvantaged socioeconomic backgrounds than the average girl in the treatment or control groups.¹³

Table 3: shows summary statistics for the control group, all treatment girls, and ARD treatment girls.¹⁴ In column 4, we present the difference between the treatment and the control group, showing that the sample is balanced at baseline. In column 5, the differences between the ARD girls and the control group are displayed. We use stars in columns 4 and 5 when differences are statistical significant.¹⁵

Table 3: Sociodemographic characteristics

	Control	Treatment		Dif T-C (4)=(2)-(1)	Dif ARD-C (5)=(3)-(1)
	(1)	All (2)	ARD (3)		
Age	14.6	14.5	14.6	-0.11	-0.02
Number of HH members	6.65	6.50	6.54	-0.15	-0.11
Ate anything today (fraction)	0.36	0.35	0.28	0.00	-0.08
Connected to electric grid (fraction)	0.29	0.27	0.21	-0.02	-0.09
Cell phone ownership by parents/family (fraction)	0.96	0.95	0.93	-0.01	-0.03
Asset index (number of items)	4.4	4.2	3.8	-0.25	-0.66**

Source: Own calculations using VVV baseline data. Standard errors are clustered at the school level. The index reflects the number of times girls answered affirmatively questions on whether their household own, in good condition, a chair, a bed, a clock, a radio, a television, a computer, a bicycle, a motorcycle, a car/truck/boat, a refrigerator, and a stove.

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

Note that values shown in column 2 include all girls in the treatment group, including ARD girls. Therefore, and because treatment assignment was randomized, we should expect results in columns 1 and 2 not to be statistically different, in general.

Girls in all groups are about the same age, 14.5 years old. The number of household members living with the girls is also similar across groups. Only about a third of the girls in both treatment and control group reported having something to eat on that day before the interview. By contrast, a smaller fraction of ARD girls (28 percent) had eaten breakfast the day of the interview, but the difference with respect to the control group is not statistically significant. Twenty-nine percent of the households in the control group are connected to the electric grid, for the treatment group this figure is 27 percent and for the ARD girls it is 21 percent. Although the difference between the control and ARD girls in the percentage of households connected to the electric grid is 8 percentage points, it is not statistically significant. Cell phone ownership by parents or family members is over 90 percent across all three groups.

The last row of Table 3: displays results for an index that summarizes household assets ownership. More specifically, the survey included items asking whether households own, in good condition, a chair, a bed, a clock, a radio, a television, a computer, a bicycle, a motorcycle, a car/truck/boat, a refrigerator, and a

¹³ The criteria teachers are to use in selecting the most vulnerable girls include: frequent absenteeism, struggling academically, victims of any type of abuse, orphan or not living with parents, shy or withdrawn most of the time, struggling to get scholastic materials, struggling to get school uniforms, comes from a known poor family that struggles to meet basic needs, known to come from a poor/normal family but shows up to school with expensive jewelry or phones etc., and behavior change that signals having or being in sexual relationship.

¹⁴ Another subgroup of interest are girls that were selected randomly to be part of the POY Club. Because to be selected randomly girls needed first to express their willingness to participate, it is possible that these girls are observationally different from average girls in control group. As we show in Annex III, for key characteristics there are no major differences between randomly selected girls in POY Clubs and girls in the control group.

¹⁵ To calculate the level of significance, the standard errors are clustered at the school level to account for within school correlations.

stove. The index reflects the number of times girls answered affirmatively to these questions. Note that ownership rates are quite low for some of these items. For example only 3 percent of the sample own a computer, 4 percent a car/truck/boat, 26 percent a television and 9 percent a refrigerator. Girls in the control group live in households that, on average, own 4.4 of these types of items; the figure for the treatment group is 4.2, so it is slightly lower than the control group's, but the difference is not significant. The value for ARD girls, on the other hand, is much lower, and the difference with respect to the control group is statistically significant. This implies that ARD girls live in households with fewer assets, suggesting they come from more disadvantaged socio-economic backgrounds with less financial means to purchase these assets.

Table 4 shows the education level of the girls' parents.¹⁶ The first three columns in the table show data for the mothers. Approximately 67 percent of mothers completed primary school and 10 percent completed secondary school in both treatment and control groups. Mothers of ARD girls seem to have less education than the average mother, as only 6.6 percent completed secondary school. Stars in the last row of the table indicate the statistical significance of the difference with respect to the control group.¹⁷ Therefore, it can be seen that for mothers' education there are no statistically significant differences in education level between treatment and control or between ARD girls and the control group. The last three columns show data for the fathers. Roughly 60 percent of fathers completed primary school and 15 percent completed secondary school in both treatment and control groups. Fathers of ARD girls have less education than the average father, given that only 9.4 percent completed secondary school. The differences between ARD girls and the control group is significant for fathers' education.

Table 4: Education level of the parents (%)

	Mother			Father		
	Control	Treatment		Control	Treatment	
	(1)	All (2)	ARD (3)	(1)	All (2)	ARD (3)
No schooling	6.3	5.6	5.5	3.3	2.4	0.9
Some primary	8.3	10.7	13.7	7.8	11.3	11.1
Complete primary	68.7	65.7	69.8	63.3	58.0	66.7
Some secondary	5.3	6.0	4.4	6.4	8.2	10.3
Complete secondary	10.5	10.8	6.6	14.5	16.2	9.4
Higher than secondary	0.9	1.3	0.0	4.8	4.0	1.7
F-test		-	-		-	*

Note: Sample size for the mothers' data is 1,399 and for the fathers' data is 1,132. Data on education level of the parents was collected only for the parents that live with girls. Standard errors are clustered at the school level.

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

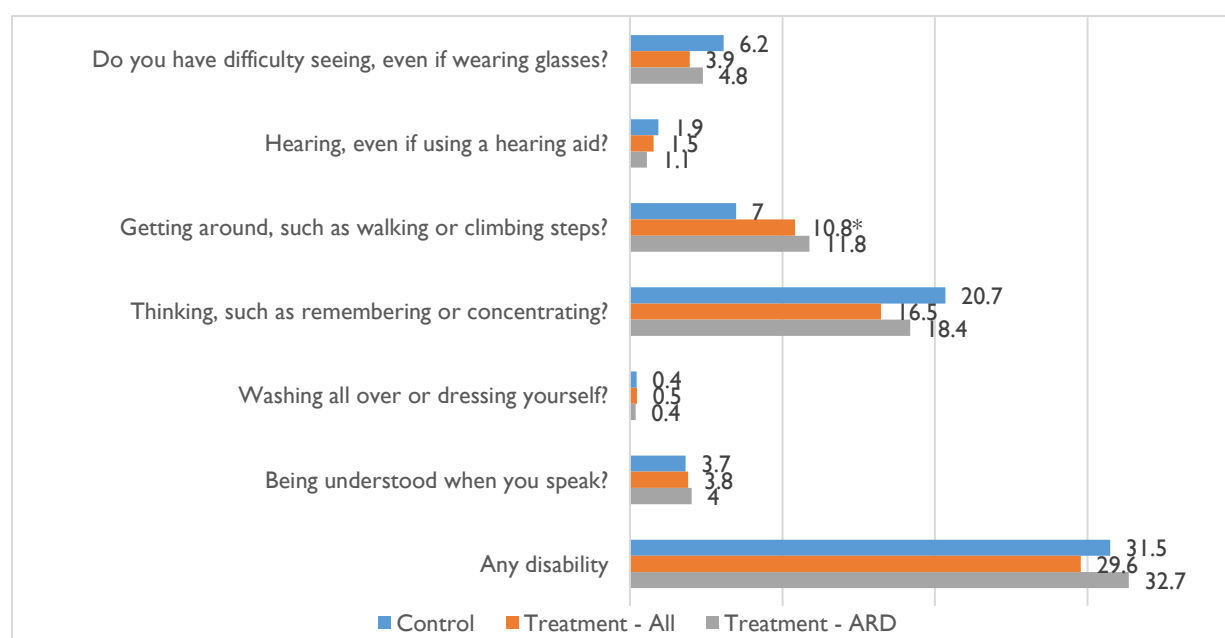
¹⁶ If the girl doesn't live with any of her parents, which was the case for 27 percent of the girls, we asked the girl the education level of the guardian.

¹⁷ To test the significance of the correlation between treatment status and a categorical variable like education level, we use the F-test of the joint significance of a regression where the dependent variable is the treatment dummy (or the ARD dummy) and the covariates are dummies for the each level of education.

The girls were surveyed about disability prevalence. Figure 2: shows the fraction of girls that reported having difficulty seeing, hearing, getting around, thinking, washing or dressing themselves, and being understood when they speak. More specifically, for each type of disability we asked girls if they had no difficulty, some difficulty, a lot of difficulty, or could not do the activity at all. The results below reflect the fraction of girls reporting having at least some difficulty. Results for the fraction of girls having any of these disabilities are also shown.¹⁸ Statistically significant differences between treatment and control groups are denoted by a star to the right of the bars for the treatment girls, and significant differences between ARD girls and girls in the control schools are denoted by stars to the right of the bars for the ARD girls.

In general, disability prevalence is low; for most of the analyzed types of disabilities, the prevalence is about 5 percent or lower. Two exceptions are having difficulty getting around, such as walking or climbing steps, and having difficulty thinking, such as remembering things or concentrating. Seven percent of girls in the control group reported having difficulty getting around, while for the treatment group this figure was 11 percent. The difference between the treatment and control group is significant. Moreover, 12 percent of ARD girls reported having difficulty getting around. The difference with respect to the control group is not significant. Twenty-one percent of girls in the control group reported having difficulty thinking while only 16 percent did so in the treatment group. Eighteen percent of ARD girls reported having difficulty thinking. The differences for having difficulty thinking are not significant in any case.

Figure 2: Disability prevalence - Self-reported (%)



Source: Own calculations using WW baseline data. Standard errors are clustered at the school level.

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

¹⁸ The Washington Group on Disability Statistics (2017) recommends estimating four disability indicators to aggregate disabilities' data. The first disability indicator corresponds to the fraction of respondents saying they have at least some difficulty at any of the mentioned six activities. This indicator is reported at the bottom of Figure 2:.. The criteria of inclusion for the other three levels of disabilities correspond to levels of difficulty higher than 'some difficulty' (e.g. a lot of difficulty, or could not do the activity at all), hence their prevalence levels are very low in this sample, which is why we are not reporting them.

Girls reported their absenteeism during the previous week. Table 5 shows that 13 percent of girls in the control group, 14 percent in the treatment group, and 18 percent ARD girls missed one or more days of school the week before data collection. However, none of the differences are statistically significant. Table 5 also shows that girls in the control group spend on average 43 minutes traveling from home to school, treatment girls spend 49 minutes, and ARD girls spend 52 minutes. Notably, both the differences between the treatment and the control groups, as well as the difference between the ARD girls and the control group, are statistically significant.

Table 5: School-related variables

	Control (1)	Treatment (2)	ARD (3)	Dif T-C (4)=(2)- (1)	Dif ARD-C (5)=(3)-(1)
Time spent traveling to school (min)	43.0	49.0	52.5	5.97*	9.41**
Missed at least one day of school	0.13	0.14	0.18	0.01	0.05

Source: Own calculations using WW baseline data. Standard errors are clustered at the school level.

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

Students in Tanzania are required to pass the Primary School Leaving Exam (PSLE) at the end of primary school (grades 1-7) 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. Official data shows that the national pass rate (grade C or higher) for this exam was 73 percent in 2017.¹⁹ Table 6 shows that less than four percent of girls received an A on this test, a little more than a third scored B, and the rest C. No statistically significant differences are found between groups.

Table 6: Primary School Leaving Exam grade (self-reported)

	Control (1)	Treatment	
		All (2)	ARD (3)
A (Passed; received a score of 161-200)	3.5	3.6	2.8
B (Passed; received a score of 121-160)	35.5	37.8	36.3
C (Passed; received a score of 85-120)	60.9	58.4	60.6
Significance test		-	-

Source: Own calculations using WW baseline data. Standard errors are clustered at the school level.

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

Finally, three fourths of the sample expect to get a tertiary education degree, which includes college/university and/or vocational education. No major differences between groups are identified in terms of how far girls expect to go in school. Data from TDHS 2015-2016 shows that only 2.9 percent of females between 25 and 29 years old had more than secondary education, suggesting that these expectations overestimate what would be observed in terms of educational achievement. This is

¹⁹ Own estimates using data from the National Examination Council of Tanzania, https://online.necta.go.tz/opendata/home/download/PSLE2017_IMPROVEMENT.xls, extracted November 26, 2018.

consistent with previous research showing that educational expectations are usually much higher than actual achievement (Guerrero et al., 2016 and Tafere, 2014).

Table 7: Expected maximum level of education

	Control (1)	Treatment	
		All (2)	ARD (3)
Finish Form 4	6.2	5.3	4.0
Finish Form 5	1.1	0.9	1.5
Finish Form 6	10.7	12.6	9.9
Some tertiary education	6.6	5.4	5.5
Finish tertiary education	75.5	75.9	79.0
Significance test		-	-

Source: Own calculations using WW baseline data. Standard errors are clustered at the school level. * p<0.05 ** p<0.01 *** p<0.001

D2. OUTCOME VARIABLES – PROTECTIVE ASSETS AND GENDER NORMS

While the ultimate goal of WW is to increase school retention, it also seeks to build agency, knowledge, and protective assets of girls in secondary school. 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 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, the baseline survey included items to measure self-esteem, self-efficacy, decision-making, communication skills, social networks, and gender norms. To construct these items, we reviewed several widely used and validated scales and adapted them to adhere to the local context and developmental stage of the girls. For brevity, not all items from the reviewed instruments were included. To analyze the collected data, we first pool all surveyed girls into one group and analyze descriptive statistics for each of the protective assets. Then, we analyze correlations between different sociodemographic characteristics and protective assets. Finally, for each protective asset, we examine if there are differences between control and treatment girls, and between control and ARD girls, similar to the analysis we showed in the previous section for sociodemographic characteristics.

DESCRIPTIVE ANALYSIS

SELF-ESTEEM

The WW program seeks to build girls' agency through empowerment and self-awareness. 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-confidence typically decreases, impacting their decision-making.

It is very common in empowerment programs to measure girls' perception of self and self-confidence using a self-esteem scale. Our baseline survey included a section aimed at measuring girls' level of self-esteem using 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).

To this end, girls were asked questions about their feelings and attitudes toward themselves. They responded using a 4-point Likert scale: strongly agree, agree, disagree, or strongly disagree.

Figure 3 shows the responses for the items related to self-esteem.

