Project Evaluation Report

Report title:	Supporting Adolescent Girls Education in Zimbabwe
Evaluator:	School-to-School International
GEC Project:	Supporting Adolescent Girls' Education (SAGE)
Country	Zimbabwe
GEC window	Leave No Girl Behind
Evaluation point:	Baseline
Report date:	December 2019

Notes:

Some annexes listed in the contents page of this document have not been included because of challenges with capturing them as an A4 PDF document or because they are documents intended for programme purposes only. If you would like access to any of these annexes, please enquire about their availability by emailing uk_girls_education_challenge@pwc.com.

Girls' Education Challenge

UKaid from the British people

Supporting Adolescent Girls Education in Zimbabwe

Baseline Report FINAL

December 2019

Contents

List of Figures

List of Tables

List of Supplementary Tables

Acronyms and Abbreviations

ALP	Accelerated Learning Programme
CBLH	Community-based Learning Hub
CoGE	Champions of Girls Education
EE	External Evaluator
EGMA	Early Grade Mathematics Assessment
EGRA	Early Grade Reading Assessment
FGD	Focus Group Discussion
FM	Fund Manager
GBV	Gender-based Violence
HoH	Head of Household
IO	Intermediate Outcomes
ISOP	Integrated Skills Outreach Programme
KAP	Knowledge, Attitudes and Practices
KII	Key Informant Interview
LNGB	Leave No Girl Behind
MEL	Monitoring, Evaluation and Learning
MoPSE	Ministry of Primary and Secondary Education
NFE	Non-formal Education
OOS	Out-of-School
SAGE	Supporting Adolescent Girls Education
SRHR	Sexual Reproductive Health and Rights
STS	School-to-School International
ToC	Theory of Change
VSLA	Village Savings and Loans Associations

1. Cover sheet

Provide the following information:

- Name of project: Supporting Adolescent Girls Education (SAGE) in Zimbabwe
- Name(s) of Authors: Hetal Thukral, Laura Conrad, Anne Laesecke, Aimee Reeves, Melyssa Sibal, Laura Zasoski
- Name of external evaluation firm: School-to-School International
- Version number: V4
- Date: 13 December 2019

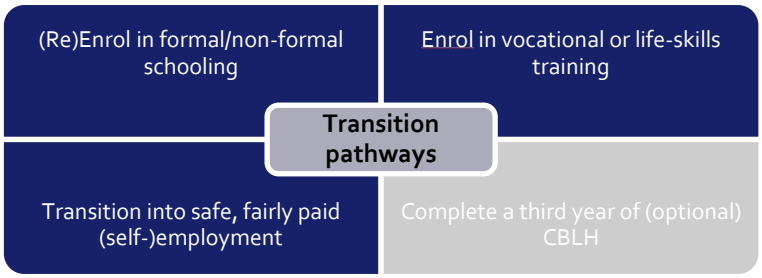
2. Executive summary

Background

Supporting Adolescent Girls Education (SAGE) is a 6-year programme funded by the United Kingdom's Department for International Development (DFID). It is being implemented by a consortium of diverse partners led by Plan International UK and supported by Apostolic Women's Empowerment Trust, Christian Blind Mission (CBM), Econet, Open University and Plan International Zimbabwe.

Working across 11 districts in Zimbabwe, SAGE aims to support 21,780 marginalised girls aged 10–19 to improve their learning outcomes and transition to further education or employment. SAGE will implement accelerated non-formal education (NFE) in 132 community-based learning hubs (CBLHs) with a focus on learning, transition and sustainability (Figure 1).

Figure 1: SAGE transition pathways



Approach

The external evaluation of the SAGE programme employs a mixed-methods, longitudinal, cross-over design. The evaluation utilises data from learning assessments, a package of quantitative and qualitative instruments and ongoing project monitoring tools. The variety of tools, respondents and methods allow data to be triangulated and linked across evaluation questions and project outcome indicators. Because SAGE will roll out activities in a cohort design across 4 cohorts, the evaluation uses subsequent cohorts as a comparison group. Girls enrolled in Grades 3, 5 and 7 in formal school were assessed and their scores were used to establish benchmarks.

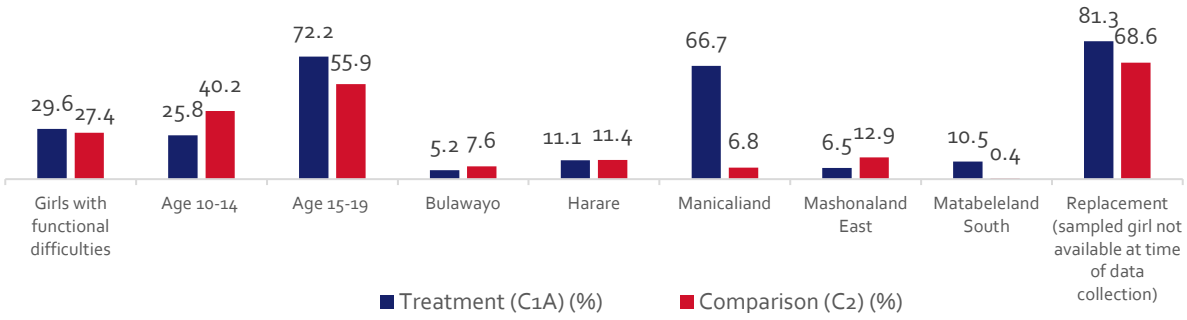
The baseline sample is drawn from cohort C1A, while the second cohort (C2) acts as the comparison group for C1A. The next planned data collection will be in 12 months and will collect data from C1A. The baseline sampling and data collection took place approximately 2 months after girls enrolled in the SAGE programme.[1] For the baseline study, quantitative data were collected from 35 CBLHs in C1A, serving as a treatment cohort, and from 12 communities in C2, serving as a comparison cohort (**Error! Reference source not found.**2).

[1] The start date for learning sessions in CBLHs varied by site; in some cases, girls continued to enrol in the weeks following the start of lessons as well as after the start of baseline data collection.

Figure 2: Baseline sample sizes

Staff from the SAGE programme pre-identified marginalised subgroups and enrolled girls from these subgroups during beneficiary selection—including girls between the ages of 10 and 19 who had less than a Grade 5 equivalent literacy and numeracy ability.[2] The demographics of girls selected for the baseline samples are described in Figure 3. The expected proportion of girls with disabilities was 6%; the baseline data suggest that over 29% of girls in the treatment cohort and over 27% of girls in the comparison cohort had at least one functional difficulty. A high replacement[3] rate in the baseline sample—4 out of 5 girls in the treatment cohort and 2 out of 3 girls in the comparison cohort—means that beneficiary identification data points were not available for the majority of the sample at the time of baseline. As a result, analyses by some subgroups were not initially possible: caregiver status, marital status, whether they had ever been to school, grade level at which they dropped out, and religion. Subsequently, SAGE staff collected data from two-thirds of girls in the treatment cohort – C1A – on these missing data points. The results of these additional analyses on the sample for whom the data are available, are provided in Annex 19.

Figure 3: Baseline sample by subgroups—functional difficulty, age, district and replacement status



Educational marginalisation analysis, barriers and analysis of projects' gender approach

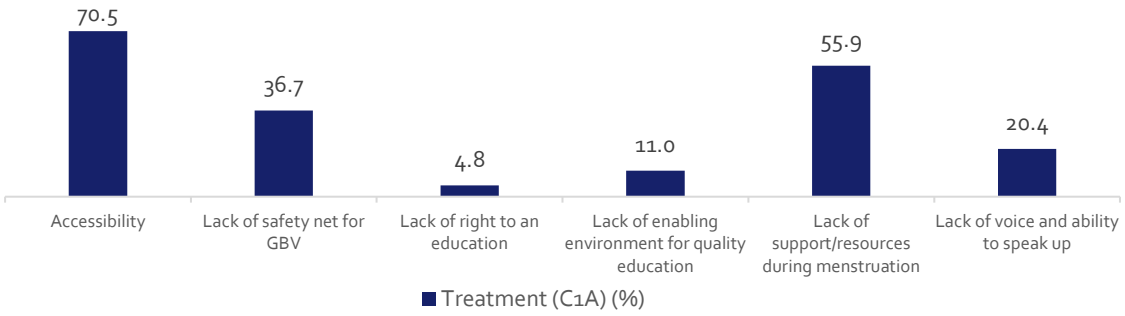
The Gender Analysis, which was completed prior to the baseline, identified key barriers that negatively impacting girls' ability to participate in schooling. These barriers included the accessibility of school, with girls referencing long distances and safety traveling to/from school as issues; gender-based violence (GBV), including early marriage, early pregnancy, sexual exploitation and violence; lack of familial approval for girls' education; lack of access to sexual reproductive health and rights (SRHR) education and low self-esteem as potential barriers to girls'

[2] Using the WRAT assessment by MoPSE.

[3] At baseline, SAGE provided STS with a list of girls enrolled in CBLHs. STS used this list to randomly select girls to participate in the baseline evaluation. During data collection, if the pre-identified girls were not available on the day of data collection, SAGE identified replacement girls to participate in the study.

education. The baseline report uses survey items to quantify some aspects of the barriers girls may face. (Figure 4). More girls in the treatment cohort reported facing accessibility as a barrier than any other; lack of support and resources during menstruation was second most common.[4]

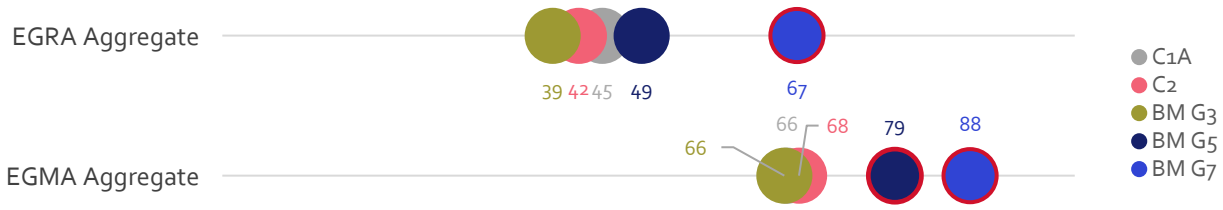
Figure 4: Percent of C1A girls experiencing barriers identified in the gender analysis and baseline evaluation



Baseline levels

***Learning assessments*[5] —** When assessed for their literacy skills using an Early Grade Reading Assessment (EGRA), girls' in both the treatment and comparison cohorts performed at a Grade 3–5 level; there were no significant differences between their scores. Girls in Grade 7 scored significantly higher than both the treatment and comparison cohorts. When assessed for numeracy using an Early Grade Mathematics Assessment (EGMA), treatment and comparison cohort girls performed at the Grade 3 level; there were no significant differences between their scores. Girls enrolled in Grades 5 and 7 had scored significantly higher than both the treatment and comparison cohorts (Figure 5).[6]

Figure 5: EGRA and EGMA aggregate mean scores for girls in C1A, C2 and Grades 3, 5 and 7



Note: Mean aggregate scores are shown for C1A (treatment cohort), C2 (comparison cohort) and benchmark grades 3, 5 and 7. The group(s) with significantly higher scores than the remaining groups is indicated with a red outline.

Figure 6 shows the distribution of the EGRA aggregate score for the treatment cohort in increments of 10. While the aggregated mean score was 44.55, as many girls (n=55) scored zero as scored between 40–50 (n=56); fewer than half that number of girls (n=27) had the highest possible score.

[4] The accessibility barrier captured girls who reported traveling more than 30 minutes to CBLH. The menstruation barrier included girls who said they do not have materials to use during menstruation, missed school because of menstruation or had no one to talk to about menstruation.

[5] Both learning assessments define learner categories as non-learners who answered 0% of questions correctly, emergent learners who answered 1–40% of questions correctly, established learners who answered 41–80% of questions correctly and proficient learners who answered 81–100% of questions correctly.

[6] The aggregate EGRA and EGMA scores are computed per FM guidance. The EGRA score is an average percentage correct for all subtasks except Oral Reading Fluency, which is on a 0–100 fluency scale of correct words per minute. The EGMA score is an average percentage correct for all subtasks. The resulting scale is 0–100 for both tests.

Figure 6: Distribution of aggregate EGRA scores, C1A

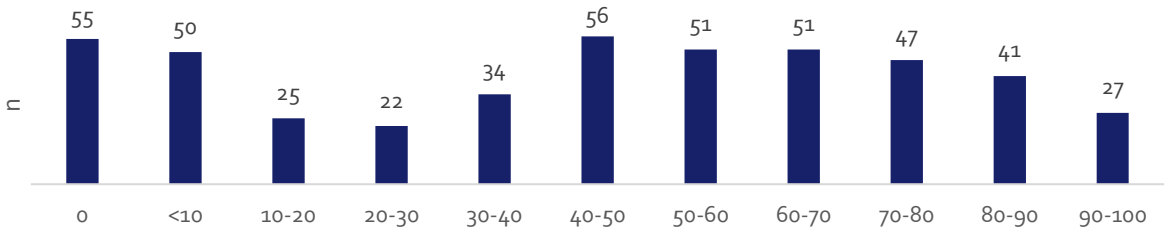


Figure 7 compares the EGRA performance for girls in the treatment cohort to those in Grade 5 using proportions of learner categories.[7] Overall, girls in the treatment cohort struggled most with the reading and listening comprehension subtasks followed by the decoding subtasks. More than one-third (41.61%) and almost half (48.15%) of girls scored as non-learners in the short and long reading comprehension subtasks respectively. More than one-third (37.69%) of girls scored as non-learners on the listening comprehension subtask, and more than one-quarter (29.98%) of girls scored as non-learners on the letter sound identification subtask. Although girls in Grade 5 also struggled with reading comprehension, a smaller proportion—only about one-quarter (19.83% and 38.02%, respectively, on the short and long passages)—scored as non-learners. Girls in the treatment cohort and in Grade 5 are most likely to be proficient on the familiar word subtask (60.78%). When asked to read a long and short passage of connected text, nearly one-quarter (23.53% and 24.62%, respectively) of girls in the treatment cohort scored as non-learners while only 3 percent of Grade 5 girls did. Conversely, more than one-third (41.18% and 35.95%, respectively) of girls in the treatment cohort and a similar proportion of girls (45.45% and 33.06%, respectively) in Grade 5 were proficient at reading the short and long passages.

[7] The average EGRA aggregate score of girls in Grade 5 was comparable to the scores of girls in Cohort 1A. As a result, the performance by learner categories of these two groups are compared at the subtask level.

Figure 7: Proportion of girls in learner categories by EGRA subtask, C1A and Grade 5

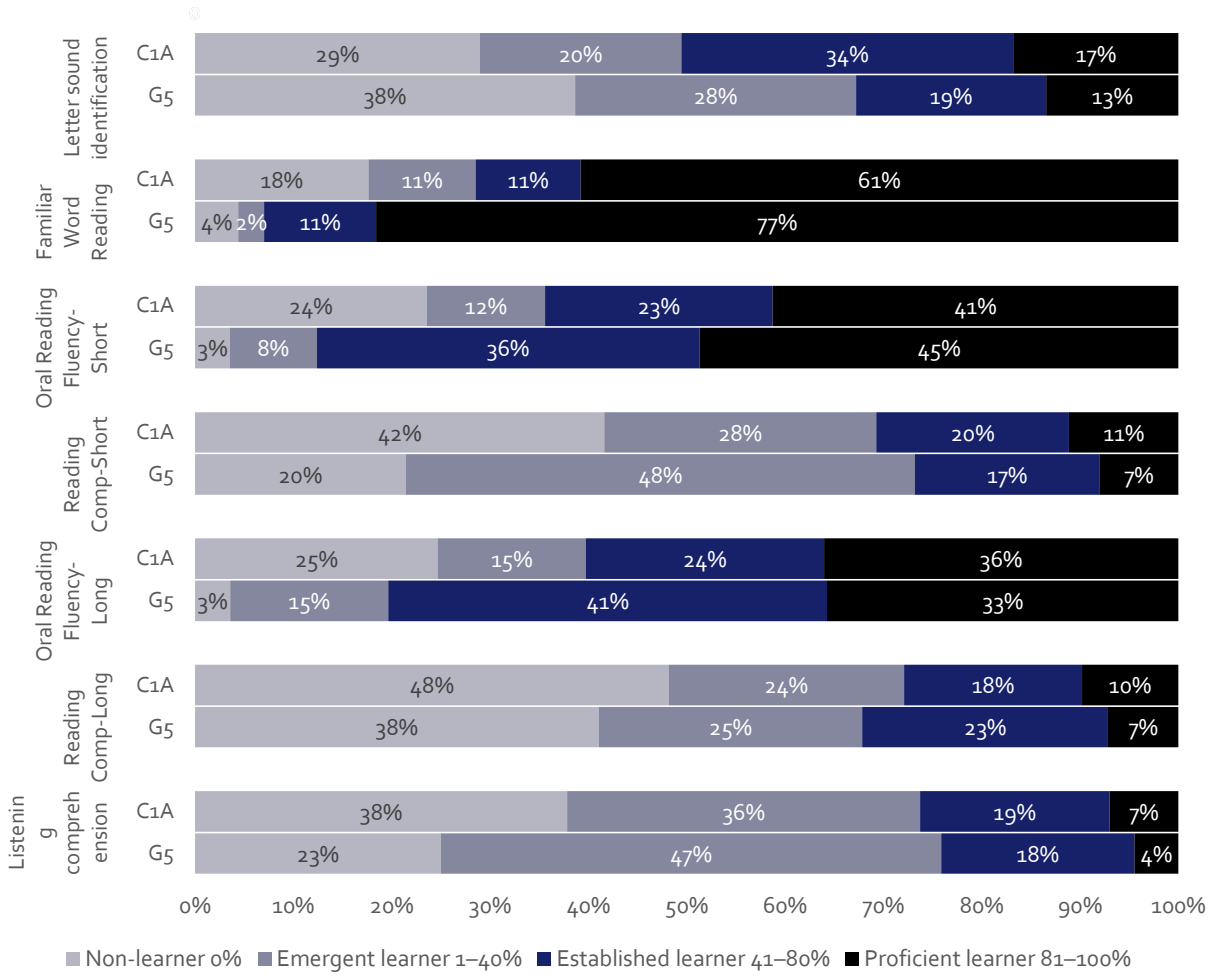


Figure 8 shows the distribution of the EGMA aggregate score for the treatment cohort of girls in increments of 10. While the mean score was 66.25, a plurality of girls (n=105) scored in the 80–90 band, followed by the 90–100 band (n=91). Few girls (n= 5) scored zero on EGMA subtasks.

Figure 8: Distribution of aggregate EGMA scores, C1A

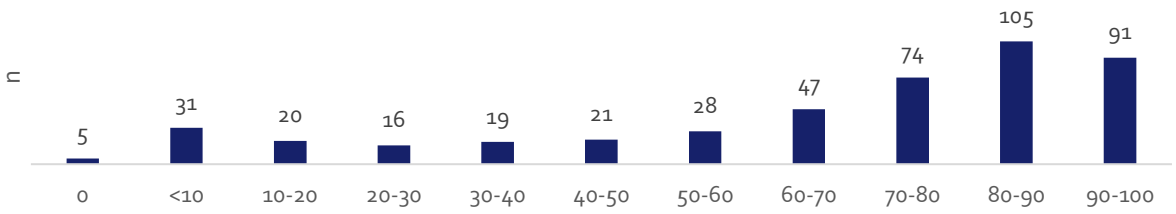
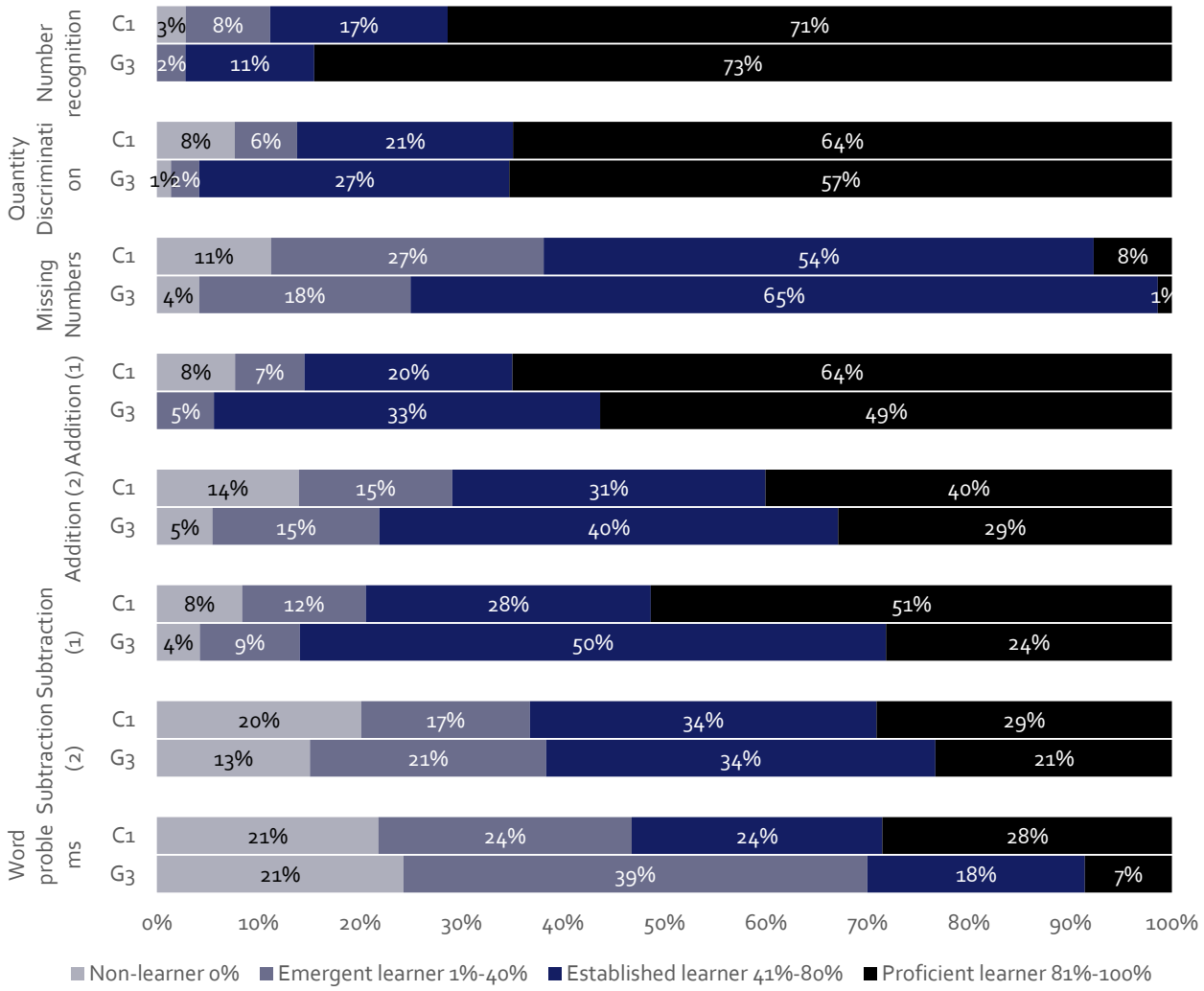


Figure 9 compares the EGMA performance for girls in the treatment cohort to those in Grade 3 using learner categories.[8] The largest proportion of girls from both the treatment cohort and Grade 3 were proficient in number recognition followed by quantity discrimination, addition level 1 and subtraction level 1. On the more difficult subtasks, just more than one-quarter of girls in the treatment cohort were proficient in subtraction level 2 (28.98%) and word problems (27.89%); these proportions were higher than those among Grade 3 girls. The missing number subtask appeared to have been the most difficult for girls; only 7.63% of the treatment group scored as proficient learners, although almost 50% scored as established learners (53.59%). Among the benchmark group, fewer than 2% of girls were proficient at the subtask.

Figure 9: Proportion of girls in learner categories by EGMA subtask, C1A and Grade 3



Given these findings, the project appears to have enrolled girls who average a Grade 5 literacy level and a Grade 3 numeracy level. However, treatment girls struggled more than their formal

[8] The average EGMA aggregate score of girls in Grade 3 was comparable to the scores of girls in Cohort 1A. As a result, the performance by learner categories of these two groups are compared at the subtask level.

education peers on the reading and listening comprehension EGRA subtasks and the missing number EGMA subtask.

The baseline report also examined learning assessments' ability to capture growth over time. Based on the baseline results, both the EGRA and EGMA appear to have ceiling effects. Two EGRA subtasks—familiar word reading and oral reading fluency-short—are most unlikely to capture growth over time. On the EGMA, only the missing number subtask does not appear to have a ceiling effect; all other subtasks are unlikely to capture growth over time. Based on discussions with the fund manager (FM), Plan will develop equated forms to compare performance at future timepoints.

Transition — Girls surveys administered at baseline show that 98.04% of C1A girls say they believed they will complete CBLH. Of those, almost half (47.62%) reported that they hope to transition to vocational training subsequently, and a similar proportion (47.12%) said they want to work in a safe, fairly paid job after CBLH. However, less than 2% of girls responded that they hoped to go to formal schooling after completing CBLH. At least 4 districts had no girls who intend to re-enter formal schooling. Most of the girls who did say they intend to return to formal schooling live in rural settings. The assumptions underlying the programme's theory of change (ToC)—particularly the proportion of girls who may re-enter formal schooling—will need to be reassessed.

Sustainability — Sustainability findings—presented for the system, community and learning space indicators—are drawn from qualitative data. The overall score on the sustainability scorecard was 1.4 out of 4, which indicates some foundation for sustainability but also substantial room for growth. System sustainability refers to education officials' knowledge about and responsiveness to marginalised girls' educational needs. While there is evidence of system-wide support for marginalised learners' education, it is unclear if the Ministry of Primary and Secondary Education (MoPSE) will have funding available to support and sustain SAGE initiatives after the project ends. Evidence of community sustainability was slightly weaker, although it was based on limited data. Among the 5 respondents, there is evidence of a perceived misalignment of the programme goals and community expectations that may hinder communities' appetite for continuing the work after the programme's completion. Evidence of sustainability of learning space was limited at baseline and drawn primarily from the Gender Analysis. Those results suggest that there are notable barriers that should be addressed in order for the quality of the learning space to be maintained after SAGE's end. However, given the limited evidence, a more nuanced understanding of learning space sustainability should be ascertained at the next evaluation.

Intermediate outcomes — In addition to the primary outcomes, SAGE outlines 5 intermediate outcomes (IO) to measure the programme's success; Figure 10 summarizes baselines values and key findings. Overall, girls in both cohorts 1A and 2 who had high levels of self-efficacy, more positive gender attitudes and high levels of SRHR knowledge scored higher on the learning assessments than did girls who had low levels, as measured by the indices for each IO. Although indicators for attendance (IO1) and life skills (IO3) were scored at 0.00 because learning sessions had not yet begun, qualitative findings highlight several considerations that SAGE should consider to ensure regular attendance, effective skills and access to financial resources. Girls' gender attitudes and knowledge of SRHR (IO2) were notably low at baseline, as were gender attitudes at the community level (IO4). The average life-skills score (outcome 3) was 29.22 on a 52-point scale, indicating that the average girl in the treatment cohort has room for growth in this area. Findings related to community support showed low levels of existing support, with room for growth over time.

Figure 10: Key baseline intermediate outcomes, C1A

Generally, assumptions in SAGE's ToC regarding subgroups and barriers appear to hold true. The most prevalent social, economic and educational barriers uncovered through the baseline are already considered in SAGE intervention planning. However, it is unclear if these assumptions may need to be adjusted once the beneficiary enrolment information is updated to include all girls in cohort 1A, including those who served as replacements in the baseline sample.

Recommendations

The following summarizes top priorities identified from the baseline evaluation. Additional and detailed recommendations can be found in section 9.

1. Collect missing demographic information for girls who were subsequently enrolled in CBLHs; disaggregate results by subgroups to determine if additional information regarding barriers can be obtained. Most importantly, when data is available, Plan can better understand learning assessment results by these subgroups as well as differences in life-skills, self-efficacy, gender perceptions and other IOs by these subgroups.
2. Review EGRA and EGMA instruments to address ceiling effects.
3. Ensure that monitoring data captures changes in enrolment.

4. Emphasize qualitative data at future evaluation points, exploring ‘why’ and ‘how’ to better understand the reasons behind observed quantitative results.
5. Review the Year 1 CBLH curriculum for literacy and numeracy to ensure that girls who are above a grade 3 level are also challenged.
6. Use monitoring data and analysis by beneficiary demographics to better understand girls’ intentions for transition and, in turn, to inform programme activities.
7. Provide training to CBLH facilitators on differentiated instruction and inclusive education strategies to meet the needs of all learners.

3. Background to project

3.1 Project context, target beneficiary groups and theory of change

Project to complete

- Please outline:
  - The main contextual factors that have influenced the project design (e.g. political, economic, social, environmental, legal and/or educational policy/system context).
  - How gender inequalities and marginalisation impact the education of girls in these areas.
  - If the context is the same or different across all the areas the project is working (e.g. is one more rural? Does one area have higher poverty, different language or education system/policy? Etc.).
  - How your project defines its direct beneficiaries. This definition should include the main characteristics girls must have to be enrolled into your project. Please also ensure you discuss if any prioritisation criteria was used to select the most marginalised direct beneficiaries and if the project was oversubscribed.
  - If applicable, how the direct beneficiaries were selected for cohort one and how future cohorts will be selected.
- Complete Table 1, 2 and 3.
- Add your Project's latest ToC diagram in this document or as an annex and briefly summarise it, including the activities, intermediate outcomes, assumptions and barriers you're aiming to overcome.

Since independence in 1980, Zimbabwe has made great strides in improving access to education. Yet in recent years, significant macroeconomic and political challenges have led to declining investment in the education sector, with corresponding impacts on learners' enrolment and performance. Marginalised learners—including girls and those with disabilities—are at particular risk of being left behind.

The SAGE programme targets communities where girls face a number of complex and interdependent barriers to accessing education: widespread poverty; long distances to the nearest school; time spent on house chores or childcare; inadequate school infrastructure in terms of accessibility, WASH, or menstrual hygiene management (MHM); gender-inadequate pedagogy; and stigma around disability. These are compounded by gender inequality, community attitudinal barriers to education—including early marriage and pregnancy—sexual violence and boys' and men's limited awareness of SRHR.

SAGE will work in the most affected areas, including Apostolic Christian communities; remote, rural areas; peri-urban informal settlements; and communities with high levels of economic migration where absent parents have left children unsupported.[9] Already economically precarious, these communities have been particularly impacted by Zimbabwe's protracted economic crisis, with visible impacts on girls' access to education.

[9] The programme has used national and regional-level statistics and its knowledge of the local context to determine the most appropriate locations for its interventions. At the community level, it has engaged with ministry staff and community leaders to identify and mobilise potential beneficiary households.

The targeted beneficiary girls come from 11 'target' districts in Zimbabwe: Bulilima, Chimanimani, Epworth, Harare South, Hatcliffe, Imbizo, Khami, Mutare Rural, Mutasa, Mutoko and Reigate.[10] The baseline assessment was conducted in 7 of the target districts, ahead of launching programme activities in the remaining 4 districts. Table 1 summarizes the proportions of Cohort 1A girls attending formal schooling, dropout status, the level at which they dropped out and age group. Data is taken from beneficiary enrollment information shared by Plan with School-to-School International (STS) at the time of baseline sampling. Table 2 summarizes the transition pathways for girls supported by SAGE, Table 3 describes the indirect beneficiary groups (See Annex 5 for additional details), and Figure 11 summarizes the programme's ToC.

Table 1: Summary of direct beneficiaries

Direct beneficiary numbers	Total figures
Total number of girls reached in C1A	4,075 (enrolled)
Total number of girls expected to reach by the end of project in all cohorts	21,780
Ever attended formal schooling	**Proportion of total direct beneficiaries (%)**
Never been to school	1,546 (37%)
Been to school but dropped out	2,526 (63%)
Dropout level	**Proportion of total direct beneficiaries (%)**
Dropped out before secondary school	1,550 (38%)
Dropped out during secondary school	976 (24%)
Age banding	**Proportion of total direct beneficiaries (%)**
10–14	461 (37%)
15–19	2,611 (63%)

[10] Epworth, Harare South and Hatcliffe are peri-urban districts of the capital Harare. Imbizo, Khami and Reigate are peri-urban areas of the country's second largest city, Bulawayo. The remaining 5 districts are rural.