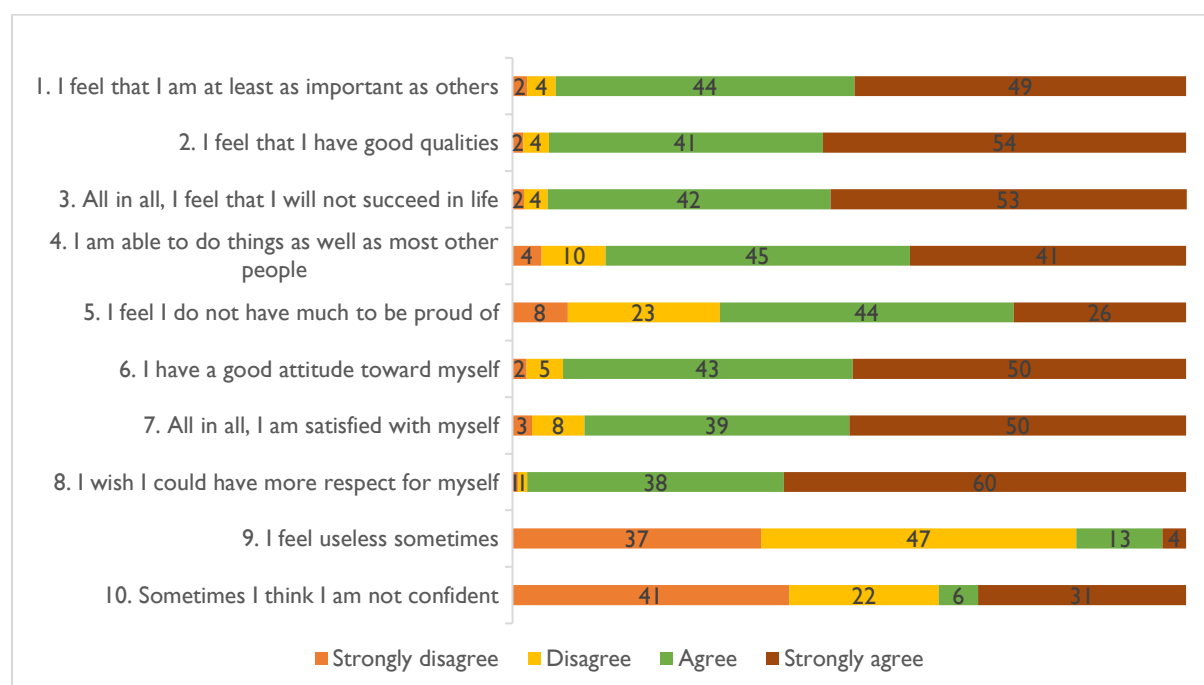
In general, girls' answers are consistent with a relatively high level of self-esteem. For example, almost half the girls report that they strongly agree with item 1 ('I feel that I am at least as important as others') and a total of 93 percent either agrees or strongly agrees with the statement. Similarly, items 2, 4, 6, and 7 have positive valence, so in principle we should expect that girls with high self-esteem would agree or strongly agree with these items, which is actually the case. A large proportion of the girls agree or strongly agree with these statements, indicating they have positive attitudes towards themselves, are able to do things as well as others, and are satisfied with themselves.

Items 3, 5, 8, 9, and 10, on the other hand, have reversed or negative valence and, therefore, we would expect girls with high self-esteem to disagree or strongly disagree with those statements. However, for these items the results are contradictory. The majority of girls disagree or strongly disagree with items 9 ('I feel useless sometimes') and 10 ('Sometimes I think I am not confident'), which is consistent with having high self-esteem. However the responses to items 3, 5, and 8 suggest a low level of self-esteem.

There are multiple considerations that could explain these results. For example, it is possible that girls do not believe they have agency in their lives around future success despite having relatively high self-esteem. This could explain why 95 percent answered either agree or strongly agree to item 3 ('All in all, I feel I will not succeed in life'). They may want to succeed in life, but do not feel they will be able to overcome the obstacles they face with the limited resources and support they have. Another possible explanation is that girls may be confused by questions that have negative valence. In a multiple-country application of the RSES, Schmitt and Allik (2005) found that, out of the 53 tested countries, in Tanzania and another four countries item 8 was probably wrongly interpreted. As these authors explain, "The item should be interpreted as indicating low self-esteem, but in some cultures, people may interpret it egotistically as though they are worthy of having even more positive views of themselves than they already have."²⁰

²⁰ Schmitt and Allik (2005), p. 627.

Figure 3: Self-esteem (%)



Source: Own calculations using WW baseline data. Sample size is 2,220.

To summarize the 10 self-esteem items, we constructed a self-esteem index. Responses to items with positive valence were coded on a 4-point scale ranging from 0 (strongly disagree) to 3 (strongly agree). Items with negative valence were reverse coded. Therefore, the minimum possible score is 0 and the maximum possible score is 30. The mean of the self-esteem score is 20.5. In a previous application of the RSES in Tanzania, Schmitt and Allik (2005) found an average score of 19.5.²¹ The distribution of the score can be consulted in Annex IV.

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. For self-esteem, the Cronbach's alpha is 65.2, which is relatively low but still within the acceptable range. In general, alphas between 65 and 80 are considered acceptable.

Table 8: Self-esteem index

	Mean	Standard Deviation	Number of items	Max possible score	Cronbach's alpha
Self-esteem	20.5	2.8	10	30	65.2

Source: Own calculations using WW baseline data. Sample size is 2,220.

SELF-EFFICACY

Girls' self-efficacy contributes to girl's empowerment (Bandura, 1997). Self-efficacy is also an important social and emotional learning skill. Bandura argued that while individuals may know what actions are

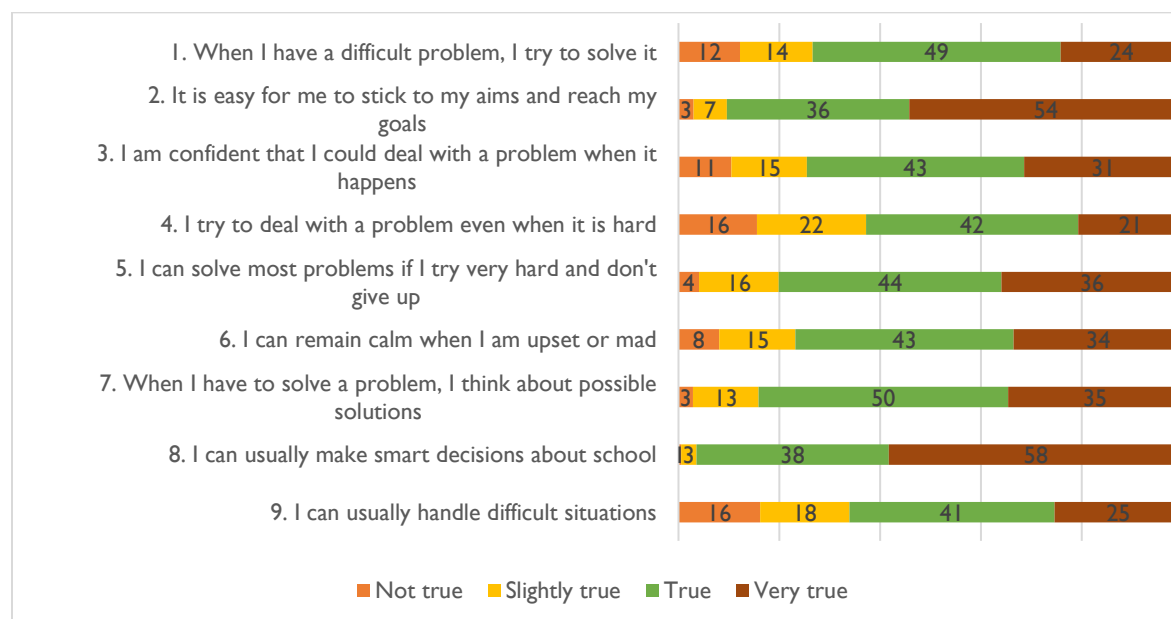
²¹ The mean documented by Schmitt and Allik (2005) is actually 29.5 but this is because their coding ranged from 1 to 4 rather than 0 to 3.

needed to accomplish what they want, this knowing is not enough. People also need confidence about their ability to carry out the behavior or complete the task. Self-efficacy is more than just having the confidence but rather the belief that you can achieve something because of that confidence. The WW program works towards helping girls build their self-efficacy competencies. When girls acquire self-efficacy competencies, they are better equipped to make healthy and smart decisions for their future. Self-efficacy skills have also been shown to play a significant role in student achievement (Alfassi, 2003). Accordingly, measuring girl's self-efficacy enables us to learn about their perceived ability to achieve a task or action.

To measure girls' self-efficacy, we used items from the Generalized Self-efficacy Scale (GSE) (Schwarzer & Jerusalem, 1995). The purpose of this rubric is to measure respondents' skills relevant to managing daily troubles and adapting to stressful events. The items focused on how to solve difficult problems, make decisions, and achieve goals.

Figure 4 shows tabulation results for items related to self-efficacy. The data suggest that girls have average to above average competencies in self-efficacy, but there is still room for improvement, as indicated by items 1, 4, and 9. Ninety percent of girls think they can stick to their aims and reach their goals (answered true or very true to item 2). Additionally, 80 percent of girls think they can solve most problems if they try very hard and don't give up (answered true or very true to item 5). In effect, most girls answered true or very true on items about problem-solving, goal-setting, and remaining calm in difficult situations. These scores indicate that girls in our sample exhibit self-efficacy skills and can employ them when needed in a difficult or challenging situation.

Figure 4. Self-efficacy (%)



Source: Own calculations using WW baseline data. Sample size is 2,220.

To construct the self-efficacy index, we coded responses to each item on a 4-point scale ranging from 0 (Not true) to 3 (Very true). Therefore, the index minimum possible score is 0 and the maximum possible score is 27. The estimated mean of the self-efficacy index is 18.5 (See Table 9). In an application

of the original rubric in 22 countries, Schwarzer & Scholz (2000) found an average score of 19.5.²² Details about the distribution of the score can be consulted in Annex IV. The Cronbach's alpha, that measures reliability, is relatively high indicating that self-efficacy items are highly interrelated.

Table 9: Self-efficacy index

	Mean	Standard Deviation	Number of items	Max possible score	Cronbach's alpha
Self-efficacy	18.5	4.5	9	27	75.7

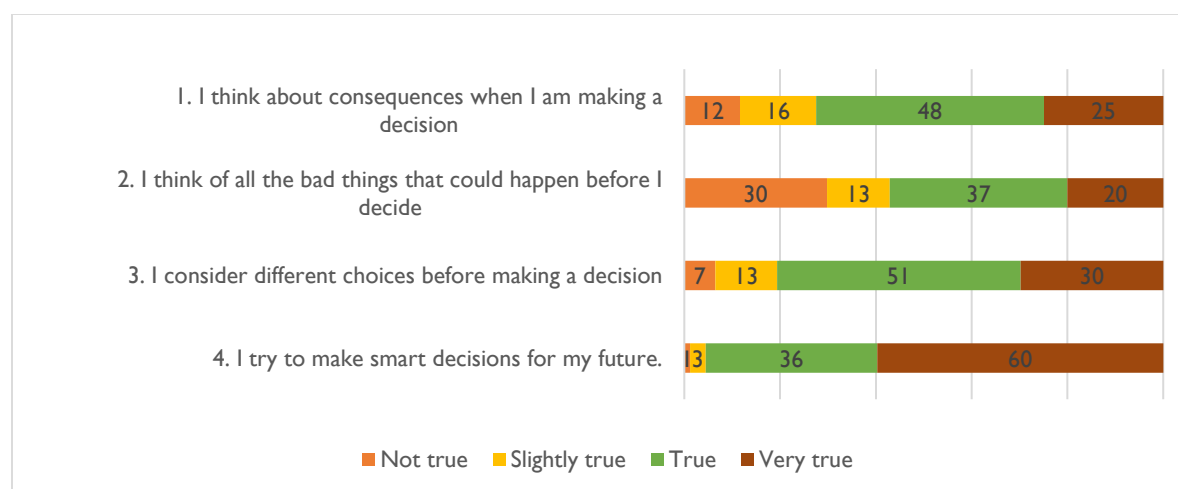
Source: Own calculations using WW baseline data. Sample size is 2,220.

DECISION MAKING

The WW POY Club curriculum includes sessions about decision-making as girls' knowledge and ability to make healthy decisions is critical for their positive development. Having decision-making knowledge and competencies will help 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). The girls surveyed have relatively strong decision-making skills. Based on the data, it appears that girls are thinking about the consequences of their decisions (73 percent said it is true or very true that they think about consequences when making a decision). Girls also appear to consider different choices when making decisions (81 percent said it is true or very true that they consider different choices before making a decision), and they want to make smart decisions for their future. Figure 5 shows tabulations for items related to decision-making.

Figure 5: Decision making (%)



Source: Own calculations using WW baseline data. Sample size is 2,220.

²² The mean documented by Schwarzer & Scholz (2000) is actually 29.5 but this is because their coding ranged from 1 to 4 rather than 0 to 3. It is also worth highlighting that these authors used the original 10-item scale, while for this application we used 9 items only. As explained before, the reviewed scales were adapted for brevity and to adhere to the local context and developmental stage of the girls.

To construct a decision making index, we coded item responses on a 4-point scale ranging from 0 (Not true) to 3 (Very true). Therefore, the minimum possible index score is 0 and the maximum possible score is 12. The mean of the decision making index is 7.9 (See Table 10). The distribution of the index is shown in Annex IV. The Cronbach's alpha is below the acceptable range. Because alphas increase with the number of items, a potential strategy to improve the measurement of this latent trait for the endline data collection is to include more items to measure it.

Table 10: Decision making index

	Mean	Standard Deviation	Number of items	Max possible score	Cronbach's alpha
Decision-making	7.9	2.3	4	12	55.7

Source: Own calculations using WWV baseline data. Sample size is 2,220.

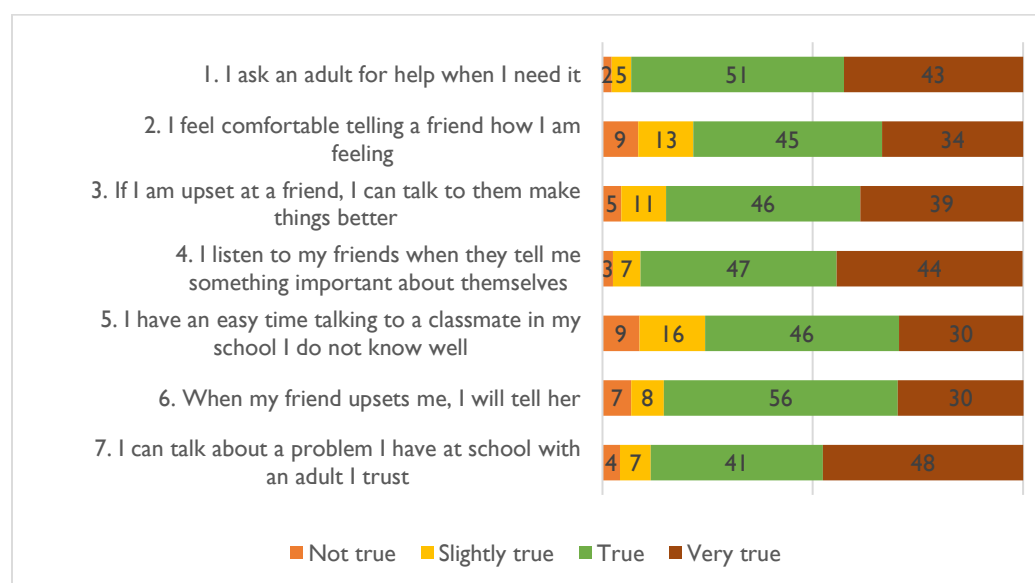
COMMUNICATION SKILLS

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. Typically, girls' education and empowerment programs focus on building girls communication skills to develop their agency and improve their social and emotional skills. 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. We developed new items because there were no appropriate existing measures focusing on communication behaviors. We wanted to understand how girls use their communication skills to build relationships, make decisions, problem solve, and practice self-management. Items were developed, adapted and pre-tested with girls during the pre-test phase of data collection. Girls were asked about their listening skills, their ability to share feelings with others, their ability to ask for help from others, and about effective communication while problem solving. Most responses were true and very true, indicating that girls' have listening and communication competencies. For example, 79 percent of girls said they feel comfortable talking about how they feel with their friends (item 2), and 89 percent said they can discuss problems at school with an adult they trust (item 7).

Figure 6: shows tabulations for items related to communication skills.

Figure 6: Communication skills (%)



Source: Own calculations using WWV baseline data. Sample size is 2,220.

To construct a communication skills index, we coded responses on a 4-point scale ranging from 0 (Not true) to 3 (Very true). Therefore, the minimum possible score for the index is 0 and the maximum possible score is 21. The mean of the communication skills index is 15.3 (See Table 11). The distribution of the index can be consulted in Annex IV. The Cronbach's alpha is 70.2, indicating that items observe an acceptable degree of correlation.

Table 11: Communication skills index

	Mean	Standard Deviation	Number of items	Max possible score	Cronbach's alpha
Communication skills	15.3	3.3	7	21	70.2

Source: Own calculations using WWV baseline data. Sample size is 2,220.

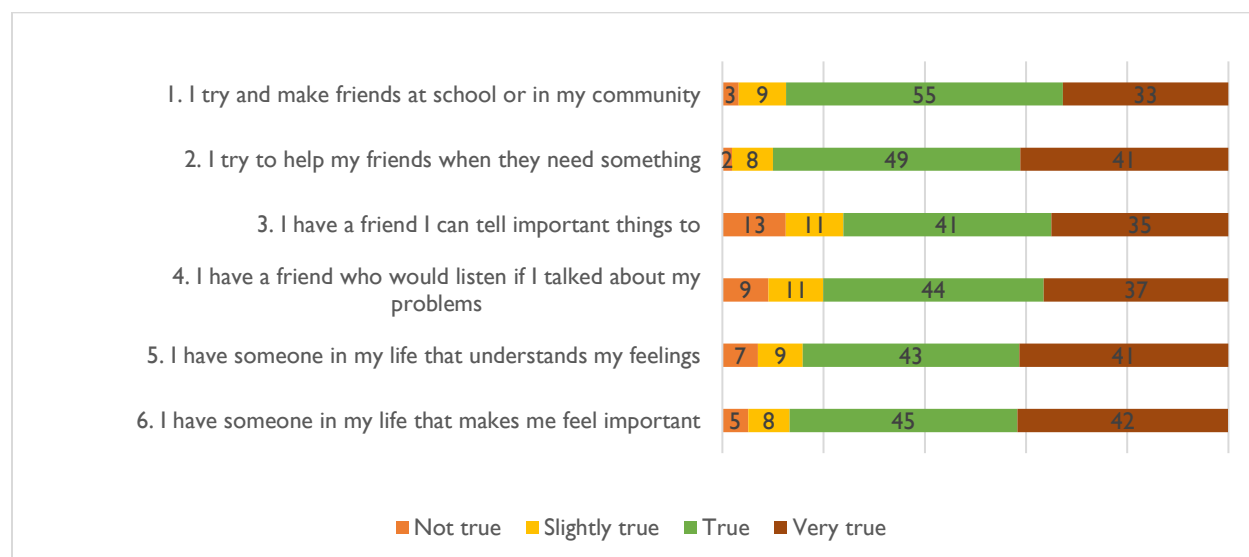
SOCIAL NETWORKS

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. Social support and networks are protective factors against stress and they are a source of wellbeing (Cohen & Wills, 1985). Belonging and connection is increased through group support processes, both through informal (friendships) social networks and formal social networks (Graczyk, 2000).

Girls were asked about their social networks with peers and adults using the Girl Empower (GE) Baseline Survey (Hallman, 2016). Figure 7 shows results for items related to social networks. The girls seem to have positive social networks with their friends and other individuals in their lives. Ninety percent of girls said they wanted to help their friends when the latter needed something. Additionally, 87 percent of girls have someone in their life that makes them feel important. This data indicates that

girls that have positive friendships with their peers, and that they have someone they trust with whom to share their feelings.

Figure 7. Social networks (%)



Source: Own calculations using WW baseline data. Sample size is 2,220.

In order to construct a social networks index, we coded responses on a 4-point scale ranging from 0 (Not true) to 3 (Very true). Therefore, the minimum possible score for this index is 0 and the maximum possible score is 18. The mean of the social networks index is 12.9 (See Table 12). The Cronbach's alpha is 70.3, indicating an acceptable level of item inter-correlation. In Annex IV we show the distribution of this index.

Table 12: Social networks index

	Mean	Standard Deviation	Number of items	Max possible score	Cronbach's alpha
Social networks	12.9	3.3	6	18	70.3

Source: Own calculations using WW baseline data. Sample size is 2,220.

ATTITUDES TOWARDS 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. These roles are assigned by social criteria rather than biological (USAID, 2007). 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.

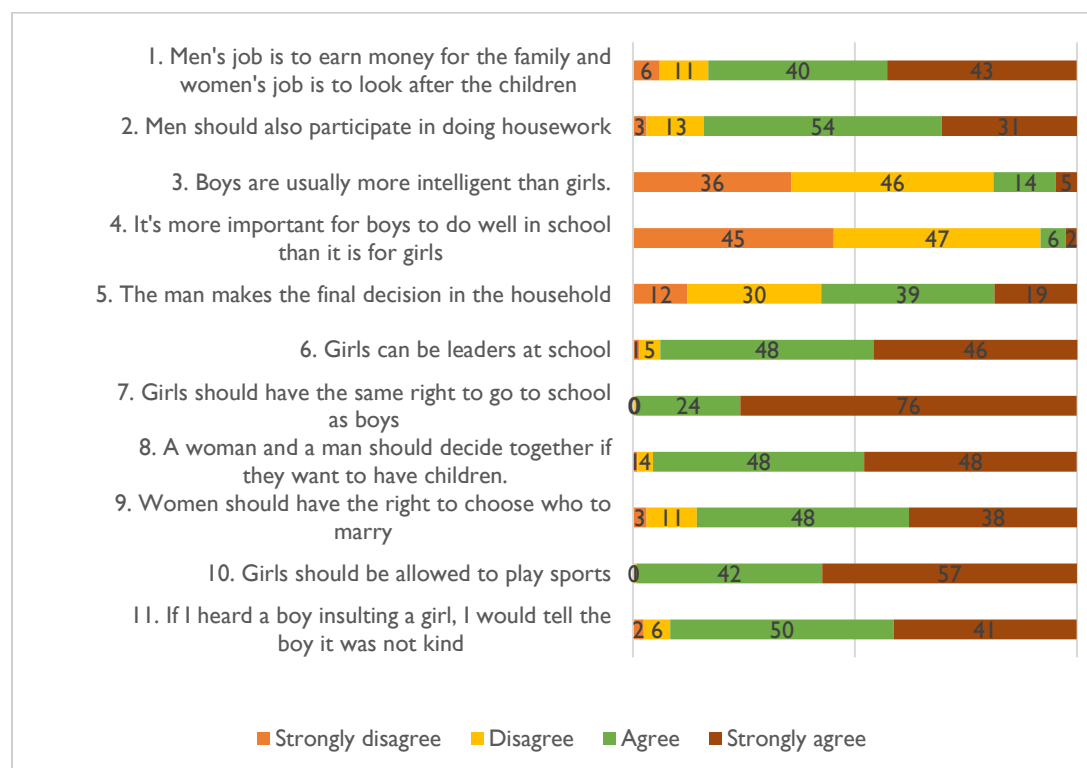
We asked girls about their beliefs on gender norms, roles, and equity using the Gender Equitable Index (GEI) (Care, 2014). They answered a series of questions about their perceptions of gender roles, gender relations, and gender equity.

Figure 8 shows results for items related to gender norms.

There are some interesting patterns worth highlighting. First, based on the data, girls seem to believe that there should be gender equity between girls and boys, especially when we asked about behaviors in school. For example the majority of girls think they can be leaders at school (item 6) and do not think boys are more intelligent than girls (item 3). Similar results can be observed for items 4, 7, 8, 9, 10, and 11.

However, for a couple of items, girls believe that there are specific male and female roles. For example, the majority of girls think a man's job is to earn money for the family and a woman's job is to look after the children (item 1). Results for item 5 ('The man makes the final decision in the household') indicate that girls do not have a strong opinion one way or another, on average. As we discuss more thoroughly in Annex V, it is possible that this reflects that girls' views on gender norms and equity are contingent on context. Girls hold more traditional views when the statements refer to gender roles within the family than when the context is their schools.

Figure 8. Attitudes towards gender norms, roles & equity (%)



Source: Own calculations using WW baseline data. Sample size is 2,220.

To construct a gender norms, roles and equity index, we coded responses on a 4-point scale ranging from 0 (Strongly disagree) to 3 (Strongly agree). Therefore, for this index, the minimum possible score is 0 and the maximum is 33. The mean of the index is 21.1 (See Table 13). We show the distribution of the index in Annex IV.

The Cronbach's alpha is below the acceptable range. This score is based on 11 items, so it is unlikely that the number of items is the reason for the low alpha in this case. Statistical analysis described in Annex V indicates that excluding two of the eleven items increases the Cronbach's alpha to .69. As we discuss in more detail in Annex V, these two items are about gender norms in the context of the girls' homes, while the majority of the other items are more relevant to the school context. If girls' beliefs and attitudes towards gender norms and stereotypes are driven by different latent traits depending on the context they are considering, using only one index that summarizes the eleven items may not be an appropriate approach. One alternative is to construct the gender norms, roles and equity index using only school-related items. For the endline report, we will document whether the impact evaluation results are sensitive to the inclusion of different items in the gender norms, roles and equity index.

Table 13: Attitudes towards gender norms, roles & equity index

	Mean	Standard Deviation	Number of items	Max possible score	Cronbach's alpha
Gender norms scale	21.1	3.3	11	33	60.8

Source: Own calculations using VVW baseline data. Sample size is 2,220.

DETERMINANTS OF PROTECTIVE ASSETS AND GENDER NORMS

Bivariate regressions between protective assets indices and selected socio-demographic characteristics are displayed in Table 14. While these coefficients have no causal interpretations, it is informative to show which socio-demographic variables are correlated with protective assets.

A few interesting patterns are worth highlighting. The first row shows coefficients for parental education. For ease of exposition rather than using the categorical variable described in the previous subsection, we constructed a dummy variable for whether the parent has more than primary education or not.²³

While parental education is positively correlated with all protective assets indices, the correlation is significant only for self-efficacy, decision making, and gender norms. Speaking Swahili at home is correlated with all six protective assets, but the correlation is significant only for self-esteem, self-efficacy, decision making, and gender norms. Interestingly, all indices are negatively correlated with age, and the correlations are significant for self-esteem, decision making, and gender norms. Perhaps not surprisingly, all indices are negatively correlated with having reported any disability, and the coefficients are significant for self-esteem and social networks.

²³ As before, if the girl is living with her two parents we use the parent with more education, and if the girl does not live with any of her parents, we use the education of her guardian.

Correlations between having had breakfast and protective assets indices do not show a clear pattern, meaning that the signs of the correlations are not the same across indices, and the coefficients are not significant in any case. Other indicators of economic wellbeing seem more clearly correlated with protective assets. Living in a house connected to the electric grid is positively correlated with all protective asset indices, and the correlations are significant for self-esteem, self-efficacy, decision-making, and gender norms. A similar pattern is observed for the household asset ownership index, except that for this sociodemographic indicator, the correlation with social networks is also significant

Table 14: Bivariate regressions between protective assets and selected socio-demographic characteristics (N=2220†)

	Self-esteem	Self-efficacy	Decision-making	Communication skills	Social networks	Gender norms
Parental ed (more than primary)	0.16 (0.13)	0.58** (0.20)	0.30* (0.13)	0.22 (0.14)	0.14 (0.16)	0.47** (0.15)
Speaks Swahili at home	1.07*** (0.13)	0.97*** (0.21)	0.62*** (0.11)	0.14 (0.14)	0.12 (0.16)	1.31*** (0.17)
Age	-0.23*** (0.062)	-0.20 (0.10)	-0.10* (0.049)	-0.046 (0.071)	-0.099 (0.076)	-0.17* (0.073)
Any disability	-0.70*** (0.14)	-0.35 (0.24)	-0.037 (0.14)	-0.24 (0.16)	-0.35* (0.17)	-0.15 (0.17)
Ate breakfast	0.26 (0.16)	0.19 (0.21)	-0.098 (0.10)	-0.043 (0.15)	0.33 (0.17)	0.011 (0.20)
Connected to the electric grid	0.85*** (0.13)	1.00*** (0.22)	0.77*** (0.11)	0.089 (0.14)	0.15 (0.16)	1.22*** (0.16)
Asset ownership index	0.21*** (0.040)	0.24*** (0.056)	0.14*** (0.027)	0.056 (0.037)	0.088* (0.040)	0.17*** (0.047)

†Sample sizes are lower for some socio-demographic characteristics due to item-specific missing data. Source: Own calculations using VVV baseline data. Standard errors are clustered at the school level. * p<0.05 ** p<0.01 *** p<0.001

It is perhaps surprising that speaking Swahili at home is positively correlated with protective assets. As Table 15: shows, speaking Swahili is also positively correlated with being connected to the electric grid and to the asset ownership index, which could explain why speaking Swahili is positively correlated with protective assets. In other words, if girls from less disadvantaged households observe higher protective assets, and are also more likely to speak Swahili, it is natural to expect that the correlation between speaking Swahili and protective assets is also positive.

We also showed that age is negatively correlated with protective assets, and as Table 15: shows, age is negatively correlated with being connected to the electric grid and the asset ownership index. It is possible that more disadvantaged households enroll their children in school later than other households, and/or that children from more disadvantaged households are more likely to have repeated a grade. Along these lines, the reason why we find a negative correlation between age and protective assets could be because age is negatively correlated with households' economic well-being, and economic well-being is positively correlated with protective assets.

Table 15: Bivariate regressions between selected sociodemographic characteristics (N=2220†)

	Speaks Swahili at home	Age	Connected to electric grid	Asset ownership index
Speaks Swahili at home	1			
Age	-0.047***	1		
Connected to the electric grid	0.42***	-0.11*	1	
Asset ownership index	0.078***	-0.048***	0.11***	1

†Sample sizes are lower for some socio-demographic characteristics due to item-specific missing data. Source: Own calculations using VVV baseline data. Standard errors are clustered at the school level.

* p<0.05 *** p<0.001

Table 16 shows bivariate regressions between protective assets indices and school-related variables. The first row shows coefficients for a binary variable that is 1 if girl expects to attend tertiary education (university or technical) and 0 if they expect to reach a lower level of education (Form 6 or lower). Notably, education expectations are positively and significantly correlated with all six protective asset indices. This indicates that there is a positive association between having higher education expectations and having higher levels of protective assets. It is important to note that this association does not imply causality, as it could be, for example, that other variables affect both aspirations and protective assets.

Missing one day of school or more during the previous week is negatively correlated with all protective asset scores, and coefficients are significant for self-esteem and gender norms.