Table 2: Proposed intervention pathways

Intervention pathway	Which girls follow this pathway?	How many girls follow this pathway for C1A?	How long will the intervention last?	How many cohorts are there?	What literacy and numeracy levels are the girls starting at?	What does success look like for learning?	What does success look like for transition?[11]
Accelerated Learning Programme (ALP) and Champions of Girls Education (CoGE)	Out-of-school (OOS) girls aged 10–14	1,461	2 years with an optional third year	4	Up to Grade 7	ALP Year 1 = Grades 1–3 ALP Year 2 = Grades 4–5 ALP Year 3 = Grades 6–7	Enrolled in formal education / NFE
ALP, CoGE and Village Savings and Loans Associations (VSLA)	OOS girls aged 15–19	2,611	As above	4	As above	As above	Enrolled in NFE / Integrated Skills Outreach Programme (ISOP) / (Self-) Employment / Entrepreneurship

Table 3: Indirect beneficiary groups

Group	Interventions received	Total number reached for C1A
Boys	ALP and CoGE sessions	1,357
Community Educators	ALP and gender-sensitive pedagogy training sessions and follow-up reflection workshops	124
Learning Assistants	As above	62
CoGE Facilitators	CoGE facilitation training	122 (70 women)

[11] It is important to note that the final decision on transition belongs to the girls themselves.

Figure 11: Summary of the theory of change

Outcomes	1 Learning	2 Transition	3 Sustainability
Assumed causal links	If girls regularly attend accessible, community-based learning sessions, staffed by well-trained and supportive community educators (CEs), they will attain desired learning outcomes. Girls' acquisition of life skills and improved self-efficacy and confidence will in turn contribute to and support their learning.	If girls have improved skills, access to financial resources, information and the ability to act on available opportunities—as well as improved financial and emotional support from their families—they will be able to transition into further formal or informal education, training or (self-) employment, as they desire.	If communities adopt more positive and supportive attitudes toward girls' education, and if the larger policy environment is made more responsive to girls' needs, girls will enjoy a more enabling environment that supports their education and project impacts will be sustained.
Intermediate Outcomes	IO1 – Highly marginalised adolescent girls regularly attend high-quality, accelerated learning sessions in CBLHs	IO3 – Highly marginalised adolescent girls and their families have improved skills and increased access to financial resources	IO4 – Communities demonstrate more positive gender attitudes and actively support and protect girls
	IO2 – Highly marginalised adolescent girls have increased self-efficacy and life skills		IO5 – Strong and active partnerships with MoPSE officials and other civil society actors actively advocate for more inclusive, gender-responsive education policies
Outputs	OP1 – Highly marginalised adolescent girls can access high-quality accelerated learning programmes through CBLHs	OP4 – Highly marginalised adolescent girls and their families are supported to participate in VSLAs and skills development opportunities	OP5 – Adolescent and adult champions of gender equality engage others in their communities in dialogue on girls' rights
	OP2 – CEs and formal sector NFE mentors are trained and supported to employ inclusive, gender-responsive teaching strategies		OP6 – Programme evidence and learning, including girls' own voices and experiences, are shared with key stakeholders at district and national level.
	OP3 – Highly marginalised adolescent girls and boys are supported to learn about and discuss life skills and their SRHR		

Assumptions

► Community engagement and mobilisation efforts are sufficient to mitigate resistance, stimulate demand for education and enable highly marginalised girls' participation in all components of the programme

► Training and incentive structures are sufficient to support and retain CEs

► Improvements in household economic security are not undermined by economic shocks and opportunities exist and are open to girls

► National- and district-level education officials remain supportive of the programme as a whole

► Relevant national- and district-level child protection officials remain supportive of efforts to work through and strengthen community-based protection mechanisms

► Political will of government to work in partnership with INGOs is maintained

► The political situation remains sufficiently stable for the project to continue its operations

4. Baseline evaluation approach and methodology

The following section presents information on the baseline evaluation approach, including details on the overall evaluation purpose and questions, quantitative and qualitative methodologies, data collection tools, enumerator training and operational baseline data collection. The baseline was conducted by the SAGE program's external evaluator (EE), STS, and a local data collection firm, Select Research.

4.1 Evaluation purpose and evaluation questions

The overall purpose of the SAGE programme baseline evaluation is to test the assumptions outlined in the programme's ToC (Figure 11). The evaluation is designed to provide meaningful and relevant findings of the programme design and its ability to meet the programme outcomes as they are related to IOs.

SAGE program's primary evaluation questions and data sources available at baseline are outlined in Table 4. Four project-level evaluation questions guide all Leave No Girl Behind (LNGB) projects, and the evaluation sub-questions align with SAGE program's ToC. These questions measure the assumptions the programme was designed on, and the results for these evaluation sub-questions are aggregated across the sample to answer primary evaluation questions.

The evaluation employed both quantitative and qualitative methods. STS and Plan worked together to ensure the findings are presented in a fair and reliable manner.

Table 4: Evaluation questions, summary of qualitative and quantitative data, analysis required to answer question

Evaluation question	Qualitative data and analysis required	Quantitative data and analysis required
GEC evaluation questions		
Process	Was the programme successfully designed and implemented according to stakeholders?	
Impact	What impact did the programme have on the learning and transition of marginalised girls, including girls living with disabilities? How and why was this impact achieved?	
Value for Money	Did the programme demonstrate good value for money approach?	
Effectiveness	What worked and what did not work to increase the learning and transition of marginalised girls (as defined by the programme)?	
Sustainability	How sustainable were the activities funded by the GEC and was the programme successful in leveraging additional interest and investment?	
SAGE evaluation questions		
1. Which activities and methodologies have been most effective in improving literacy	n/a	Learning Assessments (EGRA and EGMA)

and numeracy skills for highly marginalised girls?		
2. What impact did the programme have on the transition of highly marginalised girls into education, learning, training or work opportunities?	n/a	Data was not collected on transitions at baseline; intentions to transition were available
3. How sustainable were the programme activities? Was the programme successful in leveraging additional interest, investment and policy change?	KIIs with MoPSE district and national official, community leader, NGO and INGO leader and formal head teachers	n/a
4. What are the contributions of ALPs delivered through CBLHs towards the transition to formal or NFE by highly marginalised girls?	n/a	Data was not collected on transitions at baseline
5. What are the contributions of VSLAs and skills development opportunities for highly marginalised girls' transition to (self-) employment?	n/a	Data on VSLAs and financial skills was not collected at baseline.
6. What are the key factors needed to facilitate the transition of highly marginalised girls into education, training or employment and to increase learning? 7. What types of interventions are effective in building non-cognitive skills? 8. What were the most cost-effective and impactful activities/methodologies across the intervention?	n/a	Data was not collected on transitions at baseline
9. How successfully did the programme reduce barriers to full participation in education or vocational education for highly marginalised girls? 10. How effective were programmatic elements or adaptations at contributing to the desired change?	n/a	Girls Survey
11. To what extent are CBLH activities—both the ALP and CoGE sessions—contributing toward improvements in highly		Girls survey Household survey—boys

marginalised girls' self-esteem and social networks?		
12. How and to what extent has the programme fostered positive changes in gender attitudes and practices among different stakeholders—including girls or young women; boys or young men; mothers, fathers and other caregivers; and community and religious leaders—to create a more protective and supportive environment for highly marginalised girls? What factors have enabled or inhibited these changes?	KII with community leader	Girls survey Household survey – Parent/caregiver Household survey—boys
13. How do communities and government come together in a sustainable way to provide improved life opportunities for girls?	KIIs with government officials and community leaders	n/a
14. What impact does the programme have on the life of the girls involved? To what extent has the programme enabled changes in girls' aspirations and agency?	n/a	Girls survey
15. What is the impact of the programme on the local community through its CEs and girls who have more life-chances and associated skills?	KII with community leader	Girls survey Household survey – Parent/caregiver Household survey—Boys

4.2 Overall evaluation design

The purpose of this evaluation is to establish baseline values at the start of the implementation of the SAGE programme. In turn, these values will allow the programme to assess change over time in delivery, effectiveness, value for money and impact.

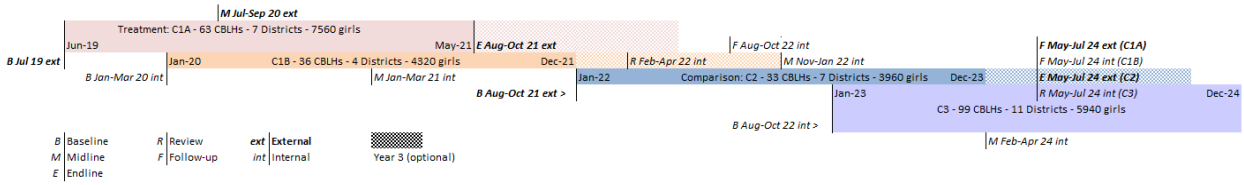
To measure the LNGB evaluation questions and the SAGE programme's evaluation questions, the baseline study uses a *mixed-method, longitudinal, cross-over design*.[12] The evaluation will utilise data from learning assessments and a package of quantitative and qualitative instruments from different respondents in subsequent timepoints. Additionally, SAGE will conduct regular monitoring of indicators outlined in the logframe. The variety of tools, respondents and methods allow for the data to be triangulated and linked across evaluation questions and indicators.

Because SAGE will initiate activities in a cohort design across 4 cohorts, the evaluation uses beneficiaries in a subsequent cohort as a comparison group and girls enrolled in formal school in Grades 3, 5 and 7 to establish benchmarks. As described in the monitoring, evaluation and learning (MEL) framework, the first cohort (C1A) will be the first treatment group, receiving the

[12] As per definition in LNGB MEL Guidance p. 143, and as noted in the MEL Framework submitted to FM December 14, 2018.

ALP intervention in 63 CBLHs across 7 districts (Figure 12). The baseline sample is drawn from the first cohort, C1A. The second cohort (C1B) will include 4 districts and will receive the SAGE intervention approximately 6 months after C1A. The third cohort (C2) acts as the comparison group and is set to receive the intervention 18 months after C1A in 33 CBLHs across the initial 7 districts. The cohorts come from different geographical areas of the 11 target districts. The districts were selected by Plan prior to baseline.

Figure 12: Phasing of cohorts for intervention and corresponding evaluation activities



A joint sampling approach was used for the SAGE evaluation. Specifically, STS and the programme collected data from girls who were randomly sampled from C1A for a treatment group and C2 for a comparison group. The team collected IO data from a smaller sample of other respondents—parents and caregivers, boys and community leaders—in the CBLHs and communities where sampled girls live. Project monitoring data on attendance is expected to be collected from all CBLHs by SAGE during the interventions and reported in subsequent evaluation reports.

The baseline evaluation design adheres to the current logframe and MEL framework. To examine the ToC's assumptions between IOs and outcomes, STS linked all data to girls' unique identifiers, allowing for analysis of the relationship between scores and outcomes. Additionally, the evaluation design is gender equality and social inclusion accommodating. The evaluation design considers gender, disability and other social differences and inequalities. These characteristics are explicitly accommodated in the selection of programme beneficiaries, evaluation tools' design and administration protocols, respondents sampling, enumerators' selection and training and evaluation results' reporting.

The baseline evaluation took place at the start of Year 1 of the SAGE intervention starting with sample selection in July and submission of the final baseline report in December 2019. Given the need to identify participants for the evaluation, sampling and data collection took place approximately 2 months after girls were enrolled in the programme. The start date for learning sessions in CBLHs varied by site; in some cases, girls continued to enrol in the weeks following the start of lessons.

The baseline evaluation report combines qualitative data collected during the Gender Analysis study, limited qualitative data collected during baseline and quantitative data collected at baseline for learning, transition and IOs reporting.

The programme's MEL framework originally outlined a census-based evaluation. Due to budget and timing constraints, this was changed, and girls were randomly sampled from SAGE CBLH sites to participate in the baseline evaluation. The sampling approach is described in further detail in section 4.4.

4.3 Evaluation ethics

STS adhered to SAGE ethics, child protection and safeguarding policies throughout the baseline process. This included providing all enumerators and Select Research staff with relevant policies

and engaging SAGE to present on the policies during enumerator trainings. Enumerators were provided with SAGE persons of contact for each district to ensure that any ethical issues that arose could be mitigated or reported. A summary of the ethical protocols and the baseline approaches to adhering to protocols is presented in Supplementary Table 1.

Supplementary Table 1: Ethical protocols and baseline approaches

Ethical issue or protocol	Baseline approach
Administrative, technical and physical safeguards to protect the confidentiality of those participating in research	All STS staff received certification on Human Subjects Research as required by institutional review boards. STS trained all staff and enumerators on the importance of confidentiality, especially for vulnerable populations. Data was uploaded electronically and stored in password protected databases. STS designed data collection logistics to ensure confidentiality of respondents is maintained to the highest extent possible. All evaluation data was saved using unique IDs to minimize the ability of respondent information to be unmasked.
Safeguards for those conducting research	SAGE's safeguarding coordinator monitored risks associated with those conducting research. The safeguarding coordinator reviewed the data collection plans and provided feedback to ensure that the plans addressed safeguarding needs.
Child-safe physical safeguards for children participating in research	SAGE's safeguarding coordinator monitored risks associated with children participating in research. STS, with support from Plan, trained enumerators on the SAGE Programme Safeguarding Strategy and Implementation Action Plan. STS designed data collection logistics to ensure proper gender and cultural sensitivities were considered during data collection.
Adherence to good practice guidelines on researching violence against women and girls	STS collected data from adult men and women as well as adolescent boys and girls to assess gender norms and awareness of GBV. Given the necessity to avoid the re-traumatisation of survivors, protect confidentiality and minimize risk, STS trained all enumerators on best practices around researching violence against women and girls. Enumerators also learned ways to make safe and sensitive referrals in the event of disclosures during data collection. This training adhered to SAGE's procedures for referrals and disclosures.
Appropriate time allocated to engage with children participating in research	During the pilot, STS tested the assessment and survey lengths and made recommendations to streamline the tools to help respondents feel comfortable and to avoid fatigue during data collection. In addition, STS trained enumerators on ways to build rapport and make respondents comfortable, as well as strategies for structuring the data collection schedule at each site to provide respondents with short

Ethical issue or protocol	Baseline approach
	breaks, particularly for the girls participating in both the learning assessments and survey. STS also ensure that additional time for assessments was allocated to children with disabilities, as per best practices.
Data protection protocols and secure maintenance procedures for personal information	All STS staff received certification on Human Subjects Research as required by institutional review boards. STS trained all staff and enumerators on the importance of confidentiality, especially for vulnerable populations. Data was uploaded electronically and stored in password-protected databases. STS designed data collection logistics to ensure confidentiality of respondents is maintained to the highest extent possible. All evaluation data was saved using unique IDs to minimize the ability of respondent information to be unmasked.
Parental consent concerning data collection from children or collation of data about children; age and ability appropriate assent processes based on reasonable assumptions about comprehension for the ages of children and the disabilities they intend to involve in the research	STS, in collaboration with Plan, ensured that, when possible, consent is sought from parents or caregivers for all respondents under the age of 18. Respondents under the age of 18 were asked for their assent to take part in the research. Where parents or caregivers were asked to consent and children did not assent, the view of the child was respected. Consent and assent protocols were administered at the start of each data collection tool. STS trained enumerators on steps to take if consent or assent was not given. STS also trained enumerators on best practices for soliciting assent from children with disabilities.
Appropriate spaces and methodologies tailored in consideration of unique needs of girls and boys, including those with disabilities	STS followed EGRA and EGMA best practices on establishing physical spaces for testing, including ensuring that respondents were assessed in a quiet and private location with no disruptions. Enumerators ensured that any portions of the surveys will be administered in private locations where responses were not be observable to outsiders. Further, STS ensured that data collection teams were composed of females due to survey content and cultural sensitivities. STS designed and implemented individualized accommodations for children with disabilities to ensure their unique needs were met in the assessment context.
Appropriate language and communication for different ages and the disabilities of children involved in the research	STS made learning assessment instructions and survey items available in Shona and Ndebele to ensure that respondents could answer questions in a language familiar to them. STS designed and implemented individualized accommodations for children with disabilities to ensure their unique needs were met in the assessment context.

Ethical issue or protocol	Baseline approach
Age-appropriate participation of girls, including in the development of data collection tools	SAGE provided a list of age-appropriate beneficiaries from whom STS selected the sample. STS distributed all data collection tools to Plan for review before and after the pilot to ensure local knowledge of age-appropriateness was considered. Further, STS adjusted data collection tools after piloting to ensure that they were appropriate for the populations to which they were administered during operational data collection.

4.4 Quantitative evaluation methodology

Quantitative evaluation tools

Five baseline evaluation surveys and 2 learning assessments were developed and used for the quantitative component of the evaluation per the MEL framework. The development of the learning assessments for SAGE is described in additional detail in the corresponding sections. STS and Plan collaboratively developed the survey tools, detailed in **Error! Reference source not found.**, prior to pretesting and data collection. They include a girls survey and 4 household surveys—boys survey, parent/caregiver survey, head of household (HoH) survey and transition-benchmark survey for girls in formal schools. The tools combined numerous domains relevant to the programme's ToC and items that corresponded to the programme's logframe indicators. Each tool uses LNGB templates as the initial source of items. Following the compilation of these items and additional programme-specific items within each tool, STS shared drafts with Plan and partners, who commented and provided revised or new items based on the project's indicators and specific implementation priorities.[13, 14] All items' sources and revisions were tracked in a master file. All surveys were shared with the FM for review and approval prior to the pre-test and operational data collection.

Table 5: Quantitative evaluation tools at baseline

Tool name	Relevant indicator(s)	Who developed the tool?	Was tool piloted?	How were piloting findings acted upon (if applicable)	Was tool shared with the FM?	Was FM feedback provided?
Girls survey	IO.2.1 IO.2.2 IO.4.1 IO.4.2 IO.4.3	STS, Plan	Yes	Minor modifications to translations and problematic items made following pilot	Yes	Yes
Boys Survey	IO.2.1 IO.2.2	STS, Plan	Yes	Minor modifications to translations and	Yes	Yes

[13] Plan provided adapted items from the Gender Norm Attitudes scale and Gender Equitable Men scale. See Nanda, Geeta. 2011. *Compendium of Gender Scales.* Washington, DC: FHI 360/C-Change.
[14] Plan provided self-efficacy items adapted from Chen, G., Gully, S.M. and Eden, D. (2001) 'Validation of a New General Self-Efficacy Scale', *Organizational Research Methods*, 4 (1): 62-83.

Tool name	Relevant indicator(s)	Who developed the tool?	Was tool piloted?	How were piloting findings acted upon (if applicable)	Was tool shared with the FM?	Was FM feedback provided?
	IO.4.1 IO.4.2 IO.4.3			problematic items made following pilot		
Head of Household Survey	IO.4.1 IO.4.2 IO.4.3	STS, Plan	Yes	Minor modification to problematic item following pilot	Yes	Yes
Parent / Caregiver Survey	IO.4.1 IO.4.2 IO.4.3	STS, Plan	Yes	Minor modification to problematic item following pilot	Yes	Yes
Transitional-benchmark Survey	O.1.1 O.1.2	STS, Plan	Yes	Minor modification to problematic item following pilot	Yes	Yes
EGRA	O.1.1 O.1.2	STS (adapted from existing tools)	Yes	Based on the pilot, revisions were made to reading passage, reading comprehension, and listening passage to align with quality guidance	Yes	Yes
EGMA	O.1.1 O.1.2	STS (adapted from existing tools)	Yes	Significant updates made to addition level 2, subtraction level 2, and word problems to align with quality guidance	Yes	Yes

It is expected that the 5 surveys should remain relatively stable across the evaluation points, with only minor revisions or additions required.[15] Additional forms of the learning assessments will be developed for future timepoints to respond to programme evaluation questions using data collected for equating.

Enumerators

STS and Select Research worked collaboratively to recruit, hire and train enumerators for the pilot and operational baseline data collection activities. STS provided Select Research with a list of key qualifications and job descriptions, and Select Research recruited local, female enumerators who fit the required qualifications. Following initial screenings, oral interviews and reference checks, Select Research selected 11 enumerators and 3 supervisors for the

[15] This assumes that the programme's ToC also remains stable across evaluation points.

quantitative activity. Two supervisors oversaw Shona-language teams, and one supervisor oversaw the Ndebele-language team. All selected enumerators had prior experience conducting surveys either on paper or electronically.

Before training commenced, all selected enumerators signed contracts with Select Research that stipulated their expected roles and expected ethical and professional conduct during training and data collection. Additionally, all enumerators underwent police security clearance checks as required by Plan as part of its child safety and protection procedures for all persons working under their programmes.

The baseline quantitative enumerator training, facilitated by STS with support from Select Research and Plan, took place from 25–28 July 2019 in Harare. During the training, all enumerators participated in large group sessions to introduce the data collection tools and procedures. The enumerators worked in pairs, by language, to practice administering the tools. Training sessions included:

- Baseline study purpose and research ethics
- Introduction to SAGE programme
- Safeguarding and child protection
- EGRAs and EGMAs
- Surveys
- Using tablets for data collection
- Team roles and responsibilities
- Accommodations for schools with disabilities
- Data collection logistics
- Supervisor roles and responsibilities

All enumerators and supervisors participated in the quantitative pilot, which took place on 29 July 2019 in Harare and Mutoko. Each enumerator administered 15 learning assessments and 5 of each quantitative survey. Enumerators provided feedback on their experience and specific components of the tools; their feedback was incorporated into the revisions presented to Plan and the FM prior to the start of operational data collection. After approval from Plan and the FM on changes from the pilot, training on the final operational tools was held on 2 August 2019.

The supervisor training day was held on 1 August 2019; it included sessions on supervisory roles and responsibilities during data collection. On the last day of training, Select Research divided the enumerators into 3 teams: 1 Shona teams with 4 enumerators and one supervisor-enumerator, and one Ndebele team with 3 enumerators and one supervisor-enumerator.

Quantitative data collection

Quantitative data collection took place from 4 August through 10 September 2019. CBLH visits varied from one to 2 days, depending on the CBLH sample targets. Shona team A was assigned to CBLHs in the Epworth, Mutasa and Mutuko districts; Shona team B to the Chimanimani and Mutare Rural districts; and Ndebele team C to Bulilima and Imbizo districts.

All data was collected electronically on Android-based tablets. The learning assessments were administered to girls using Tangerine®, and surveys were administered using SurveyCTO. At the end of each day of data collection, supervisors uploaded all data from their team's tablets to the software servers. STS's quality control team downloaded and securely stored all raw data on a password-protected server for daily review, cleaning and analysis. After data collection was completed, Select Research ensured that the software and data were permanently deleted from the tablets and that any paper documents with identifying information were discarded.

Data quality was assured through several strategies. The use of tablets for electronic data capture mitigated data entry errors and helped ensure data quality, consistency and collection efficiency. Records were linked across tools using SAGE's unique beneficiary IDs, which were programmed into all tools and populated into the dataset. During community visits, supervisors completed tracking sheets to keep a record of girls who had been assessed; girls who completed the girls survey; and parents, caregivers, heads of households or boys who completed any surveys. As a result, it was possible for STS's quality control team to know which and how many tools were completed daily, determine any data quality issues and ensure that the correct girls were sampled. Any issues or challenges were recorded into a data collection tracker, and STS's quality control team coordinated directly with team supervisors through WhatsApp to reconcile any quality issues.

Quantitative data cleaning and storage

STS stores all raw data on a password-protected server. Raw datasets are subject to 3 levels of data cleaning based on a standard protocol. During the first level, final raw data are labelled and reviewed to ensure the data was uploaded within the data collection period; any duplicates were removed; the number of records per CBLH was checked against the expected sample; and consent was received for all respondents. In the second level, disposition codes taken from the quality control team's data collection tracker are integrated and applied to the data to identify, remove or adjust cases based on issues uncovered during the data collection.[16] Afterwards, analysts again reviewed datasets for duplicates, missing data, and inconsistencies. Finally, at the third level, analysts compute learning assessment subtask scores, aggregate literacy and numeracy scores and survey composite scores. Outliers are identified and examined for inconsistencies. At the end of the 3 levels of cleaning, datasets are merged to complete the analysis.

Quantitative data analysis

All quantitative data were analysed using Stata and IBM SPSS® software platforms. The learning assessment analysis included girls who were sampled and had unique ID numbers that matched the SAGE enrolment database. For girls in the original sample, these unique IDs were provided prior to data collection by Plan. For the replacements, STS and Plan matched the girls in the data set with the ID numbers in the enrolment database. The raw learning assessment data includes 1,167 records. The final analytical learning assessment file contains 1,061 records.

Similarly, the girls survey analysis included girls who were sampled and had a unique ID number that matched the enrolment database, as well as replacement girls where the unique IDs were matched following data collection. Raw data from the girls survey includes 737 records. The final girls survey analysis file contains 664 records.

[16] Disposition codes are STS's internal system for data cleaning. Specifically, disposition codes are used to indicate the type of issue in a record or data point and the proposed resolution. During the cleaning process, disposition codes assist the analyst to determine the extent of discrepancies in a specific record or a specific variable and make appropriate decisions about the data quality and cleaning.

Household survey analysis includes parents, caregivers and heads of households of girls who were sampled and had a unique ID number that matched the enrolment database, as well as replacement girls in which case, unique IDs were matched following data collection. The surveys also included boys who are related to or residing in the same house as girls in the sample. At each site, 7 caregivers, 2 heads of household and 2 boys participated in the evaluation to provide community-level data. The raw parent/caregiver survey data file contains 346 records from the sample and replacement girls' households; the final parent/caregiver survey analytical file contains 346 records. The raw HoH survey data file contains 127 records from sample and replacement girls' households; the final HoH survey analytical file contains 100 records. The raw boys survey data file contains 114 records from sample and replacement girls' households; the final boys survey analytical file contains 100 records.

The survey datasets were merged to enable an analysis of marginalisation characteristics and barriers to education. Finally, these datasets were merged with the learning assessment dataset.

All results use the unit of analysis that most accurately reflects the way in which the data was collected, and items were structured. For all learning data, results are presented across girls, as the unit of analysis is the individual learner. For survey data, the unit of analysis varies. For indexes related to aspects of the community, the unit of analysis is respondents but is described as the community.

For the learning assessment, scores and learning bands were computed and reported per LNGB guidance. Guidance for aggregate scoring in subsequent evaluation points may be revised to account for fluency rates on timed subtasks, instead of reporting only percentage correct.[17]

STS created composites—or indexes—for IO indicators by mapping survey items to indicators. The mapping of items to indicators and the construction of composites was shared with Plan and reviewed with the FM prior to analysis. Relevant but non-overlapping items from the girls and household surveys were included in indices constructed for each indicator.[18] Although the majority of indexes were constructed based on the theory underlying the survey construction, the reliability of each composite was also checked by computing Cronbach's alpha (Annex 14).[19]

Learning tests

SAGE's learning assessments were adapted from existing EGRAs and EGMAs that had been previously administered in Zimbabwe.[20] Both the learning assessments instructions were administered in Shona and Ndebele, with the subtasks measuring performance in English. Shona and Ndebele were selected because they are the primary languages in the districts where the baseline was administered.

Details on both learning assessments' subtasks are included in Supplementary Table 2. Most subtasks included autostops—early stop rules—meaning that if learners do not correctly answer a predetermined set of items, the subtask would automatically stop, and enumerators would move to the next subtask. These were established to allow learners to efficiently move through the assessment and to prevent learners from spending a long period testing skills that they do not have. This allowed for respondents with low learning levels to forgo attempting all items on each subtask. The length of time allocated for each timed subtask is discussed in Supplementary Table

[17] The FM will provide additional guidance on scoring at midline based on conversations with the funder.
[18] Only respondents who answered 25% or more of the underlying items were included in the index calculation.
[19] Cronbach's alpha is a measure of internal consistency or scale reliability. It measures how closely related a set of items are within a defined group.
[20] The learning assessments were developed in 2015 by World Vision under the GEC-funded Improving Girls' Access through Transforming Education (IGATE) programme and under the Malawi National Reading Program funded by USAID in 2018.

2. For similar reasons, learners who did not correctly answer any items on the addition or subtraction level 1 subtasks were not asked items from the corresponding level 2 subtasks.

Supplementary Table 2: Learning assessments

Tool name	Subtask	Purpose	Administration	Scoring
EGRA	Letter sound identification	Alphabet knowledge	Timed—2 minutes; autostop after first 10 items	Correct letter sounds per minute; 100 items total
	Familiar word reading	Sight-word recognition and decoding	Timed—2 minutes; autostop after first 5 items	Correct familiar words per minute; 50 items total
	Oral reading fluency (short story)	Decoding and reading fluency	Timed—2 minutes; autostop after 6 items	Correct words per minute; 65 items total
	Reading comprehension (short story)	Reading comprehension	Untimed; number of questions asked corresponds to how many items read in oral reading fluency passage	Correct out of 5
	Oral reading fluency (long story)	Decoding and reading fluency	Timed – 3 minutes; autostop after 6 items	Correct words per minute; 93 items total
	Reading comprehension (long story)	Reading comprehension	Untimed; number of questions asked corresponds to how many items read in oral reading fluency passage	Correct out of 5
	Listening comprehension	Oral language comprehension and vocabulary	Untimed; all questions asked of all respondents	Correct out of 5
EGMA	Number recognition	Numerals and numericities identification	Timed—2 minutes; no autostop	Correct per minute; 20 items total
	Quantity discrimination	Numerical magnitudes comparisons	Untimed; autostop after 4 consecutive incorrect items	Correct out of 10
	Missing numbers	Number patterns identification	Untimed; autostop after 4 consecutive incorrect items	Correct out of 10

Tool name	Subtask	Purpose	Administration	Scoring
	Addition (level 1)	Arithmetic skills	Timed—2 minutes; no autostop	Correct per minute; 20 items total
	Addition (level 2)	Arithmetic skills	Untimed; no autostop; only administered if respondent correctly answered at least one item correct on addition level 1	Correct out of 5
	Subtraction (level 1)	Arithmetic skills	Timed—2 minutes; no autostop	Correct per minute; 20 items total
	Subtraction (level 2)	Arithmetic skills	Untimed; no autostop; only administered if respondent correctly answered at least one item correct on subtraction level 1	Correct out of 5
	Word problems	Conceptual and real-word mathematics understanding	Untimed; autostop after 4 consecutive incorrect items	Correct out of 6

Quantitative sample selection

After Plan developed the MEL framework, STS drafted the inception report. During this phase, STS, Plan and the FM carried out discussions regarding the sample size, specifically resource constraints for conducting a census level baseline study. Based on those conversations—which are documented in the inception report and a subsequent memo—the final sampling approach approved in collaboration with the FM was a 3-stage clustered random sampling approach that adheres to the sampling standards set forth by the FM (Supplementary Table 3). The 3-stage approach accounts for the clustering of girls in CBLHs. To determine the sample size, the sampling methodology first used the requisite sampling parameters to determine the number of girls in the sample, assuming a simple random selection (Supplementary Table 4). Next, this sample size was increased to account for the attrition rate of 30%. Finally, the sample was increased again to account for the design effect of a 3-stage sampling approach (Supplementary Table 5).

Upon receiving the sampling frame, the sample for the comparison group was again adjusted because the number of available CBLHs in the comparison cohort was significantly lower than expected. Specifically, when the sampling frame was received, STS identified 3 major differences from the original intended population:

1. The proportions of the population between treatment and comparison cohorts were assumed to be 0.52 from previously available data, but in the final sampling frame is 0.40.

2. The number of eligible girls in each CBLH varied widely when applying the eligibility criteria, including:

    i. Dropped out before secondary OR

    ii. Newly enrolled and under 16 years old OR

    iii. Identified as a girl with disabilities OR

    iv. Performed below a Grade 5 level on the Wide Range Achievement Test

    The final eligible population was 846 girls in the comparison group and 4,568 in the treatment group.

3. The number of communities with small enrolments were high in the comparison group. A threshold of at least 20 enrolees was used for the treatment group to ensure that the requisite number of girls were available for the study. For comparison communities, when applying the threshold of 20 or more, only 10 communities were available. Instead, a threshold of 15 girls enrolled was applied to increase the number of eligible CBLHs to 12. All 12, therefore, were selected for the sample.

Based on these parameters, the resulting design effect was 3.5. The change in the population proportion, design effect and the smaller number of eligible communities in the comparison group resulted in having to recompute the power calculations. The resulting sample is shown in Supplementary Table 6.

Supplementary Table 3: Proposed sample standards

Minimum detectable effect	0.4 standard deviations (this was approved with the FM on 29 April 2019)
Level of significance	5%
Power	80%
Attrition buffer	30%
ICC[21]	0.1

Supplementary Table 4: Proposed power calculations[22]

Test family	t tests
Statistical test	Means: difference between 2 independent means (2 groups)
Type of power analysis	A priori: Compute required sample size—given α, power and effect size

[21] UK Aid Girls' Education Challenge, *LNGB MEL Guidance July 2018*, 117.
[22] Computed using G-Power.

Input parameters		Output parameters	
Tail(s)	2	Non-centrality parameter δ	2.8123106
Effect size d	0.4	Critical t	1.9709056
α err prob	0.05	Df	218
Power (1-β err prob)	0.8	Sample size group 1	145
Allocation ratio N2/N1	0.52	Sample size group 2	75
		Total sample size	220
		Actual power	0.7995291

Supplementary Table 5: Proposed learning sample sizes

Group	Total number of CBLHs	Treatment learners per CBLH	Total number of learners to be assessed	Design Effect	Effective number of learners (SRS-equivalent n)
Treatment	35	12	420	2.1	200
Comparison	28	16	448	2.5	179

Supplementary Table 6: Actual learning sample sizes feasible in the sampling frame

Group	Total number of CBLHs	Learners per CBLH	Total number of learners to be assessed	Design Effect	Effective number of learners (SRS-equivalent n)
Treatment	35	12	375	2.1	179
Comparison	12	20	250	3.5	72

The power calculations were reviewed with Plan and the FM, who agreed to a two-stage sampling approach. Based on these power calculations, the C1A sample included a first-stage random selection of 35 CBLHs, proportional to the total number of CBLHs in C1A by district (see Supplementary Table 6). The C2 sample, which serves as a comparison, included 28 CBLHs. However, when the final sampling frame was received from Plan, only 12 communities identified to participate in the C2 intervention met the criteria for inclusion in the baseline data collection as comparison sites.

The final sample for the baseline study included 420 girls for the treatment group and 250 girls for the comparison group. In line with the joint sample approach, all of the girls selected for

learning assessments also responded to the girls survey. Sample sizes for the remaining 3 household surveys were determined based on resources. Specifically, there were sufficient resources for 895 surveys to be administered to other respondent groups. As such, STS and Plan distributed these surveys as follows: 70% to be conducted with parents/caregivers, 15% to be conducted with heads of household and 15% to be conducted with boys. Using a mapping of items to indicators, surveys with parents/caregivers were determined to have greater importance to respond to the programme logframe.

To achieve this sample size, and to ease logistics and administration, the number of respondents to test or survey was determined per site based on the proportional quotas. At treatment sites, teams were mandated to collect a quota of 12 learning assessments, 12 girls surveys, 7 parent/caregiver surveys, 2 HoH surveys and 2 boys surveys. At comparison sites, a minimum number of learning assessments and girls surveys to be administered was determined based on the size of the CBLH. At the comparison sites, enumerators collected data from 15–26 girls, 7 parents/caregivers, 2 heads of households and 2 boys. At schools in the communities selected for benchmarking data, 45 learning assessments were mandated in Grades 3, 5, and 7 as well as 45 benchmarking surveys. In all, the baseline sample yielded a one-to-one ratio of learning assessments to girls surveys. All other respondent groups data are available at the community level.

Following the selection of the CBLHs per district, STS conducted the second stage of the sampling procedure and randomly selected 12 girls and 5 replacements from each selected C1A CBLH. In C2 comparison sites, STS randomly selected 15–26 girls per site depending on the size of the site. All girls selected for the learning assessments also participated in the girls survey. At each CBLH, 7 parent/caregivers, 2 heads of household and 2 boys comprised the sample for the household surveys. As a result, the findings presented are aggregated for parent/caregivers, heads of household and boys.

CBLH facilitators and SAGE staff were responsible for mobilizing the girls and their caregivers, heads of household and boys to the assessment site for data collection. If selected girls were unavailable, the CBLH facilitator contacted the 5 replacements and their caregivers. If the quotas were still unmet, CBLH facilitators mobilized any other eligible girls to the assessment site to participate in the baseline. The names of replacement girls were entered manually into Tangerine and SurveyCTO. Supervisors also created paper-based replacement lists to serve as a back-up.

Quantitative sample sizes

Table 6: Quantitative sample sizes

Tool name	Sample size agreed in MEL framework – treatment	Sample size agreed in MEL framework – comparison	Actual sample size – treatment	Actual sample size – comparison	Remarks on why anticipated and actual sample sizes are different
EGRA / EGMA	420	250	459	264	The MEL framework and the Inception report suggested a census administration with a 1:1 ratio for the girls, parents/caregivers, HoH and boys surveys. However,

Tool name	Sample size agreed in MEL framework – treatment	Sample size agreed in MEL framework – comparison	Actual sample size – treatment	Actual sample size – comparison	Remarks on why anticipated and actual sample sizes are different
					during the writing of the inception report, a cluster-based sample was found to be more appropriate given the constraints of the budget. At the time of writing the inception report, the sample of girls was expected to be 895. The budget allowed for 895 non-girl surveys to be administered, and therefore, this allocation was distributed among the 3 remaining surveys as follows: 70% parent/caregiver, 15% HoH and 15% boys. In some cases, CBLHs were not able to mobilize sufficient respondents at each site. In these instances, SAGE staff mobilized additional girls at other sites in the district to reach the overall target. Due to this over-sampling at some sites, the final totals were greater than the original sample.
Girls Survey	420	250	416	248	In some instances, girls were not able to complete the survey. Data collection notes indicate that some girls had to leave the CBLH prior to completing the survey, explaining the difference in the total number of learning assessments compared to the survey.
Parent / Caregiver Survey	420	250	257	89	Because of budgetary constraints, the one-to-one ratio could not be adhered to

Tool name	Sample size agreed in MEL framework – treatment	Sample size agreed in MEL framework – comparison	Actual sample size – treatment	Actual sample size – comparison	Remarks on why anticipated and actual sample sizes are different
					for girls to survey respondents. Instead the proportion used to arrive at the number of caregiver surveys was 70%. To ease the logistics of data collection, this equated to 7 surveys per CBLH. This also enabled the data collection team to adhere to the data collection schedule, which called for one day per C1A CBLH and up to 2 days per C2 CBLH.
Head of Household Survey	420	250	77	23	Because of budgetary constraints, the one-to-one ratio could not be adhered to for girls to survey respondents. Instead, the proportion used to arrive at the number of HoH surveys was 15%. To ease the logistics of data collection, this equated to 2 surveys per CBLH.
Boys Survey	420	250	72	28	Because of budgetary constraints, the one-to-one ratio could not be adhered to for girls to survey respondents. Instead, the proportion used to arrive at the number of boys surveys was 15%. To ease the logistics of data collection, this equated to 2 surveys per CBLH.

Representativeness of the sample

Demographics of the baseline sample are presented in Table 7 through Table 10. The representativeness of the baseline sample has been assessed by comparing these tables with the tables in **Error! Reference source not found.**. Overall, the baseline sample was drawn to be

representative of the total beneficiary population. However, with almost 80% replacement of the sample during data collection, the EE could not determine the representativeness of the baseline data.

Table 7: Sample breakdown by intervention pathways

Intervention pathway (girls reported intentions after CBLH completion)	Sample proportion of intervention group (%)
Re-entry into formal education	2.76%
Vocational training	47.62%
Employment/self-employment	47.12%
Get married / other / don't know / refused	2.51%[23]

Table 8: Sample breakdown by regions

Region	Sample proportion of C1A beneficiaries (%)[24]	Sample proportion of C2 beneficiaries (%)[25]	Proportion of girls in baseline survey dataset (%) – intervention	Proportion of girls in baseline survey dataset surveyed (%) – comparison
Bulawayo	4.00%	2.00%	5.23%	7.58%
Harare	18.00%	53.00%	11.11%	11.36%
Manicaland	60.00%	34.00%	66.67%	67.80%
Mashonaland East	6.00%	8.00%	6.54%	12.88%
Matabeleland South	12.00%	3.00%	10.46%	0.38%
Source	beneficiary database, C1A	beneficiary database, C2	cleaned data analysis file	cleaned data analysis file
N =	N = 4568	N = 876	N = 459	N = 264

Table 9: Sample breakdown by age

Age (adapt as required)	Sample proportion of C1A beneficiaries (%)[26]	Sample proportion of C2 beneficiaries (%)[27]	Proportion of girls in baseline survey dataset of intervention (%)	Proportion of girls in baseline survey dataset surveyed of comparison (%)
Aged <10 (%)	12.28%	6.22%	0.67%	3.45%
Aged 10 (%)	4.68%	4.68%	4.23%	6.13%

[23] The proportion of girls in the treatment cohort by sub-category are: 0.3% said 'get married and care for my family', 0.5% said 'other', and 1.8% said 'don't know'.
[24] Proportions are based on the beneficiaries provided by Plan in the baseline sampling frame.
[25] Proportions are based on the beneficiaries provided by Plan in the baseline sampling frame.
[26] Proportions are based on the beneficiaries provided by Plan in the baseline sampling frame.
[27] Proportions are based on the beneficiaries provided by Plan in the baseline sampling frame.

Age (adapt as required)	Sample proportion of C1A beneficiaries (%)[26]	Sample proportion of C2 beneficiaries (%)[27]	Proportion of girls in baseline survey dataset of intervention (%)	Proportion of girls in baseline survey dataset surveyed of comparison (%)
Aged 11 (%)	3.44%	3.44%	2.67%	3.07%
Aged 12 (%)	4.07%	4.07%	3.56%	6.13%
Aged 13 (%)	5.93%	5.93%	5.12%	11.49%
Aged 14 (%)	26.27%	26.27%	10.24%	13.41%
Aged 15 (%)	7.20%	7.20%	7.80%	12.26%
Aged 16 (%)	6.50%	6.50%	8.24%	11.11%
Aged 17 (%)	4.95%	4.95%	11.80%	11.88%
Aged 18 (%)	5.65%	5.65%	16.93%	8.43%
Aged 19 (%)	7.88%	7.88%	27.39%	12.26%
Aged 20 + (%)	0.09%	0.00%	1.34%	0.38%
Unknown	15.74%	15.74%		
Source: N =	Beneficiary database, C1A N = 4568	Beneficiary database, C2 N = 876	Cleaned data analysis file N = 449	Cleaned data analysis file N = 261

Table 10: Sample breakdown by disability

Domain of difficulty	Proportion of girls in baseline survey dataset of intervention (%)	Proportion of girls in baseline survey dataset surveyed of comparison (%)	Guidance (record as true if they meet the criteria below)
Seeing	6.04%	2.02%	If CF1=1 AND (CF2=3 OR CF2=4) **OR** If CF1=2 AND (CF3=3 OR CF3=4)
Hearing	2.17%	1.22%	If CF4=1 AND (CF5=3 OR CF5=4) **OR** If CF4=2 AND (CF6=3 OR CF6=4)
Walking	2.66%	2.42%	If CF7=1 AND (CF8=3 OR CF8=4) OR (CF9=3 OR CF9=4) **OR**

Domain of difficulty	Proportion of girls in baseline survey dataset of intervention (%)	Proportion of girls in baseline survey dataset surveyed of comparison (%)	Guidance (record as true if they meet the criteria below)
			If CF7=2 AND (CF12=3 OR CF12=4) OR (CF13=3 OR CF13=4)
Self-care	0.48%	0.81%	CF14=3 OR CF14=4
Communication	3.38%	2.02%	CF15=3 OR CF15=4 **OR** CF16=3 OR CF16=4
Learning	6.33%	4.03%	CF17=3 OR CF17=4
Remembering	8.01%	4.84%	CF18=3 OR CF18=4
Concentrating	2.68%	2.86%	CF19=3 OR CF19=4
Accepting Change	2.94%	6.17%	CF20=3 OR CF20=4
Controlling Behaviour	2.93%	3.66%	CF21=3 OR CF21=4
Making Friends	4.84%	2.42%	CF22=3 OR CF22=4
Anxiety	7.93%	9.27%	CF23=1
Depression	6.01%	4.03%	CF24=1
Girls with disabilities overall	29.57%	27.42%	**Note**: The percentage of girls with disabilities (functional difficulty) is represented by those for whom **at least one domain** is coded 3 or 4 [1 for Anxiety or Depression] (true) as shown above. This is the total proportion meeting at least one of the criteria outlined above. When reporting this, please ensure you do it accurately do not take the sum of the %s above as it will result in double counting.
Source: N =	Girls survey N = 416	Girls Survey N = 248	

Challenges in baseline data collection and limitations of the evaluation design

STS and SAGE faced several challenges during the quantitative data collection and analysis:

- Many girls and caregivers initially selected into the sample were unavailable during data collection. The overall replacement rate was 76.9%, with only 167 girls from the original sample available for testing and surveying. Among the intervention cohort, 81.3% of the sample was replaced; among the comparison cohort, 68.6% of the sample was replaced. The highest replacement rate for the girls survey was among girls aged 15–19 and in the Manicaland district (377 replacement girls). Most girls recruited as replacements are programme beneficiaries – their participation was confirmed by Plan as of this writing – and therefore were retained in the baseline sample. Several girls indicated they were enrolled in school; SAGE staff followed up and determined these girls are enrolled in CBLHs, not in formal schools. Due to the high level of replacements, key demographic information collected in the enrolment database was not available at baseline for disaggregation, including girls' caregiver status, marital status, whether they had ever been to school, grade level at which they dropped out and religion. Although the parent/caregiver survey asks key demographic questions, these were not available for all girls in the survey, and the enrolment database remains the only source of this demographic information. At future timepoints, Plan may consider adding the demographic questions to the girls surveys to ensure the data can be disaggregated by these key characteristics.
- Of the girls who participated in the learning assessments, 8.20% were missing survey responses. This means that the final data set does not include a complete one-to-one ratio of learning assessment data to girls survey data.
- The team attempted to match replacement girls to the girl in the original sample according to key demographics and barriers. During data collection, enumerators noted in the daily tracking sheets the matches between the original sampled girl and her replacement. It was not always feasible to find a match with all key demographics and barriers given time constraints and data collection schedule. The SAGE team used the age and enrolment requirements for replacement girls to participate in the baseline, but certain demographics or barriers may not have been applicable to each replacement girl.
- Due to limited time and budget, STS and SAGE did not pilot survey items prior to the operational baseline data collection. Instead, surveys were pretested with a limited number of respondents to assess the length of the surveys, appropriateness of Shona and Ndebele translations of instructions and relevance of items for the target population. Without sufficient sample sizes, it was not possible to test the reliability of items before operational baseline data collection. At future evaluation points, additional items may be added to the indices to improve the index reliability measure.
- Although STS trained enumerators on accommodating girls with disabilities during the assessment and provided notes on which girls would require accommodations based on programme screening data, only 32 girls (4.40%) used the large-print stimuli accommodation and no girls (0.00%) used assistive devices such as glasses, magnifiers or hearing aids. This was likely because the number of girls selected into the sample who were identified by SAGE disability screening partners as needing assistive devices was small, and because the programme had not yet distributed devices at the time of baseline data collection. Additionally, enumerators knew the accommodations required for girls who had been pre-selected on the sample and replacement lists; however, they would not have

known any special needs for girls who were identified on the day of data collection. If any girls from the baseline sample are provided with assistive devices during the intervention, they will no longer be able to be in the sample because the girls must use—or not use—the same devices at each evaluation time point. It would be unethical not to allow a girl to use an assistive deceive just to ensure comparability.

- STS assumed a 6% prevalence rate of girls with disabilities (GWD) based on initial programme targets. Their screenings were only conducted in the treatment communities. This proportion is significantly lower than the proportion of girls who reported having some or a lot of difficulty on the child functioning questions of the Washington Group questions in the baseline survey.

Cohort tracking and next evaluation point

To facilitate tracking the same girls from the baseline into subsequent evaluation points, STS captured the names and unique IDs of all girls and any parents, boys, caregivers or heads of households sampled. Identifiers are available and should be verified by the project for replacement girls so that the same girls can be identified in future evaluation points as well as project monitoring data with evaluation data. The EE will need to rely on SAGE staff and CBLH facilitators to locate sampled girls at the next evaluation point to ensure adherence to the longitudinal design of the evaluation. The second midline and the endline evaluations will provide an opportunity for Plan to follow-up with girls on their transition pathways.[28]

4.5 Qualitative evaluation methodology

Qualitative data collection tools

The qualitative tools that were administered at baseline are detailed in Table 12. Qualitative baseline findings were supplemented by data collected through the Gender Analysis, as sample sizes for baseline were small due to budget limitations. In addition to providing findings related sustainability, qualitative tools also examined baseline status of O2 Transition, IO4 Communities demonstrate more positive gender attitudes and IO5 Strong and active partnerships with MoPSE officials and other civil society actors actively advocate for more inclusive, gender-responsive education policies.

Table 11: Qualitative evaluation tools

Tool name	Relevant indicator(s)	Who developed the tool?	Was tool piloted?	How were piloting findings acted upon (if applicable)	Was FM feedback provided?
KII with MoPSE officials[29]	O3.5 O3.6a O3.6b IO5.1 IO5.2	STS, Plan	No	n/a	n/a

[28] Attrition buffers were incorporated into sample size calculations to account for girls from the baseline sample who cannot be tracked and assessed in year 3 and year 5 evaluation points. See **Error! Reference source not found.**.
[29] Includes district and national officials.

Tool name	Relevant indicator(s)	Who developed the tool?	Was tool piloted?	How were piloting findings acted upon (if applicable)	Was FM feedback provided?
KII with community leaders	O3.1 O3.2 IO4.1 IO4.2 IO4.3	STS, Plan	No	n/a	n/a
KII with formal school head	O2.1 O3.1 O3.2 IO4.1 IO4.2 IO4.3	STS, Plan	No	n/a	n/a

Qualitative sample selection and sample sizes

The baseline qualitative sample was developed in consideration of budgetary limitations, and the sample selection was conducted purposively. All respondents were pre-selected by SAGE staff. Sample sizes by type of key informant interview (KII) are included in Table 12. All sample sizes agreed upon in the inception report were met during operational data collection.

Table 12: Qualitative sample sizes

Tool	Beneficiary group	Sample size agreed in MEL framework [30]	Actual sample size	Remarks on why there are major differences between anticipated and actual sample sizes (if applicable)
KII with MoPSE officials	District-level officials	4	4	n/a
	National-level officials	1	1	n/a
KII with community leaders	n/a	2	2	n/a
KII with formal school head	n/a	3	3	n/a

Qualitative field researchers

Similar to the selection and hiring process for the quantitative enumerators, STS and Plan worked collaboratively to recruit, hire and train qualitative field researchers for the operational baseline data collection activities. Based on previous experience with qualitative research, 3 data collection

[30] Sample sizes were proposed in Inception Report.

team supervisors were selected for the KII qualitative survey data collection. The selected field researchers had extensive prior experience with qualitative research—including administering focus group discussions (FDGs) and KIIs with adolescents on SRHR and GBV topics. One researcher was fluent in Ndebele, and 2 were fluent in Shona. Before training commenced, the selected field researchers signed contracts with Select Research that stipulated their expected roles and expected professional conduct during training and data collection.

The baseline qualitative researcher training, facilitated by STS with support from Select and SAGE, took place on 1 August in Harare. Training sessions covered the objectives of the SAGE study and the qualitative component, qualitative research practices and an overview and practice of each KII. All 3 researchers were trained in facilitation and note-taking to enable them to rotate roles during the data collection.

Qualitative data collection

Qualitative data collection took place from 5 August to 19 September 2019. STS drafted a schedule of for each qualitative activity, and SAGE developed the final schedule for the KIIs. Qualitative researchers contacted SAGE district staff prior to their visit to reconfirm the schedule of activities and ensure respondent participation.

All KIIs were administered in English and Shona or Ndebele. Researchers took detailed field notes and reflections during the activities. Researchers were required to securely submit a debrief form and expanded notes in English at the end of each day. Researchers supplemented their expanded notes with audio-recordings. Although STS requested that these were submitted each night, researchers ultimately completed these within a 2- to 3-day time period. Delays were due to the rigour of the data collection schedule and the quantity of qualitative data collected each day.

STS reviewed documents daily for completeness, outstanding questions, concerns or clarifications. STS and the qualitative researchers communicated during data collection by WhatsApp, following up with questions about the data and quotas as well as any logistical challenges that may have been encountered.

Qualitative data handling and analysis

Qualitative researchers managed transcription and translation per STS guidance. The notetaker took handwritten field notes during KIIs.[31] Utilizing the handwritten field notes as references, the notetaker and facilitator collaboratively completed an expanded notes template in English for each KII. The most pertinent quotes were also typed up verbatim in the language of the interview, translated to English and included in the expanded field notes. Qualitative researchers did not complete verbatim transcripts and translations; however, their expanded field notes and translations of key quotes from local language to English were reviewed and cross-checked by the facilitator to ensure quality and accuracy.

Qualitative researchers uploaded all data—including expanded field notes—to STS's secured, password-protected server. All raw qualitative data and materials were returned to the local data collection firm, Select Research, after the completion of data collection.

Finalized expanded field notes were coded and analysed systematically in Microsoft Word. All coding was completed by a single user. The qualitative data analysis methodology incorporated an iterative approach and included content analysis and constant comparison of narrative data to

[31]This included quotes, key points and themes that emerged for each question, non-verbal activity or body language, as well as any big ideas, thoughts or take-aways from the note-taker.

identify and validate emerging themes. A list of themes was developed based on the LNGB baseline report template and SAGE's outcomes and IOs. Qualitative data were coded according to these themes. While observations by researchers were included in the qualitative analysis, reflections and recommendations are clearly distinguished from the raw data and findings.

Challenges in baseline qualitative data collection, handling and analysis and limitations of the qualitative aspects of the evaluation design

Due to time, budget and logistical constraints, STS utilized detailed field notes in place of fully translated transcriptions. Expanded field notes produced by the note-taker enabled a quicker turnaround that was less labour intensive and fit within the budget constraints the baseline evaluation. However, the discussions, reflections and insights from KIIs may be limited due to a lack of full transcriptions and translations.

The number of KIIs that could be conducted at baseline was limited due to budget constraints. District officials, community leaders and formal school heads were interviewed from a subset of SAGE's intervention areas. Qualitative findings from KIIs should be understood as only representing a portion of the programme's districts.

5. Key characteristic subgroups and barriers of baseline samples

The following section examines the main characteristics of the subgroups of interest and the barriers to learning and transition that they face. This section also examines the intersection between the main barriers and the characteristics to help determine how appropriate the SAGE programme activities are for these subgroups and if the ToC is appropriate. Barriers were drawn from the qualitative study, and STS used surveys to quantify barrier prevalence to the extent feasible.

5.1 Educational marginalisation

SAGE identified the characteristic subgroups presented in Table 13, which are a critical part of girls' enrolment marginalisation criteria. Proportions of girls by disability are presented in Table 10.

Table 13: Characteristic subgroups

Characteristic	Proportion of sample with this characteristic	N of subgroup	How characteristic was calculated
Girl with disabilities	29.57%	123 out of 416	Calculated using the Washington Group Functional Difficulty Questions. Girls counted as having a disability if they have one or more functional difficulty.
Age Group: 9 and under	0.67%	3 out of 449	Used age given by girl on Learning Assessment. If no age given there, used age given by girl on girls survey.
Age group: 10–14	25.84%	116 out of 449	
Age group: 15–19	72.16%	324 out of 449	
Age group: 20 and older	1.34%	6 out of 449	

SAGE identified the following additional characteristics for programme beneficiaries in the programme's TOC and collected these data during beneficiary selection: high poverty, high chore burden, married or about to get married, has or is expecting a child, religion and chronic illness. These characteristics are maintained in the enrolment database for beneficiaries on an ongoing basis. Given the high number of replacement girls in the sample, this information is not available for all girls in the final baseline sample. As a result, EE could not disaggregate baseline learning outcomes by these characteristics.

The key barriers to learning and transition available for disaggregation are listed in **Error! Reference source not found.**. To populate these barriers, STS used a mixed-methods approach. First, STS analysed the Gender Analysis report (Annex 12) to identify the key barriers

mentioned by adolescent girls in the FGDs.[32] The key barriers to education identified through qualitative data included challenges related to the accessibility of schools, including long distances to school and safety travelling to or from school; GBV, including early marriage, early pregnancy, sexual exploitation and violence; families' lack of approval for girls' education, lack of access to learning about SRHR and low self-esteem.

The results presented in this section cover quantitative data that is not included in the IOs section.[33] Quantitative items in the girls survey captured specific aspects of the barriers identified in the Gender Analysis. Therefore, this section is not intended to replace the broader findings in the Gender Analysis but to provide subgroups for further disaggregating outcomes.

For each barrier identified in the Gender Analysis, related items from the quantitative girls survey were mapped and used for disaggregation. In the Gender Analysis and the girls survey, accessibility was defined as long distances to school and safety concerns when travelling to school, as well as the physical accessibility of the school infrastructure. To measure this barrier with the quantitative data, one item on the girls survey tracked girls who reported travelling more than 30 minutes to CBLH. The quantitative analysis did not explicitly examine perceptions of and experiences around GBV. Rather, the quantitative data examined whether girls reported having a safety net for GBV, including a safe space in the community outside of their home and knowing where to go if they experience violence. Using 5 items in the girls survey, the barrier focuses on girls' perceived lack of safety net for GBV. The quantitative data also examined girls' perceptions around the concept of a girl's right to an education. This barrier was measured using 5 items in the girls survey that addressed if a girl perceived education as a child's right. At baseline, the quantitative data indicated a lack of voice and ability to speak up as an additional barrier girls face. This was defined as girls who are not able to talk to a parent or caregiver about issues that are important to them, who cannot speak up for girls' rights in the community and who lack the confidence to work with others to help girls access education. This barrier was measured using an index of 3 items on the girls survey. Table 14 lists the barriers for which quantitative data were available and the proportion of the sample within that subgroup.

Table 14: Barriers

Barriers: Identified in quantitative surveys	Proportion of sample affected by this barrier	N of subgroup	How barrier was calculated using the survey data
Accessibility	70.53%	280 out of 397	Girls who reported traveling more than 30 minutes to CBLH on girls survey
Lack safety net for GBV	36.71%	152 out of 414	Girls who report at least two of the 3 criteria from the girls survey: not having a safe place in community, not having somewhere safe to go outside the home, and not knowing where to go for support if they experience violence

[32] Girls included in the Gender Analysis were not necessarily included in the baseline, as the Gender Analysis took place several months prior to the baseline. Nevertheless, sample selection criteria for the Gender Analysis was similar to enrolment, and girls' experiences from the Gender Analysis are likely similar to those girls included in the baseline sampling frame.

[33] Once items that were intended to report on IO's were analyzed, remaining items were used to describe the sample by subgroups and barriers to avoid any overlap of items with IO indices.

Barriers: Identified in quantitative surveys	Proportion of sample affected by this barrier	N of subgroup	How barrier was calculated using the survey data
Lack of right to an education	4.79%	22 out of 459	Girls who perceive that at least two of the following criteria from the girls survey are true: that children do not have the right to go to school and CBLH; that girls do not have the right to go to school and CBLH; that boys do not have the right to go to school and CBLH; that children with disabilities do not have the right to go to school and CBLH.
Lack of enabling environment for quality education	11.03%	45 out of 408	Girls who reported 'no' or strongly disagree/disagree' to at least 3 of the following 8 items on the girls survey are facing barriers: school has books, computers, drinking water facilities, seats, toilet to use, and CE makes students feel welcome, treats boys and girls differently and often absent for class
Menstruation	55.89%	204 out of 365	Any girl who says she does not have materials to use during period, misses school because of period OR has no one to talk to about period is classified as facing the barrier
Lack of voice and ability to speak up	20.35%	81 out of 398	Any girls who does not feel able to talk to parents / caregivers / spouses about issues that are important to them; to speak up for girls' rights in community; or feel confident to work with others to make sure other girls can access education

Findings indicate that accessibility to school is the most frequently experienced barrier—70.53% of girls experienced this barrier at baseline. Menstruation and a lack of safety net for GBV were also mentioned by girls—55.89% of girls mentioned barriers around menstruation, and 36.71% indicated lacking a safety net.

5.2 Intersection between key characteristics subgroups and barriers

The intersections between characteristic subgroups and barriers are presented in Table 15.[34] Girls in 4 age groups are included—two are the focus age groups for SAGE interventions. The youngest age group includes 3 girls who were age 9 or under, and the oldest age group includes 5 girls who were age 20 or older. Subsequent analyses by age group exclude these two groups due to the small sample sizes and because they are outside of the focus of SAGE interventions.

Accessibility is the most common barrier faced by girls—affecting at least two-thirds of girls with disabilities, girls aged 10–14 and girls aged 15–19. Across all subgroups, one-third of girls lack a safety net for GBV. Girls who faced barriers related to menstruation tended to be aged 10–14, followed by girls with disabilities. At least one-quarter of girls with disabilities and a slightly smaller proportion of girls aged 10–14 lack of voice or have an inability to speak up.

[34] Statistical (chi-squared) tests are not included on relationships between marginalisation characteristics and barriers, as the sample was not powered to be large enough to make generalisations within subgroups.

Table 15: Key barriers to education by characteristic subgroups

	Girl with disabilities (n=115)	Age group: 9 and under (n=3)	Age group: 10–14 (n=103)	Age group: Age 15–19 (n=279)	Age group: 20 and older (n=5)
Accessibility	31.43% of girls who walk 30 minutes or more to CBLH have a functional difficulty (88 of 280). 76.52% of girls with a functional difficulty walk 30 or more minutes to CBLH (88 of 115).	0.73% of girls who walk 30 or more minutes to CBLH are aged 9 and under (2 out of 274). 66.67% of girls who are aged 9 and under walk 30 or more minutes to CBLH (2 out of 3.)	25.55 % of girls who walk 30 or more minutes to CBLH are aged 10–14 (70 out of 274). 67.96% of girls who are aged 10–14 walk 30 or more minutes to CBLH (70 out of 103.)	72.63% of girls who walk 30 or more minutes to CBLH are aged 15–19 (199 out of 274). 71.33% of girls who are aged 15–19 walk 30 or more minutes to CBLH (199 out of 279.)	1.09% of girls who walk 30 or more minutes to CBLH are aged over 20 (3 out of 274). 60.00% of girls who are aged over 20 walk 30 or more minutes to CBLH (3 out of 5.)
Lack of safety net for GBV	29.61% of girls who face barriers around a lack of a safety net for GBV have a functional difficulty (45 of 152). 36.89% of girls who have a functional difficulty face	0.67% of girls who face a lack of a safety net for GBV are aged 9 and under (1 out of 150). 33.33% of girls who are aged 9 and under face barriers around a lack of a safety net for GBV (1 out of 3.)	26.67 % of girls who face barriers around a lack of a safety net for GBV are aged 10–14 (40 out of 150). 35.71% of girls who are aged 10–14 face barriers around a lack of a safety	70.00 % of girls who face barriers around a lack of a safety net for GBV are aged 15–19 (105 out of 150). 36.71% of girls who are aged 15–19 face barriers around a lack of a safety net for GBV (105 out of 286.)	2.67 % of girls who face barriers around a lack of a safety net for GBV are aged 20 and older (4 out of 150). 80.00% of girls who are aged 20 and up face barriers around a lack of a safety net for GBV (4 out of 5.)