The third and last row shows results for a binary variable that is 1 if the grade in the Primary School Leaving Exam was A or B, and 0 if it was C. There is no clear pattern across protective assets, and while (perhaps counterintuitively) the coefficients are negative for four of the six protective assets, the parameters are not significant in any case.

Table 16: Bivariate regressions between protective assets and school-related variables (N=2220†)

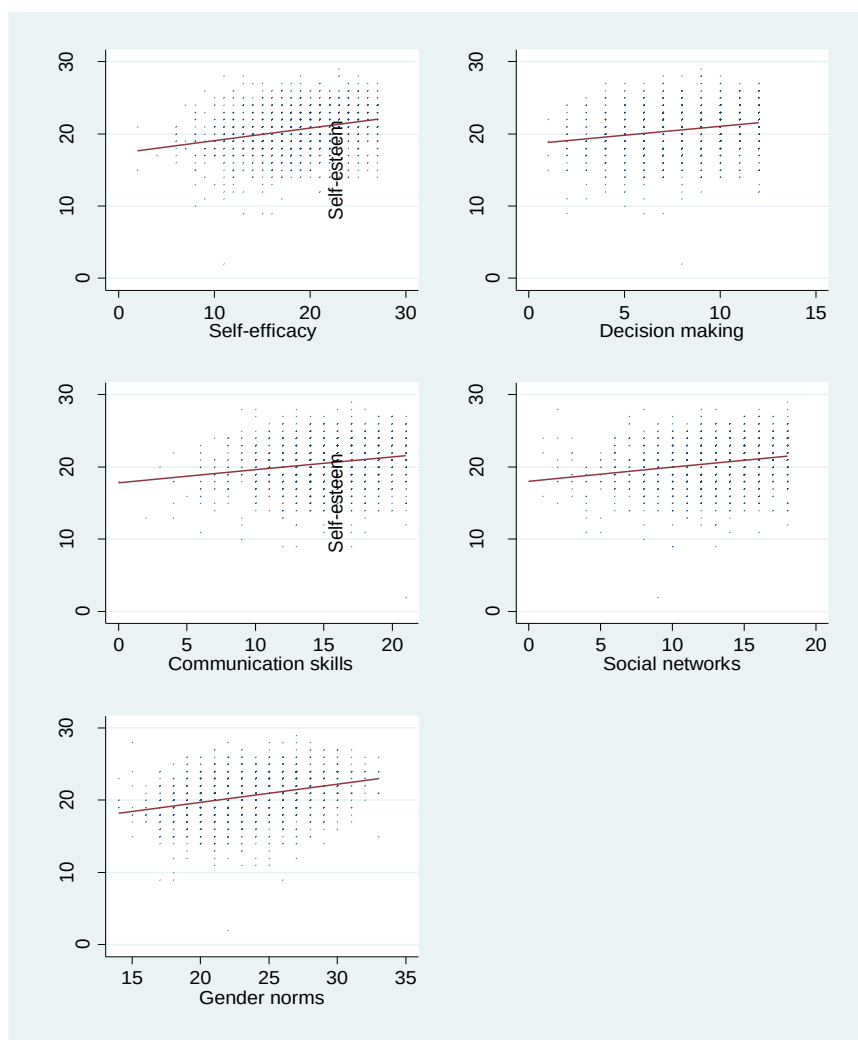
	Self-esteem	Self-efficacy	Decision-making	Communication skills	Social networks	Gender norms
Expects attending tertiary education	0.58** (0.18)	1.02*** (0.26)	0.76*** (0.12)	0.72*** (0.19)	0.42* (0.18)	1.08*** (0.18)

	Self-esteem	Self-efficacy	Decision-making	Communication skills	Social networks	Gender norms
Missed school last week	-0.64*** (0.17)	-0.39 (0.27)	-0.24 (0.13)	-0.26 (0.23)	-0.34 (0.21)	-0.55** (0.20)
PSLE grade	-0.12 (0.12)	0.15 (0.21)	-0.065 (0.11)	-0.0018 (0.16)	-0.17 (0.17)	0.15 (0.19)

†Sample sizes are lower for some socio-demographic characteristics due to item-specific missing data. Source: Own calculations using WW baseline data. Standard errors are clustered at the school level. * p<0.05 ** p<0.01 *** p<0.001

It is also worth mentioning that protective assets indices are highly correlated with each other, reflecting that girls who observe higher values for one of these traits are likely to observe high values for other traits as well. Figure 9 shows correlations between the index of self-esteem and the other constructed indices. Annex VI shows bivariate regressions for all possible pairs of indices. All correlations are positive and statistically significant.

Figure 9. Correlations between self-esteem score and other protective assets scores

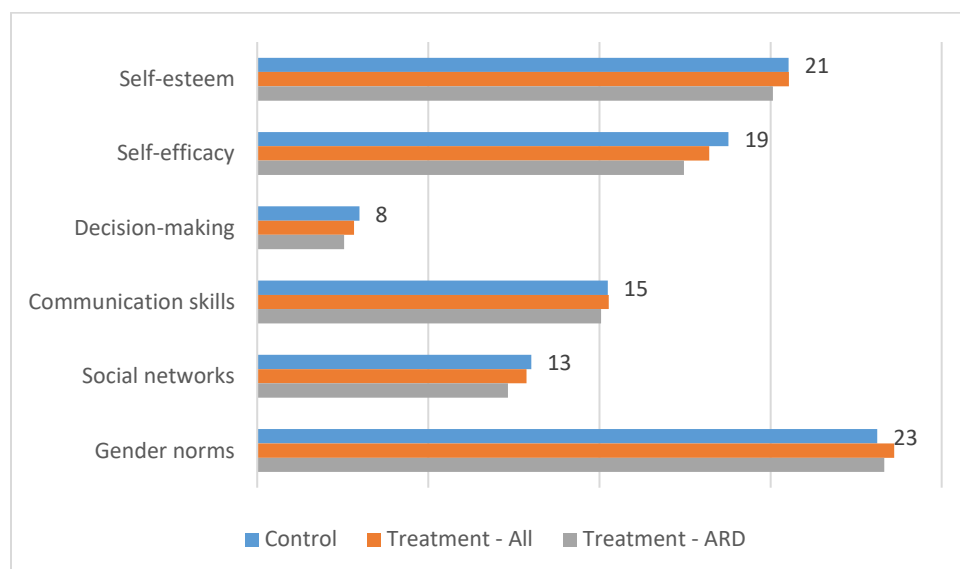


Source: Own calculations using WW baseline data.

BALANCE OF PROTECTIVE ASSETS AND GENDER NORMS

To check for balance between treatment and control groups, and ARD girls and girls in the control group, Figure 10 shows mean scores for protective assets for these three groups of girls. Statistically significant differences between treatment and control groups are denoted by stars on top of the bars for the treatment girls, and significant differences between ARD girls and girls in the control schools are denoted by stars on top of the bars for the ARD girls. Mean values for treatment and control groups are pretty balanced for all indices except for self-efficacy, where we find a significant difference in favor of the control group. For most indices, ARD girls' mean values are lower than for the control group, and the differences are significant for self-efficacy, decision making, and social networks. The fact that ARD girls are observationally different from average girls implies that using girls in the control group as a comparison group could provide a biased estimate of the impact of the program for ARD girls. As explained in the EDR, our strategy for the impact evaluation will be to model the ARD selection process using baseline data, use this model to select girls in the control group that would have been selected as ARD girls if their school had been a treatment school, and evaluate whether there is a difference in the outcomes of interest between these selected girls in the control schools and ARD girls.

Figure 10. Mean indices for protective assets for girls in control schools, girls in treatment schools and ARD girls



Note: Statistically significant differences between treatment and control groups are denoted by stars on top of the treatment bars, and significant differences between ARD girls and girls in the control schools are denoted by stars on top of the bars for the ARD girls. Standard errors are clustered at the school level.

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

D3. OUTCOME VARIABLES – SRGBV

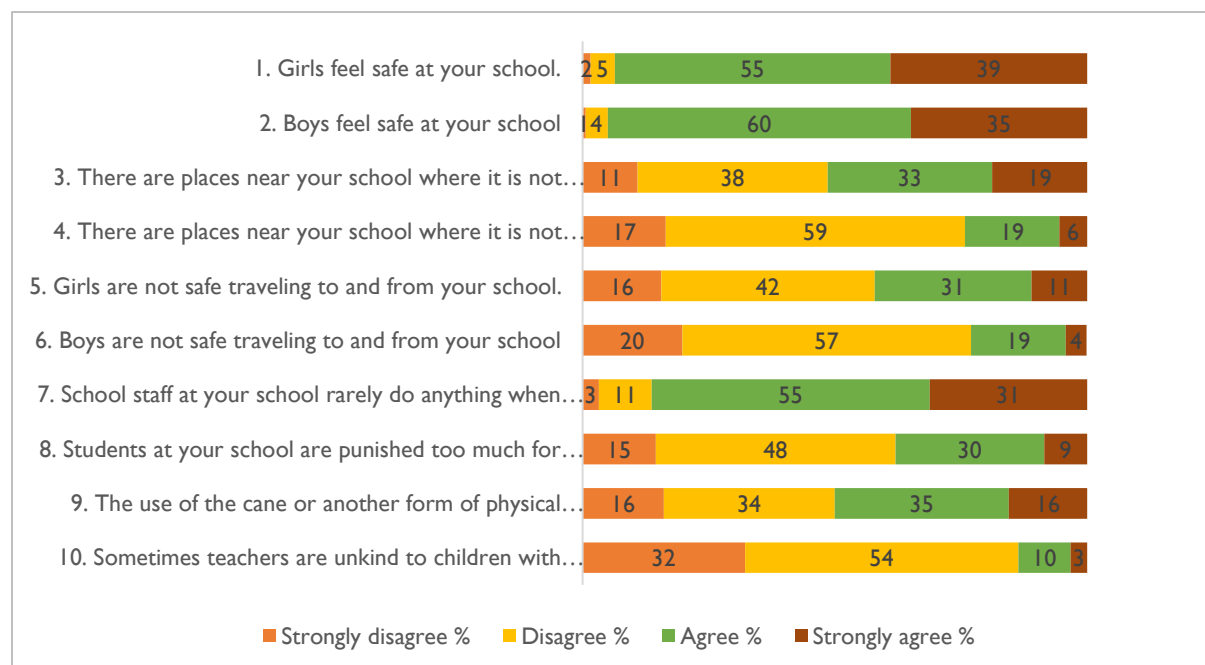
Incidents of SRGBV have been recorded in many settings, cultures, and schools around the world. Although both boys and girls experience this violence, typically girls are targeted at higher rates. When schools are not safe for students, their ability to learn and develop suffers. Girls' access to education and learning is impacted when the classroom and school climate is not safe and supportive. SRGBV can take many forms, including emotional abuse and bullying, physical violence, sexual violence, and corporal punishment (RTI International, 2016). Perpetrators of this gender-based violence can be peers, teachers,

and other adults in a girl's life. WW program aims to provide girls with a safe and supportive learning environment by providing SRGBV training to students, parents/caregivers, and teachers.

Figure 11 shows results for a series of statements related to perceived frequency and nature of SRGBV, as well as safety more broadly defined. To construct these items, we reviewed instruments from RTI (2016). Most girls feel safe in school. Ninety-four percent of girls agree or strongly agree with the statement 'Girls feel safe at your school' (item 1). When girls were asked if boys feel safe at their school (item 2), the answers were relatively similar, suggesting that there are no major differences on how safe girls and boys feel in school, at least in the girls' opinion. The results are different when we look at environments outside the school. Roughly half the girls agree or strongly agree 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 agree or strongly agree with the statement. In other words, in the girls' opinion, there are more risks for girls near school than for boys. A similar result can be derived when the survey asked about safety traveling to and from school (items 5 and 6).

Teachers and school staff do not seem to be a strong deterrent of violence in schools. Eighty-six percent of girls agree or strongly agree with the statement 'School staff at your school rarely do anything when students hurt other students' (item 7). Thirty-nine percent of girls agree or strongly agree with the statement 'Students at your school are punished too much for little things' (item 8), and 51 percent agree or strongly agree with the statement 'The use of the cane or another form of physical discipline is common in your school' (item 9). Items 8 and 9 suggest that teachers in surveyed schools use harsh discipline methods. Finally, most girls do not think teachers are unkind to children with disabilities.

Figure 11. Perceptions of SRGBV



Source: Own calculations using WW baseline data. Sample size is 2,220.

To measure attitudes towards SRGBV, we presented girls with different hypothetical scenarios where girls like them were being bullied or being victims of other forms of abuse including emotional abuse, corporal punishment, physical violence, sexual violence by a teacher or school staff, and sexual violence by boys at school. To construct these scenarios and item questions, we reviewed instruments from RTI (2016) and NORC (2017).

After each scenario was described, we asked girls if they had witnessed such a situation. Then we asked girls if they believed someone should be treated the way the person in the scenario was treated. The girls were provided with a four point response scale of strongly disagree, disagree, agree, and strongly agree.

Table 17 shows findings for the questions on girls' attitudes towards SRGBV. The table presents summarized versions of the stories for context, but the full stories used in the interviews are available in Annex I. Not surprisingly, most girls reject any form of abuse or violence in school. For example, 99 percent of respondents do not think girls should be subject to emotional abuse (answered disagree or strongly disagree to item 1), and 95 percent do not think that pupils should be slapped by a teacher or school staff (answered disagree or strongly disagree to item 3). Interestingly, there seems to be an imbalance between treatment and control schools; although most responses in both groups are either disagree or strongly disagree, girls in the treatment schools are more likely to answer strongly disagree than in the control schools, and girls in the control schools are more likely to answer (just) disagree than in treatment schools. This pattern is observed across all five questions and the differences are significant across all five questions as well. Girls in control schools and ARD girls seem much more comparable, as the differences between these two groups are significant only for the first question.

The fact that attitudes towards SRGBV girls in treatment and control schools are not observationally equivalent at baseline implies that, for these outcomes, the evaluation should consider these baseline differences. At endline, different options will be considered to incorporate this imbalance in the analysis, for example by controlling for baseline differences in the outcome regressions.

Table 17: Attitudes towards SRGBV

	Control (1)	Treatment (2)	ARD (3)
1. Emotional abuse: <i>Neema just started Form 1. A girl in Neema's class sometimes call her mean names such as "stupid" and "ugly" a few times a week during class.</i>			
Do you believe that pupils should be treated like Neema in or around school?			
Strongly disagree	46.7	52.1	50.4
Disagree	52.8	46.2	47.5
Agree	0.5	0.7	1.1
Strongly agree	0.0	0.9	1.1
F-test		***	**
2. Corporal punishment: <i>Amidah has just joined Form 1. One day Amidah answered a question incorrectly and the teacher hit her on the head with his hand. The next day she gave the wrong answer to a question. As punishment she was required to carry heavy buckets of water from the school to the teacher's house every day for two weeks.</i>			
Do you believe that teachers or school staff should treat pupils with harsh punishment?			
Strongly disagree	41.0	46.0	40.6

	Control (1)	Treatment (2)	ARD (3)
Disagree	55.0	49.4	55.8
Agree	3.9	3.6	2.9
Strongly agree	0.1	1.0	0.7
F-test		***	-
3. Physical violence ^(a) : Do you believe that pupils should be slapped with a hand really hard by a teacher or school staff?			
Strongly disagree	40.6	45.9	40.9
Disagree	54.9	48.9	53.3
Agree	4.1	4.2	5.4
Strongly agree	0.5	1.0	0.4
F-test		*	-
4. Sexual violence: Fatma walks to school every day with other girls. One day one of the boys came up to Fatma and slapped her on her bottom. All the boys laughed. From then on, Fatma tried to hide when she saw that boy, but he still picked on her.” Do you believe that pupils should be treated the way Fatma was treated by a boy at school?			
Strongly disagree	54.1	60.4	58.0
Disagree	45.9	39.3	41.7
Agree	0.0	0.2	0.4
Strongly agree	0.1	0.1	0.0
F-test		**	-
5. Sexual violence ^(b) : Do you believe that pupils should be treated the way Fatma was by a teacher or school staff person at school?			
Strongly disagree	57.3	62.6	57.3
Disagree	42.6	37.0	42.4
Agree	0.1	0.3	0.4
Strongly agree	0.0	0.1	0.0
F-test		*	-

^(a)A story was not included for this question. ^(b)This item referred to the same story as the previous one.