	Girl with disabilities (n=115)	Age group: 9 and under (n=3)	Age group: 10–14 (n=103)	Age group: Age 15–19 (n=279)	Age group: 20 and older (n=5)
	barrier around a lack of a safety net for GBV (45 of 122).		net for GBV (40 out of 112.)		
Lack of right to an education	31.82% of girls lacking a perceived right to an education have a functional difficulty (7 out of 22). 5.69% of girls with functional difficulties lack a perceived right to an education (7 out of 123).	None of the 3 girls aged 9 and under lack a perceived right to an education.	36.36% of girls who lack a perceived right to an education are aged 10–14 (8 out of 22). 6.90% of girls aged 10–14 lack a perceived right to an education (8 out of 116).	63.32% of girls who lack a perceived right to an education are aged 15–19 (14 out of 22). 4.32% of girls aged 15–19 lack a perceived right to an education (14 out of 324).	None of the 6 girls aged 20 and older lack a perceived right to an education.
Lack an enabling environment for quality education	42.22% of girls lack an enabling environment for quality education have a functional difficulty (19 out of 45).	2.22% of the girls who lack an enabling environment for quality education are aged 9 and under (1 out of 45). 33.33% of the girls	17.78% of the girls who lack an enabling environment for quality education are aged 10–14 (8 out of 45).	73.33% of the girls who lack an enabling environment for quality education are aged 15–19 (33 out of 45). 11.66% of the girls aged 15–19 lack an	6.67% of the girls lack an enabling environment for quality education are aged 20 and up (3 out of 45). 60.00% of the girls aged 20 and up lack

	Girl with disabilities (n=115)	Age group: 9 and under (n=3)	Age group: 10–14 (n=103)	Age group: Age 15–19 (n=279)	Age group: 20 and older (n=5)
	15.70% of girls with a functional difficulty lack an enabling environment for quality education (19 out of 121).	aged 9 and under lack an enabling environment for quality education (1 out of 3).	7.27% of the girls aged 10–14 lack an enabling environment for quality education (8 out of 110).	enabling environment for quality education (33 out of 283).	an enabling environment for quality education (3 out of 5).
Menstruation	28.43% of girls facing barriers around menstruation have a functional difficulty (58 out of 204). 55.24% of girls with a functional difficulty face barrier around menstruation (58 out of 105.)	1.50% of girls facing barriers around menstruation are aged 9 and under (3 out of 200). 100% of girls aged 9 and under face barriers around menstruation (3 out of 3).	26.00% of girls facing barriers around menstruation are aged 10–14 (52 out of 200). 74.29% of girls aged 10–14 face barriers around menstruation (52 out of 70).	70.00% of girls facing barriers around menstruation are aged 15–19 (140 out of 200). 50.18% of girls aged 15–19 face barriers around menstruation (140 out of 279).	2.50% of girls facing barriers around menstruation are aged 20 and older (5 out of 200). 83.33% of girls aged 20 and up face barriers around menstruation (5 out of 6).
Lack of voice and the ability to speak up	35.80% of girls who lack voice and the ability to speak up have functional	1.25% of girls who lack voice and the ability to speak up are aged 9 or under (1 out of 80).	23.75% of girls who lack voice and the ability to speak up are aged 10–14 (19	75.00% of girls who lack voice and the ability to speak up are aged 15–19 (60 out of 80).	None of the 6 girls aged 20 and older lack voice and the ability to speak up.

	Girl with disabilities (n=115)	Age group: 9 and under (n=3)	Age group: 10–14 (n=103)	Age group: Age 15–19 (n=279)	Age group: 20 and older (n=5)
	difficulties (29 out of 81). 24.58% of girls with disabilities lack voice and the ability to speak up (29 out of 118).	33.33% of girls aged 9 and under lack voice and the ability to speak up (1 out of 3).	out of 106). 17.92% of girls aged 10–14 lack voice and the ability to speak up (19 out of 106).	21.82% of girls aged 15–19 lack voice and the ability to speak up (60 out of 80).	

5.3 Appropriateness of project activities to the characteristic subgroups and barriers identified

Using the available baseline data, there do not appear to be any unanticipated characteristic subgroups that were not considered in intervention planning. Appropriateness of project activities should be re-examined when enrolment data are collected for the large number of replacement girls. Due to the limited nature of the data available in the baseline surveys, greater emphasis should be placed on the Gender Analysis findings for a description of subgroups and barriers. Furthermore, once beneficiary data for the replacement sample is available, subgroup analyses should be conducted.

To measure the prevalence of all barriers and characteristics in the Gender Analysis and enrolment database, additional information needs to be collected for all sampled beneficiaries. Given the transient population SAGE is targeting, it would be useful to triangulate this information between Plan's ongoing monitoring, beneficiary selection database and quantitative surveys at each evaluation point to more fully understand the prevalence and trends of these barriers and characteristics.

The programme interventions appear to address key barriers for key characteristic subgroups as identified in the Gender Analysis. Accessibility is a major barrier identified through the baseline survey and should be monitored routinely for girls. The programme should also ensure that girls, especially older girls, have a safety net for reporting and discussing issues around GBV. The programme should also ensure that girls age 15–19 receive support to develop their sense of voice and the ability to speak up. Additionally, the programme should closely monitor and support girls with disabilities and functional difficulties as 76.52% of girls with a functional difficulty walk more than 30 minutes to CBLH, 36.84% report a lack of safety net for issues around GBV and 55.24% reported barriers around menstruation.

Assumptions in the programme's ToC regarding subgroups and barriers hold true based on the findings of the Gender Analysis and limited relevant quantitative data available in the baseline. Further analysis by SAGE may be warranted once the beneficiary selection database is updated to include the replacement girls in the baseline sample.

Project to complete

- The project should respond to the external evaluators' comments on the above questions. In particular the project should respond to:
    - Why the projects theory of change may not correspond with some of the key barriers or characteristic subgroups identified.
    - Whether the project plans to review some aspects of their Theory of change in light of these findings.

See management response in Annex 18.

6. Outcome findings

Baseline results for the following SAGE outcomes are presented in this section:

- O1: Number of highly marginalised girls supported by GEC with improved learning outcomes[35]
- O2: Number of marginalised girls who have transitioned through key stages of education, training or employment
- O3: Project can demonstrate that the changes it has brought about which increase learning and transition through education cycles are sustainable

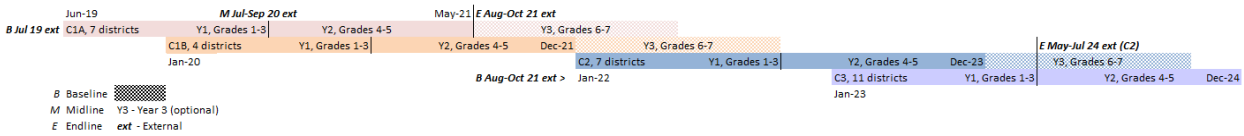
6.1 Learning outcomes

SAGE's first outcome is improved learning outcomes. This section will present findings on the following indicators:

- O1.1: Number of highly marginalised girls supported by GEC with improved literacy outcomes
- O1.2: Number of highly marginalised girls supported by GEC with improved numeracy outcomes

Baseline findings for the third learning outcome—*O1.3 Number of highly marginalised girls supported by GEC with improved life skills outcomes*—are detailed in Section 7.2. This third learning outcome was added in discussions with the FM who requested it be included in this baseline report.

The below diagram outlines the learning levels girls are starting with and their expected level by the subsequent evaluation points:



According to the programme's selection criteria, beneficiaries were eligible if they performed below Grade 5 equivalent on the Wide Range Assessment Tests. These tests were conducted post-enrolment on 2,612 girls, or 64% of the total beneficiaries in C1A. The test was not possible to conduct with the entire cohort due to the limited capacity of the MoPSE team to conduct the assessments. As a result, not all girls who were in C1A were fully screened and deemed eligible prior to enrolment. Similarly, this data is not available for all girls in the baseline.

The first year of the CBLH aims to ensure girls' literacy and numeracy levels are on par with Grade 2 and 3 literacy and numeracy curricula, while the second and optional third year aim for a Grade 5 and 7 equivalent, respectively. Two years of instruction in the CBLH should allow girls to re-enter formal school at Grades 5 or 7, should they wish to take that transitional pathway. To

[35] Baseline results for O1.3 Number of highly marginalised girls supported by GEC with improved life skills outcomes are presented in section 7.2.

compare the current learning levels of girls in the baseline sample, beneficiaries' scores were compared to those of girls in Grades 3, 5 and 7 in formal state schools in the target districts. While the comparison in this section focuses on girls in the sample compared to girls in formal schooling, the study does not assess whether girls in grades 3, 5 and 7 are performing at the grade-level expectations for their grade.

The second evaluation point in 2020, as stated in the MEL and confirmed by Plan, will be at the midline point for C1A and will re-assess the girls from C1A after they complete one year of CBLH.

Headline results

Girls in the treatment cohort, C1A, are compared to girls in formal schools, or the benchmark group. On the EGRA, girls in the treatment cohort performed below a Grade 7 level and at a Grade 3–5 level, on average.[36] On the EGMA, girls in the treatment cohort performed below a Grade 5 level and at a Grade 3 level, on average.[37]

Girls in both the treatment cohort and the benchmark group appeared to have stronger performance in mathematics than in literacy, as evidenced by the higher proportion of girls in the 'proficient' learner category. In general, fewer girls from either group were unable to answer a single item correctly on a subtask in mathematics than did so in literacy. Furthermore, the relationship between EGRA and EGMA performance shows that girls with higher overall EGRA scores tended to have higher EGMA scores. The strongest relationships were observed between EGRA overall score and the missing numbers, subtraction levels 1 and 2 and word problems subtasks.

Baseline findings by cohort and benchmark grade are presented in Supplementary Table 7. There was no statistically significant difference between the treatment cohort's scores and the comparison cohort's scores, meaning that the two cohorts are comparable at baseline.

Supplementary Table 7: Baseline findings by cohort and benchmark

	C1A—Treatment	C2—Comparison	Significant differences between C1A and C2	Benchmark Grade 3	Benchmark Grade 5	Benchmark Grade 7	Significant differences between C1A and Benchmark
EGRA Aggregate Score (overall score across 7 subtasks)	44.55	41.82	No difference	38.75	49.18	67.40	BM Grade 7 significantly higher than C1A & C2

[36] Girls in Grade 7 have significantly higher EGRA scores than girls in the treatment cohort but there is no significant difference in the performance of girls in Grades 3 or 5 and the treatment cohort.
[37] Girls in Grade 5 and 7 have significantly higher EGMA scores than girls in the treatment cohort but there is no significant difference in the performance of girls in Grade 3 and the treatment cohort.

	C1A—Treatment	C2—Comparison	Significant differences between C1A and C2	Benchmark Grade 3	Benchmark Grade 5	Benchmark Grade 7	Significant differences between C1A and Benchmark
EGMA Aggregate (overall score across 8 subtasks)	66.25	67.65	No difference	65.93	78.90	87.74	BM Grade 5 and 7 significantly higher than C1A & C2

Literacy

Girls' baseline literacy findings are first presented in two ways: first using learner categories provided in the report template, and second using mean percentage correct scores.

First, the proportions of girls in each of the 4 learner categories is compared across the treatment cohort and the comparison cohort (Figure 13) and with the benchmark group Grade 5 (Tables 17a and 17b) by subtask.[38] Learner categories are defined as non-learners who answered 0% of questions correctly, emergent learners who answered 1–40% of questions correctly, established learners who answered 41–80% of questions correctly and proficient learners who answered 81–100% of questions correctly.

Then, the mean percentage correct scores are compared across the treatment cohort and the same two groups: first, the comparison cohort and second, the Grade 5 benchmark group. Since the average aggregate literacy score for girls in the treatment cohort is below that of girls in Grade 7 but comparable to that of girls in Grades 3 and 5 benchmark groups, the analysis focuses specifically on comparisons to Grade 5 girls within the benchmark group. With both comparisons, statistically significant differences are identified (Supplementary Table 7).

Results by Learner Categories

Figure 13 and Tables 17a and 17b present the proportions of girls in each learner category by EGMA subtask for C1A and Grade 3 girls, respectively. Overall, girls in the treatment group struggled most with comprehension as a skill followed by decoding. The highest proportion of girls were classified as non-learners in reading and listening comprehension subtasks and the letter sound identification subtask.

The highest proportion of non-learners were observed in the reading comprehension subtasks, both on the short and the long passage. Specifically, 41.61% and 48.15% of girls were unable to

[38] Because girls in the treatment cohort performed comparably to girls in Grades 3 and 5, no comparisons are made to girls in Grade 7.

answer a comprehension question correctly on the short and long passages, respectively.[39] Similarly, 37.69% of girls were non-learners on the listening comprehension subtask. By comparison, in the benchmark group (Grade 5), 19.83% and 38.02% girls were non-learners on the reading comprehension short and long passages, respectively, and 23.14% on the listening comprehension (Table 17b).

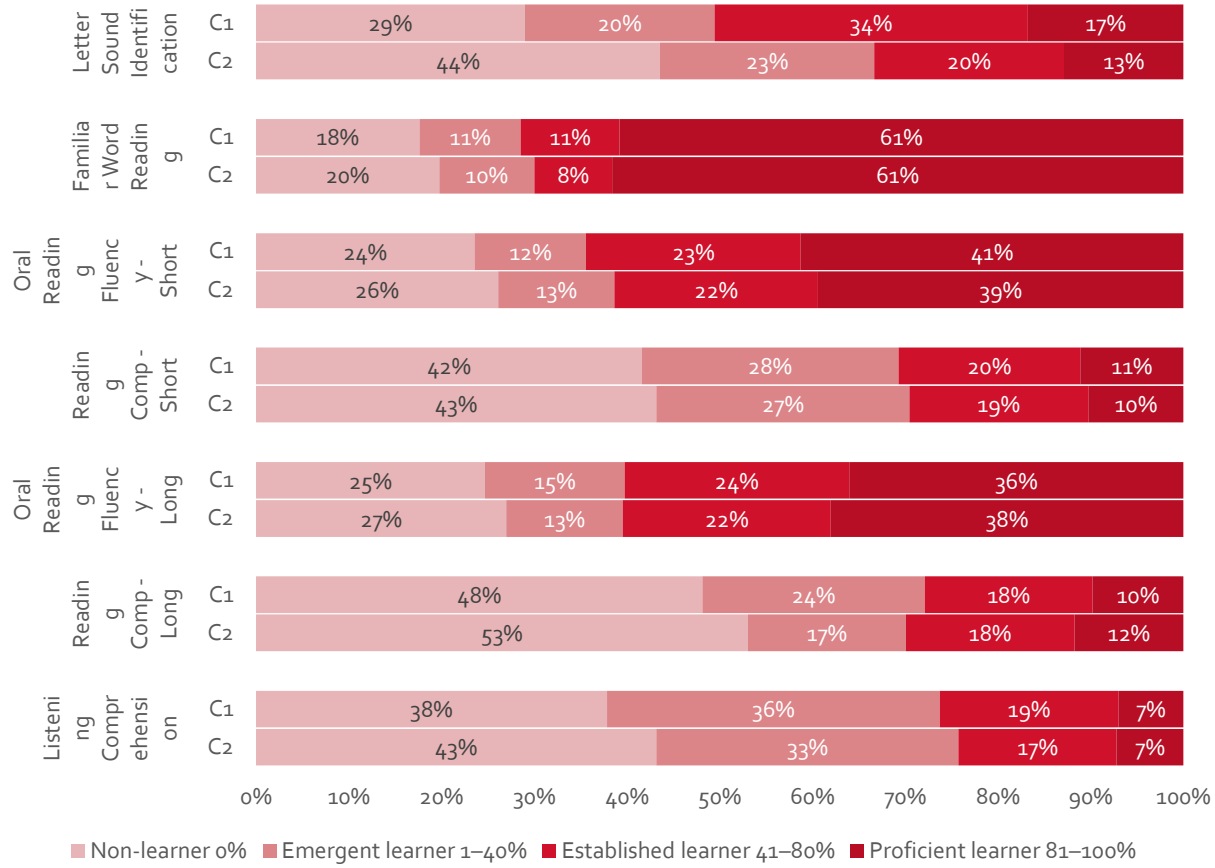
When examining the proportion of learners who were proficient on the comprehension subtasks, approximately one in 10 girls in the treatment cohort was classified as proficient in reading comprehension—11.11% on the short passage and 9.80% on the long passage. In listening comprehension, the lowest proportion of girls on any subtask was classified as proficient—6.97%. Compared to girls in Grade 5, the proportion of girls who were in the proficient category on the short and long reading comprehension subtasks was 7.44% and 6.61% respectively; on the listening comprehension subtask, 4.13% were in the proficient category (Table 17b).

On letter sound identification and familiar word reading, more girls in the treatment cohort are proficient in whole word reading than in decoding. Specifically, 16.78% and 60.78% of girls in the treatment cohort were proficient on letter sound identification and familiar word reading, respectively. At the same time, almost one-third of treatment girls were non-learners on the letter sounds subtask and almost one in 5 were non-learners on the familiar word reading subtask—suggesting that there are girls who continue to struggle with both tasks. When compared to girls in the benchmark group, the trend was similar to that observed among the treatment cohort. Specifically, 38.02% of Grade 5 girls were non-learners on letter sound and 4.13% on familiar word reading while 13.22% of girls were proficient in letter sounds and 76.86% were proficient in familiar word reading (Table 17b).

When asked to read two passages of connected text—one short text consisting of 65 words and one long text consisting of 93 words—one-quarter of girls in the treatment cohort were non-learners while one-third were proficient—41.18% on the short passage and 35.95% on the long passage. Of the girls in Grade 5, less than 5% were non-learners on either reading passage (3.31%) and almost half were proficient on the short passage (45.45%) and one-third on the long passage (33.06%) (Table 17b).

[39] Reading comprehension zero scores are comprised of girls who were not given the opportunity to answer any questions due to receiving a zero score on the oral reading fluency subtask; and girls who were asked comprehension questions but did not get any correct.

Figure 13: Proportion of girls in literacy learner categories by subtask, C1A and C2



Literacy Results by Mean Percentage Correct Scores

Statistical significance tests on the percentage correct scores were conducted to compare the performance of girls in the treatment cohort to the two other groups—girls in the comparison cohort and girls in the benchmark group. Results of these significance tests are shown by subtask in Supplementary Table 7. The results show that girls in the treatment cohort had, on average, statistically significantly higher mean percentages correct than did girls in Grade 5 on the letter sound subtask, but the opposite was true for the familiar word subtask. For all other EGRA subtasks, there were no statistically significant differences between the treatment cohort and Grade 5 girls on the average percentage correct scores. Compared to girls in the comparison cohort, girls in the treatment cohort had comparable scores at baseline on all EGRA subtasks.

Supplementary Table 8: Mean literacy scores by subtask and significance results between C1A, C2 and Grade 5

	C1A (treatment)		C2 (comparison)		Benchmarking Grade 5		Significant differences between C1A and C2	Significant differences between C1A and G5
	n	Mean	n	Mean	n	Mean		
Mean Aggregate Literacy Score	459	44.55	264	41.82	119	49.18	No difference	No difference
Letter Sound Avg. Percentage Correct	459	40.43	264	28.44	119	29.92	No difference	C1>BM G5
Familiar Word Avg. Percentage Correct	459	67.70	264	67.08	114	87.67	No difference	C1<BM G5
Listening Comprehension Avg. Percentage Correct	457	30.28	264	27.73	112	31.96	No difference	No difference
Oral Reading Fluency Short Passage (Correct Words Per Minute)	458	65.13	264	76.70	113	85.19	No difference	No difference
Reading Comprehension Short Passage Avg. Percentage Correct	459	32.81	264	31.89	112	36.43	No difference	No difference
Oral Reading Fluency Long Passage (Correct Words Per Minute)	458	65.61	263	66.47	112	87.69	No difference	No difference
Reading Comprehension Long Passage Avg. Percentage Correct	459	29.41	264	28.94	112	32.86	No difference	No difference

Note: Fluency scores for Oral Reading Fluency short and long passages show the mean fluency score before the fluency ranges were capped at 100 for inclusion in the aggregate EGRA score.

Ceiling effects appear at baseline on the familiar word reading, oral reading fluency short passage and oral reading fluency long passage subtasks. More than one-third of girls in the treatment cohort were proficient on the reading passages. The same trends are observed for girls in the benchmark Grade 5 group, suggesting that more difficult passages would be needed to capture the reading levels of girls in the benchmark group as well. Of these 3 subtasks, the subtask most unlikely to capture girls' growth in reading skills at midline is the familiar word subtask, followed by oral reading fluency short passage and finally the long passage. The intention of including two passages of varying lengths was to mitigate the high proportion of proficient readers on the short passage observed in the pilot. However, it seems the long passage is inadequate to capture girls' learning at midline.

Given these findings, the programme appears to have enrolled girls who, on average, have relatively high overall EGRA scores, but have low scores in subtasks such as reading and listening comprehension. Because girls have literacy skills comparable to those of girls in Grade 5 in formal schooling, the first year of intervention materials may be misaligned for some girls, as the materials are intended to match a Grade 3 level. As mentioned before, the current analysis does not include a comparison of girls in formal schooling and the expectations of students at Grades 3, 5 and 7 based on the curriculum in each grade. As such, it is outside of the scope of this evaluation to determine whether the materials developed for Year 1 of the programme are at the appropriate learning level for girls in the project; this exercise should be conducted by Plan following the baseline.

Indicator O1.1 will measure improved literacy outcomes of girls participating in the programme after one year of CBLH participation. Alignment of learning materials to expectations in Grade 3 and 5, for those girls who chose this transition pathway, should be examined prior to the next time point. However, due to the high levels of performance of the girls in the treatment group, at least 3 of the current literacy subtasks—familiar word reading, oral reading short passage, and oral reading long passage—appear to be inadequate to capture the growth in girl's literacy skills.

Numeracy

Girls' baseline numeracy findings are presented in the same way as literacy findings. First, the proportions of girls in each of the 4 learner categories is compared across treatment cohort with the comparison cohort (Figure 14) and with the benchmark Grade 3 group (Table 16a and 16b) by subtask. Then, the mean percentage correct scores are compared between the treatment cohort and two groups. Because the average aggregate numeracy score is below that of girls in Grades 5 and 7 but comparable to that of girls in Grade 3, numeracy analysis focuses on comparisons to Grade 3 girls. With both comparisons, statistically significant differences are identified (Supplementary Table 8).

Results by Learner Categories

Figure 14 and Tables 16a and 16b present the proportions of girls in each learner category by EGMA subtask for C1A and Grade 3 girls, respectively. Overall, the majority of girls in the treatment cohort are proficient in number recognition, quantity discrimination, addition level 1 and subtraction level 1, while the majority are established learners on the missing number identification subtask. Similar to the treatment group, the missing numbers subtask was the most difficult for Grade 3 girls, only 1.22% of girls scored as proficient learners.

The largest proportion of girls in the treatment cohort scored as proficient on the number recognition subtask—almost 3 out of 4 girls (71.02%), while 2.83% received zero scores. The second-largest proportion of girls scored as proficient learners on the quantity discrimination and addition level 1 subtask—64.49% and 64.27%, respectively. Among Grade 3 girls, the largest proportion of proficient learners was also observed on the number recognition subtask—73.17% of Grade 3 girls were proficient—followed by the quantity discrimination subtask—57.32% (Figure 14).

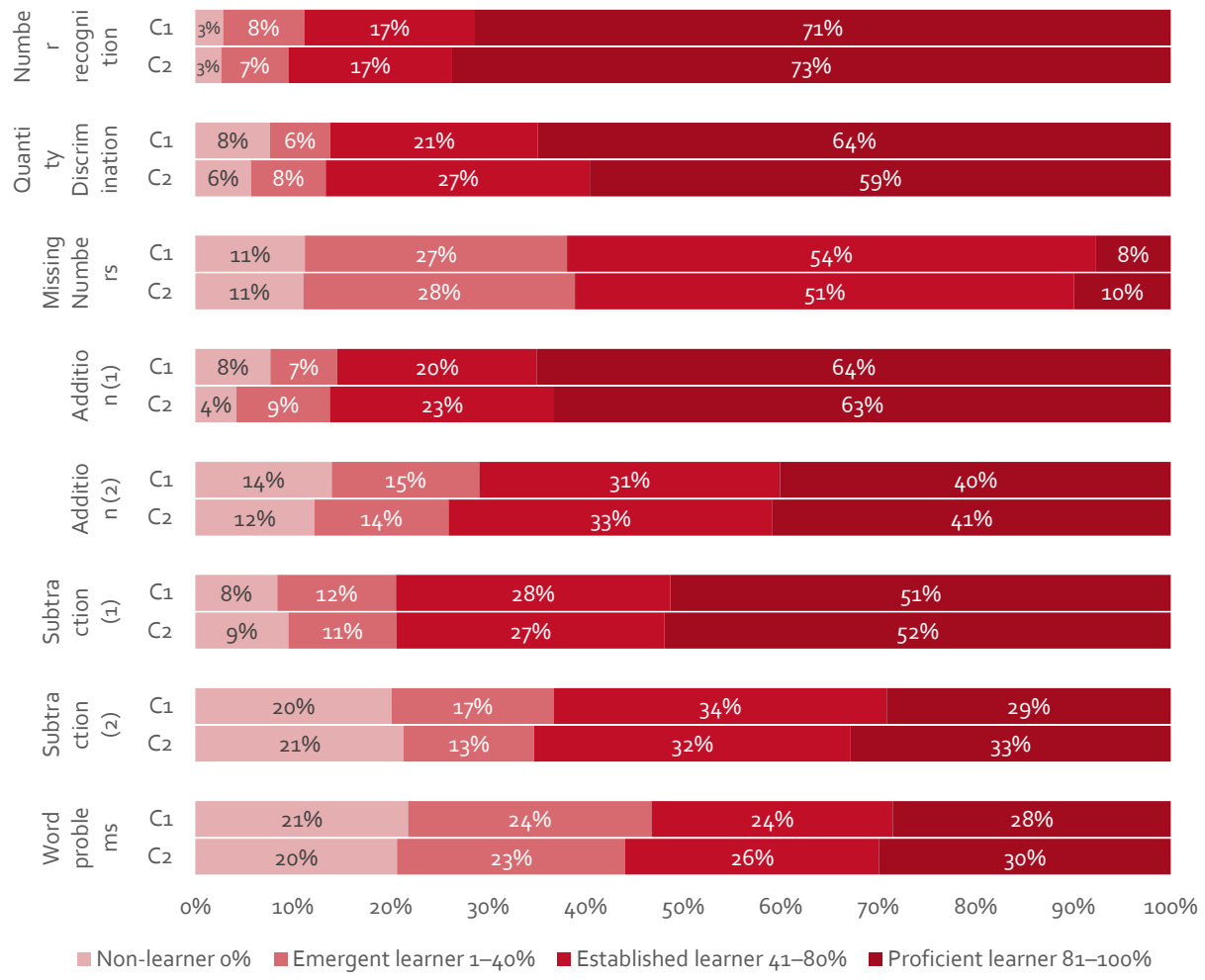
On addition and subtraction, the proportion of non-learners in the treatment group was comparable—with 7.63 and 8.28%. respectively, of students unable to answer any addition or subtraction items correctly. Of those who answered at least one item correctly in the level 1 subtask, 6.32% were unable to answer any addition level 2 items correctly and 11.76% were unable to answer any subtraction level 2 items correctly. The proportion of non-learners on the level 2 subtasks in the treatment cohort were higher than the proportion of non-learners on these subtasks in Grade 3 girls—13.94% versus 4.88%in addition level 2 and 20.04% versus 13.41% on subtraction level 2 (Table 16b).

On the word problems subtask, the proportion of girls who were non-learners was comparable across the treatment cohort (21.35%) and Grade 3 girls (20.73%). However, the proportion of girls in the treatment cohort who were proficient was statistically significantly higher than in the Grade 3 group. Almost one-third of girls in the treatment cohort were proficient in word problems (27.89%) while only one in 10 girls in Grade 3 was proficient (7.32%) (Table 16b).

Numeracy Results by Mean Percentage Correct Scores

Statistical significance tests on the percentage correct scores were conducted to compare the performance of girls in the treatment cohort to two other groups: first, girls in the comparison cohort and second, girls in formal schooling in Grade 3 from the benchmark group. Results of these significance tests are shown by subtask in Supplementary Table 5. The results show that girls in the treatment cohort had, on average, significantly higher mean scores than girls in the benchmark Grade 3 group on the word problems subtask. There were no statistically significant differences on the average percentage correct score between treatment and benchmark (grade 3) groups on the remaining EGMA subtasks. Further, there were no significant differences in the average percentage correct score between treatment and comparison cohorts on any EGMA subtasks.

Figure 14: Proportion of girls in numeracy learner categories by subtask, C1A and C2



Supplementary Table 9: Mean numeracy scores by subtask and significance results between C1A, C2 and Grade 3

	C1A (treatment)		C2 (comparison)		Benchmarking Grade 3		Significant differences between C1A and C2	Significant differences between C1A and G3
	n	Mean	n	Mean	n	Mean		
Mean Aggregate Numeracy Score	457	66.25	262	67.65	73	65.93	No difference	No difference
Number Recognition Avg. Percentage Correct	457	83.97	262	84.87	71	91.97	No difference	No difference

	C1A (treatment)		C2 (comparison)		Benchmarking Grade 3		Significant differences between C1A and C2	Significant differences between C1A and G3
	n	Mean	n	Mean	n	Mean		
Quantity Discrimination Avg. Percentage Correct	456	80	262	79	72	85	No difference	No difference
Missing Number Avg. Percentage Correct	454	50	262	51	72	54	No difference	No difference
Addition Level 1 Avg. Percentage Correct	454	77.14	261	77.72	71	81.62	No difference	No difference
Addition Level 2 Avg. Percentage Correct	457	67	262	69	73	71	No difference	No difference
Subtraction Level 1 Avg. Percentage Correct	452	69.68	262	70.23	71	67.18	No difference	No difference
Subtraction Level 2 Avg. Percentage Correct	457	59	262	61	73	58	No difference	No difference
Word Problems	449	46.03	261	49.23	70	29.29	No difference	C1A> G3

Note: Fluency scores for Oral Reading Fluency short and long passages show the mean fluency score before the fluency ranges were capped at 100 for inclusion in the aggregate EGRA score.

As with the EGRA, ceiling effects do appear to be a concern when examining the fluency rates for timed subtasks and percentage correct scores for untimed subtasks on the EGMA. The only subtask that appears to have fewer than 10% of girls in the proficient category is missing numbers. For all other subtasks, at least one-quarter and as many as three-quarters of girls in the treatment cohort were proficient learners.[40] As a result, the current EGMA subtasks may not adequately capture girls' growth over time.

Given these findings, the programme appears to have targeted girls with relatively high overall EGMA score but lower scores on the missing numbers subtask. Since girls' have numeracy skills comparable to those of girls in Grade 3 in formal schooling, the first year of intervention materials should be reviewed.

Indicator O1.2 will measure improved numeracy outcomes of girls participating in the programme; but given the high risk of ceiling effects, additional items or subtasks need to be added to capture numeracy improvement over time.

[40] While the results for the benchmark group suggest that the word problem subtask may not have a ceiling effect, the results from the treatment cohort show that the tool may have a ceiling effect.

Table 16a: Foundational numeracy skills, C1A

Categories	Subtask 1 Number Recognition	Subtask 2 Quantity Discrimination	Subtask 3 Missing Numbers	Subtask 4 Addition (1)	Subtask 5 Addition (2)	Subtask 6 Subtraction (1)	Subtask 7 Subtraction (2)	Subtask 6 Word problems
Non-learner 0%	2.83%	7.63%	11.11%	7.63%	13.94%	8.28%	20.04%	21.35%
Emergent learner 1–40%	8.28%	6.10%	26.58%	6.75%	15.03%	11.98%	16.56%	24.40%
Established learner 41–80%	17.43%	21.13%	53.59%	20.26%	30.72%	27.67%	33.99%	24.18%
Proficient learner 81–100%	71.02%	64.49%	7.63%	64.27%	39.87%	50.54%	28.98%	27.89%
Source: N= 459	100%	99%	99%	99%	100%	98%	100%	98%

Table 16b: Foundational numeracy skills, Grade 3[41]

Categories	Subtask 1 Number Recognition	Subtask 2 Quantity Discrimination	Subtask 3 Missing Numbers	Subtask 4 Addition (1)	Subtask 5 Addition (2)	Subtask 6 Subtraction (1)	Subtask 7 Subtraction (2)	Subtask 6 Word problems
Grade 3								
Non-learner 0%	0.00%	1.22%	3.66%	0.00%	4.88%	3.66%	13.41%	20.73%
Emergent learner 1–40%	2.44%	2.44%	18.29%	4.88%	14.63%	8.54%	20.73%	39.02%

[41] Results for Grades 5 and 7 can be found in Annex 13.