Source: Own calculations using VVV baseline data. Significance levels use the F-test for the joint significance of a regression where the dependent variable is the treatment dummy (or the ARD dummy) and the covariates are dummies for the categorized variable. Standard errors are clustered at the school level. Sample size is 2,220.

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

D4. CONCLUSIONS

In this subsection we summarize the main findings of the report, and provide recommendations when appropriate.

First, the analysis on socio-demographic characteristics showed that:

- Surveyed girls come from relatively disadvantaged backgrounds. Only a third of the girls had eaten before we conducted the interview, and only 28 percent of girls live in households connected to the electric grid. Moreover, parental education is relatively low; only 20 percent of fathers and 12 percent of mothers completed secondary education or more.
- Disability prevalence is low. Only 5 percent of girls reported having difficulties seeing, and similar or lower fractions were found for hearing, washing or dressing themselves, and being understood when

they speak at home. Having difficulty getting around and having difficulty thinking observed higher prevalence rates.

- Self-reported absenteeism is relatively high. On average, 13 percent of girls reported having missed one day of school or more the week before data collection.

We calculated protective asset indices to measure latent traits that we expect to be correlated with school retention. We also showed results for each of the items that constitute these indices, but is important to keep in mind that the evaluation will focus on the impacts of the program on the indices, rather than on the specific items. We studied the properties of the calculated indices in terms of their inter-item correlations and correlation with other variables. The main conclusions of these exercises are as follows:

- Responses reflect an average or above average level of protective assets. For example, for self-esteem, we calculated an average score of 20.5, which is close to the average of 19.5 estimated by Schmitt and Allik (2005) in a previous application of the RSES in Tanzania. The mean of the self-efficacy index is 18.5, which is relatively close to results found by Schwarzer & Scholz (2000) for 22 countries. For other protective asset indices, it is more difficult to establish comparisons with the existing literature because we adapted scales for brevity and our specific project context. However, we found that girls are thinking about the consequences of their decisions, they have listening and communication competencies, and positive social networks with their friends and other individuals in their lives.
- For gender norms, we found that girls seem to believe that there should be gender equity between males and females, especially when we asked items contextualized in school. The results are more nuanced when items are contextualized in their homes. In the latter case, girls seem to hold more traditional views.
- One important limitation for all these indices is that the items used to construct them may be affected by desirability bias (i.e., answering in a manner that is likely to be accepted by others). Previous research indicates this may be a minor concern in this type of context (Dhar et al., 2018); however, a test for the validity of scores will be whether they are correlated with observed school retention. We will analyze this correlation at endline.
- For all indices, we calculated the Cronbach's alpha. The Cronbach's alpha measures the extent to which items in a specific scale are correlated with each other, therefore is considered a measure for scales' reliability. Cronbach's alphas for most indices are within the commonly acceptable range (.65 - .80). Two exceptions are decision making (.55) and gender norms (.61). For decision making, we only included four items in the instrument; expanding the number of items should be considered for the endline data collection. For gender norms, the reason for low inter-item reliability may be that the eleven items that constitute the index could be affected by more than one latent trait. For the endline report, we will document whether the impact evaluation results on gender norms are sensitive to the inclusion of different items in the gender norms score.
- Except for self-esteem, we did not find much evidence that protective asset indices are correlated with parental education. The results are a little different for households' asset ownership; in this case, we found that 4 of 6 protective asset index scores have a positive and significant correlation with the household asset index.
- All protective assets are positively correlated with education expectations, and these correlations are all statically significant. All protective asset indices are negatively correlated with having missed

one day of school or more the week before data collection, although the coefficients are significant for self-esteem and gender norms only.

- Most girls feel safe in school, but roughly half of them do not feel safe around school or on their way to school.
- Teachers and school staff do not seem to be a strong deterrent of violence in schools, and non-negligible fractions of girls believe their discipline methods are harsh.
- Most girls reject any form of SRGBV, as recorded by their attitudes to different scenarios we proposed in the survey.

In terms of balance between treatment and control schools, and ARD girls and girls in control schools, two main conclusions can be derived from the baseline results:

- Treatment and control groups are well-balanced. For the majority of outcomes, we did not find significant differences between treatment and control. One exception is attitudes towards SRGBV, where we found that girls in the treatment schools more strongly reject all surveyed forms of abuse. For the impact evaluation, we will control for these baseline differences.
- ARD girls come from more disadvantaged backgrounds than average girls in the treatment or the control group. ARD girls scored lower in the asset ownership index, spend more time traveling to school, and their parents have less education than the parents of the girls in the control group. Moreover, ARD girls observe significant differences with respect to girls in control schools for self-efficacy, decision making, and social networks. The fact that ARD girls are observationally different from average girls implies that using girls in the control group as a comparison group could provide a biased estimate of the impact of the program for ARD girls. As explained in the EDR, to tackle this problem for the endline impact evaluation, we will model the ARD selection process, use this model to select girls in the control group that would have been selected as ARD if their school had been a treatment school, and use these girls in control schools as the comparison group for ARD girls. This method is similar to other matching methods used in the empirical literature, except that usually matching methods are applied to observational data. In the present case, the comparison group is drawn from an experimental control group.

REFERENCES

- Acevedo, Paloma, Guillermo Cruces, Paul Gertler, Sebastian Martinez (2017). Living Up to Expectations: How Job Training Made Women Better Off and Men Worse Off. NBER Working Paper No. 23264
- 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*.
- 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.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY, US: W H Freeman/Times Books/ Henry Holt & Co.
- Benard, B., (2004). *Resiliency, What We Have Learned*, WestEd, San Francisco, CA.
- Buchmann, Nina, Erica Field, Rachel Glennerster, Shahana Nazneen, Svetlana Pimkina, Iman Sen (2017). *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., & 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.
- Drago, Francesco (2011). Self-esteem and earnings *Journal of Economic Psychology* 32 (2011) 480–488.
- 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 Young Lives, Department of International Development at the University of Oxford.
- Graczyk, P.A. (2000). *Adolescent peer relationships and their association with emotional and physical wellness* (Doctoral dissertation, Northern Illinois University, 2000). *Dissertation Abstracts International*, 60, 12-B.
- 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), pp. 1057-1094

- Meeks-Gardner, J., Guerra, N. G., Williams, K. R., & Walker, I. (2011). The Jamaica Youth Survey: Assessing core competencies and risk for aggression among Jamaican youth. *Caribbean Child Development Quarterly*, 57, 35-53.
- 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, USA.
- NORC at the University of Chicago (2017). USAID/Uganda Performance and Impact Evaluation For Literacy Achievement And Retention Activity (LARA).
- NORC at the University of Chicago (2018). USAID/Tanzania Waache Wasome Evaluation Design Report.
- Population Council (2015). Building Assets toolkit: Developing positive benchmarks for adolescent girls.
- 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*. Washington, DC: U.S. Agency for International Development.
- 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. *Journal of Personality and Social Psychology* Copyright 2005 by the American Psychological Association. 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.
- The Washington Group on Disability Statistics (2017). *Analytic Guidelines: Creating Disability Identifiers Using the Washington Group Short Set (WG-SS) SPSS Syntax*.
- 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.

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.

ANNEX I. INSTRUMENTS



USAID
FROM THE AMERICAN PEOPLE

LET GIRLS LEARN –TANZANIA BASELINE SURVEY- QUESTIONNAIRE

DECEMBER 2017

FIELD CONTROL

District _____

School _____

Name _____

Confirm_name: Confirm that this interview is for [current.GirlID]

Girl_name: Girl name and surname

Start_gps: GPS of interview

Start_timev1: start time of interview

IntIDv1: interviewer Name

Start_timev2: start time of interview

IntIDv2: interviewer Name

Par_consent: Signed parental/caregiver consent (only for interviewers)

Resp_av: Is the respondent available? (only for interviewers)

Consent

Hello, I work for EDI, a research organization, based in Bukoba, and we want to learn about your ideas and beliefs about what girls experience at school. I am here today to ask you some questions that will help us learn how to help girls stay in school. We need your help so we can learn more. It is completely your choice to answer the questions. If you choose not to, we will not tell your teacher, or anyone else, and it will not affect your grade. If you agree to answer the questions, and any of the questions make you uncomfortable, we can skip them. You can also choose to stop answering questions at any time. This is not a test and it will not affect your grade/standing at school. Do you want to answer the questions today?

- Yes
- No [Go to END_TEXT]

Great! I will ask you the same questions in two years, which is why we need to know your name. We will not share your names with anyone at school but it will help us remember who you are when we need to ask you questions in two years. Is that fine with you? Your responses will be kept anonymous and nothing that you say will be linked to your name. The only time we would have to share your answer is if you tell us someone has hurt you. We want to make sure that you are safe at school. It is completely your choice to answer these questions.

DEMOGRAPHICS

Enumerator Read: Now, we are going to ask you some questions so we can learn more about you.

I.1. How old are you?

__ __ years

- Don't know (-98)
- Refused (-99)

I.2. In what grade are you currently?

Form _

- Don't know (-98)
- Refused (-99)

I.2b In which grade where you enrolled last year?

- Standard VII (1)
- Form 1 (2)
- Form 2 (3)
- Don't know (-98)
- Refused (-99)

I.3 Generally, how do you travel to and from school? MARK ALL THAT APPLY

- By foot (1)
- By bicycle or motorcycle (2)
- By car (3)
- By public transportation (eg. boda boda bus) (4) Other, specify _____ (96)

I.4 How far do you travel from home to school, in minutes? _____ minutes [PROG: min 0; max 300]

I.5. What is the primary language that you speak at home?

1	Arusha	19	Jita	37	Nyiramba
2	Barabaig	20	Kara	38	Ogiek
3	Bemba	21	Kerewe	39	Pare
4	Bena	22	Kisankasa	40	Rangi
5	Chagga	23	Kurya	41	Rui
6	Datooga	24	Kwaya	42	Safwa
7	Digo	25	Luguru	43	Simbiti
8	Gogo	26	Maasai	44	Sonjo
9	Hadzabe	27	Makonde	45	Sukuma
10	Haya	28	Mang'ati	46	Sumbwa
11	Hehe	29	Meru	47	Swahili
12	Ikizu	30	Mburu	48	Turu
13	Ikoma	31	Ngasa	49	Yao
14	Iramba	32	Ngoreme	50	Zanaki
15	Iraqw	33	Nyakyusa	51	Zinza
16	Isanzu	34	Nyamwezi	-99	Don't know
17	Isenya	35	Nyaturu	-96	Other (specify)
18	Jaluo	36	Nyika	-97	Refused

I.6. What do you consider to be your ethnic group or community?

1	Arusha	22	Kerewe	43	Pare
2	Barabaig	23	Kine	44	Rangi
3	Bemba	24	Kiroba	45	Rui
4	Bena	25	Kisankasa	46	Ruri
5	Chagga	26	Kurya	47	Safwa
6	Datooga	27	Kwaya	48	Simbiti
7	Digo	28	Luguru	49	Sizaki
8	Gogo	29	Maasai	50	Sonjo
9	Hadzabe	30	Makonde	51	Sukuma
10	Haya	31	Mang'ati	52	Sumbwa
11	Hehe	32	Mburu	53	Surwa
12	Ikizu	33	Meru	54	Swahili
13	Ikoma	34	Nata	55	Turu
14	Iramba	35	Ngasa	56	Yao
15	Iraqw	36	Ngoreme	57	Zanaki
16	Isanzu	37	Nyakyusa	58	Zinza
17	Isenya	38	Nyamwezi	-99	Don't know
18	Jaluo	39	Nyaturu	-96	Other (specify)
19	Jita	40	Nyika	-97	Refused
20	Kabwa	41	Nyiramba		
21	Kara	42	Ogiek		

I.7. Did you eat anything today?

- Yes (1)
- No (2)
- Don't know (-98)
- Refused (-99)

I.8 Do you have a boyfriend? That is, are you in a romantic or intimate relationship?

- Yes (1)
- No (2)
- Don't know (-98)
- Refused (-99)

I.8.1 If yes, how old is the person? [PROG: If I.8=Yes]

_____ years old [PROG: Min 0, max 90]

I.9. How far in school do you expect to go?

- Finish Form 1 (1)
- Finish Form 2 (2)
- Finish Form 3 (3)
- Finish Form 4 (4)
- Finish form 5 (5)
- Finish Form 6 (6)
- Some vocational or technical school after secondary school (7)
- Finish vocational or technical school after secondary school (8)

- Some college or university (9)
- Finish college or university (10)

I.10. If less than secondary school, why? [PROG: If I.9 is less than 4][ENUMERATOR: Probe: Are there any other reasons?][PROG: Select all that apply].

- I am not interested in school (1)
- I will be helping at home (2)
- I will be learning a trade (3)
- I will be working for pay (4)
- It is unsafe for me to walk to school (5)
- I am concerned with safety at school (6)
- School is too far from my home (7)
- School costs too much (e.g. books, uniforms, school supplies, transport) (7)
- Members of my household do not want me to attend school (10)
- Other, specify _____ (96)
- Don't know (-98)
- Refused (-99)

I.11 How many days did you miss school last week?

_____ Days [PROG: Min is 0; Max is 5]

- Don't know (-98)
- Refused (-99)

I.12. If you missed one or more days, what was the main reason? [PROG: If I.11 is greater than 0]

- I was sick (1)
- I had housework/farm work to do (2)
- I had to take care of a child or sick relative (3)
- I had to go to work for money (4)
- I was too tired (5)
- I had plans with friends (6)
- I did not want to go to school (7)
- I was menstruating (8)
- I had school fees or tuition outstanding (9)
- Other, specify _____ (96)
- Don't know (-98)
- Refused (-99)

I.13. What was your score in the Primary School Leaving Exam? *Please provide your best estimate.*

- A (Passed; received a score of 161-200) (1)
- B (Passed; received a score of 121-160) (2)
- C (Passed; received a score of 85-120) (3)
- D (Did not pass; received a score of 45-84) (4)

- E (Did not pass; received a score of (0-44) (5)
- Don't know (-98)
- Refused (-99)

I.14 Do you have difficulty seeing, even if wearing glasses?

- No - no difficulty (1)
- Yes – some difficulty (2)
- Yes – a lot of difficulty (3)
- Cannot do at all (4)
- Don't know (-98)
- Refused (-99)

I.15 Do you have difficulty hearing, even if using a hearing aid?