Categories	Subtask 1 Number Recognition	Subtask 2 Quantity Discrimination	Subtask 3 Missing Numbers	Subtask 4 Addition (1)	Subtask 5 Addition (2)	Subtask 6 Subtraction (1)	Subtask 7 Subtraction (2)	Subtask 6 Word problems
Established learner 41–80%	10.98%	26.83%	64.63%	32.93%	40.24%	50.00%	34.15%	18.29%
Proficient learner 81–100%	73.17%	57.32%	1.22%	48.78%	29.27%	24.39%	20.73%	7.32%
Source: N= 82	87%	88%	88%	87%	89%	87%	89%	85%

Table 17a: Foundational literacy skills, C1A

Categories	Subtask 1 Letter Sound	Subtask 2 Familiar Word Reading	Subtask 3 Oral Reading Fluency – Short	Subtask 4 Oral Reading Comprehension – Short	Subtask 5 Oral Reading Fluency – Long	Subtask 6 Oral Reading Comprehension – Long	Subtask 7 Listening Comprehension
Non-learner 0%	28.98%	17.65%	23.53%	41.61%	24.62%	48.15%	37.69%
Emergent learner 1–40%	20.48%	10.89%	11.98%	27.67%	15.03%	23.97%	35.73%
Established learner 41–80%	33.77%	10.68%	23.09%	19.61%	24.18%	18.08%	19.17%
Proficient learner 81–100%	16.78%	60.78%	41.18%	11.11%	35.95%	9.80%	6.97%
Source: N=459	100.00%	100.00%	99.78%	100.00%	99.78%	100.00%	99.56%

Table 17b: Foundational literacy skills, Grade 5[42]

Categories	Subtask 1 Letter Sound	Subtask 2 Familiar Word Reading	Subtask 3 Oral Reading Fluency – Short	Subtask 4 Oral Reading Comprehension – Short	Subtask 5 Oral Reading Fluency – Long	Subtask 6 Oral Reading Comprehension – Long	Subtask 7 Listening Comprehension
Grade 5							
Non-learner 0%	38.02%	4.13%	3.31%	19.83%	3.31%	38.02%	23.14%
Emergent learner 1–40%	28.10%	2.48%	8.26%	47.93%	14.88%	24.79%	47.11%
Established learner 41–80%	19.01%	10.74%	36.36%	17.36%	41.32%	23.14%	18.18%
Proficient learner 81–100%	13.22%	76.86%	45.45%	7.44%	33.06%	6.61%	4.13%
Source: N= 121	98.35%	94.21%	93.39%	92.56%	92.56%	92.56%	92.56%

Results for life skills, Outcome 1.3, are presented in section 7.2.

6.2 Characteristic subgroup analysis of the learning outcome

Subgroup and barriers analysis

Literacy and numeracy aggregate scores by subgroup and barrier are presented in Table 18: for girls in the treatment and comparison cohorts.[43] There were no statistically significant differences in girls' average aggregate literacy or numeracy scores in the treatment group by province. There was a weak but statistically significant correlation between age and average aggregate literacy and numeracy scores. The correlation between age and the overall EGRA score was 0.34, and the correlation between age and the overall EGMA score was 0.32. These correlations suggest that although older girls perform better, there was high variability in performance despite age. Using the age groups provided by Plan, girls who are 15–19 years old had higher literacy and numeracy aggregate scores than girls who are 10–14 years old.

[42] Results for Grades 3 and 7 can be found in Annex 13.
[43] Information for subgroup analysis was not collected from girls in the benchmark sample.

The majority of girls were from rural areas. However, here were no statistically significant differences in girls' performance between urban, peri-urban and rural areas. Finally, there were no differences in girls' performance by the language in which instructions were provided, Shona or Ndebele.

Girls who had at least one disability—based on the Washington Group Child Functioning questions—are classified as having a functional disability.[44] As noted previously, 29.57% of girls in the treatment cohort have a functional disability as do 27.42% of girls in the comparison cohort. Girls in the treatment cohort who have at least one functional disability have statistically significantly lower literacy and numeracy performance than do girls in the treatment cohort who do not have functional disabilities. Among the comparison cohort, however, there were no differences in the aggregate scores between girls with functional difficulties and those without. In the treatment cohort, girls who had functional difficulties with seeing, walking or communicating had lower numeracy scores than did girls who do not have those functional disabilities. Girls who had communication disabilities in the intervention cohort had lower literacy scores than do girls who did not have a functional disability in communication. There were no statistically significant differences in these subgroups among the comparison cohort.

By subgroup, girls in the treatment cohort who face a barrier in accessibility—defined as long distances to school or CBLH—had statistically significantly lower literacy and numeracy aggregate scores than do girls in the treatment cohort who did not face this barrier. Similarly, girls in the treatment cohort who perceived that they lack the right to education had lower literacy and numeracy aggregate scores than do girls in the treatment cohort who did not have this perception. Finally, girls in the treatment cohort who lack of a voice and ability to speak up had lower EGMA performance than do girls in the treatment cohort who did not face this barrier. At baseline, there were no differences between girls who did face the following barriers and those who did not: lack of enabling environment for quality education and barriers related to menstruation.

Table 18: Scores by key characteristic subgroups and barriers

	Treatment Group (C1A)				Comparison Group (C2)			
	n	Avg. EGRA score	Avg. EGMA score	Significant Difference	n	Avg. EGRA score	Avg. EGMA score	Significant Differences
All girls	459	44.55	66.25		264	41.82	67.65	
Age group 1: 10–14 years old	111	32.19	57.89	Grp 2 > grp 1	102	32.05	50.90	Grp 2 > grp 1
Age group 2: 15–19 years old	316	49.29	69.97		146	59.13	76.97	
Province								
Bulawayo	24	44.44	64.98	none	20	41.53	61.98	
Harare	51	52.52	75.78		30	52.45	77.95	

[44] In the treatment cohort, 61 girls had 1 functional disability, 36 had 2 functional disabilities, 13 had 3, 6 had 4, 5 had 5 and 2 had 6. In the comparison cohort, 42 girls had 1 functional disability, 18 had 2, 3 had 3, 3 had 4 and 1 had 6.

	Treatment Group (C1A)				Comparison Group (C2)			
	n	Avg. EGRA score	Avg. EGMA score	Significant Difference	n	Avg. EGRA score	Avg. EGMA score	Significant Differences
Manicaland	306	46.02	72.89		179	37.84	64.90	Mash. E EGRA > Manicaland
Mashonaland East	30	52.52	75.78		34	53.27	75.90	
Matabeleland South	48	46.02	72.89		1	51.45	77.71	
District								
Bulilima	48	46.02	72.89	EGRA: Chim > Mut. Rural; EGMA: Bul, Chim, Mutasa, Mutoko > Mut. Rural	1	51.45	77.71	EGRA, EGMA: ep, Mutasa, Mutoko > Chim;
Chimanimani	64	52.77	72.91		69	28.25	54.97	
Epworth	51	43.04	67.12		30	52.45	77.95	
Imbizo	24	36.24	55.36		20	41.53	61.98	
Mutare_Rural	130	38.30	57.21		89	40.79	69.60	
Mutasa	112	46.82	69.47		21	56.88	78.05	
Mutoko	30	52.52	75.78		34	53.27	75.90	
Area								
Urban	3	14.70	57.99	none	21	42.97	64.37	none
Peri-urban	77	43.28	65.94		27	48.06	76.24	
Rural	336	45.36	67.41		200	41.12	67.68	
Language in which instructions were given on assessment								
Shona	387	44.88	66.10	none	243	41.81	68.08	none
Ndebele	72	42.76	67.04		21	42.00	62.73	
Disability subgroup								
Seeing	25	34.01	55.80	EGMA: has < does not	5	41.52	77.88	None
Hearing	9	41.71	65.51	None	3	5.89	36.25	EGRA, EGMA: has < does not
Walking	11	29.98	46.69	EGMA: has < does not	6	43.97	63.54	None

	Treatment Group (C1A)				Comparison Group (C2)			
	n	Avg. EGRA score	Avg. EGMA score	Significant Difference	n	Avg. EGRA score	Avg. EGMA score	Significant Differences
Self-care	2	44.74	69.79	none	2	0.79	5.00	EGMA: has < does not
Communication	14	20.49	35.65	EGRA, EGMA: has < does not	5	20.28	51.21	none
Learning, Remembering and Concentrating	46	11.28	34.27	none	19	39.89	60.59	none
Accepting Change, Controlling Behaviour and Making Friends	39	23.34	48.42	none	23	35.71	66.88	none
Mental Health (Anxiety and Depression)	40	33.72	57.61	none	27	36.70	61.45	none
Subgroup								
Girls with at least 1 functional disability	123	29.75	52.77	EGRA, EGMA: has < does not	68	39.37	65.82	none
Accessibility—long distances to school	280	42.79	65.67	EGRA, EGMA: has < does not	9	33.62	61.48	none
Lack safety net for GBV	152	43.42	65.10	none	95	33.93	63.55	EGRA, EGMA: has < does not
Lack of right to an education	22	45.51	50.68	EGRA, EGMA: has < does not	7	32.26	59.76	None
Lack of enabling environment for quality education	45	45.51	65.64	None	0	0	0	n/a

	Treatment Group (C1A)				Comparison Group (C2)			
	n	Avg. EGRA score	Avg. EGMA score	Significant Difference	n	Avg. EGRA score	Avg. EGMA score	Significant Differences
Logistic barriers during menses	204	43.65	66.09	None	110	38.62	63.65	EGRA, EGMA: has < does not
Lack of voice and ability to speak up	81	39.92	59.61	EGMA: has < does not	50	31.13	60.61	EGRA, EGMA: has < does not

Note: Differences between girls who face the barrier and those who do not are statistically compared within the treatment cohort and within the comparison cohort. Significant differences at the p<0.05 level are indicated.

Intermediate Outcomes Analysis

To understand the relationships between different levels of the SAGE ToC, average literacy and numeracy scores are presented by IO indicator scores in Supplementary Table 10. IOs and the indices used to report against each of the IO are described in detail in section 7. IOs that were measured at the girl-level are used to disaggregate learning outcomes in this section. Furthermore, results for the treatment and comparison cohort are presented and statistically significant differences within each cohort are discussed by subgroup.

Overall, girls in both cohorts who had high levels of self-efficacy, more positive gender attitudes, and high levels of SRHR knowledge had higher literacy and numeracy scores than did girls who had low levels of these IOs, as measured by the indices for each IO. The perception of safety mattered among the treatment cohort, where girls who had low levels of perceived safety had lower EGMA aggregate scores but comparable EGRA scores. No differences in learning outcomes were observed among the comparison cohort by perceptions of safety index. Finally, there were no differences between girls with a high level of perceived community support for education and those with a low level of perceived community support, in either the treatment or comparison cohorts.

With the life-skills outcome, girls who had a high level of life-skills had higher literacy and numeracy scores than did girls with low levels. This was true for both the treatment and the comparison cohort.

Supplementary Table 10: Learning scores by intermediate outcomes, C1A and C2

	Treatment cohort				Comparison cohort			
	n	Average literacy score (aggregate)	Average numeracy score (aggregate)	significant differences	n	Average literacy score (aggregate)	Average numeracy score (aggregate)	significant differences
All girls	459	44.55	66.25		264	41.82	67.65	
Low Self-Efficacy (IO2.1)	60	33.27	52.45	EGRA, EGMA: low< high	38	28.77	51.98	EGRA, EGMA: low< high
High Self-Efficacy (IO2.1)	356	46.69	69.53		210	44.43	71.29	
Low gender attitudes (IO 2.2 gender)	403	44.03	66.74	EGRA, EGMA: low< high	244	41.37	67.85	EGRA, EGMA: low< high
High gender attitudes (IO 2.2 gender)	12	72.77	83.65		4	82.34	97.66	
Low SRHR (IO2.2 SRHR)	370	42.34	64.87	EGRA, EGMA: low< high	232	40.50	67.38	EGRA, EGMA: low< high
High SRHR (IO2.2 SRHR)	45	65.61	86.63		16	64.20	82.19	
Low community gender attitudes (IO 4.1)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a
High community gender attitudes (IO 4.1)	n/a	n/a	n/a		n/a	n/a	n/a	
Low perceived safety (IO4.2)	188	41.56	63.11	EGMA: low< high	238	42.28	68.51	no difference

	Treatment cohort				Comparison cohort			
	n	Average literacy score (aggregate)	Average numeracy score (aggregate)	significant differences	n	Average literacy score (aggregate)	Average numeracy score (aggregate)	significant differences
High perceived safety (IO4.2)	228	47.39	70.33		10	35.97	64.04	
Low community support for girls' education (IO4.3)	148	44.31	65.43	no difference	237	42.55	68.89	no difference
High community support for girls' education (IO4.3)	268	45.00	67.97		11	30.77	56.29	
Low life-skills (Outcome 1.3)	388	43.37	66.00	EGRA, EGMA: low< high	241	40.92	67.66	EGRA, EGMA: low< high
High life-skills (Outcome 1.3)	27	66.32	84.87		7	80.35	91.46	

Note: significant differences are indicated at the p<0.05 level. Differences in learning outcomes by community-level gender attitudes (IO 4.1) were not possible since the items comprising this index were administered to a sub-sample of boys, heads-of-households and parent/caregivers within each community and not associated with individual girls' learning outcomes. Low and high groups for each index were defined in collaboration with Plan and the FM—high scores are defined as at or above 75% of the score range, low scores are defined as below 75% of the score range. See Table 7 for descriptions of each IO and the construction of the indices.

Additionally, the relationships between the IOs and the learning outcomes were examined using the correlation between the index scores and learning outcome scores, shown in Supplementary Table 11. In contrast to the previous analysis—which uses high and low groups to explore the proportions of girls in each group and the differences in learning outcomes by these groups—an analysis by correlations tells how scores are directly related.[45] Results show that higher literacy

[45] Relationships between 2 scores are typically examined using the Pearson Correlation, with a range of 0 to 1, with 0 indicating no relationship between the 2 scores and 1 indicating perfect relationship between the 2 scores. Correlations that are large—above 0.7 or more—indicate that there is a strong relationship between the 2 variables, suggesting that a change in one score is likely to be

and numeracy scores were moderately associated with higher life-skills scores and higher SRHR knowledge scores and weakly associated with perceived safety score, self-efficacy and gender attitudes.

Supplementary Table 11: Pearson correlations between learning scores and IO scores

Learning Outcome	IO 2.1 Self-Efficacy Score	IO 2.2 Gender attitudes Score	IO 2.2 SRHR Score	IO 4.2 Perceived Safety Score	IO 4.3 Community Support for Education Score	Outcome 2. Life-skills Score
EGRA Aggregate Score	0.185**	0.362**	0.524**	0.152**	0.068	0.539**
EGMA Aggregate Score	0.274**	0.357**	0.557**	0.121**	0.046	0.576**

Note: 2 asterisks (**) indicates that the correlation is significant at the 0.01 level.

6.3 Transition outcome

SAGE's second outcome is a transition through key stages of education, training or employment. This section will present baseline findings that relate to the following indicators:

- O2.1a: Percentage of highly marginalised girls who have transitioned into formal/non-formal schooling[46]
- O2.1b: Percentage of highly marginalised girls who have transitioned into training (vocational training, life-skills training)
- O2.1c: Percentage of highly marginalised girls who have transitioned into fairly paid employment or self-employment

Project to complete

- Complete the table overleaf by outlining the transition pathways for your main intervention pathway groups.

associated with a change in the other variable. Correlations between 0.4 and 0.7 are considered moderate, and correlations below 0.4 are considered weak. Note that these relationships indicate associations, and not causation between the 2 scores. The relationships described below are for all girls in the treatment and comparison cohort together, since the trends in relationships observed above were similar in the 2 cohorts. At subsequent evaluation points, however, the variability in the correlations between treatment and comparison cohorts will be of interest to explore independently.

[46] This does not include continued participation in SAGE activities, but with formal/non-formal options outside of the SAGE activities.

Table 19: Transition pathways

Intervention pathway tracked for transition	Please describe the possible transition pathways for this group	Aim for girls' transition for next evaluation point	Aim for girls' transition level by the time project stops working with cohort
Girls age 10–14	Continue into formal schooling Enter vocational training5 Employment or self-employment	• Enrols into school • Enters third (optional year) of CBLH • Enters vocational training • Positive Employment or self-employment If above fails, girl completes at least 2 years of CBLH but does not transition into school, years 3 CBLH, training or employment	• Enrols into school or continues to be in school and progressing through the relevant grades • Enters vocational training (after the age of 15) • Positive Employment or self-employment If above fails, girl completes at least 2 years of CBLH but does not transition into school, years 3 CBLH, training or employment
Girls age 15–19	Continue into formal schooling Enter vocational training (after the age of 15) Employment or self-employment	• Enrols into school • Enters third (optional year) of CBLH • Enters vocational training • Employment or self-employment If above fails, girl completes at least 2 years of CBLH but does not transition into school, years 3 CBLH, training or employment	• Enrols into school or continues to be in school and progressing through the relevant grades • Enters vocational training • Employment or self-employment If above fails, girl completes at least 2 years of CBLH but does not transition into school, years 3 CBLH, training or employment

Pathway analysis

The pathway analysis is appropriate for girls enrolled in SAGE activities in C1A, but girls' responses regarding their intentions to transition suggest that formal schooling may be an under-utilized option. As stated in the ToC, girls are expected to participate in 2 years of CBLH—with an optional third year. They then have the opportunity to transition into formal schooling, vocational training or employment/self-employment. Based on girls' responses when asked their intentions at baseline, the majority of beneficiaries aim to transition into vocational training or employment/self-employment; few aim to transition into formal schooling.

At baseline, girls' transition pathways are estimated based on girls' intentions to transition. Girls were asked if they believe they will finish CBLH; 98.04% said yes, 0.49% said no, and 1.47% said they did not know. Girls who said they intend to finish CBLH were then asked about their hopes for themselves after CBLH. First, results by subgroup are presented for girls in the treatment cohort in Supplementary Table 5, followed by a comparison between the treatment cohort and the benchmark group in Supplementary Table 12.

Supplementary Table 12: Percentage of girls' hopes after completing CBLH, C1A

Group (transition)	N	Formal education	Vocational training	Employment or self-employment	Get married, other, don't know, Refused[47]
All girls	399	2.76%	47.62%	47.12%	2.51%
District					
Bulilima	48	8.51%	44.68%	44.68%	2.13%
Chimanimani	64	0.00%	71.11%	28.89%	0.00%
Epworth	51	0.00%	13.95%	86.05%	0.00%
Imbizo	24	0.00%	33.33%	52.38%	14.29%
Mutare Rural	130	4.39%	50.88%	40.35%	4.4%
Mutasa	112	1.98%	51.49%	45.54%	0.99%
Mutoko	30	0.00%	50.00%	50.00%	0.00%
Area					
Urban	3	0.00%	100%	0.00%	0.00%
Peri-urban	69	0.00%	27.54%	68.12%	4.35%
Rural	327	3.36%	51.38%	43.12%	2.14%
Language of instruction of assessment					
Shona	331	2.11%	48.94%	47.13%	1.81%

[47] The proportion of girls in the treatment cohort by sub-category are: 0.3% said 'get married and care for my family', 0.5% said 'other', and 1.8% said 'don't know'.

Group (transition)	N	Formal education	Vocational training	Employment or self-employment	Get married, other, don't know, Refused[47]
Ndebele	68	5.88%	41.18%	47.06%	5.88%
Age Group					
Age 10–14	101	6.93%	33.66%	55.45%	3.97%
Age 15–19	272	1.10%	54.04%	43.38%	1.47%
Barriers					
Girls with at least 1 functional disability	117	2.56%	37.61%	56.41%	3.41%
Accessibility—long distances to school	273	2.20%	50.55%	44.69%	2.56%
Lack safety net for GBV	147	2.04%	49.66%	45.58%	2.72%
Lack of right to an education	22	9.09%	45.45%	40.91%	4.55%
Lack of enabling environment for quality education	43	4.65%	37.21%	51.16%	6.98%
Logistic barriers during menses	196	4.08%	47.96%	43.37%	4.59%
Lack of voice and ability to speak up	79	1.27%	45.57%	48.10%	5.06%

Of the 399 girls in C1A who believed they would complete CBLH, almost half reported that they hoped to go to vocational training and the other half reported that they hoped to go into employment or self-employment. Only 2.76% reported that they would re-enter formal education and 2.51% reported that they would get married and take care of their family, other, do not know or refused to answer.

At least 4 districts had no girls reporting that they intend to re-enter formal schooling. In Chimanimani and Mutasa, girls reported that they want to go into vocational training while the majority of girls in Epworth and Imbizo reporting that they want to go to employment or self-employment. None of the 72 girls in urban and peri-urban areas reported that they want to re-enter formal schooling.

By age group, a larger proportion of girls in the 10–14 group (6.93%) reported that they hoped to return to primary school after completing CBLH; however, the majority still preferred vocational training (33.66%) or employment/self-employment (55.45%). A larger proportion of girls in older age groups reported that they hope to go to vocational training (54.04%) and only a few reported they wanted to return to formal schooling (1.10%).

Headline analysis

The post-CBLH hopes of girls in the treatment cohort are compared to the post-school hopes of girls in the benchmark group (Supplementary Table 13). The majority of girls in the benchmark group hope to continue in formal education as would be expected while the majority of girls in the treatment cohort hope to go into vocational training or employment. In the programme's logframe an estimated 70% of girls are expected to transition into formal or non-formal schooling, 30% into vocational training and 10% into self-employment or employment. It appears that these estimates are not aligned with the intentions expressed by girls in the treatment cohort in the baseline survey. However, the baseline surveys do not examine the intersection between the barriers with missing data—ethnicity, stated religion of household, school experience, including drop-out status, and carer status. Qualitative data from the Gender Analysis (Annex 12) along with further analysis once missing data are obtained can together provide a nuanced understanding of how the transition pathways supported under the project may need to be adjusted, if at all.

Furthermore, 7.93% or 32 girls in the treatment cohort reported at the time of the surveys that they were enrolled in formal school. SAGE followed-up with these girls and all 32 girls confirmed they are not enrolled in formal schools and are therefore eligible to participate in CBLHs. Data for girls who have never been to school or who have been to school but dropped out were not available to report in Table 20 due to the high rate of replacement in the sample—almost 80%. As of this writing, Plan does not have enrolment information from the replacement girls in order to report against these categories for the sample.

Supplementary Table 13: Percentage of girls' hopes after completing CBLH, C1A and Grades 3, 5 and 7

Age group (transition)	N	Formal education[48]	Vocational training	Employment or self-employment	Get married, other, don't know, Refused[49]
Treatment Cohort	399	2.76%	47.62%	47.12%	2.51%
Benchmark group—all Grades (3,5,7)	212	95.79%	0.00%	0.93%	3.27%
Grade 3	44	95.45%	0.00%	0.00%	4.55%
Grade 5	84	95.24%	0.00%	2.38%	2.38%
Grade 7	82	98.78%	0.00%	0.00%	1.22%

[48] Among the benchmarking group, students were also able to select 'go to an ALP'; only 1 student in grade 5/6 selected this option.
[49] The proportion of girls in the treatment cohort by sub-category are: 0.3% said 'get married and care for my family', 0.5% said 'other', and 1.8% said 'don't know'.

Age group (transition)	N	Formal education[48]	Vocational training	Employment or self-employment	Get married, other, don't know, Refused[49]

Note: girls in the Treatment Cohort were asked about their hopes after completing CBLH. Girls in the benchmark group, who are currently enrolled in formal schooling, were asked "thinking about next year (January 2020) what do you expect you will be doing?"

Table 20: Status at baseline, C1A and C2

Status	Treatment Cohort (%)	Comparison Cohort (%)
Never been to school (%)	**Not available**	**n/a**
Been to school, but dropped out	**Not available**	**n/a**
Not currently enrolled in formal school	92.07%	89.11%
Currently enrolled in formal school	0	10.89% (n=27)
Currently employed	**Not available**	**n/a**
Source: N =	**416**	**248**

Note: following this analysis, STS shared the list of girls in the intervention group who said they were currently enrolled in formal school with Plan for follow-up and confirmation of their enrolment. At the time of this writing, no further information was available to update these data.

6.4 Sustainability outcome

Baseline evidence on O3 Sustainability is presented in the following section for system, community and learning space indicators and primarily draws upon qualitative data.

System

The EE conducted KIIs with 5 government officials—4 district-level officials and one national-level official—whowork within SAGE intervention areas. These interviews focused on understanding current support for marginalised girls' education, support for NFE programmes and alignment of SAGE with government priorities. All interviewees had been familiarised with the programme through workshops held in Harare or in their district. All mentioned that they had been tasked with some form of oversight of programme activities. These oversight activities included supervising teaching and learning activities, selecting schools and teachers to support the programme, and planning activities that the programme will undertake. One official noted that running CBLHs out of formal schools increased integration. Another official was dissatisfied with the collaboration with SAGE to-date, noting that, although SAGE had agreed to design programme activities collaboratively, there had not been close coordination, and programme activities had been designed without MoPSE input.

All government officials who were interviewed stated that the MoPSE put strengthening support and programming for marginalised girls' education at a high priority at the national and district levels. One respondent stated that the government did not discriminate by gender, and that

government policies were targeted at improving education for all learners—not just girls. Officials also noted the high level of alignment between MoPSE priorities and policies and the SAGE programme. The MoPSE currently has mandates to reach zero illiteracy and to have NFE programmes in all primary and secondary schools, as well as to build vocational skills, which officials noted are scaffolded by SAGE programming. One official said that the MoPSE initiatives and SAGE are complementary programmes working together to reach the same goal.

Evident through interviews with government officials is that funding remains a significant barrier to the sustainability of the SAGE programme. Nearly all respondents mentioned that the government's NFE initiatives had been unsuccessful due to the lack of remuneration provided to the teachers. Specifically, although the MoPSE had passed a policy that all primary and secondary schools should run NFE programmes, the government had informed district officials that no funds would be made available to pay teachers who run the programmes. Additionally, one district official exhibited scepticism over the ability of a nongovernmental programme to be sustainable, as their funding is not long-term. A respondent mentioned that MoPSE generally refers donors to support education in marginalised areas, as there are not sufficient government funds for these areas.

When asked about ways to motivate MoPSE staff to more actively support marginalised girls' education, officials suggested that community engagement is critical. One official said that it is an obstacle if the community environment is not friendly to an initiative and that the MoPSE alone cannot do anything but instead needs the community to support a programme. One official specifically mentioned that village heads and *kraals* should be encouraged to assist. A common theme across responses was the need for educator buy-in; specifically, that teachers and communities should not view the programming for marginalised girls as extra work. One respondent said that the MoPSE would not be able to support a SAGE-type activity more than part-time. Other suggestions included holding forums to allow like-minded people to meet and create solutions, to conduct workshops with communities, and to develop success stories.

Given this evidence, the proposed system sustainability score at baseline using the sustainability scorecard is 2.00. There is evidence of system-wide support for marginalised learners' education, and the SAGE programme closely aligns with MoPSE priorities. SAGE has also engaged key district officials in planning and oversight of the programme. However, it is unclear if the MoPSE will have funding available to support and sustain SAGE activities after the end of the programme.

Community

Two community leaders responded to KIIs focused on exploring engagement with the SAGE programme, communities' relationship with the programme, possible positive and negative outcomes of the programme and potential for community-led sustainability. Both community leaders reported that SAGE had engaged with them at the development and enrolment stages. One respondent said they had been tasked with recruitment and engagement at the beginning of the programme, and the other said they had provided input into potential barriers to participation. When asked about their community's relationship with the programme to-date, one respondent said that the community was responding well and that most potential beneficiaries—including the respondent's daughter—had enrolled. The other respondent said that the relationship had been good so far but that community members had doubts about whether the programme would be successful and that there had been misinformation spread about the programme's activities, including that it would provide food and goods provisions. One noted that many men were initially

against the programme, as they believed that the police would become involved and punish them for cases of early marriage.

Three school heads within communities involved in the SAGE programme participated in KIIs. They acknowledged being engaged in the programme through community outreach meetings and recruitment activities, and one said that a teacher from his school served as a mentor for the programme.

Expected outcomes of the programme as explained by community leaders included employment creation for young women, reduction in adolescent pregnancy, increased literacy and reduction in child marriage. The community leaders also stated that they hoped that beneficiaries would finish the programme with basic life skills, go to tertiary and higher learning institutions and be able to make a living, which would, in turn, help the community grow in a positive direction. Another community leader said that there were no problems so far, but that the programmes should actively engage boys in the community so they do not feel left out. A school head said that the presence of SAGE might decrease interest in formal school.

One school head expressed that the programme should consider an ALP that is shorter than 2 years so that adolescent girls can finish their learning sessions and quickly earn a living. Another school head said that, although he prefers to focus on the positives of the programme, an increase in adolescent pregnancies might be possible because of the inclusion of boys in the programme. Further, findings from community leaders, parents and young men during the Gender Analysis indicated that some believed that women up to age 30 could benefit from the programme, as there is high demand and need to support those who missed out on their education due to early marriage and financial constraints.

Both community leader respondents said that the community could run CBLHs after the end of the programme under conditions. One said that the community would run the programme only when community members see positive results of the programme. The other respondent said that they would need materials and tutors to be able to run the programme. One school head emphasized that the programme needs to invest in good training of staff so that it can sustain itself and not be an end in itself. He said that it must be integrated into government structures so that there is sustainability. Another school head emphasized the need for support with resources —books, infrastructure, classrooms—in order for the programme to be sustained.

Given the evidence, the proposed community sustainability score at baseline is 1.00. Although there is evidence of programme engagement, there appears to be potential misalignment in programme goals and community expectations and understanding of the programme, which may hinder communities' appetite for sustainability after the programme's completion. Furthermore, there appear to be positive and potentially negative perceptions regarding the benefits of the interventions to girls, suggesting that shifting community perceptions and ensuring a clear understanding of the programme's goals are a pre-requisite for sustainability at the community level. Baseline evidence was limited, however, as it only drew from 3 school heads and 2 community leaders. STS recommends that a wider range of community actors are engaged in data collection at the next evaluation point to better understand the enabling environment for sustainability at the community level.

Learning space

Limited data were available to evaluate conditions for sustainability in the learning space at baseline. Instead, sustainability at the learning space will be highly dependent on the programme's ability to address potential barriers to attendance and learning that girls experienced prior to involvement with the SAGE programme. One of the barriers uncovered through the Gender Analysis,[50] which captured girls' pre-existing feelings about learning environments and were not specific to CBLH, included low quality of instruction. Specifically, girls expressed concerns that the SAGE programme may not incorporate child-friendly, inclusive or gender-responsive teaching and learning strategies. For example, when discussing barriers at the school, girls mentioned teachers being unfriendly, fearing teachers, being beaten, fearing being beaten by teachers, and teachers' manipulating girls into sexual relationships. Additional barriers detailed in the Gender Analysis included dirty learning facilities, bullying, stigma and discrimination of programme participants and long distances to learning centres and safety in transit. These barriers, though specific to girls' previous learning experiences, should be taken into consideration, as they previously limited girls' participation in education. Failure to mitigate these barriers could limit the programme's sustainability at the learning space.

An additional factor affecting sustainability may be the lack of resources available to support CBLH teachers. Government officials, community leaders and school heads all expressed concerns over remuneration to teachers in charge of running NFE programmes. Based on feedback provided from respondents, without appropriate incentives, CBLHs may face challenges in recruiting and maintaining quality educators after the end of the programme.