- No - no difficulty (1)
- Yes – some difficulty (2)
- Yes – a lot of difficulty (3)
- Cannot do at all (4)
- Don't know (-98)
- Refused (-99)

I.16 Do you have difficulty getting around, such as walking or climbing steps?

- No - no difficulty (1)
- Yes – some difficulty (2)
- Yes – a lot of difficulty (3)
- Cannot do at all (4)
- Don't know (-98)
- Refused (-99)

I.17 Do you have difficulty thinking, such as remembering or concentrating?

- No - no difficulty (1)
- Yes – some difficulty (2)
- Yes – a lot of difficulty (3)
- Cannot do at all (4)
- Don't know (-98)
- Refused (-99)

I.18 Do you have difficulty washing all over or dressing yourself?

- No - no difficulty (1)
- Yes – some difficulty (2)
- Yes – a lot of difficulty (3)
- Cannot do at all (4)
- Don't know (-98)

- Refused (-99)

I.19 Do you have difficulty being understood when you speak [prim_lang] at home?

- No - no difficulty (1)
- Yes – some difficulty (2)
- Yes – a lot of difficulty (3)
- Cannot do at all (4)
- Don't know (-98)
- Refused (-99)

Enumerator Read: Now, I would like to ask you some questions about the people who live with you.

I.20. Do you live with your mother?

- Yes (1)
- No (2)
- Don't know (-98)
- Refused (-99)

I.20.1 What is the highest level of education of your mother? [PROG: If I.20=Yes]

- No schooling (1)
- Some primary (2)
- Complete primary (3)
- Some secondary (4)
- Complete secondary (5)
- Higher than secondary (6)
- Other, specify _____ (96)
- Don't know (-98)
- Refused (-99)

I.20.2 Can your mother read and write? [PROG: If I.20=Yes]

- Yes (1)
- No (2)
- Don't know (-98)
- Refused (-99)

I.21. Do you live with your father?

- Yes (1)
- No (2)
- Don't know (-98)
- Refused (-99)

I.21.1 What is the highest level of education that your father? [PROG: If I.21=Yes]

- No schooling (1)
- Some primary (2)
- Complete primary (3)
- Some secondary (4)
- Complete secondary (5)
- Higher than secondary (6)
- Other, specify _____ (96)
- Don't know (-98)
- Refused (-99)

I.21.2 Can your father read and write? [PROG: If I.21=Yes]

- Yes (1)
- No (2)
- Don't know (-98)
- Refused (-99)

I.22. Who is your guardian? Your guardian is the adult who takes care of you the most. [PROG: If I.20=No AND I.21=No]

- Step-mother (8)
- Step-father (9)
- Grandmother (1)
- Grandfather (2)
- Aunt (3)
- Uncle (4)
- Sister (5)
- Brother (6)
- Cousin (7)
- Neighbor (10) Other, specify _____ (96)
- Don't know (-98)
- Refused (-99)

I.22.1. What is the highest level of education that your guardian completed? [PROG: If I.20=No AND I.21=No]

- No schooling (1)
- Some primary (2)
- Complete primary (3)
- Some secondary (4)
- Complete secondary (5)
- Higher than secondary (6)
- Other, specify _____ (96)
- Don't know (-98)
- Refused (-99)

I.22.2. Can your guardian read and write? [PROG: If I.20=No AND I.21=No]

- Yes (1)
- No (2)
- Don't know (-98)
- Refused (-99)

I.23 How many people are currently living at your house? These are the people who live with you at your home, not people who visit for a short time. Please include yourself.

_____ people (including myself) [PROG: Min=1; Max=30]

- Don't know (-98)
- Refused (-99)

Enumerator Read: The next questions are about items in your house.

I.24 Does your household have...

Q		Yes (1)	No (2)	Don't know (- 98)	No response (-99)
a)	a chair in good condition?				
b)	a bed in good condition?				
c)	a clock in good working condition?				
d)	a radio in good working condition?				
e)	a television in good working condition?				
f)	a computer in good working condition?				
g)	a bicycle in good working condition?				
h)	a motorcycle in good working condition?				
i)	a car, truck, or boat with engine in good working condition?				
j)	a refrigerator in good working condition?				
k)	a stove in good working condition? (excluding firewood)				

I.25 Is your household connected to the electric grid?

- Yes (1)
- No (2)
- Don't know (-98)
- Refused (-99)

I.26 Do you or someone in your household have a working cellphone?

- Yes (1)

- No (2)
- Don't know (-98)
- Refused (-99)

PROTECTIVE ASSETS

Enumerator Instructions: The following questions are for the girl to learn how to answer the questions. Please do not rush through these questions, make sure they understand this before proceeding.

Enumerator Read: During the time we are going to be together, I am going to read some statements, please indicate whether each statement is Not true, Slightly true, True or Very true. Before this starts, we will first try some practice questions so you know how to give your answers.

Enumerator Instructions: Introduce the scale, and show the scale to the girl with the gradation in different options.

Enumerator Read: There is no good or bad answer, we just want to know how you feel and think about some things. If you don't understand a question you should always tell me, and I can explain more. If you don't want to answer, just tell me and I'll go to the next question.

I like candy. Is this not true, slightly true, true, or very true? [PROG: No response options needed]

I like spending time with my friends. Is this not true, slightly true, true, or very true? [PROG: No response options needed]

SELF-ESTEEM

Enumerator Read: I am going to read some statements that could possibly apply to you. Please tell me whether you strongly disagree, disagree, agree, or strongly agree.

If you have any questions about the statements, please let me know. I can always repeat the statement more than once, if needed.

[NOTE ON SCORING:

Positive items: For items 1, 2, 4, 6, and 7: Strongly agree = 3; Agree = 2; Disagree = 1; Strongly disagree = 0

Negative items: For items 3, 5, 8, 9, and 10 (which are reversed in valence):

Strongly agree = 0; Agree = 1; Disagree = 2; Strongly disagree = 3]

Q		Strongly Disagree	Disagree	Agree	Strongly agree	Don't know (-98)	No response (-99)
2.1	I feel that I am at least as important as others						
2.2	I feel that I have good qualities.						
2.3	All in all, I feel that I will not succeed in life.						

Q		Strongly Disagree	Disagree	Agree	Strongly agree	Don't know (-98)	No response (-99)
2.4	I am able to do things as well as most other people.						
2.5	I feel I do not have much to be proud of. (For interviewers: IF THE RESPONDENT IS CONFUSED ON THE MEANING OF PROUD EXPLAIN WHAT IT MEANS I am proud that I am kind to others. I am proud that I study hard for my exams. I am proud that I am a good friend.)						
2.6	I have a good attitude toward myself.						
2.7	All in all, I am satisfied with myself.						
2.8	I wish I could have more respect for myself.						
2.9	I feel useless sometimes.						
2.10	Sometimes I think I am not confident (ENUMERATOR: IF RESPONDENT IS CONFUSED ON THE MEANING OF PROUD, EXPLAIN WHAT IT MEANS: For example, I am proud that I am kind to others. I am proud that I study hard for exams. I am proud that I am a good friend.)						

SELF-EFFICACY

Enumerator Read: I am going to read some statements that could possibly apply to you. For each statement, please tell me if this is Not True, Slightly True, True, or Very True.

If you have any questions about the statements, please let me know. I can always repeat the statement more than once, if needed.

[NOTE ON SCORING:

Positive: For items 1, 2, 3, 4, 5, 6, 7, 8, and 9: Very true= 3; True = 2; Slightly true = 1; Not true = 0
Self-efficacy score was generated by adding each item]

Q		Not true (0)	Slightly true (1)	True (2)	Very true (3)	Don't know (- 98)	No response (-99)
5.1	When I have a difficult problem, I try to solve it.						
5.2	It is easy for me to stick to my aims and reach my goals.						
5.3	I am confident that I could deal with a problem when it happens all of a sudden.						
5.4	I try to deal with a problem even when it is hard.						
5.5	I can solve most problems if I try very hard and don't give up						
5.6	I can remain calm when I am upset or mad						
5.7	When I have to solve a problem, I think about possible solutions that will be helpful.						
5.8	I can usually make smart decisions about school						

Q		Not true (0)	Slightly true (1)	True (2)	Very true (3)	Don't know (- 98)	No response (-99)
5.9	I can usually handle difficult situations. (ENUMERATOR: If the respondent is confused, say: If a close relative, like a grandmother, dies I will feel sad but I am still able to go to school and focus on my lessons.)						

DECISION MAKING

Enumerator Read: I am going to read some statements that could possibly apply to you. For each statement, please tell me if this strongly disagree, disagree, agree or strongly agree.

If you have any questions about the statements, please let me know. I can always repeat the statement more than once, if needed.

[NOTE ON SCORING:

For items 1, 2, 3, and 4: Very true= 3; True = 2; Slightly true = 1; Not true = 0
Decision-making score was generated by adding each item]

Q		Not true (0)	Slightly true (1)	True (2)	Very true (3)	Don't know (- 98)	No response (-99)
4.1	When I am making up my mind about something (a decision) I think about all the things that could happen						

Q		Not true (0)	Slightly true (1)	True (2)	Very true (3)	Don't know (-98)	No response (-99)
4.2	When I make a decision, I think of the negative things that could happen (ENUMERATOR: If the respondent is confused ON THE MEANING OF DECISION AND NEGATIVE THINGS, say: For example, If I study for my exam, I will have good marks. If I don't study for my exam, I will not have good marks)						
4.3	I consider different choices before making up my mind about something (ENUMERATOR: If the respondent is confused ON THE MEANING OF DIFFERENT CHOICES, say: For example, If I go to vocational after secondary, I can choose to be a tailor or a cook)						
4.4	I try to make smart decisions for my future.						

COMMUNICATION SKILLS

Enumerator Read: I am going to read some statements that could possibly apply to you. For each item, please tell me whether it is not true, slightly true, true, very true.

If you have any questions about the statements, please let me know. I can always repeat the statement more than one, if needed.

[NOTE ON SCORING:

For items 1, 2, 3, 4, 5, 6, and 7: Very true= 3; True = 2; Slightly true = 1; Not true = 0
Communication skills score was generated by adding each item]

Q		Not true (0)	Slightly true (1)	True (2)	Very true (3)	Don't know (- 98)	No response (-99)
6.1	I ask an adult for help when I need it.						
6.2	I feel comfortable telling a friend how I am feeling.						
6.3	If I am upset at a friend, I can talk to them make things better.						
6.4	I listen to my friends when they tell me something important about themselves						
6.5	I have an easy time talking to a classmate in my school I do not know well.						
6.6	When my friend upsets me, I will tell her						
6.7	I can talk about a problem I have at school with an adult I trust.						

SOCIAL NETWORKS & RELATIONSHIP SKILLS

Enumerator Read: *I am going to read some statements that could apply to you. For each item, please tell me if this is Not True, Slightly True, True, or Very True.*

If you have any questions about the statements, please let me know. I can always repeat the statement more than once, if needed.

[NOTE ON SCORING:

*For items 1, 2, 3, 4, 5, and 6: Very true= 3; True = 2; Slightly true = 1; Not true = 0
Social networks and relationship skills score was generated by adding each item]*

Q		Not true (0)	Slightly true (1)	True (2)	Very true (3)	Don't know (-98)	No response (-99)
7.1	I try and make friends at school or in my community.						
7.2	I try to help my friends when they need something.						
7.3	I have a friend I can tell important things to.						
7.4	I have a friend who would listen if I talked about my problems.						
7.5	I have someone in my life, like my friends, parents, relatives, teachers, that cares about my feelings.						
7.6	I have someone in my life (i.e. friends, parents, relatives, teachers...) that makes me feel important.						

7.7 Is there any person who usually knows where you are during the day?

- Yes (1)
- No (2)
- Don't know (-98)
- Refused (-99)

7.8 Who usually knows where you are during the day? [PROG: If 4.14=Yes]

- Mother (1)
- Father (2)
- Step-mother (11)
- Step-father (12)
- Grandmother (3)
- Grandfather (4)
- Aunt (5)
- Uncle (6)

- Sister (7)
- Brother (8)
- Cousin (9)
- Friend (10)
- Other: _____(96)
- Don't know (-98)
- Refused (-99)

GENDER EQUITY & NORMS

Enumerator Read: The next statements are about how do you think/feel about girls, boys, men, and women should behave. For each item, please tell me whether you strongly disagree, disagree, agree, or strongly agree.

If you have any questions about the statements, please let me know. I can always repeat the statement more than once, if needed.

[NOTE ON SCORING:

Positive: For items 2, 6, 7, 8, 9, 10, and 11: Strongly agree = 3; Agree = 2; Disagree = 1; Strongly disagree = 0

Negative: For items 1, 3, 4, and 5 (which are reversed in valence):

Strongly agree = 0; Agree = 1; Disagree = 2; Strongly disagree = 3

Gender equity and norms score was generated by adding each item]

Q		Strongly Disagree	Disagree	Agree	Strongly Agree	Don't know (-98)	No response (-99)
8.1	Men's job is to earn money for the family and women's job is to look after the family						
8.2	Men should also participate in doing housework						
8.3	Boys are usually more intelligent than girls						
8.4	It's more important for boys to do well in school than it is for girls						
8.5	The man makes the final decision in the household						

Q		Strongly Disagree	Disagree	Agree	Strongly Agree	Don't know (-98)	No response (-99)
8.6	Girls can be leaders at school.						
8.7	Girls should have the same right to go to school as boys						
8.8	A woman and a man should decide together if they want to have children						
8.9	Women should have the right to choose who to marry						
8.10	Girls should be allowed to play sports						
8.11	If I heard a boy insulting a girl, I would tell the boy it was not kind						

9. SRGBV: KNOWLEDGE AND SKILLS

Enumerator Read:

Some questions we ask may seem upsetting, or hard to answer, but try and answer them as best you can. This is not a test. There is no right or wrong answer. If at any point you feel too uncomfortable to continue, we can stop or skip a question. Your answers and the information you share will not be shared with anyone unless you share something that will make you unsafe. We will not tell your parents, teachers or friends your answers.

The next statements are about what happens to students at school or traveling to and from school. For each item, please tell me whether you strongly disagree, disagree, agree, or strongly agree. If you have any questions about the statements, please let me know.

I can always repeat the statement more than once, if needed.

[NOTE ON SCORING:

Positive: For items: 1 and 2: Strongly agree = 3; Agree = 2; Disagree = 1; Strongly disagree = 0

Negative: For items 3, 4, 5, 6, 7, 8, 9, and 10 (which are reversed in valence):

Strongly agree = 0; Agree = 1; Disagree = 2; Strongly disagree = 3

SRGBV knowledge and skills score was generated by adding each item]

	Strongly Disagree	Disagree	Agree	Strongly Agree	Don't know (-98)	No response (-99)
9.1 Girls feel safe at your school						
9. 2 Boys feel safe at your school						
9. 3 There are places near your school where it is not safe for girls to go alone						
9.4 There are places near your school where it is not safe for boys to go alone						
9.5 Girls are not safe traveling to and from your school						
9.6 Boys are not safe traveling to and from your school						
9.7 Teachers and school staff at your school rarely do anything when students hurt other students						
9.8 Students at your school are punished too much for little things.						
9.9 The use of the cane or another form of physical discipline (e.g., pulling ears, kicking, slapping, standing in the sun) is common in your school						

	Strongly Disagree	Disagree	Agree	Strongly Agree	Don't know (-98)	No response (-99)
9.10 Sometimes teachers are unkind to children with disabilities in your school						

EMOTIONAL ABUSE & BULLYING

Enumerator Read: Now I would like to read you a story. The story is not real. We will discuss the story, and then will go on to the second part of the interview. I will ask you about experiences that are similar to the ones in the story. I will ask whether you heard of any specific things that happened to classmates or a friend.