Given the evidence, the proposed learning space sustainability score at baseline is 0.00. There are significant existing barriers to sustainability that should be addressed for the quality of the learning space to be maintained after the end of the programme. Interviews with community educators and community members at the next evaluation point will provide a more nuanced understanding of sustainability conditions at the learning space.

Table 21: Sustainability indicators

	System	Community	Learning space
Indicator 1:	% of relevant MoPSE officials who support Girls' National Education Forum / other relevant initiatives **Results:** All interviewed MoPSE officials said that girls' education initiatives were high priority.	Average allocation of resources to the education of girls **Results:** NA	% of Community Educators who feel they are able to fulfil their roles **Results:** NA

[50] The Gender Analysis took place prior to the start of the programme, and the research focused on understanding girls' previous experiences with learning environments. Though not specifically applicable to CBLH, SAGE should consider the learning space findings of the Gender Analysis as it designs and rolls out the programme.

	System	Community	Learning space
Indicator 2:	# of district-level education officials that have participated in Hub monitoring visits (Midline) **Results:** NA	# of community leaders / Community Educators reporting that CBLHs will continue to function after project end (Endline) **Results:** Both community leaders expected CBLHs to be run by community after the end of SAGE, with conditions.	% of Hub Development Committees that are functional **Results:** NA
Indicator 3:	# of new initiatives taken by MoPSE officials aligned to Girls' National Education Forum joint advocacy goals (Endline) **Results:** NA		
Baseline Sustainability Score (0–4)	2.00	1.00	0.00
Overall Sustainability Score (0–4, average of the 3-level scores)	1.40 (weighted score)		
Note: Weighted score is based on weights assigned by Plan in the logframe as follows: 60% system, 20% community, 20% learning space			

Project to complete

Complete the table below by answering the questions in the table. Once completed, provide narrative analysis of the points raised in the table to explain the change the project intends to achieve. Ensure your analysis reflects the scores your external evaluator rated for each of your sustainability indicators.

Table 22: Changes needed for sustainability

Questions to answer	System	Community	CBLH	Family / /household	Girl
Change: what change should happen by the end of the implementation period	CBLHs are recognised and become part of the mainstream education system; SAGE materials approved for use by all schools offering NFE; budget allocated for CBLHs in other communities	Communities acknowledge CBLHs as a valuable education opportunity and develop a sense of ownership with relation to the hubs	Community Educators (CE) and learning assistance (LA) are driven by the desire to make a difference in their communities; school heads and NFE buddies see CEs as part of the education system and record data in EMIS; HDCs develop and adopt action plans to improve hubs	Families acknowledge CBLHs as a valuable education opportunity; recognise that girls have a right to be educated as much as boys	Girl learners believe that attaining fluency in literacy and numeracy is not academic but critical life skills components that will see them move from one step of their life to the next
Activities: What activities are aimed at this change?	Activities under Outcome 3 and Output 6 (logframe)	Activities under Outcomes 3, 4 and 5 (logframe)	Activities under Outcomes 1 and 2 (logframe)	Activities under Outcomes 4 and 5 (logframe)	Activities under Outcomes 1, 3 and 4 (logframe)
Stakeholders: Who are the relevant stakeholders?	MoPSE, incl. District School Inspectors and Lifelong Learning Coordinators	Community leaders, parents / husbands, employers, CoGE facilitators	Community Educators, Learning Assistants, school heads and NFE buddies, HDC members	Parents, caregivers, husbands	Highly marginalised adolescent girls
Factors: what factors are hindering or	Lack of funds for NFE at the national level	Community attitudes	Attitudes towards certain	Attitudes towards girls' education,	Attitudes towards education

Questions to answer	System	Community	CBLH	Family / /household	Girl
helping achieve changes? Think of people, systems, social norms etc.	due to the economic crisis affecting Zimbabwe	towards NFE, poverty	categories of girls, outdated teaching standards, lack of funds, attitudes towards NFE	conservative social norms and attitudes towards gender and SRHR, poverty	(NFE in particular), attitudes towards gender and SRHR, poverty

7. Key Intermediate outcome findings

Baseline results for the following SAGE IOs are presented in this section:

- IO1: OOS adolescent girls regularly attend accelerated learning sessions in CBLHs
- IO2: OOS adolescent girls have increased self-efficacy and life skills
- IO3: Adolescent girls and their families have improved skills and increased access to financial resources
- IO4: Communities demonstrate more positive gender attitudes and actively support and protect girls
- IO5: Strong and active partnerships with MoPSE officials and other civil society actors actively advocate for more inclusive, gender-responsive education policies

For each IO, qualitative findings from the Gender Analysis are also summarized where appropriate, in addition to any available quantitative findings from the baseline study. Additionally, tables show the results for the girls in the treatment and comparison groups by subgroup; results for girls in the treatment group comparing those facing barriers and not facing barriers are provided in the narrative below each table. Two results are presented at baseline—the mean score on an index and the proportion of girls categorized as having high scores, defined as at or above 75% of the score on an index. This cut-off at 75% of the index score was established based on Plan's guidance for high scores and is applied to all indices created for SAGE programme indicators.

7.1 Key Intermediate outcome findings

IO1: Attendance

SAGE's first IO is attendance. Specifically, the programme ToC assumes that improved attendance to sites of learning is a prerequisite for better learning, transition and sustainability for marginalised girls. At baseline, since programming had just begun at CBLHs, the baseline attendance level is set at zero and is based on CBLH attendance records.

IO1 indicators and relevant baseline information are detailed in Table 23: IO1 Attendance indicators, C1A

. Qualitative findings for this IO will provide critical feedback to the programme about how to support attendance over the years of the programme.

Table 23: IO1 Attendance indicators, C1A

IO	IO indicator	Sampling and measuring technique used	Who collected the data?	Baseline level	Target for next evaluation point	Will IO indicator be used for next evaluation point? (Y/N)
IO1: Attendance	IO1.1: % of girls regularly	Girls survey;	Select Research	Girls – 0%	Girls – 60%%	Yes

IO	IO indicator	Sampling and measuring technique used	Who collected the data?	Baseline level	Target for next evaluation point	Will IO indicator be used for next evaluation point? (Y/N)
	attending session in CBLHs	HoH survey; Caregiver survey				

Main qualitative findings

- The FGDs indicated there are several household level barriers to attendance including a lack of parental or family support for girls attending school, including a high chore burden making it difficult to attend school chores and a preference to send boys to school over girls.
- The FGDs also highlighted several community-level barriers to attending formal schools, including long distances to schools, issues of safety while girls are in transit and low-quality instruction in schools, including poor instructional techniques and corporal punishment. FGDs did not explore girls' perceptions of CBLHs, as CBLHs had not yet been established.
- The FGDs also highlighted some challenges related to the attendance of girls with disabilities. The discussions suggested that in some instances, girls with disabilities are not able to physically access schools.

Main findings

The qualitative and quantitative data collected at baseline provided substantial insight into girls' perceived barriers to attendance in the SAGE programme. Because these beneficiaries had recently enrolled in CBLHs at the time of baseline, the barriers identified are likely informed by their previous experiences with the formal school system. Barriers are grouped by individual-level and community-level and include, but are not limited to, the barriers described in **Error! Reference source not found.** of section 5.

While attendance records were not available for the baseline report, among the treatment cohort, 98.60% of girls say they are enrolled in CBLH and among the comparison cohort, 5.20% say they are currently enrolled in CBLH.[51, 52] It is unclear why girls in the comparison cohort would report that they are currently enrolled as selection for this cohort has yet to commence. In the event girls from the comparison group are in fact enrolled in CBLH during the first year, the comparison at midline should exclude these girls from the comparison cohort and reassign their baseline data to the treatment cohort.

[51] Of the remaining six girls in the intervention group who said they were not currently enrolled in CBLH, 5 were replacement girls. These girls did not provide any reasons as to why they were not enrolled in CBLH.
[52] Thirteen comparison girls reported being enrolled in CBLH. These girls were from Chinamano, Mutoko Central, Chitekwe, and Gandai A. Of these girls, 7 were replacements.

Other girls survey questions provide context around girls' perceptions of CBLHs and their learning environments. Of the 416 treatment respondents who reported that they are enrolled in CBLH, 97.75% agree or strongly agree that their CE makes them feel welcome in the classroom and 17.21% agree or strongly agree that their CE is often absent for class. The majority (99.04%) of girls in the treatment cohort said that CBLH is important for what they want to do when they grow up. Similarly, 98.80% of girls in the treatment cohort think that it is important for children to go to school and CBLH, and 98.31% think that they have a right to go to school and CBLH. However, when asked about their ability to choose whether they can attend or stay in education, 63.45% of girls agreed or strongly agreed that they could not choose.

The qualitative data from the Gender Analysis suggests that girls perceive a variety of barriers to their attendance in school. Because the Gender Analysis took place prior to the beginning of CBLH learning, these perceptions are likely based on their experience in formal schools and can provide formative information to the SAGE treatment. On the girls survey, the most commonly cited reason for not enrolling in formal school was an inability to afford school fees, with 91.12% of girls in the treatment cohort stating this was the reason they were not enrolled. Other reasons for not enrolling included being married or about to be married (6.79%); having or expecting a child (4.96%); disability (1.57%); and chronic illness (0.78%). Similar to the girls survey, girls in the Gender Analysis reported challenges to attendance when there is a lack of family support for education. They cited household chores and responsibilities, as well as prioritizing boys' education as reasons for their lack of attendance in school. Both the girls survey and the Gender Analysis also indicated that low-quality education opportunities influence whether a girl attends school. In the Gender Analysis, girls' responses suggested that poor instruction and corporal punishment in school are reasons for not attending. Finally, distance to schools, as well as a girl's sense of safety travelling to and from school are potential barriers to school attendance.

Reflections

Because the baseline took place at the start of CBLH learning sessions treatment, data collection for IO1 focused primarily on identifying potential obstacles or barriers to access to, attendance at and completion of CBLH through quantitative and qualitative methods. The findings under IO1 at baseline should be used to provide formative feedback to the programme in order to reduce or eliminate obstacles and barriers that learners may confront in the coming years of the programme.

At baseline, perceptions of barriers to attendance from quantitative and qualitative data were used to report against this IO. At the next evaluation point, several new tools or items will be introduced to track indicators under IO1. The quantitative tools or items that STS suggests adding include

- For IO1.1:
    - Classroom observations including headcount attendance at evaluation points and programme monitoring data
    - Classroom attendance records, if available
    - Additional girls survey items regarding frequency of attendance at CBLHs
    - Additional household survey items regarding frequency of attendance of girls at CBLHs

Between evaluation periods, SAGE should ensure that attendance records from CBLHs are consistently tracked and collated. The monitoring data on attendance should be combined with data collected at the next evaluation point to report on trends in attendance across the sample during the intervening months.

Targets

SAGE established IO1 targets based on specific contextual knowledge. Baseline results did not provide any data to contradict current assumptions of possible attendance rates at the next evaluation period.

Project Checks on Intermediate Outcomes

Ensure that the IO analysis reflects the links between different levels in the logframe and informs the validity of the Theory of Change. This includes checking whether the EE (?) have:

- Measured and analysed all IO indicators presented in logframe.
- Disaggregated the data according to the logframe.
- Used both the qualitative and quantitative analysis stated in the logframe.
- Related the IO analysis to the analysis of Outcomes.

See the management response section in Annex 18.

IO2: Adolescent girls have increased self-efficacy and life skills

SAGE's second IO is that adolescent girls have increased self-efficacy and life skills. Specifically, the programme ToC assumes that improved knowledge and understanding of self-efficacy and life skills are prerequisites for better learning, transition and sustainability outcomes for marginalised girls. The indicators in this section measure and report on girls' self-efficacy as well as SRHR and gender KAPs, while results for life skills are reported separately in Section 7 as it has been requested by the FM and Plan to include this as an additional outcome for SAGE. In the analysis for this outcome, STS first compared the overall mean scores of the girls in the treatment and comparison cohorts to evaluate if there are statistically significant differences. Next, STS compared treatment girls' mean scores by subgroup —for example, girls facing barriers around menstruation compared to those not facing such barriers—in order to better understand what factors affect girls' scores.

IO2 indicators and relevant baseline information are detailed in Table 24: IO2 Self-efficacy and life skills indicators

. Item-level frequencies are available in Annex 15. Two results are presented at baseline—the mean score on an index, and the proportion of girls categorized as having high scores, defined as at or above 75% of the score on an index.[53] This cut-off at 75% of the index score was based on Plan's guidance for high scores and is applied to all indices created for SAGE programme indicators. At midline, the number of girls with improved mean scores over baseline should be reported, as well as the proportion of girls in the high score category.

[53] The index for self-efficacy, IO 2.1, included a scoring range from 0-3. For IO2.2, the gender attitudes index used a scoring range from 0-2, and the SRHR index used a scoring range from 0-30.

Table 24: IO2 Self-efficacy and life skills indicators

IO	IO indicator	Sampling and measuring technique used	Who collected the data?	Index name	Baseline level for Treatment Cohort	Target for next evaluation point	Will IO indicator be used for next evaluation point? (Y/N)
IO2: OOS adolescent girls have increased self-efficacy and life skills	IO2.1: % of marginalised girls demonstrating improved self-efficacy	Girls survey	STS	Self-Efficacy	Mean score – 2.67 on an index; 85.58% of girls considered having a high score at baseline	90% of girls have a high score at midline	Yes
IO2: OOS adolescent girls have increased self-efficacy and life skills	IO2.2: % of marginalised girls demonstrating improved knowledge, attitudes and practices on gender and SRHR	Girls survey	STS	Gender Attitudes	Mean score – 1.05 on index; 2.89% of girls considered having a high score at baseline	20% above baseline	Yes
	IO2.2: % of marginalised girls demonstrating improved knowledge, attitudes and practices on gender and SRHR	Girls survey	STS	SRHR Knowledge	Mean score – 14.91 on index; 10.84% of girls considered having a high score at baseline	20% above baseline	Yes
	Main qualitative findings						

IO	IO indicator	Sampling and measuring technique used	Who collected the data?	Index name	Baseline level for Treatment Cohort	Target for next evaluation point	Will IO indicator be used for next evaluation point? (Y/N)
	• FGDs suggest that low self-esteem and lack of opportunities to learn about SRHR are barriers girls' self-esteem and life skills.						

Main findings

IO2.1 % of marginalised girls demonstrating improved self-efficacy

At baseline, girls responded to items on a self-efficacy instrument related to overcoming challenges, achieving goals, perceptions of personal capabilities, and perceptions of individual performance on tasks.54 The self-efficacy index for IO2.1 contained a set of 6 items, with response options scaled from 0–3.Girls in both treatment and comparison groups had high self-efficacy scores. In the treatment group, girls have a mean self-efficacy score of 2.67, while girls in the comparison group have a mean self-efficacy score of 2.63.

Because the indicator for IO2.1 measures improvement in scores, impossible to show on the baseline, STS also grouped girls into those with high scores and low scores to better understand where improvement might happen. The cut off for high scores were provided by Plan at 75% of the scale or a score of 2.25. Girls who score 2.25 or higher are categorized as high scores, and girls who score less than 2.25 are categorized as low scores. The proportions of girls in high and low score categories by subgroups and barriers are presented in Supplementary Table 14, which provides the mean scores on the self-efficacy index, as well as the percentage of girls with high and low scores. Overall, the majority of both treatment and comparison girls fall into the high score category. As a result, it appears that high sense of self-efficacy, overall, is a characteristic of the population of intended beneficiaries.

Supplementary Table 14: IO2.1 Self-efficacy results by subgroup and barrier

Characteristic	Treatment vs Comparison	Number of respondents in subgroup	Mean Score	Category	Proportion of total in subgroup
All girls	Treatment	416	2.67	Low score	14.42%
				High score	85.58%
	Comparison	248	2.63	Low score	15.32%
				High score	84.68%

[54] The self-efficacy items were adapted from Chen, G., Gully, S.M. and Eden, D. (2001) 'Validation of a New General Self-Efficacy Scale', *Organizational Research Methods*, 4 (1): 62-83.

Characteristic	Treatment vs Comparison	Number of respondents in subgroup	Mean Score	Category	Proportion of total in subgroup
Barrier: Accessibility/Long distance to School	Treatment	280	2.69	Low score	14.29%
				High score	85.71%
	Comparison	9	2.59	Low score	11.11%
				High score	88.89%
Barrier: Lack of safety net for GBV	Treatment	152	2.73	Low score	13.16%
				High score	86.84%
	Comparison	95	2.54	Low score	16.84%
				High score	83.16%
Barrier: Perceived lack of right to education	Treatment	22	2.61	Low score	9.09%
				High score	90.91%
	Comparison	7	2.69	Low score	14.29%
				High score	85.71%
Barrier: Lack of enabling environment for quality education*	Treatment	45	2.55	Low score	20.00%
				High score	80.00%
	Comparison	0	n/a	Low score	0.00%
				High score	0.00%
Barrier: More barriers due to menstruation*	Treatment	204	2.63	Low score	15.20%
				High score	84.80%
	Comparison	110	2.58	Low score	19.09%
				High score	80.91%
	Treatment	81	2.54	Low score	20.99%

Characteristic	Treatment vs Comparison	Number of respondents in subgroup	Mean Score	Category	Proportion of total in subgroup
Barrier: Lack of voice/ability to speak up***				High score	79.01%
	Comparison	50	2.38	Low score	28.00%
				High score	72.00%
Barrier: Girl has functional difficulty*	Treatment	123	2.57	Low score	19.51%
				High score	80.49%
	Comparison	68	2.54	Low score	22.06%
				High score	77.94%
Girls 10–14 years old	Treatment	107	2.65	Low score	15.89%
				High score	84.11%
	Comparison	95	2.62	Low score	17.89%
				High score	82.11%
Girls 15–19 years old	Treatment	279	2.69	Low score	13.98%
				High score	86.02%
	Comparison	141	2.68	Low score	12.06%
				High score	87.94%
Bulawayo**	Treatment	24	2.35	Low score	29.17%
				High score	70.83%
	Comparison	19	2.57	Low score	21.05%
				High score	78.95%
Harare**	Treatment	48	2.76	Low score	14.58%
				High score	85.42%

Characteristic	Treatment vs Comparison	Number of respondents in subgroup	Mean Score	Category	Proportion of total in subgroup
	Comparison	25	2.66	Low score	16.00%
				High score	84.00%
Manicaland**	Treatment	269	2.65	Low score	15.61%
				High score	84.39%
	Comparison	169	2.62	Low score	14.79%
				High score	85.21%
Mashonaland East**	Treatment	28	2.81	Low score	7.14%
				High score	92.86%
	Comparison	34	2.67	Low score	14.71%
				High score	85.29%
Matabeleland South**	Treatment	47	2.80	Low score	4.26%
				High score	95.74%
	Comparison	1	2.50	Low score	0.00%
				High score	100.00%

Note: One asterisk (*) indicates a difference in treatment girls' mean scores between those in a subgroup versus those not in a subgroup at p<0.05. Three asterisks indicate a difference in treatment girls' mean scores between those in a subgroup versus those not in a subgroup at p<0.001.

These high scores seem to contradict findings from the FGDs and Gender Analysis, in which OOS and highly marginalized girls, especially young mothers and girls with disabilities who face high levels of stigma and discrimination within learning environments from peers and/or teachers — frequently cited low self-esteem as a barrier to improving self-efficacy and life skills. Therefore, to better understand what factors may affect self-efficacy scores for girls in the treatment group, statistical tests were used to evaluate if the differences in mean scores between subgroups within the treatment group at baseline are significant (Supplementary Table 14).[55] Asterisks in

[55] A t-test is a type of inferential statistic used to determine if there is a significant difference between the means of 2 groups (for example, treatment and comparison groups). T-tests were used to test the differences in means between girls categorized as facing a barrier compared to those not facing the barrier. ANOVA is an inferential statistic that determines if there is a significant difference between the means of 3 or more groups, and was used to test differences in means for girls in different age groups and from different regions.

Supplementary Table 14 indicate that for girls in the treatment group, the difference in mean scores between those who face barriers and those who do not is statistically significant. Significant differences were observed for girls who lack an enabling environment for quality education; girls with barriers around menstruation; girls who lack voice or an ability to speak up; and girls with at least one functional difficulty. Girls lacking an enabling environment for quality education have an average self-efficacy score of 2.55 compared to those who do not face this barrier (2.70). Girls with barriers around menstruation have an average self-efficacy score of 2.63 compared to those who do not face this barrier who average 2.74. Girls lacking voice or an ability to speak up have a significantly lower average self-efficacy score (2.54) compared to those who do not have low self-esteem (2.73). Additionally, girls with functional difficulty have a significantly lower average self-efficacy score (2.57) compared to those who do not have functional difficulty (2.72). The data also shows significant differences between the average scores of girls in the treatment cohort from different regions, with girls from Bulawayo's mean score of 2.35 being statistically lower than girls from all other regions. Girls from Harare have a mean score of 2.76; girls from Manicaland have a mean score of 2.65, girls from Mashonaland East have a mean score of 2.81; girls from Matabeleland South have a mean score of 2.80.

IO2.2 % of marginalised girls demonstrating improved knowledge, attitudes and practices on gender and SRHR

Results for IO2.2 are reported as scores on 2 indices. Girls were asked 17 items to assess their gender KAP, addressing themes such as women in the workplace, girls' education and gender roles in the home.[56] Additionally, girls responded to questions about sexual and reproductive health topics such as sexually transmitted diseases, examples of sexual and reproductive health rights, and practices around SRHR topics. Because this indicator measures 2 different KAP categories, 2 separate indices were created for this indicator. Items were reviewed and revised by SAGE to ensure alignment with the curriculum they will deliver over the life of the programme.

Gender index

The gender KAP measure is based on the Gender Equitable Men Scale[57] and supplemented with customized survey items (see Supplementary Table 15 for a list of items).[58] Girls are scored on a scale from 0.00 to 2.00. Those scoring above a 1.50, the 75% mark on the scale, are considered to have a high score on the gender items. A very small percentage of girls—2.89%—had high scores on the gender KAP index. Overall, girls in the treatment group have a statistically significantly higher mean gender KAP score than girls in the comparison group. Treatment girls score 1.05 on average, while comparison girls score 0.99 on average. A comparison of treatment and comparison girls by subgroup is presented in Supplementary Table 15. Furthermore, the scale was subdivided by thematic groups and mean scores are presented in Supplementary Table 16. On a scale of 0–2, girls had the highest scores—indicating positive perceptions—on aspirations (1.73) followed by gender stereotypes (1.48) and violence and blame (1.17) and the lowest scores—indicating negative perceptions—on masculinity (0.56), and sexual relationships (0.69). When compared to girls in the comparison cohort, girls in the treatment cohort had higher mean scores on gender stereotypes, masculinity and domestic roles.

[56] all negatively worded items were recoded in reverse, resulting in a scale with higher scores indicating more positive perceptions.

[57] Nanda, Geeta. "Compendium of Gender Scales." Compendium of Gender Scales, September 2011. https://www.c-changeprogram.org/content/gender-scales-compendium/pdfs/4. GEM Scale, Gender Scales Compendium.pdf.

Supplementary Table 15: IO2.2 Gender attitudes and practices

Characteristic	Treatment vs Comparison	Number of respondents in subgroup	Mean Score	Category	Proportion of total in subgroup
All girls^	Treatment	415	1.05	Low score	97.11%
				High score	2.89%
	Comparison	248	0.99	Low score	98.39%
				High score	1.61%
Barrier: Accessibility/Long distance to School	Treatment	280	1.06	Low score	96.79%
				High score	3.21%
	Comparison	9	1.09	Low score	100.00%
				High score	0.00%
Barrier: Lack of safety net for GBV	Treatment	152	1.05	Low score	98.03%
				High score	1.97%
	Comparison	95	0.93	Low score	98.95%
				High score	1.05%
Barrier: Perceived lack of right to education*	Treatment	22	0.94	Low score	100.00%
				High score	0.00%
	Comparison	7	0.87	Low score	100.00%
				High score	0.00%
Barrier: Lack of enabling environment for quality education*	Treatment	45	1.13	Low score	93.33%
				High score	6.67%
	Comparison	0	n/a	Low score	0.00%
				High score	0.00%
Barrier: More barriers due to menstruation *	Treatment	203	1.03	Low score	97.04%
				High score	2.96%
	Comparison	110	0.98	Low score	100.00%
				High score	0.00%
Barrier: Lack of voice/ability to speak up	Treatment	81	1.04	Low score	100.00%
				High score	0.00%
	Comparison	50	0.89	Low score	100.00%
				High score	0.00%
Barrier: Girl has functional difficulty*	Treatment	123	1.00	Low score	96.75%
				High score	3.25%
	Comparison	68	1.00	Low score	95.59%
				High score	4.41%
Girls 10–14 years old***	Treatment	107	0.96	Low score	98.13%
				High score	1.87%
	Comparison	95	0.94	Low score	100.00%
				High score	0.00%

Characteristic	Treatment vs Comparison	Number of respondents in subgroup	Mean Score	Category	Proportion of total in subgroup
Girls 15–19 years old***	Treatment	278	1.09	Low score	97.84%
				High score	2.16%
	Comparison	141	1.02	Low score	97.16%
				High score	2.84%
Bulawayo	Treatment	24	0.98	Low score	87.50%
				High score	12.50%
	Comparison	19	0.99	Low score	100.00%
				High score	0.00%
Harare	Treatment	48	1.05	Low score	95.83%
				High score	4.17%
	Comparison	25	1.10	Low score	92.00%
				High score	8.00%
Manicaland	Treatment	268	1.06	Low score	98.51%
				High score	1.49%
	Comparison	169	0.96	Low score	100.00%
				High score	0.00%
Mashonaland East	Treatment	28	1.04	Low score	100.00%
				High score	0.00%
	Comparison	34	1.03	Low score	94.12%
				High score	5.88%
Matabeleland South	Treatment	47	1.05	Low score	93.62%
				High score	6.38%
	Comparison	1	0.94	Low score	100.00%
				High score	0.00%

Note: One asterisk (*) indicates a difference in treatment girls' mean scores between those in a subgroup versus those not in a subgroup at p<0.05. Three asterisks indicate a difference in treatment girls' mean scores between those in a subgroup versus those not in a subgroup at p<0.001. One caret (^) indicates a difference between treatment and comparison girls at p<0.001.

Supplementary Table 16: IO2.2 Gender attitudes and practices by thematic group

Thematic Group	Items Included	Mean Score Treatment Cohort	Mean Score Comparison Cohort
Gender Stereotypes*	Boys are naturally better than girls in maths and sciences. Girls and women can be good leaders.	1.48	1.37
Masculinity*	Boys lose respect if they cry or talk about their problems.	.56	.43

Thematic Group	Items Included	Mean Score Treatment Cohort	Mean Score Comparison Cohort
	If someone insults a man, he should defend his reputation with force if he has to. A man should have the final word about decisions in his home.		
Aspirations	Higher education is just as important for girls as it is for boys. Young women should have the same opportunities to work outside the home as young men.	1.73	1.68
Domestic Roles*	Men and women should share equal responsibility for household chores and childcare. A woman should obey her husband in all things.	.93	.82
Sexual Relationship	If a girl says no to sex, her partner should respect that. It is a girl's responsibility to avoid getting pregnant. A girl should be able to decide who and when she marries.	.69	.70
Violence and Blame	It is not a girl's fault if a male student or teacher sexually harasses her, it is the fault of the male involved. Girls wearing short dresses provoke boys. A girl or woman never deserves to be beaten. A woman should not tolerate violence to keep her family together. A man using violence against his wife or girlfriend is a private matter that shouldn't be discussed outside the couple.	1.17	1.12

Note: all negatively worded items were recoded in reverse, resulting in a scale with higher scores indicating more positive perceptions. One asterisk (*) indicates a difference in treatment girls' mean scores and comparison girls' mean scores at $p<0.05$.

The scores of girls in the treatment cohort differ significantly on the gender KAP by certain subgroups (Supplementary Table 15). Girls who perceive the lack of a right to education have a significantly lower average gender norms score (0.94) compared to those who do perceive a right

to an education (1.06). The data also suggests that girls lacking an enabling environment for quality education have significantly higher gender scores (1.13) than those who do not face this barrier (1.05). Girls who face barriers with menstruation also demonstrate significantly lower gender norms scores (1.03) than girls who do not experience that barrier (1.10). Girls with one or more functional difficulties have a mean score of 1.00, significantly lower than the mean score of 1.07 for girls who do not have functional difficulties. Finally, there are statistically significant differences in the mean scores of girls by age group. Girls aged 10–14 have a mean score of 0.96; girls 15 and older have a mean score of 1.09.[59]

SRHR Index

The SRHR KAP index is made up of 30 items (see Supplementary Table 17 for list of items) and scored on a scale from 0–30. Girls scoring above 22.50, the 75% mark on the scale, are considered to have a high score. Around one-tenth of treatment girls—10.84%—have high scores on the SRHR KAP index. Overall, girls in the treatment cohort again show statistically significantly higher mean scores than comparison girls on the SRHR index. Treatment girls score 14.91 on average on the index, while comparison girls score 13.81. Differences in treatment and comparison girls' mean scores and proportions with high scores are outlined in Supplementary Table 17.

Supplementary Table 17: IO2.2 SRHR knowledge attitudes and practices

Characteristic	Treatment vs Comparison	Number of respondents in subgroup	Mean Score	Category	Proportion of total in subgroup
All girls^	Treatment	415	14.91	Low score	89.16%
				High score	10.84%
	Comparison	248	13.81	Low score	93.55%
				High score	6.45%
Barrier: Accessibility/Long distance to School	Treatment	280	14.81	Low score	89.29%
				High score	10.71%
	Comparison	9	16.44	Low score	88.89%
				High score	11.11%
Barrier: Lack of safety net for GBV	Treatment	152	14.99	Low score	89.47%
				High score	10.53%

[59] Because of the high percentage of replacement girls surveyed at baseline, responses were collected from girls outside of SAGE eligibility criteria—namely girls aged 7-9 and 20+. The results for these girls are presented in Annex 11.

Characteristic	Treatment vs Comparison	Number of respondents in subgroup	Mean Score	Category	Proportion of total in subgroup
	Comparison	95	12.34	Low score	95.79%
				High score	4.21%
Barrier: Perceived lack of right to education	Treatment	22	13.23	Low score	90.91%
				High score	9.09%
	Comparison	7	12.00	Low score	85.71%
				High score	14.29%
Barrier: Lack of enabling environment for quality education	Treatment	45	15.10	Low score	91.11%
				High score	8.89%
	Comparison	0	n/a	Low score	0.00%
				High score	0.00%
Barrier: More barriers due to menstruation***	Treatment	203	14.33	Low score	90.64%
				High score	9.36%
	Comparison	110	12.10	Low score	96.36%
				High score	3.64%
Barrier: Lack of voice/ability to speak up*	Treatment	81	13.70	Low score	92.59%
				High score	7.41%
	Comparison	50	12.14	Low score	100.00%
				High score	0.00%
Barrier: Girl has functional difficulty	Treatment	123	12.78	Low score	95.12%
				High score	4.88%
	Comparison	68	13.37	Low score	97.06%

Characteristic	Treatment vs Comparison	Number of respondents in subgroup	Mean Score	Category	Proportion of total in subgroup
				High score	2.94%
Girls 10–14 years old***	Treatment	107	9.57	Low score	99.07%
				High score	0.93%
	Comparison	95	10.39	Low score	100.00%
				High score	0.00%
Girls 15–19 years old***	Treatment	278	17.06	Low score	84.89%
				High score	15.11%
	Comparison	141	16.49	Low score	90.07%
				High score	9.93%
Bulawayo	Treatment	24	14.25	Low score	87.50%
				High score	12.50%
	Comparison	19	12.58	Low score	89.47%
				High score	10.53%
Harare***	Treatment	48	10.75	Low score	100.00%
				High score	0.00%
	Comparison	25	14.44	Low score	100.00%
				High score	0.00%
Manicaland	Treatment	268	15.44	Low score	88.43%
				High score	11.57%
	Comparison	169	13.31	Low score	94.67%
				High score	5.33%

Characteristic	Treatment vs Comparison	Number of respondents in subgroup	Mean Score	Category	Proportion of total in subgroup
Mashonaland East	Treatment	28	14.88	Low score	82.14%
				High score	17.86%
	Comparison	34	16.25	Low score	88.24%
				High score	11.76%
Matabeleland South***	Treatment	47	16.48	Low score	87.23%
				High score	12.77%
	Comparison	1	23.00	Low score	0.00%
				High score	100.00%

Note: One asterisk (*) indicates a difference in treatment girls' mean scores between those in a subgroup versus those not in a subgroup at p<0.05. Three asterisks indicate a difference in treatment girls' mean scores between those in a subgroup versus those not in a subgroup at p<0.001. One caret (^) indicates a difference between treatment and comparison girls at p<0.05.

STS further analysed differences in treatment girls' scores comparing girls in barrier subgroups with those not in barrier subgroups. Girls facing barriers around menstruation have a significantly lower score (14.33) compared to girls who do not face these barriers (17.39). Girls with low self-esteem have significantly lower scores (13.70) compared to girls who did not (15.37). There are also significant differences in scores on the SRHR index between girls who had a functional difficulty (12.78) and those who do not (15.81). As might be expected, there are statistically significant differences in the average SRHR scores for girls of different age groups as well, with younger girls having a lower average score than older girls on both indices. Region is a factor in treatment girls' SRHR scores, with significant differences between treatment girls from Harare (10.75) and Matabeleland South (16.48) (Supplementary Table 17).

Reflections

The high mean score of 2.67 and very high percentage of girls with high scores in indicator 2.1—self-efficacy—indicate that girls in both cohorts 1 and 2 have a high sense of self-assurance and the belief that they can take action and bring about change. This is a strong asset that the programme should leverage. However, these scores mean that there may not be room for growth on this measurement scale. In future evaluation points, STS recommends supplementing the baseline scale, which mostly uses questions about general beliefs of self, with questions about practical scenarios to capture a more nuanced perspective of girls' self-efficacy. This will require enhanced qualitative tools that ask girls to respond to situations and demonstrate self-efficacy rather than their agreement with concepts around self-efficacy.

In contrast to the high percentage of girls with self-efficacy scores in IO2.1, much lower percentages of girls had high scores on the gender KAP and SRHR KAP indices for IO2.2. Both

indices for indicator 2.2 had relatively high-reliability coefficients (0.63 for the gender KAP; 0.82 for the SRHR KAP), thus are proving to be effective scales measuring the desired domains or KAPs. The low percentage of girls with high scores may be a result of the types of questions asked. For example, the SRHR KAP asks girls to indicate knowledge about circumstances under which a girl can get pregnant. This may be the knowledge that younger girls have not yet learned or that girls are not yet comfortable responding to. As the Gender Analysis points out, only 1% of girls aged 10–14 years old are sexually active, but the overwhelming majority (99.1%) lack comprehensive knowledge around pregnancy.[60] Furthermore, girls age 15–19 are more likely to have had their first sexual experience (25%), and a very large majority (93%) still lack comprehensive knowledge around pregnancy.[61] Thus, the low percentages of girls with high scores represent an opportunity for growth in these areas that the SAGE curriculum will cover.

Furthermore, since the addition of life-skills as the third outcome for SAGE, this IO should be revised to exclude life-skills and instead be reworded as follows: 'adolescent girls have increased self-efficacy, gender and SRHR knowledge, attitudes and practices'.

Targets

The programme logframe sets the target for the next evaluation point at a 20% increase from baseline levels. This would mean that 100% of girls receive a high self-efficacy score; 22.14% of girls receive a high gender score, and 29.15% of girls receive a high SRHR score. However, because a 20% increase for the self-efficacy indicator would mean more than 100% of girls are categorized as having high self-efficacy, STS suggests setting the target to 90% of girls categorized as having high scores.

These targets, set at changing social norms and attitudes in addition to knowledge and practices, are quite ambitious. It should be noted that frequently with this type of change, self-reported rates of knowledge or self-efficacy may at first stagnate or even decrease as learners' frames of reference expand and change with new knowledge. Should such a pattern emerge, SAGE might consider including retrospective pre-test type questions on the endline to capture responses aligned to girls' changed frames of reference.[62]

Project Checks on Intermediate Outcomes

Ensure that the IO analysis reflects the links between different levels in the logframe and informs the validity of the Theory of Change. This includes checking whether the EE have:

- Measured and analysed all IO indicators presented in logframe.
- Disaggregated the data according to the logframe.
- Used both the qualitative and quantitative analysis stated in the logframe.
- Related the IO analysis to the analysis of Outcomes.

See management response in Annex 18.

[60] Republic of Zimbabwe Ministry of Health and Child Care. *Zimbabwe National Adolescent Fertility Study*. [SAGE Proposal Language]
[61] Ibid.
[62] https://archive.globalfrp.org/evaluation/the-evaluation-exchange/issue-archive/evaluation-methodology/the-retrospective-pretest-an-imperfect-but-useful-tool

IO3: Adolescent girls and their families have improved skills and increased access to financial resources

SAGE's third IO is improved skills and increased access to financial resources. Specifically, the programme ToC assumes improved skills and increased access to financial resources are a prerequisite for better learning, transition and sustainability outcomes for marginalised girls.

Since formal financial instruction had not yet begun at CBLH centres at baseline, the focus of data collection for this indicator was to identify current access to Village Savings and Loan Associations (VSLAs) and financial resources. SAGE should use this information to help inform the structure and focus of treatments for CBLH facilitators.

Table 25: IO3 Improved skills and increased access to financial resources

IO	IO indicator	Sampling and measuring technique used	Who collected the data?	Baseline level	Target for next evaluation point	Will IO indicator be used for next evaluation point? (Y/N)
IO3: Adolescent girls and their families have improved skills and increased access to financial resources	IO3.1: % of highly marginalised girls aged 15–19 who have accessed a VSLA reporting improved access to financial resources	Girls survey, HoH survey, Caregiver survey	STS	0	N/A	Yes[63]
IO3: Adolescent girls and their families have improved skills and increased access to financial resources	IO3.2: % of adult female VSLA participants reporting increased capacity to invest in education	NA at baseline	NA at baseline	0	TBC	Yes

[63] The VSLA groups will be established by the next evaluation point. These groups will support mothers of girls participating in SAGE programmes.

IO	IO indicator	Sampling and measuring technique used	Who collected the data?	Baseline level	Target for next evaluation point	Will IO indicator be used for next evaluation point? (Y/N)
IO3: Adolescent girls and their families have improved skills and increased access to financial resources	IO3.3: Confidence in vocational skills score of highly marginalised girls aged 15–19 who participated in ISOP training	N/A at baseline	N/A at baseline	0	TBC	TBC

Main qualitative findings

- At the individual level, the FGDs suggest girls have limited access to vocational and livelihood skills training. Girls responses indicated that, particularly in rural areas, where girls reported that these types of centres are more often located in towns and cities and are difficult to access from rural areas. Additionally, girls reported that a lack of resources, such as the ability to pay school fees or lack of access to capital for businesses are also barriers to financial skills and resources.

Main findings

Because VSLA sessions had not yet begun, quantitative baseline data for indicators IO3.1, IO3.2 and IO3.3 were not collected. However, some relevant information was gathered from the HoH survey, providing contextual information about their access to financial services. Of the 100 respondents, only 13.00% report having access to financial services. These 13.00% were from Harare, Manicaland, and Matabeleland South. The services they list having access to include

- Village savings and loans (53.85%)
- Savings groups (23.08%)
- Banks and financial services cooperatives (15.38%)
- Other services (30.77%)

Beyond this, the Gender Analysis provides insight into the availability of livelihood trainings and limited resources to improve financial resources. For example, young mothers in Mutare Rural noted that there are no training centres near their communities since all the centres are in town. Moreover, although they would like to enrol in the available training centres in Mutare Rural, they cannot afford the bus fare for travel. Similarly, in the FGD on girls with disabilities in Imbizo, it was mentioned that the girls' parents lack a source of income and, therefore, cannot afford to pay their daughters' school fees. Similar sentiments around parents' lack of access to sources of income and employment opportunities were expressed in FGDs in Mutare Rural.

Reflections

Because learning sessions, as well as VSLAs and ISOP training, had not yet started at the time of data collection, data collection for IO3 focused primarily on current skills and access to financial resources. As a result, limited findings are presented under IO3 at baseline.

At the next evaluation point, several new tools or items should be introduced to track indicators under IO3. These include:

- For IO3.1, additional items should be added to the girls survey around financial literacy and access to financial services, based primarily on the programme curriculum.
- For IO3.2:
    - Additional girls survey items and/or focus group questions that specifically report against the indicator "girls' capacity to invest in education."
    - Develop a qualitative FGD for girls, specifically examining financial services and how girls interpret investments in education.
- For IO3.3: Evaluators will use program records to disaggregate self-efficacy and other relevant IO scores by those who participated in the ISOP trainings compared to those who did not. The next evaluation point should also include questions in the girls survey based on programme curriculum around vocational skills.

Targets

SAGE should work with the FM to establish IO3 targets based on specific contextual knowledge. Baseline results are limited and do not provide sufficient context to estimate realistic yet aspirational targets.

Project Checks on Intermediate Outcomes

Ensure that the IO analysis reflects the links between different levels in the logframe and informs the validity of the Theory of Change. This includes checking whether the EE (?) have:

- Measured and analysed all IO indicators presented in logframe.
- Disaggregated the data according to the logframe.
- Used both the qualitative and quantitative analysis stated in the logframe.
- Related the IO analysis to the analysis of Outcomes.

See management response in Annex 18.

IO4: Communities demonstrate more positive gender attitudes and actively support and protect girls

SAGE's fourth IO is an improvement in community members' use of more positive gender attitudes and active support and protection for girls. Specifically, the programme ToC assumes that improved gender attitudes and support are prerequisites for better learning, transition and sustainability outcomes for marginalised girls.

IO4 indicators and relevant baseline information are detailed in Table 26: IO4 Improvement in community members' understanding and use of support mechanisms for marginalised girls' indicators

. Indicator IO4.1: *Percentage of community members demonstrating improved gender attitudes,* is reported at the community level. At baseline, the proportion of communities with a high score on the gender attitudes index is zero. The mean Community Gender Attitude index score is 25.52 on a scale of 1 to 53 for treatment communities and 25.15 for comparison communities. At midline, the proportion of communities with improved scores over baseline should be reported for this indicator. Item-level frequencies are available in Annex 15.

Indicators IO4.2 and IO4.3 are reported as mean scores at the individual girl level. At midline, the proportion of girls with improved scores over baseline should be reported for both of these indicators. In the analysis for these 2 indicators, STS first compared the overall mean scores of the treatment and comparison girls to evaluate if there were statistically significant differences. Next, STS compared treatment girls' mean scores by subgroup (for example, girls facing barriers around menstruation compared to those not facing such barriers) in order to better understand what factors affect girls' scores. For IO4.2: *Perception of safety and security amongst girls in the community*, treatment girls at baseline have a mean score of 3.56 out of 5.00. 54.81% of all girls have a high score, defined as at or over 3.75 out of 5. For IO4.3: *Percentage of marginalised girls who feel they are given appropriate support to stay in school / learning environment*, treatment girls have a mean score of 7.81, and 64.42% of girls have a high score—defined as at or over 7.50 out of 10.

Table 26: IO4 Improvement in community members' understanding and use of support mechanisms for marginalised girls' indicators

IO	IO indicator	Sampling and measuring technique used	Who collected the data?	Baseline level	Target for next evaluation point	Will IO indicator be used for next evaluation point? (Y/N)
IO4: Communities demonstrate more positive gender attitudes and actively support and protect girls	IO4.1: % of community members demonstrating improved gender attitudes (demonstrating increased mean / median attitudes on selected scales)	Boys survey, Caregiver survey, HoH survey		Mean community score at baseline was 25.52. 0% of communities had a high score.	20% of communities have a high score	Yes

IO	IO indicator	Sampling and measuring technique used	Who collected the data?	Baseline level	Target for next evaluation point	Will IO indicator be used for next evaluation point? (Y/N)
IO4: Communities demonstrate more positive gender attitudes and actively support and protect girls	IO4.2: Perception of safety and security amongst girls in the community	Girls survey		Mean score of 3.56. 54.81% of girls had a high score.	N/A	Yes
IO4: Communities demonstrate more positive gender attitudes and actively support and protect girls	IO4.3: % of marginalised girls who feel they are given appropriate support to stay in school / learning environment	Girls survey		Mean score of 7.81. 64.42% of girls had a high score.	20% above baseline	Yes

Main qualitative findings

- At the individual level, FGDs suggest that girls often have a higher chore burden at home, making it more difficult to find the time for schooling and education-related activities.
- At the household level, FGDs indicate that families may prioritize boys' education over girls' education.
- At the community level, FGDs suggest that GBV and harmful community practices, such as early marriage and early pregnancy, as well as religious beliefs and practices, can serve as barriers to girls accessing education.

Main findings

IO4.1 Percentage of community members demonstrating improved gender attitudes (demonstrating increased mean / median attitudes on selected scales)

Quantitative data was collected to understand community attitudes and perceptions towards gender. Questions from the boys survey, the caregiver survey and the HoH survey, based on the Gender Equitable Men scale and the Gender Norm Attitudes scale[64], were used to understand community member perceptions of gender norms, girls' education, SRHR, and GBV, and an index was developed for each tool. For boys in the community, an index was developed using 18 items to understand their gender attitudes and perceptions. The mean score for boys on this index was 9.26 and the maximum score was 12 out of 18. For caregivers, an index of 12 items was developed, and for heads of household, an index of 12 items was developed, also measuring gender attitudes and perceptions. The mean score for caregivers on this index was 8.23 and the maximum score was 11 out of 12. The mean score for heads of household was 8.09 and the maximum score was 10.50 out of 12. There were no significant differences in the respondent-level scales between treatment and comparison cohorts.

These respondent-level indices were aggregated at the community level to create a single index, ranging from 0–42.[65] Aggregation at the community-level was necessary due to the limited sample size at each CBLH. Based on this Community Gender Attitudes index, 0% of communities have high scores for gender attitudes and practices, defined as a score at or above 31.50, 75% of the scale. No statistically significant differences are found in the mean score by treatment versus comparison community, where treatment communities have a mean of 25.52 and comparison communities have a mean of 25.15. Similarly, there are no significant differences by region, as outlined in table 26.[66]

Table 27: IO4.1 Community gender attitude scores, by region

Characteristic	Treatment vs Comparison	Number of respondents in subgroup	Mean Score	Category	Proportion of total in subgroup
All Communities	Treatment	35	25.52	Low score	100.00%
				High score	0.00%
	Comparison	13	25.15	Low score	100.00%
				High score	0.00%
Bulawayo	Treatment	2	25.64	Low score	100.00%
				High score	0.00%
	Comparison	1	20.93	Low score	100.00%

[64] Nanda, Geeta. "Compendium of Gender Scales." Compendium of Gender Scales, September 2011. https://www.c-changeprogram.org/content/gender-scales-compendium/pdfs/4. GEM Scale, Gender Scales Compendium.pdf.
[65] Responses for each item were first reduced to a 0-1 scale; negatively worded items were reverse coded. For each respondent group, the total score was computed. At the community level, the 3 respondent-level scores were added together resulting in an index of 42 items on a range of 0-42.
[66] The subgroups presented in other IO tables were calculated based on girls' individual responses. No community-level barriers were calculated and thus are not presented here.

Characteristic	Treatment vs Comparison	Number of respondents in subgroup	Mean Score	Category	Proportion of total in subgroup
				High score	0.00%
Harare	Treatment	5	23.22	Low score	100.00%
				High score	0.00%
	Comparison	1	23.00	Low score	100.00%
				High score	0.00%
Manicaland	Treatment	22	25.86	Low score	100.00%
				High score	0.00%
	Comparison	9	25.54	Low score	100.00%
				High score	0.00%
Mashonaland East	Treatment	2	25.92	Low score	100.00%
				High score	0.00%
	Comparison	2	26.58	Low score	100.00%
				High score	0.00%
Matabeleland South	Treatment	4	26.21	Low score	100.00%
				High score	0.00%
	Comparison	0	n/a	Low score	0.00%
				High score	0.00%

Results from FGDs presented in the Gender Analysis provide some insight into these rates on the Community Gender Attitudes index. Young men expressed their perception that domestic violence in the community was linked with poverty, thus indicating that meeting basic needs is perceived as a higher priority than treating women equitably. For example, one young man in Mutare Rural commented ‘if we get enough food to eat some of our problems will go away including domestic violence’. Within KIIs, the low or lesser value placed on girls’ education was often associated with Mutare Rural and Apostolic communities—and was mentioned in both KIIs

with SAGE partners, as well as Apostolic and non-apostolic community leaders in Mutare Rural. A common reason for this were beliefs that girls will marry into other families, and therefore, paying for her education would be a waste of resources. Moreover, preference to invest in boys' education over girls more broadly, especially when facing financial shortages or constraints, was also noted. This further supports the low scores on the Community Gender Attitudes index, as many questions ask about these values directly.

IO4.2 Perception of safety and security amongst girls in the community

Results for IO4.2 are reported at baseline as a girls' mean scores on a Perceived Safety index. The Perceived Safety index consists of 5 items on the girls survey asking about girls' knowledge of safe places in their community and perception of community safety. The index is scored by adding each girls' responses resulting in a scale of 0.00 to 5.00, with a higher score indicating greater perceived safety. At baseline, the average score for treatment girls is 3.56 out of 5, with 54.81% of girls receiving a high score—defined as at or above 3.75, 75% of the scale. This is significantly higher than the mean for comparison girls, who scored an average of 1.69 on the index.

Table 28: IO4.2 Girls' perceptions of community safety

Characteristic	Treatment vs Comparison	Number of respondents in subgroup	Mean Score	Category	Proportion of total in subgroup
All girls^	Treatment	416	3.56	Low score	45.19%
				High score	54.81%
	Comparison	248	1.69	Low score	95.97%
				High score	4.03%
Barrier: Accessibility/Long distance to School	Treatment	280	3.64	Low score	42.50%
				High score	57.50%
	Comparison	9	4.11	Low score	22.22%
				High score	77.78%
Barrier: Lack of safety net for GBV***	Treatment	152	2.53	Low score	100.00%
				High score	0.00%
	Comparison	95	0.62	Low score	100.00%
				High score	0.00%
Barrier: Perceived lack of right to education	Treatment	22	3.32	Low score	54.55%
				High score	45.45%
	Comparison	7	1.71	Low score	85.71%

Characteristic	Treatment vs Comparison	Number of respondents in subgroup	Mean Score	Category	Proportion of total in subgroup
				High score	14.29%
Barrier: Lack of enabling environment for quality education	Treatment	45	3.49	Low score	51.11%
				High score	48.89%
	Comparison	0	n/a	Low score	0.00%
				High score	0.00%
Barrier: More barriers due to menstruation	Treatment	204	3.55	Low score	44.12%
				High score	55.88%
	Comparison	110	1.60	Low score	97.27%
				High score	2.73%
Barrier: Lack of voice/ability to speak up	Treatment	81	3.42	Low score	53.09%
				High score	46.91%
	Comparison	50	1.24	Low score	98.00%
				High score	2.00%
Barrier: Girl has functional difficulty	Treatment	123	3.51	Low score	46.34%
				High score	53.66%
	Comparison	68	1.87	Low score	89.71%
				High score	10.29%
Girls 10–14 years old	Treatment	107	3.52	Low score	49.53%
				High score	50.47%
	Comparison	95	1.39	Low score	97.89%
				High score	2.11%
Girls 15–19 years old	Treatment	279	3.59	Low score	42.65%
				High score	57.35%
	Comparison	141	1.92	Low score	95.04%
				High score	4.96%
Bulawayo***	Treatment	24	3.00	Low score	70.83%

Characteristic	Treatment vs Comparison	Number of respondents in subgroup	Mean Score	Category	Proportion of total in subgroup
				High score	29.17%
	Comparison	19	1.16	Low score	100.00%
				High score	0.00%
Harare***	Treatment	48	3.50	Low score	50.00%
				High score	50.00%
	Comparison	25	1.92	Low score	88.00%
				High score	12.00%
Manicaland***	Treatment	269	3.51	Low score	46.84%
				High score	53.16%
	Comparison	169	1.53	Low score	98.82%
				High score	1.18%
Mashonaland East***	Treatment	28	4.25	Low score	17.86%
				High score	82.14%
	Comparison	34	2.62	Low score	85.29%
				High score	14.71%
Matabeleland South***	Treatment	47	3.81	Low score	34.04%
				High score	65.96%
	Comparison	1	1.00	Low score	100.00%
				High score	0.00%

Note: One asterisk (*) indicates a difference in treatment girls' mean scores between those in a subgroup versus those not in a subgroup at p<0.05. Three asterisks indicate a difference in treatment girls' mean scores between those in a subgroup versus those not in a subgroup at p<0.001. One caret (^) indicates a difference between treatment and comparison girls at p<0.001.

Analysis of treatment girls' mean scores by barrier subgroup revealed significant differences in girls' score according to whether or not they face barriers related to a safety net for GBV. Girls lack a safety net related to GBV have a mean Perceived Safety score of 2.53, while girls who do not have a mean score of 4.18. Additionally, there are significant differences in girls in the treatment cohort's mean scores by region. The mean score of girls from Bulawayo is lowest, at 3.00. Girls from Harare have a mean score of 3.50; girls from Manicaland have a mean score of 3.51; girls from Matabeleland South have a mean score of 3.81; and girls from Mashonaland East have the highest scores with a mean of 4.25 (Table 28).

IO4.3 % of marginalised girls who feel they are given appropriate support to stay in school / learning environment

Results for IO4.3 are reported at baseline as a mean score on the Support for Education Index, which measures girls' perception of learning facilities through a CBLH sub-index and perceived caregiver support for education on a Caregiver Support sub-index. The Support for Education index uses 10 items from the girls survey and is scored on a scale of 0.00 to 10.00, with a higher score indicating that the girl has better quality CBLH learning facilities and perceives more support from her caregiver.

There is a statistically significant difference in mean scores between treatment girls and comparison girls, with treatment girls showing a mean Support for Education score of 7.81 while comparison girls show a mean score of 4.72. Table 28 provides more details about girls' mean index scores by subgroup.

Table 29: IO4.3 Girls' perceptions of support for education

Characteristic	Treatment vs Comparison	Number of respondents in subgroup	Mean Score	Category	Proportion of total in subgroup
All girls^	Treatment	416	7.81	Low score	35.58%
				High score	64.42%
	Comparison	248	4.72	Low score	95.56%
				High score	4.44%
Barrier: Accessibility/Long distance to School	Treatment	280	7.90	Low score	32.14%
				High score	67.86%
	Comparison	9	8.44	Low score	22.22%
				High score	77.78%
Barrier: Lack of safety net for GBV	Treatment	152	7.88	Low score	36.84%
				High score	63.16%
	Comparison	95	4.44	Low score	97.89%
				High score	2.11%
Barrier: Perceived lack of right to education	Treatment	22	7.77	Low score	31.82%
				High score	68.18%
	Comparison	7	5.14	Low score	85.71%
				High score	14.29%
Barrier: Lack of enabling environment for quality education***	Treatment	45	6.33	Low score	86.67%
				High score	13.33%
	Comparison	0	n/a	Low score	0.00%

Characteristic	Treatment vs Comparison	Number of respondents in subgroup	Mean Score	Category	Proportion of total in subgroup
				High score	0.00%
Barrier: More barriers due to menstruation	Treatment	204	7.71	Low score	37.25%
				High score	62.75%
	Comparison	110	4.56	Low score	97.27%
				High score	2.73%
Barrier: Lack of voice/ability to speak up***	Treatment	81	7.35	Low score	45.68%
				High score	54.32%
	Comparison	50	4.70	Low score	96.00%
				High score	4.00%
Barrier: Girl has functional difficulty	Treatment	123	7.71	Low score	32.52%
				High score	67.48%
	Comparison	68	4.97	Low score	89.71%
				High score	10.29%
Girls 10–14 years old	Treatment	107	8.07	Low score	24.30%
				High score	75.70%
	Comparison	95	4.65	Low score	97.89%
				High score	2.11%
Girls 15–19 years old	Treatment	279	7.74	Low score	38.35%
				High score	61.65%
	Comparison	141	4.79	Low score	94.33%
				High score	5.67%
Bulawayo	Treatment	24	7.38	Low score	50.00%
				High score	50.00%
	Comparison	19	3.16	Low score	100.00%
				High score	0.00%
Harare	Treatment	48	8.21	Low score	16.67%

Characteristic	Treatment vs Comparison	Number of respondents in subgroup	Mean Score	Category	Proportion of total in subgroup
				High score	83.33%
	Comparison	25	5.28	Low score	88.00%
				High score	12.00%
Manicaland	Treatment	269	7.62	Low score	43.49%
				High score	56.51%
	Comparison	169	4.69	Low score	98.82%
				High score	1.18%
Mashonaland East	Treatment	28	8.25	Low score	17.86%
				High score	82.14%
	Comparison	34	5.44	Low score	82.35%
				High score	17.65%
Matabeleland South	Treatment	47	8.40	Low score	12.77%
				High score	87.23%
	Comparison	1	1.00	Low score	100.00%
				High score	0.00%

Note: One asterisk (*) indicates a difference in treatment girls' mean scores between those in a subgroup versus those not in a subgroup at p<0.05. Three asterisks indicate a difference in treatment girls' mean scores between those in a subgroup versus those not in a subgroup at p<0.001. One caret (^) indicates a difference between treatment and comparison girls at p<0.001.

To better understand the factors affecting the scores of girls in the treatment cohort, the analysis compared the mean scores of treatment cohort girls facing specific barriers with those who do not face these barriers. Treatment girls' Support for Education scores differ significantly between girls depending on if they lack an enabling environment for quality education; lack voice or the ability to speak up; and what region they are from. Girls lacking an enabling environment for quality education have an average score of 6.33, significantly lower than girls not facing this barrier (8.08). Girls lacking voice or the ability to speak up have an average score of 7.35, again significantly lower compared to other girls' average of 7.96. Finally, girls from different regions have significant differences in scores. Girls from Bulawayo have the lowest mean scores (7.38); girls from Manicaland have a mean score of 7.62; girls from Harare have a mean score of 8.21; girls from Mashonaland East have a mean score of 8.25; and girls from Matabeleland South have the highest scores with 8.40 (Table 29).

Analysis of the sub-indices also provides some insight into the mean scores of girls in the treatment cohort. The CBLH sub-index, on a scale of 0.00 to 5.00, measures girls' perceptions of the quality of CBLH facilities, and thus uses items only asked of girls currently attending CBLH.

Girls in the treatment cohort have a mean score of 3.21 out of 5, above the midpoint of the scale, indicating that they find the CBLH facilities to be a supportive learning environment. The Caregiver Support index, also measured from 0.00 to 5.00, uses items asked of all girls. Girls in the treatment cohort showed a very high mean score of 4.56 of 5 on this sub-index.

Reflections

It is likely that the treatment cohort has significantly higher scores than the comparison cohort since girls in the comparison cohort have yet to enrol in CBLH, and at least half of the items were in reference to support to stay in school and learning environment. Overall, there is significant room for growth on indicator IO4.1, given that no communities received high Community Gender Attitudes scores. At the next evaluation point, it would be useful to compare girls' gender scores (IO2.2) to Community Gender Attitudes scores from members in their households, if a one-to-one ratio of girls surveys and household surveys can be followed. Caregivers of individual girls would provide responses that can then be matched to individual girls.

The Perceived Safety scores of girls in the treatment cohort support qualitative findings from the Gender Analysis: the programme should focus initial efforts on girls who lack a safety net around GBV as well as girls in Bulawayo, where scores were significantly lower than other regions. Girls in this category are those who report not having a safe place in the community, somewhere safe to go outside the home or where to go for support if they experience violence. Recommendations from the Gender Analysis include creating increasing awareness and community sensitisation of safe, accessible and confidential locations and procedures to report incidents of GBV, as well as how to assist victims and refer cases to appropriate authorities while closely monitoring for any perceived backlash. Additionally, qualitative tools should be enhanced to collect more nuanced information about girls' perceptions of safety in the community and at CBLH and/or going to/from the hubs.

To better understand mismatches between girls' perceived caregiver support for education, and actual caregiver support for education, it would be useful to collect caregivers' levels of support for girls' education at the midline. In addition to items added for IO4.1, STS will supplement the caregiver survey with items for IO4.2 and match caregiver responses to individual girls' responses.

Targets

The programme logframe sets ambitious targets for the indicators in IO4 at the next evaluation point that should be carefully reviewed by SAGE. For I04.1—*Percentage of community members demonstrating improved community gender attitudes*, the logframe target is set at a 20% increase above the baseline. This would mean 20% of communities show an increase in score since the indicator is not calculated at the individual level. STS recommends SAGE review this target given community-level change often requires more time than individual-level change. Additionally, the logframe does not provide a target for IO4.2—*Perceptions of safety and security amongst girls in the community.* Finally, the target for IO4.3—*Percentage of marginalised girls who feel they are given appropriate support to stay in school / learning environment* is also set at a 20% increase at midline. Given that this indicator is partially driven by how supportive CBLH facilities are, this should also be reviewed by SAGE in consideration with priorities to support these facilities in the first stage of the programme.

Project Checks on Intermediate Outcomes

Ensure that the IO analysis reflects the links between different levels in the logframe and informs the validity of the Theory of Change. This includes checking whether the EE (?) have:

- Measured and analysed all IO indicators presented in logframe.
- Disaggregated the data according to the logframe.
- Used both the qualitative and quantitative analysis stated in the logframe.
- Related the IO analysis to the analysis of Outcomes.

See management response section in Annex 18.

IO5: Strong and active partnerships with MoPSE officials and other civil society actors actively advocate for more inclusive, gender-responsive education policies

SAGE's fifth IO is strengthened district and national leadership and engagement in marginalised adolescent girls' education. Specifically, the programme ToC assumes that stronger governmental engagement in marginalised adolescent girls' education is a prerequisite for better learning, transition and sustainability outcomes for marginalised girls.

IO5 indicators and relevant baseline information are detailed in **Error! Reference source not found.**9. Item-level frequencies are available in Annex 15. Baseline data for IO5 was comprised of qualitative findings from KIIs and desk research completed during the gender-equality-and-social-inclusion analysis.

Table 30: IO5 Strengthened district and national leadership and engagement in marginalised adolescent girls' education indicators

IO	IO indicator	Sampling and measuring technique used	Who collected the data?	Baseline level	Target for next evaluation point	Will IO indicator be used for next evaluation point? (Y/N)
IO5: Strong and active partnerships with MoPSE officials and other civil society actors actively advocate for more inclusive, gender-	IO5.1: # of SAGE-supported materials on inclusive and gender-responsive education approved by MoPSE	KIIs; Gender Analysis	STS	10	TBC	Yes

IO	IO indicator	Sampling and measuring technique used	Who collected the data?	Baseline level	Target for next evaluation point	Will IO indicator be used for next evaluation point? (Y/N)
responsive education policies						
IO5: Strong and active partnerships with MoPSE officials and other civil society actors actively advocate for more inclusive, gender-responsive education policies	IO5.2: Increased resources allocated by MoPSE to support NFE	KIIs; Gender Analysis	STS	0	TBC	TBC

Main qualitative findings

- KIIs with government officials suggest that officials have a good understanding of the SAGE programme and support NFE for OOS adolescent girls. Qualitative data suggest that MoPSE priorities and policies are in alignment with the SAGE intervention.
- KIIs indicate funding will remain a barrier to the sustainability of non-formal accelerated learning for adolescent girls.

Main findings

Baseline data collection tools did not focus on indicators under IO5 given the national scope of this outcome. However, findings from the Gender Analysis identify previously unidentified potential champions for girls' education in civil society and at the local level that may contribute to this outcome. For example, OOS adolescent girls identified the police and headmen as potential allies they might turn to for advice, as well as community child workers and churches.