Do you have any questions? Let's continue.

Enumerator Read: We will continue with me reading the first short story. Listen carefully to the story. After I read it, we will discuss the story together, and I will ask a few questions to you.

Let's Begin!

"Neema and Leyla are good friends. Neema just started Form 1. At first, Neema liked going to school but now she does not like school very much. Every day Neema tells Leyla how mean the girls in her class are. A girl pupil in Neema's class sometimes call her mean names such as "stupid" and "ugly" a few times a week during class. Today, Neema told Leyla that the other girls will not let her play with them at break and just ignored her. Neema cried when she told her friend Leyla about this."

What is happening in this story?

INSTRUCTION 3 : Talk about the story together. This is an icebreaker and an opportunity to build trust. You can prompt the respondent about specific events in the story, ask simple questions, and read the whole story again if you believe that it is important to do so. Before going to the next questions, summarize the story. To summarize, you say, "Some girls in Neema's class are bothering her. They call her names, grab her school bag, and will not let her play. Neema is sad."

10.1 Did you understand the story?

- Yes (1)
- No (2)

10.2: Have you seen a student being treated the way Neema was, or have you seen pupils being mean to other girls at your school?

- Yes (1)
- No (2)

- Don't know (-98)
- No Response (-99)

If YES: Please tell me more about this. ENUMERATOR: GENUINELY LISTEN BUT DO NOT WRITE DOWN RESPONSE.

I0.4: Do you believe that pupils should be treated like Neema in or around school?

- Strongly disagree (1)
- Disagree (2)
- Agree (3)
- Strongly agree. (4)
- Don't know (-98)
- No Response (-99)

CORPORAL/HARSH PUNISHMENT

Enumerator Read: Now, we will talk about things that may have happened to pupils/classmates/students and friends at school. I will read a short story. Listen carefully and after I read it we will discuss the story together and I will ask you a few questions.

If you have any questions about the statements, please let me know. I can always repeat the statement more than once, if needed

Amidah has just joined FI. One day in her new class, Amidah answered a question incorrectly and the teacher hit her on the head with his hand. This made Amidah afraid of her teacher. The next day, Amidah was working with a classmate on a task and the teacher heard her talking. The teacher shouted at her and told her to stand in the corner for the rest of the day without any break. Amidah was embarrassed to stand in the corner in front of her new classmates in school. The next day she gave the wrong answer to a question. As punishment she was required to carry heavy buckets of water from the school to the teachers house every day for two weeks. Amidah is no longer happy to go to school.

I1.1 Did you understand the story?

- Yes (1)
- No (2)

I1.2: Have you seen teachers punish pupils like the story describes or in similar ways, like pulling ears, kicking, or slapping, at your school?

- Yes (1)
- No (2)
- Don't know (-98)
- No Response (-99)

I1.3: Do you believe that teachers or school staff should treat pupils with harsh punishment?

- Strongly disagree (1)

- Disagree (2)
- Agree (3)
- Strongly agree. (4)
- Don't know (-98)
- No Response (-99)

PHYSICAL VIOLENCE & SAFETY

Enumerator Read: *Now, we will talk about things that may have happened to pupils/classmates/students and friends at school or when they were walking to and from school.*

If you have any questions about the statements, please let me know. I can always repeat the statement more than once, if needed

12.1 Do you feel safe at school?

- Yes (1)
- No (2)
- Don't know (-98)
- No Response (-99)

12.2: In the last school term have you seen of any girls at your school that were physically hurt, like getting hit, slapped, or punched, by another pupil on their way to and from school?

- Yes (1)
- No (2)
- Don't know (-98)
- No Response (-99)

12.3 During the last school term, have you seen of any students/pupils or friends getting slapped with a hand really hard by a teacher or school staff?

- Yes (1)
- No (2)
- Don't know (-98)
- No Response (-99)

12.4 Do you believe that pupils should be slapped with a hand really hard by a teacher or school staff?

- Strongly disagree (1)
- Disagree (2)
- Agree (3)
- Strongly agree. (4)
- Don't know (-98)
- No Response (-99)

SEXUAL VIOLENCE (PEER TO PEER, TEACHER/STAFF TO STUDENTS)

Enumerator Read: Now we will discuss another short story. Just like the first stories, listen carefully and after I read it we will discuss the story together and I will ask you a few questions.

“Fatma walks to school every day with other girls. One day one of the boys came up to Fatma and slapped her on her bottom. All the boys laughed. From then on, Fatma tried to hide when she saw that boy, but he still picked on her.”

I3.1 Did you understand the story?

- Yes (1)
- No (2)

I3.2 In the last school term, have you seen of any boys treating girls the same way Fatma was treated at school?

- Yes (1)
- No (2)
- No Response (-99)

I3.3 In the last school term, have you seen of a girl being treated this way by a teacher or school staff person at school?

- Yes (1)
- No (2)
- No Response (-99)

I3.4 Do you believe that pupils should be treated the way Fatma was treated by a boy at school?

- Strongly disagree (1)
- Disagree (2)
- Agree (3)
- Strongly agree (4)
- Don't know (-98)
- No Response (-99)

I3.5 Do you believe that pupils should be treated the way Fatma was by a teacher or school staff person at school?

- Strongly disagree (1)
- Disagree (2)
- Agree (3)
- Strongly agree. (4)
- Don't know (-98)
- No Response (-99)

Enumerator Read: That's the very last question I have for you. Thank you so much for taking the time to complete this interview, I realize that some of the questions were difficult. Do you have any questions now that we are finished?

Often, in answering questions such as these, people find that there are issues they would like to discuss further with someone who can give them additional information or support. We don't want to leave anyone with concerns that we might be able to help with. Before I leave, I want to let you know that you can contact me or other members of the research team at any time with questions or concerns. You can also speak with a counselor or go to a local health and psychosocial care providers to discuss the issues you talked about with me today.

NOTE: (There is a counselor here who can speak to her and provide a referral list and resources in the community. Please remember to ask the girl again at the end of the survey if she wants to speak to a counselor and follow the NOTE if the girl is required to speak to a counselor because she reported an incident)

14. CONTACT INFORMATION

Now I would like to ask you some information that could help me remember who you are. The reason I am asking for this information is that I would like to ask you these questions at school again in two years.

Phone_num. What is your phone number? (For Interviewers: +255 (0) - DO NOT WRITE THE LEADING 0. Enter -95 if no phone number can be provided.)

Owner_phone Who is the owner of this phone number?

- Self (1)
- Mother/Female caregiver (2)
- Father/Male caregiver (3)
- Grandparents (4)
- Brother/Sister (5)
- Neighbor (6)
- Aunt/Uncle (7)
- Cousin (8)
- Village leader (9)
- Friend/Schoolmate (10)
- Teacher / School head teacher (11)
- Other, specify (-96)

Name_phone: What's the full name of the owner of the phone number you gave?

(SECOND PHONE NUMBER)

Sec_phone_num. What is your phone number? (For Interviewers: +255 (0) - DO NOT WRITE THE LEADING 0. Enter -95 if no phone number can be provided.)

Sec_owner_phone Who is the owner of this phone number?

- Self (1)

- Mother/Female caregiver (2)
- Father/Male caregiver (3)
- Grandparents (4)
- Brother/Sister (5)
- Neighbor (6)
- Aunt/Uncle (7)
- Cousin (8)
- Village leader (9)
- Friend/Schoolmate (10)
- Teacher / School head teacher (11)
- Other, specify (-96)

Sec_name_phone: What's the full name of the owner of the phone number you gave?

(THIRD PHONE NUMBER)

Third_phone_num. What is your phone number? (For Interviewers: +255 (0) - DO NOT WRITE THE LEADING 0. Enter -95 if no phone number can be provided.)

Third_owner_phone Who is the owner of this phone number?

- Self (1)
- Mother/Female caregiver (2)
- Father/Male caregiver (3)
- Grandparents (4)
- Brother/Sister (5)
- Neighbor (6)
- Aunt/Uncle (7)
- Cousin (8)
- Village leader (9)
- Friend/Schoolmate (10)
- Teacher / School head teacher (11)
- Other, specify (-96)

Third_name_phone: What's the full name of the owner of the phone number you gave?

Email: Can you give me an electronic email address?

- Yes (1)
- No (2)

Electronic email address (if email=YES)

To interviewers: PLEASE, RECORD HERE THE INFORMATION ON THE LOCATION OF THE RESPONDENT'S FAMILY HOUSE.

Urb_rur: Is your home located in urban or rural?

- Urban (1)
- Rural (2)

Region_home: What is the Region?

- List of Tz region

Other_region: Other region, specify

District_home: What is the District?

- List of Tz districts

Other_district: Other district, specify

Ward_home: What is the Ward?

- List of Tz Ward

Other_Ward: Other Ward, specify

Village_home: What is the Village? (if urb_rur=Rural 2)

- List of Tz Village

Other_Village: Other Village, specify

Mtaa_home: What is the Mtaa? (if urb_rur=Urban 1)

- List of Tz Mtaa

Other_Mtaa: Other Mtaa, specify

Subvillage_home: What is the Sub-village? (if urb_rur=Rural 2)

- List of Tz sub-village

Other_subvillage: Other sub-village, specify

Direction_home: Direction to reach the house (for interviewers)Please write directions on how to find the home of the respondent. Start with a location that is well known in the area as a reference point. In the directions, include landmarks, roads, and any other detailed information where relevant. References to specific businesses or homes "(ask for the home of...)" should also be included where possible.

Mail_home: Does your household have an address where you collect mails?

- Yes (1)
- No (2)

Number_post: If yes, Postal address number: (if mail home=yes)

Road_home: If yes, Road name / Neighborhood: _____

For interviewers: RECORD HERE THE NAME OF THE MOTHER (FEMALE CAREGIVER) AND FATHER (MALE CAREGIVER) OF RESPONDENT, OR OF THE PERSON WHO GAVE CONSENT.

Mother_name: What is the name of your mother / female caregiver? (for interviewers: RECORD THE FULL NAME. WRITE NONE IF THERE IS NO MOTHER/FEMALE CAREGIVER)

Father_name: What is the name of your father / male caregiver? (for interviewers: RECORD THE FULL NAME. WRITE NONE IF THERE IS NO FATHER/MALE CAREGIVER)

Consent_name: Who came today for giving the consent?

- Mother / female caregiver (1)
- Father/ male caregiver (2)
- Granmother/grandfather (3)
- Aunt/uncle (4)
- Sister / Brother (5)
- Cousin (6)
- Other family members (specify) (7)
- Teacher (8)
- 9 Other, specify (-96)

Cosent_other: Specify the name of the person (if consent_name not in 1,2).

NOTE: If the respondent (girl) has disclosed an incident about herself that is unsafe and is indicated on the risk and response protocol guide, READ the following to the respondent (girl): “You had shared with me an incident that happened to you at school that is unsafe. The counselor is going to come and talk to you to make sure you ok. She is very nice and I am going to go get her. “

NOTE: All girls are allowed to speak to the counselor if they want to at the end of the survey voluntarily. If the girl is not required to see a counselor because she reported an incident (see above). Please remind her she can talk to a counselor at the end of the survey.

ENUMERATOR QUESTIONS

14.1: Did the respondent appear to understand the questions?

- Never (1)
- Seldom (2)
- Some of the time (3)
- Most of the time (4)

I4.2: Overall, how was the respondent's interest in the interview?

- Very low (1)
- Below average (2)
- Average (3)
- Above average (4)
- Very high (5)

I4.3: Does the respondent need a referral to any service providers? (Note: this is just a follow up question so that it is noted on the form for the researchers)

- Yes (1)
- No (2)

I4.4: Did the respondent (girl) need to speak to the counselor?

- Yes (1)

No (2)

Enumerator, please provide any comments that might be helpful to our understanding the data collected on this survey (e.g. if you feel the girl was not being truthful, if the girl seemed upset or agitated, if you felt e.g. if you feel the girl was not being truthful, if the girl seemed upset or agitated, if you felt uncomfortable in anyway, etc.).

COMMENTS:

END OF SURVEY, GREAT WORK!

ANNEX II. 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 the reader with a brief overview of the fieldwork set-up, the sample, and data collection protocols.

QUESTIONNAIRE PRE-TESTING

Testing the appropriateness of the survey instruments to local conditions, as well as verifying the translation of the questionnaire, is essential. Testing was split into two activities; the first was desk testing of the tools which took place on the 17th – 18th January 2018. Interviews were carried out in EDI Bukoba office, with EDI personnel not involved in the project participating as interviewees. The second consisted of a field pre-test of the girls' quantitative survey between the 29th – 31st January 2018, in three schools, in Kagera region, Bukoba Municipal (MC) and District (DC). The schools used for the pre-testing were selected based on their similarity to the schools in the sampled districts – that is, schools serving students of poorer demographics in urban and rural areas. The goal of the pre-test was to ensure that the instruments are capable of capturing the data required by NORC, as well as identifying the need for any additional consistency checks, and testing the routing, length, and general flow of the questionnaire.

REPORTING AND SENSITIZATION TO DISTRICTS

EDI carried out a reporting and sensitization mission to the four districts in the sample. This activity took place between 23rd - 31st January, after the visit of WEI to districts to randomly run the sample to select the 61 schools and identify the treatment and control schools. The mission aimed at explaining the Waache Wasome data collection and program, as well as introducing the difference in the activities implemented in treatment and control schools and clarify the importance of having treatment and control schools for analysis purposes.

CHILD PROTECTION PROTOCOLS

Children and adolescents can experience distress from some of the questions concerning SRGBV. Although the questionnaire did not ask about direct experiences with SRGBV, they could remind the girl of past events, either a firsthand experience or witnessing a violent act. For this reason, each survey team included trained counselors to be available at each interview site to work with girls who might become distressed because of the interview.

In addition, NORC's gender expert worked with the District Social Welfare Office in Bukoba to adapt and finalize a referral protocol for cases where respondents disclosed being at risk. This protocol outlined the steps to refer a child to the appropriate services if necessary. Counselors were trained by NORC's gender expert on how to navigate the protocol as well as on other tools to deal with disclosure of violence or risk situations during interviews.

TRAINING

Training included a supervisors training and the main interviewers training. Supervisor training took place between the 25th – 27th January 2018 (three days) at the EDI office in Bukoba and was conducted

by the EDI coordination team. The training consisted of an introduction to the project, training on the girls' questionnaire, and on the random selection questionnaire.

The main interviewers training took place between the 1st – 9th February 2018 (nine days), and was attended by 38 female interviewers, six supervisors and seven female counsellors, and was conducted by the survey manager, field coordinator and data manager from EDI, and by representatives of NORC at the University of Chicago.

The training introduced trainees to the project, and included in-depth training on:

- Waache Wasome project and program;
- Girls' questionnaire;
- Working in school: best practices;
- Confidentiality and privacy;
- Risk and response protocol and guide;
- Introduction to surveybe and data transfer;
- A detailed look at the field protocol and selection of girls, interview results;
- Data security; and,
- Administration and preparation to field.

During the training program, mock interviews were carried out, as well as two outdoor practices in schools. Finally, on the last day of training 36 of the 38 trainees, and six out of the seven counsellors were hired, with the remaining trainees and counsellor designated as replacements.

Questions were added and removed, response lists and question texts were better contextualized, translation improved, and the surveybe instruments fine-tuned. Questionnaire length and interviewing time were tested at this stage in order to check that the time necessary to complete an interview was suitable to accomplish the planned field work schedule.

FIELDWORK SUMMARY

Data collection took place between 12th February – 12th March 2018, with the treatment schools visited first then the control schools. The schools were spread over the four districts of Karatu DC in Arusha region and Musoma DC, Musoma MC, and Tarime DC in Mara region. In total, 2,200 girls were interviewed.

One week prior to the fieldwork, pre-visits took place: a team composed by EDI staff pre-visited the 61 schools in order to randomly select a maximum of 46 girls enrolled in Form I per school. During the pre-visits, the team introduced the scope of the survey, drew the sample, and delivered invitation letters and consent forms to the sampled girls.

Two days before the visit of field teams to the schools, the supervisors of each team contacted the Head Teachers in order to remind them about the visit to the school. Then, during the two days in the schools, the teams carried out interviews after collecting the parental signed consent forms and recording the information on the parents.

TEAM SETUP

The project was led by field and technical oversight, a survey oversight, a field coordinator, and a data processing officer. In addition, interviews were conducted by 36 interviewers, divided in six teams, each

team headed by a supervisor, and supported by a counsellor. Finally, a pre-visit team composed by two staff from EDI visited the schools one week prior to the data collection in order to randomly select the girls in each school.

FIELDWORK SUMMARY (RANDOM SELECTION QUESTIONNAIRE)

The pre-visit team visited the schools one week before the field-team in order to randomly select 46 girls in each school (38 to be interviewed and 8 for replacement).

The random selection questionnaire was administered to Head Teachers, teachers, teachers on duty or any other informed people, and it includes information on the total number of students enrolled in Form I in the school and on the number of girls and boys in each school. In addition, the questionnaire randomly selected a maximum of 46 girls per school, through the generation of random numbers. Using the attendance registers, the EDI staff were counting the number of girls enrolled in Form I in the school and assigning them a number from 1 to n , where n is the total number of girls in Form I in that school. Then, after inserting the total number, the questionnaire returned (maximum) 46 empty rows. Next to each ID, the EDI staff were filling the information (name and surname) on the girl, as reported in the attendance register. Finally, they delivered the invitation letters to the sampled girls to invite the parents to the school meeting as well as the informed consent forms (to be signed at home).

ELIGIBILITY CRITERIA AND REPLACEMENT

The girls are considered eligible if:

- They are enrolled in Form I;
- They are randomly selected from surveybe during the pre-visits;
- There is the signed parental consent form; and
- The girls consent to be interviewed and are available for the interview.

If any of the above criteria is not met, then the girl is considered as not eligible, and the interview results will reflect the reasons. Therefore, the non-eligible girls will be replaced, using any of the eight replacement IDs sampled during the pre-visits.

FIELDWORK SUMMARY (GIRLS QUESTIONNAIRE)

The field teams visited the 61 sampled schools between February 12th – March 12th. The treatment schools were visited before than the control schools. The field teams spent two days in schools: on day one the parental meeting took place, the signed parental consent forms were collected, and interviews with girls were carried out; on day two, additional consent forms were collected (if any), and the remaining girls were interviewed. In some schools, the teams spent only one day as the target of completed interview files was reached.

In total, 2,200 girls were interviewed and the target of completed interview files was reached in the schools. The following section summarizes the interview results.

ETHICS REVIEW

All instruments were reviewed and approved by the Tanzania Commission for Science and Technology (December 20th, 2017), as well as by NORC's IRB (November 15, 2017).

INTERVIEW RESULTS

A total of 2,431 interviews were attempted, 2,220 of which were completed. The main reason why sampled girls were not interviewed was that no parental consent was secured. This does not mean that parent refused, only that no consent letter was received and parents did not attend the meeting. The following table shows the baseline data collection results.

Table 18. Baseline data collection results

Interview Result	N
Completed	2,220
Parental consent was not received	167
Parent refused	10
Absent from School	29
Refused	3
Other	2
Total	2,431

Source: WW Baseline data

QUALITY CONTROL

EDI has applied a range of data quality protocols in the collection, checking, and cleaning, and translation of the data to ensure the data meets EDI's high quality standards.

DATA COLLECTION USING CAPI

All data for this project were collected using EDI's CAPI software surveybe. Interviews were administered in the form of electronic survey questionnaires on android tablets. Automated routing and many built-in consistency checks allowed the identification of errors and missing fields during the interview, while interviewers still had the opportunity to correct the response with the respondent at the source.

ANNEX III. BALANCE BETWEEN RANDOMLY SELECTED GIRLS FOR POY CLUBS AND CONTROLS

Table 19 compares sociodemographic characteristics and protective asset indices between girls randomly selected for POY Clubs and girls in the control group. There are no major differences between these two groups of girls, and only one case, namely the gender norms index, the difference is statistically significant.

Table 19. Balance between POY Club-randomly selected girls and girls in the control group

	Control (1)	Randomly assigned (2)	Diff (3)=(1)-(2)
Sociodemographic characteristics			
Age	14.6	14.5	0.1
Number of household members	6.6	6.5	0.2
Connected to electric grid	0.4	0.3	0.1
Cell phone ownership by parents/family	0.3	0.3	0
Did you eat breakfast today?	1.0	1.0	0
Asset index	4.4	4.4	0.1
Protective asset indices			
Self-esteem index	20.5	20.8	-0.3
Self-efficacy index	18.8	18.2	0.6
Decision-making index	8.0	7.9	0.1
Communication skills index	15.2	15.5	-0.3
Social networks index	13.0	12.9	0.1
Gender norms index	23.1	24.1	-1.0**

Source: Own calculations using WW baseline data. Standard errors are clustered at the school level.

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

ANNEX IV. DISTRIBUTIONS OF PROTECTIVE ASSET INDICES AND GENDER NORMS

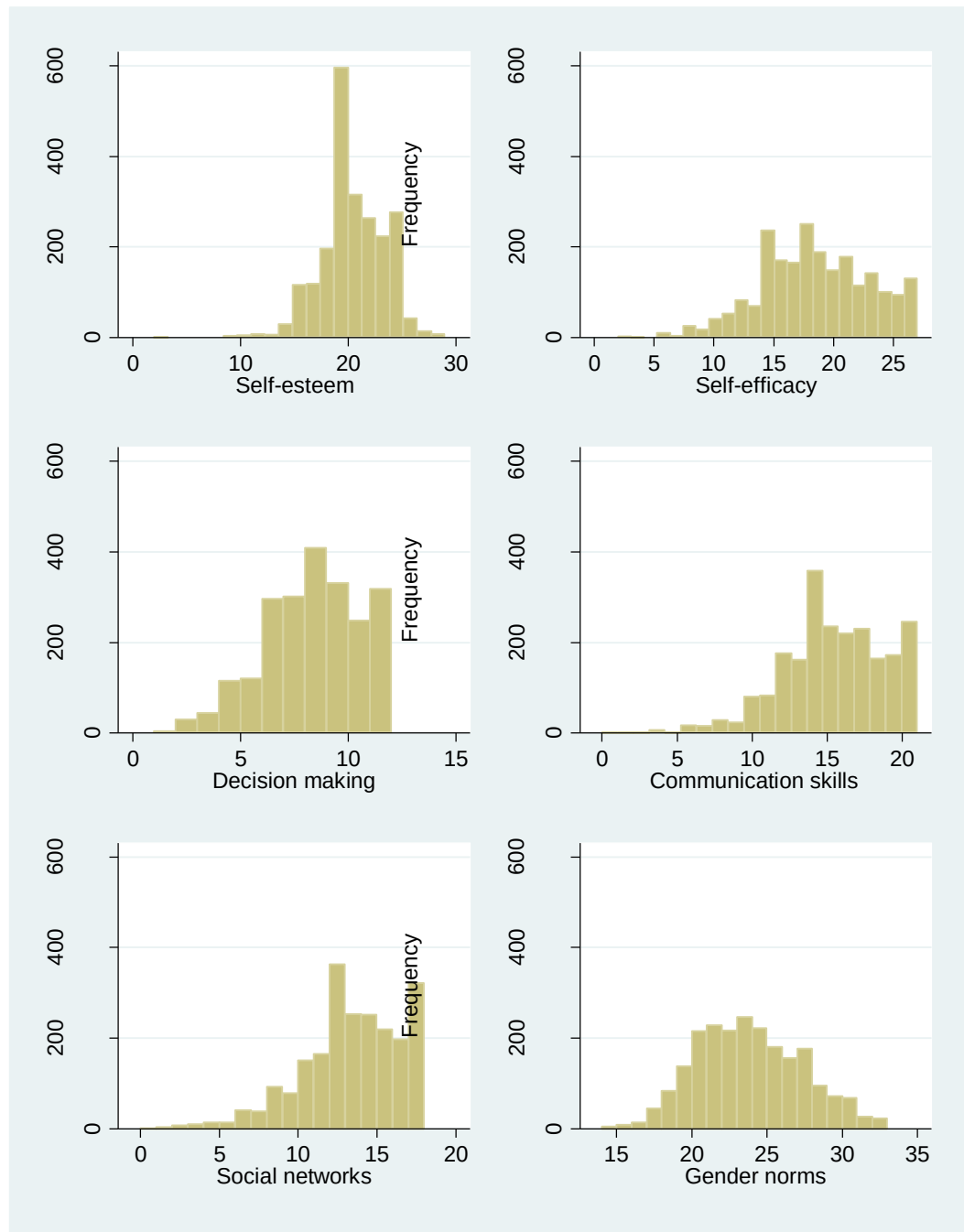
In this section we analyze the distributions of the calculated protective asset scores and the gender norms indices.

Figure 12 shows the observed distribution of the scores. For self-esteem, and gender norms the indices are relatively symmetric, with most of the mass around the mean value and thin tails at both sides of the distribution, meaning that there are relatively few girls with very high or very low levels of self-esteem or gender norms scores.

For self-efficacy, decision-making, communication skills and social networks, on the other hand, the indices seem to be truncated at the right tail. For these protective assets a non-negligible fraction of girls scored the maximum possible score. This suggests that the indices fail to discriminate variability between girls with high scores as several are reaching the maximum. If the latent traits that we are trying to measure have a normal distribution (or other distribution with thin tails at both sides) the fact that these scores are truncated to the right suggests that the surveyed items are not capturing the entirety of the variation of the trait we were trying to measure. Including more items to measure these latent traits should be considered for the endline data collection.

However, it is important to keep in mind the number of items we include in the endline survey is restricted by survey fatigue. The average duration of the baseline interview was 55 minutes. Adding more items will extend the interview duration, which could deteriorate the quality of the responses. Along these lines, rather than adding more items perhaps a more appropriate option is to turn the score into a binary variable to classify girls with 'high' and 'low' level of decision making, using a predetermined threshold (e.g. the median score).

Figure 12. Protective indices distributions



Source: Own calculations using WW baseline data.

ANNEX V. ANALYSIS ON GENDER NORMS ITEMS

In this section we conduct factor analysis on the 11 items included in the girls' survey to measure gender norms. The Cronbach's alpha for the items is only 61, and the eigenvalue of the second factor is lower than one but relatively high (.89). This may be reflecting the fact that more than one factor is needed to model these eleven items.²⁴ Table 18 shows the factor loadings for the four retained factors using the principal factor method. Items 1 and 5 are the only ones that observe a negative loading with the first factor.

Table 20. Factor loadings for the 11 items on gender norms

	Factor1	Factor2	Factor3	Factor4
Item 1	-0.11	0.36	0.14	0.06
Item 2	0.33	-0.04	0.03	0.07
Item 3	0.33	0.54	-0.07	-0.05
Item 4	0.42	0.49	-0.09	-0.05
Item 5	-0.02	0.39	0.12	0.06
Item 6	0.49	-0.13	0.03	0.06
Item 7	0.54	-0.03	-0.06	0.11
Item 8	0.59	-0.15	0.04	-0.09
Item 9	0.47	-0.07	0.15	-0.10
Item 10	0.58	-0.10	-0.03	0.08
Item 11	0.45	-0.14	0.01	-0.04

The eleven survey questions were related to gender norms, attitudes and roles in different contexts. Table 19 shows each question the girls were asked. Most items are asking about gender norms and roles of boys and girls in a school context. While a few items are more associated with gender norms and roles of adults (men/women/mother/father). Items 1 and 5, together with item 2, refer to gender norms in the context of a man and woman's role in the home. On the other hand, items 8 and 9 are not referring to the school or the home girls are currently living, but more to gender relationship norms and decisions girls may face in the future.

Table 21. Context for gender norms questions

	Home	School	Future
1. Men's job is to earn money for the family and a women's job is to look after the children	X		
2. Men should also participate in doing housework	X		
3. Boys are usually more intelligent than girls.		X	
4. It's more important for boys to do well in school than it is for girls		X	
5. The man makes the final decision in the household	X		
6. Girls can be leaders at school		X	
7. Girls should have the same right to go to school as boys		X	

²⁴ It is worth mentioning that for most of the other domains documented in the report the factor loadings for the first factor are all positive. One exception is self-esteem that has two items that observe negative loadings.

	Home	School	Future
8. A woman and a man should decide together if they want to have children.			X
9. Women should have the right to choose who to marry			X
10. Girls should be allowed to play sports		X	
11. If I heard a boy insulting a girl, I would tell the boy it was not kind		X	

If girls' beliefs and attitudes towards gender norms and roles are driven by different latent traits depending on the context they are considering, using only one score that summarizes the eleven items may not be an appropriate approach. As the focus of the analysis in this evaluation is the school, one alternative is to construct the gender norms index using only school-related items. For the endline report we will document whether the impact evaluation results are sensitive to the inclusion of different items in the gender norms index. For example, we will conduct the evaluation analysis using all the 11 items on gender norms, and also using an alternative version of the index where we drop home-related items (i.e. 1, 2 and 5).

That being said, it is also important to keep in mind that WW may affect gender norms across different contexts (not only the school), so including all items in the analysis may provide a better outlook of the program effect, even if the inter-correlation of the items is low. Moreover, if the program has an impact on gender norms across different contexts, we should probably observe an increase on the inter-correlation of items for the treatment group between baseline and endline. This and other related analyses will be conducted at endline.

ANNEX VI. BIVARIATE REGRESSIONS

Table 22 shows bivariate regressions for each protective asset index with the rest of indices. All coefficients are positive and significant.

Table 22: Bivariate regressions between protective assets

	Self-esteem	Self-efficacy	Decision-making	Communication skills	Social networks	Gender norms
Self-esteem	1					
Self-efficacy	0.17***	1				
Decision-making	0.25***	1.04***	1			
Communication skills	0.18***	0.72***	0.30***	1		
Social networks	0.20***	0.66***	0.28***	0.69***	1	
Gender norms	0.25***	0.36***	0.21***	0.28***	0.23***	1

Source: Own calculations using VVV baseline data. Standard errors are clustered at the school level.

*** p<0.001

U.S. AGENCY FOR INTERNATIONAL DEVELOPMENT

1300 Pennsylvania Avenue, NW

Washington, DC 20523