IO5.1: Number of SAGE-supported materials on inclusive and gender-responsive education approved by MoPSE

SAGE received a letter of support from the Ministry of Education confirming their support to the programme on material development. Specifically, the material development process works with

the MoPSE curriculum development and technical services to provide direction to materials and ensure alignment with the national curriculum.

IO5.2 Increased resources allocated by MoPSE to support non-formal education

As presented in the Gender Analysis, STS conducted a total of 12 KIIs with 13 respondents, including 10 men and 3 women, to understand perspectives on NFE.[67] Six KIIs were conducted with community leaders, 4 with parents or caregivers and 2 with SAGE partners.[68] According to KII respondents, funding and resource allocation for NFE remain a challenge. For example, while policy mandates that schools implement NFE programs, the government is not able to provide additional funds to the districts for the teachers delivering the programs. It is unclear if the MoPSE will have funding available to support and sustain SAGE activities after the end of the programme.

Reflections

It is unclear how much the indicators as stated will be able to capture the strengthening of government support in marginalised girls' education. While SAGE activities are aligned with government goals, funding and resource support for NFE are limited. SAGE should evaluate their strategy for strengthening government support and ensure that it has the potential to lead to the changes being measured in their selected indicators.

At the next evaluation point, the following actions should be taken to ensure that indicator data is adequately collected:

- Government official survey to capture quantitative data on government involvement and support for NFE
- Investigate the strategy for strengthening government support for SAGE activities

Qualitative data for IO5 indicators will be captured from the district- and national-level government officials. Efforts should be made to target interviews to officials with the greatest interaction and knowledge of SAGE and marginalised girls' education initiatives.

Project Checks on Intermediate Outcomes

Ensure that the IO analysis reflects the links between different levels in the logframe and informs the validity of the Theory of Change. This includes checking whether the EE (?) have:

- Measured and analysed all IO indicators presented in logframe.
- Disaggregated the data according to the logframe.
- Used both the qualitative and quantitative analysis stated in the logframe.
- Related the IO analysis to the analysis of Outcomes.

See the management response section in Annex 18.

[67] One KII with Apostolic religious community leaders included 2 male respondents.
[68] KIIs with community leaders included 3 targeted subgroups: (1) community leaders, such as traditional authorities, chiefs or religious leaders; (2) local girls' rights leaders and advocates; and (3) Apostolic religious leader.

7.2 Life skills

Life skills are a key component of the outcomes targeted in IO 2; however, it is reported separately here as a key skill. Plan developed an index to measure girls' life skills, comprised of domains specifically related to the SAGE curriculum, and provided this to STS for analysis. The index also builds on IOs lower in the programme's ToC. Specifically, the life skills index contains 52 items from the following domains already measured and reported under the IOs: attitudes towards education, self-esteem, self-confidence, SRHR KAPs, child protection knowledge and attitudes, and attitudes towards GBV.

To calculate baseline levels of life skills, each girl's mean score on the life skills index was computed as a sum of her index responses, with higher scores indicating a better grasp of the life skills the programme aims to build. Then, girls' scores were categorized as high and low—high life skills scores were defined as scores over 39.75, that is 75% of the life skills index. 5.13% of all girls received a high life skill score on the baseline. Item-level frequencies are available in Annex 15.

Main findings

Findings for life skills are presented in Supplementary Table 18. Overall, girls in the treatment cohort have significantly higher mean life skills scores compared to girls in the comparison cohort. Girls in the treatment cohort have a significantly higher mean score of 29.22, while comparison girls have a mean score of 27.46. In sum, 6.51% of treatment girls received high scores, compared to 2.82% of comparison girls.

Supplementary Table 18: Life skills index results by subgroup and barrier

Characteristic	Treatment vs Comparison	Number of respondents in subgroup	Mean Score	Category	Proportion of total in subgroup
All girls^	Treatment	415	29.22	Low score	93.49%
				High score	6.51%
	Comparison	248	27.46	Low score	97.18%
				High score	2.82%
Barrier: Accessibility/Long distance to School	Treatment	280	29.21	Low score	93.93%
				High score	6.07%
	Comparison	9	30.67	Low score	88.89%

Characteristic	Treatment vs Comparison	Number of respondents in subgroup	Mean Score	Category	Proportion of total in subgroup
				High score	11.11%
Barrier: Lack of safety net for GBV	Treatment	152	29.37	Low score	96.05%
				High score	3.95%
	Comparison	95	25.30	Low score	98.95%
				High score	1.05%
Barrier: Perceived lack of right to education	Treatment	22	26.55	Low score	90.91%
				High score	9.09%
	Comparison	7	24.82	Low score	100.00%
				High score	0.00%
Barrier: Lack of enabling environment for quality education	Treatment	45	29.71	Low score	88.89%
				High score	11.11%
	Comparison	0	n/a	Low score	0.00%
				High score	0.00%
Barrier: More barriers due to menstruation***	Treatment	203	28.34	Low score	92.61%
				High score	7.39%
	Comparison	110	25.56	Low score	98.18%
				High score	1.82%

Characteristic	Treatment vs Comparison	Number of respondents in subgroup	Mean Score	Category	Proportion of total in subgroup
Barrier: Lack of voice/ability to speak up*	Treatment	81	27.69	Low score	97.53%
				High score	2.47%
	Comparison	50	24.50	Low score	100.00%
				High score	0.00%
Barrier: Girl has functional difficulty***	Treatment	123	26.41	Low score	96.75%
				High score	3.25%
	Comparison	68	26.97	Low score	100.00%
				High score	0.00%
Girls 9 years old and under***	Treatment	0	n/a	Low score	0.00%
				High score	0.00%
	Comparison	0	n/a	Low score	0.00%
				High score	0.00%
Girls 10–14 years old***	Treatment	107	23.02	Low score	99.07%
				High score	0.93%
	Comparison	95	23.64	Low score	100.00%
				High score	0.00%
Girls 15–19 years old***	Treatment	278	31.73	Low score	92.09%

Characteristic	Treatment vs Comparison	Number of respondents in subgroup	Mean Score	Category	Proportion of total in subgroup
				High score	7.91%
	Comparison	141	30.59	Low score	96.45%
				High score	3.55%
Girls 20 years old and older***	Treatment	0	n/a	Low score	0.00%
				High score	0.00%
	Comparison	0	n/a	Low score	0.00%
				High score	0.00%
				High score	3.57%
Bulawayo**	Treatment	24	27.16	Low score	87.50%
				High score	12.50%
	Comparison	19	26.26	Low score	89.47%
				High score	10.53%
Harare**	Treatment	48	25.19	Low score	100.00%
				High score	0.00%
	Comparison	25	29.20	Low score	100.00%
				High score	0.00%
Manicaland**	Treatment	268	29.79	Low score	94.03%

Characteristic	Treatment vs Comparison	Number of respondents in subgroup	Mean Score	Category	Proportion of total in subgroup
				High score	5.97%
	Comparison	28	26.70	Low score	98.22%
				High score	1.78%
Mashonaland East**	Treatment	47	29.36	Low score	92.86%
				High score	7.14%
	Comparison	34	30.32	Low score	94.12%
				High score	5.88%
Matabeleland South**	Treatment	47	31.02	Low score	87.23%
				High score	12.77%
	Comparison	1	36.25	Low score	100.00%
				High score	0.00%

Note: One asterisk (*) indicates a difference in treatment girls' mean scores between those in a subgroup versus those not in a subgroup at p<0.05. Two asterisks indicate a difference in treatment girls' mean scores between those in a subgroup versus those not in a subgroup at p<0.01. Three asterisks indicate a difference in treatment girls' mean scores between those in a subgroup versus those not in a subgroup at p<0.001. One caret (^) indicates a difference between treatment and comparison girls at p<0.01.

To better understand factors driving the scores of girls in the treatment cohort, the analysis compared the scores of girls in the treatment cohort facing barriers with those not facing barriers. Results show that girls in the treatment cohort with more barriers due to menstruation have a mean score of 28.34, statistically significantly lower than those without as many barriers due to menstruation (32.24). The same trend is seen for girls who feel they lack voice or cannot speak up. Girls who feel they lack voice have a mean score of 27.69, statistically significantly lower than girls who do not feel this way (29.86). Similarly, treatment girls with functional difficulties have a mean score of 26.41, again statistically significantly lower than their counterparts without functional difficulties (30.40) (Supplementary Table 18).

Reflections

Baseline levels of life skills scores mirror those of the IOs that support this index. Very few girls met the cut-off for high life skills scores, indicating that the programme is well-positioned to support girls in this area of development, specifically girls who face barriers around menstruation, girls with low self-esteem, girls with functional difficulties, younger girls, and girls from Harare and Bulawayo regions.

Because the current life skills index is based on several items that are used for other IOs as well, there is an opportunity to supplement these items with additional items that support other domains of the SAGE curriculum not currently included in the index at the midline. Finally, rewording IO2 to exclude life skills is recommended, since the life skills outcome was requested to be added during baseline analysis.

8. Conclusions

This baseline report presents comprehensive, mixed-method evidence of the current status of outcomes and IOs for SAGE's C1A and C2 beneficiaries. A summary of the findings and implications for the planned interventions are detailed in this section.

Learning outcomes

Overall, girls in the treatment cohort performed comparably to girls in Grades 3 and 5 in formal schooling on the learning assessments. Because the selection criteria of beneficiaries included girls who were not proficient at the Grade 5 level based on the equivalent Wide Range Assessment Tests score, these results corroborate the selection criteria. The average aggregate EGRA score among girls in the treatment cohort is 44.55, in the comparison cohort it is 41.82, and in the benchmark group of Grade 5 girls it is 49.18. There were no statistically significant differences between the aggregate EGRA scores of girls in the treatment and comparison groups. Girls in the treatment cohort struggled the most with comprehension and decoding, as demonstrated by the high proportion of girls who were classified as non-learners in both reading and listening comprehensions subtasks and the letter sound identification subtask.

On the EGMA, girls in the treatment cohort performed at a Grade 3 level overall. The average aggregate EGMA score among girls in the treatment cohort is 66.25; it is 67.65 in the comparison cohort and 65.93 in the Grade 3 benchmark group. As with the EGRA, there were no statistically significant differences between the aggregate EGMA scores of girls in the treatment cohort and those in the comparison cohort. The majority of girls in the treatment cohort are proficient in number recognition, quantity discrimination and addition level 1, while the majority were emergent on the missing number identification subtask. Similar to the SAGE's cohort, girls in Grade 3 had the most difficulty with the missing numbers subtask.

SAGE identified girls as having a functional disability if they ahd at least one disability based on the Washington Group Child Functioning questions. By this definition, girls in the treatment cohort with at least one functional disability had statistically significantly lower literacy and numeracy performance than girls without any functional disabilities. There was not a statistical difference in the comparison group.

Transition outcomes

SAGE identified 3 potential pathways for girls following their participation in CBLHs. These include enrolment in the formal school system, enrolment into vocational training and employment. The pathway analysis is appropriate for the beneficiary girls; however, girls' responses indicate that transition to formal schooling may be an under-utilized option. The majority of girls indicated that they aim to transition into vocational training or employment after completing the CBLH. At baseline, 98.04% of beneficiaries believe they will finish CBLH with 2.76% indicating they hope to enrol in formal education, 47.62% hope to enrol in vocational training, 47.12% hope to be employed and 2.51% either hope to get married or did not know what they plan to do.

The programme's logframe estimates 70% of girls will transition into formal or non-formal schooling, 30% into vocational training and 10% into self-employment or employment. These estimates do not align with C1A girls' intentions and need to be revisited so that programme -

supported pathways can accommodate girls intending to go into vocational training or employment.

Sustainability outcomes

Sustainability findings at baseline—presented for school, system and community indicators—were drawn from qualitative data via KIIs. To measure system-level indicators, the baseline included KIIs with 5 government officials—4 district-level officials and one national-level official—who work within SAGE intervention areas. KIIs focused on understanding current support for marginalised girls, as well as support for NFE programs. The government officials indicated that they are familiar with SAGE, gaining this familiarity from attending SAGE workshops in Harare and the districts. Additionally, all officials had been tasked with a level of oversight for SAGE activities. All respondents indicated that SAGE specifically, and non-formal girls' education generally, remain high priorities for MoPSE; however, all also felt that funding for these programs remains a challenge. At the community-level, 2 community leaders were interviewed via KIIs to explore community engagement with the SAGE programme. Both leaders participated in the development and enrolment stages of the SAGE programme. At the school level, 3 school heads within communities involved in the SAGE programme participated in KIIs and reported on their engagement in community outreach and recruitment activities.

Overall, the KIIs suggest that challenges will persist in sustaining the SAGE programme. Respondents indicated that material provision, training for teachers and staff, integration within the government system and a lacking community support for girls' education all challenge the programme moving forward. Given the evidence, the weighted SAGE sustainability score at baseline is 1.40 out of 4.00.

Theory of change

Assumptions in the programme's ToC regarding subgroups and barriers appear to hold true. The most prevalent social, economic and educational barriers uncovered through the baseline are already considered in SAGE's intervention planning and were explored during the Gender Analysis. These include support for girls' SRHR—specifically menstrual health— GBV and community support for girls' education. However, it is unclear if and how these assumptions may need to be adjusted once the beneficiary enrolment information is updated to include all girls enrolled in the first cohort, including those who served as replacements in the baseline sample.

Risks

Given the high level of sensitivity surrounding SAGE's beneficiaries, the programme should be aware of any heightened stigma or security threats that arise for the girls who attend CBLHs. Because girls and their caregivers noted safety and security at and on the way to school as barriers, the programme should closely monitor any threats faced by participants as a result of their attendance. Given mentions of physical and sexual violence against girls, the programme should also ensure training on proper safeguarding to ensure that staff are aware of abuse signs and reporting mechanisms.

9. Recommendations

This section reflects on baseline findings and provides recommendations to SAGE's staff.

Monitoring, evaluation and learning of the project

- Considering the high scores at baseline on self-efficacy questions, STS recommends adding questions to the girls survey to provide more detail on girls' experiences related to self-efficacy. SAGE should consider adding practical scenario questions to gather more nuanced data on girls' perceptions and experiences related to self-efficacy.
- The programme should develop additional, equated learning assessment forms to be used in future evaluation time points. These equated forms should accommodate the ceiling effects observed for both the EGRA and EGMA. SAGE should develop more complex items for the familiar word reading and oral reading fluency EGRA subtasks and all EGMA subtasks except missing number identification. The addition of items will have implications for piloting and comparability to baseline results.
- Data on whether respondents are young mothers, members of the apostolic community or engaged in labour was not collected in the girls survey at baseline. Given the large number of replacement girls, this data was also unavailable in the enrolment database. To triangulate this information and ensure the data can be disaggregated by these subgroups, STS recommends adding additional items to the girls survey to capture this information.
- Baseline data suggests regional disparities around GBV and girls' access to related safety resources. Additional items should be added at the midline assessment to explore these regional differences and better understand how to support girls in less supported areas.
- To better understand the nuances of girls' gender scores and communities' gender scores, SAGE should consider conducting one-to-one girls surveys and parent/caregiver surveys. This would allow implements to directly match girls' gender scores with the community gender attitudes scores of their parent/caretaker.
- To better report on attendance indicators for girls enrolled in CBLHs, SAGE should consider collecting additional attendance information, including headcounts during classroom observations, classroom attendance records and data from items on the girls survey and household surveys that capture attendance frequency.
- To better understand girls' financial literacy and access to financial services, SAGE should consider adding qualitative data—including FGDs—with girls participating in CBLHs and the financial literacy curriculum.
- To better understand girls' perceptions of financial resources and investment in education, SAGE should consider disaggregating self-efficacy scores by girls who participate in ISOP trainings and those who did not.
- To better understand the nuances of the quantitative data, SAGE should consider adding additional qualitative tools to provide additional information and insight to support survey findings.
- Given the qualitative findings that reported challenges around quality learning spaces, SAGE should continue to include questions about learning spaces in future evaluation

points to unearth whether these concerns relate to previous experience in schools or persist in the CBLHs.

- Qualitative data suggested potential challenges to recruiting and maintaining quality educators in CBLHs. SAGE should explore and monitor these challenges through routine M&E.
- SAGE should monitor enrolment and attendance data in CBLHs, confirming that girls who participated in a baseline communities have not enrolled in a neighbouring CBLH.
- To better understand ways to strengthen girls' enrolment and attendance in CBLHs, SAGE should consider adding additional qualitative questions to parents, caregivers, and community leaders to identify potential strategies for addressing these challenges.
- Based on findings relating to the perceptions of safety questions, SAGE should consider adding additional qualitative tools to gather more nuanced data about girls' perceptions of safety within the community.

Programme design

- Given the range of scores on the EGMA at baseline, SAGE should continue to incorporate differentiated approaches to mathematics instruction to support all beneficiaries, including the high proportion of girls who scored proficient at baseline and may benefit from more complex mathematics instruction.
- SAGE should consider reviewing the mathematics curriculum to ensure the modules provide opportunities for girls to continue to build on their mathematics skills and knowledge. Because many girls are already performing at a Grade 3 level—which is the target for Year 1 in CBLH instruction—the current modules may not have enough new and complex content to ensure the girls continue to build on their mathematics understanding and skills.
- In mathematics, 7.63% of girls scored proficient on the missing number subtask. SAGE should investigate whether this is a misalignment in how girls were previously taught early multiplication skills or whether the SAGE modules incorporate skip counting and repeated addition to building these foundational multiplication skills.
- In general, respondents appeared to struggle on the reading and listening comprehension EGRA subtasks as well as in decoding. SAGE should consider reviewing the curriculum to ensure classroom instruction provides sufficient opportunities for girls to build and practise these skills. SAGE may also determine whether girls' higher proficiency in familiar word reading is a consequence of their prior instruction and how, or if, decoding skills have previously been taught.
- Given the programme's aim to provide the equivalent of a Grade 3 education at the end of Year 1, the programme may consider focusing recruitment and enrolment to girls with less than Grade 3 schooling.
- There may be a mismatch between the programme's transition pathways for beneficiaries and the intended transition pathways girls reported. At baseline, transition pathways are estimated based on girls' stated intentions to transition following CBLH. 98.04% of beneficiaries believe they will finish CBLH with 2.76% indicating they hope to enrol in formal education, 47.62% hoping to enrol in vocational training, 47.12% planning for employment/self-employment and 2.51% plan to get married or did not know what they

plan to do following CBLH. It appears that Plan's estimates for the proportions of girls who will enter into each pathway are not aligned with the intentions expressed by girls. However, the baseline surveys do not examine the intersection between the barriers with missing data—ethnicity, stated religion of household, school experience, including drop out status and carer status. Qualitative data from the Gender Analysis along with further analysis once missing beneficiary data is obtained may provide a nuanced understanding of how the transition pathways supported under the programme could need to be adjusted. If current trends remain, the programme should work to deconstruct existing perceptions of access to formal education and ensure beneficiaries do not assume vocational training and employment are their only option.

- At baseline, girls identified as having a functional disability on the Washington Group Child Functioning Questions had statistically significantly lower literacy and mathematics scores. More than one-quarter of the baseline sample were girls with one or more functional difficulties. To meet the needs of this population, SAGE should provide training to CBLH facilitators on differentiated instruction and inclusive education strategies to meet the needs of all learners.
- Based on quantitative data from the girls survey and FGDs with girls in the Gender Analysis, many girls—particularly in Bulawayo—had low scores on perceived safety, with many girls reporting a lack of a safety net for GBV. SAGE should consider adding interventions to provide support to girls who report lacking a safety net for GBV. Based on FGDs, this support could take the form of increased community awareness and community sensitisation for safe, accessible and confidential locations and procedures to report GBV.
- Based on the girls survey data, approximately 75% of girls who lack voice and the ability to speak up were 15–19 years old. SAGE should continue to explore this age gap through the Champions of Girls Education (CoGE) programme.
- Because programme activities around financial services had not started at the time of the baseline, SAGE should consider using the data collected from the household surveys to inform the design and implementation of VSLAs and ISOPs to better target IO3—improved skills and increased access to financial resources.
- Because a low percentage of girls received high scores on the Gender KAP and SRHR KAP survey items, SAGE should review the programme's curriculum and prioritize opportunities to increase knowledge on SRHR issues, such as the ways in which a girl can get pregnant.

Sustainability

- SAGE should consider increasing the number of community leaders who participate in future KIIs. By incorporating additional perspectives, SAGE can gain a better understanding of the environment that will enable sustainability at the community level.
- Based on findings in the KIIs with MoPSE officials, SAGE should evaluate their strategy for strengthening government support and capacity to lead SAGE-like programmes in the future and focus on building shared accountability for aspects of the programme that align with the MoPSE.
- SAGE should consider adding a quantitative survey for government officials to provide data on the government's involvement in NFE.

10. Annexes

Annex 1: Baseline evaluation submission process

Please submit all baseline reports and accompanying annexes to your respective evaluation officer. Please note, some annexes can be sent for FM review separately and before the baseline report analysis is completed. We advise programmes and EEs to follow the sequence outlined below to speed up the review process and avoid unnecessary back and forth. Where possible, we also advise that programmes and EEs do not begin their baseline report analysis until annex 8 is signed off by the FM.

Annexes to submit for FM review any time before the baseline report is completed:

- Annex 3: Cohort approach evaluation
- Annex 4: Beneficiaries table (sample data)
- Annex 5: Beneficiaries table (Project mapping data)
- Annex 5: MEL framework
- Annex 6: External evaluator's inception report (where applicable)
- Annex 7: Data collection tools used for baseline
- Annex 8: Datasets, codebooks and programs
- Annex 9: Learning test pilot and calibration
- Annex 10: Sampling framework

Annexes to finalise after annex 11 'Datasets, codebooks and programs' is signed off by the FM:

- Annex 2: Logframe
- Annex 11: External evaluator declaration
- Annex 12: Project management response

Annex 2: Logframe

Annex 2_Logframe.xlsx

Annex 3: Cohort approach evaluation

Project to complete

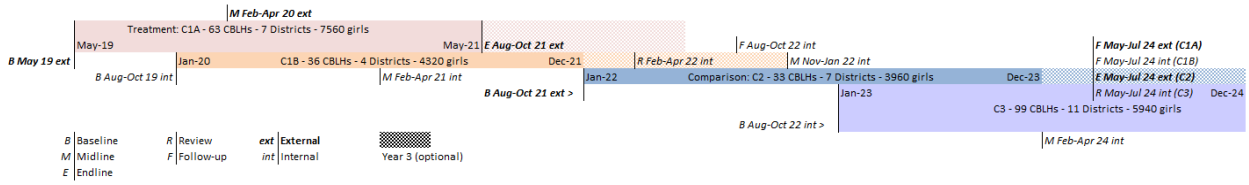
- Please outline if and how you will evaluate learning and, if applicable, transition and any key intermediate outcomes for your other cohorts (i.e. will some be evaluated internally etc.? If so, how).
- Please explain the logic for your approach. For instance, why were certain cohorts prioritised to be externally evaluated over others?

Please note, this is only required if projects have multiple cohorts and are not commissioning your External Evaluator to evaluate all cohorts.

The SAGE programme uses an adapted cross-over design[69] looking at both participant and programme analysis across 4 cohorts. Each of the 4 cohorts will be receiving the intervention for 2 years, with an optional third year available to both the treatment (pathfinder) and comparison groups.

The first cohort (C1A in Figure 15 below) will be our treatment group, receiving the ALP intervention in 63 CBLHs across 7 districts, while the second cohort (C1B) attending another 36 CBLHs in the remaining 4 districts, will receive the intervention from 6 months later. This staggered approach is necessary in order for the deployment of the intervention in new districts to be done in a manageable way. The third cohort (C2), which will act as our comparison group, will receive the intervention in 33 CBLHs across the initial 7 districts, with the final cohort (C3) starting a year later across all 11 districts. The 4 cohorts will come from different geographical areas of the 11 target districts, identified before the baseline.

Figure 15: Phasing of cohorts for intervention and corresponding evaluation activities



We will carry out evaluations at baseline, midline and endline for all 4 cohorts. The C1A and C2 baseline and endline will be externally led, as will the C1A midline and the C1A follow-up. The timing of the C1A endline could be subject to change—for instance, if at the midline point it becomes clear that most girls intend to stay on for the optional third year of teaching, the C1A endline could be pushed back to the end of the third year (however, this will need to be weighed against the cost implications). Moreover, due to budget constraints, the remaining evaluations will be internal 'light-touch' reviews and follow-ups led by consortium MEL staff, with potential input / support from other INGO actors in the Zimbabwe context. These exercises will be looking at Outcomes 1 and 2 (IOs 1 to 3) only and will be conducted on smaller samples using only the tools

[69] As per definition in LNGB MEL Guidance p. 143.

relevant to those IOs. During the endline, when some of them will coincide with an externally led process, they will be treated as separate exercises.

The learning and transition performance of the comparison group will enable us to understand whether improvements in our programming have contributed to better learning and transition outcomes for our target beneficiaries. This design will also allow for assessing the sustainability of our intervention for both the treatment and comparison groups at endline, as this will usually happen after the first 3 cohorts have completed their ALPs (3 years later in the case of cohort C1A).

However, our analysis will be complicated by the fact that, even within the treatment group, there may be different lengths of treatment for different individuals. For example, girls who have had some prior experience of primary or secondary education might be able to transition into a school in one year, before the full ALP is undertaken. Likewise, movement into either of the other alternatives can occur at any time within the period the programme is running, while new girls could join a learning hub later in the year. Thus, the analysis needs to consider for all learners the length of their experience of the intervention.

Annex 4: Beneficiaries table (sample data)

The table below provides data on the characteristic's subgroups and barriers.

Table 31: Characteristic subgroups and barriers of sample for portfolio level aggregation and analysis

Characteristic/Barrier	Proportion of baseline sample (%)
Disability	29.57%
Age 9 and under	0.67%
Age 10–14	25.84%
Age 15–19	72.16%
Age 20 and over	1.34%
Barrier: Accessibility	70.53%
Lack safety net for GBV	36.71%
Lack of right to an education	4.79%
Lack enabling environment for quality education	11.03%
Barriers around menstruation	55.89%
Lack voice and ability to speak up	20.35%
Ethnicity	Not available
Household stated religion	Not available
School experience	Not available
Drop-out status	Not available

Characteristic/Barrier	Proportion of baseline sample (%)
Carer status	Not available

Annex 5: Beneficiaries table (project mapping data)

Project to complete

- Please fill in the tables below and overleaf. In the first instance, use your project monitoring data. If you haven't collected the relevant data, use your sample data to extrapolate to your whole beneficiary population. If you do not have data from your beneficiary data or sample, please put 'NA' in the relevant cell.
- Describe the methodology used for calculating the number of direct and indirect beneficiaries for cohort one and, if applicable, the assumptions you have made for calculating the number you expect to reach by the end of the intervention.
- Comment on the number of direct beneficiaries that you estimate as still meeting your definition of educational marginalisation and how you've verified this.
- If any direct beneficiaries do not meet your definition or are outside the age criteria (<10 and >20), are already in formal school or have already completed the grade level your project is aiming to get the girls up to, please outline your rationale for this and why they were selected as a beneficiary.
- If the direct and indirect beneficiary numbers of girls meeting your definition of educational marginalisation is different to the numbers outlined in your original proposal, please comment on the reasons why.
- How accurate you feel your data is on the age of beneficiaries. For instance, did you collect birth certificates or just rely on the girls' self-reported data?

Table 32: Direct beneficiaries by age

Age (adapt as required)	Proportion of C1A direct beneficiaries (%)	Data source – Project monitoring data, data from sample used in external evaluation or assumption?
Aged <10	3 (0.7%)	Initial beneficiary identification exercise data, updated with enrolment data
Aged 10	141 (3.4%)	
Aged 11	90 (2.2%)	
Aged 12	110 (2.7%)	
Aged 13	144 (3.5%)	
Aged 14	976 (24%)	
Aged 15	988 (24%)	
Aged 16	232 (5.8%)	

Age (adapt as required)	Proportion of C1A direct beneficiaries (%)	Data source – Project monitoring data, data from sample used in external evaluation or assumption?
Aged 17	282 (6.9%)	
Aged 18	414 (10.2%)	
Aged 19	698 (17%)	
Aged 20 +	695 (17%)	

N = 4,075 (100%)

Note: The above data relies on girls' self-reported age data as 3,708 (90%) of the enrolled girls did not have any form of ID. We are also aware that in some regions the heads of household would report some girls to be 18 years of age even though they seemed younger, as they feared being reported to the police for under-age marriage. N =

Table 33: Target groups by out-of-school status

Status	Proportion of C1A direct beneficiaries (%)	Data source – Project monitoring data, data from sample used in external evaluation or assumption?
E.g. Never been to formal school	1,546 (37%)	Same as for Table 31
E.g. Been to formal school, but dropped out	2,526 (63%)	

N = 4,075

Table 34: Direct beneficiaries by drop-out grade

Level of schooling before dropping out (adapt wording as required)	Proportion of C1A direct beneficiaries (%)	Data source – Project monitoring data, data from sample used in external evaluation or assumption?
Grade 0 / Early Childhood Development (ECD) A&B	17 (0%)	Same as for Table 31
Grade 1	32 (0.8%)	
Grade 2	102 (2.5%)	
Grade 3	132 (3.3%)	
Grade 4	156 (3.8%)	
Grade 5	176 (4.3%)	

Level of schooling before dropping out (adapt wording as required)	Proportion of C1A direct beneficiaries (%)	Data source – Project monitoring data, data from sample used in external evaluation or assumption?
Grade 6	226 (5.5%)	
Grade 7	709 (17.6%)	
Form 1	166 (4.1%)	
Form 2	266 (6.5%)	
Form 3	288 (7.1%)	
Form 4	247 (6.1%)	
Form 5	5 (0.1%)	
Form 6	4 (0.1%)	
N = 4,075		

Table 35: Other selection criteria

Selection criteria	Proportion of C1A direct beneficiaries (%)	Data source – Project monitoring data, data from sample used in external evaluation or assumption?
Disabled	54 (1.3%)	Formal disability assessment data
Married	805 (19.7%)	Same as for Table 31
Young mothers (incl. expectant)	921 (23%)	
Apostolic religion	1,351 (33%)	
N = 4,075		

Table 36: Other beneficiaries

Beneficiary type	Total project number for C1A	Total number by the end of the project.	Comments	Data source – Project monitoring data, data from sample used in external evaluation or assumption?
Learning beneficiaries (boys) – as above, but specifically counting boys who will get the same exposure and therefore be also expected to achieve learning gains, if applicable.	1,357	N/A	Ongoing recruitment	CoGE registration forms
Community Educators – number of teachers/tutors who benefit from training or related interventions. If possible /applicable, please disaggregate by gender and type of training, with the comments box used to describe the type of training provided.	124	196	The remaining 72 CEs will be recruited before the ALP starts in the 4 districts of cohort C1B	SAGE employment records
Learning Assistants	62	36	As above	As above
CoGE Facilitators	122	196	As above	CoGE volunteering and training records
Female community members	0	N/A	No VLSA activities are scheduled until 2020	–

These numbers are based on the beneficiary selection database. Given the high rate of replacement girls during baseline data collection, an updated beneficiary database will be needed.

Annex 5: MEL framework

Annex 5_MEL Framework.pdf

Annex 6: External evaluator's inception report (where applicable)

Annex 6_EE Inception Report.pdf

Annex 7: Data collection tools used for baseline

Annex 7_Data Collection Tools.docx

Annex 8: Datasets, codebooks and programs

See attached documents.

Annex 9: Learning test pilot and calibration

Annex 9_Learning test pilot and calibrati

Annex 10: Sampling framework

Annex 8_Sampling Framework.xlsx

Annex 11: Intermediate outcome significance test table

Annex 11_IO significance test anne

Annex 12: Gender Analysis

SAGE Gender Analysis Report_REVIS

Annex 13: Expanded results tables

Table 37: Expansion of Table 16b: Foundational numeracy skills, Grades 3, 5 and 7

Categories	Subtask 1 Number Recognition	Subtask 2 Quantity Discrimination	Subtask 3 Missing Numbers	Subtask 4 Addition (1)	Subtask 5 Addition (2)	Subtask 6 Subtraction (1)	Subtask 7 Subtraction (2)	Subtask 6 Word problems
Grade 5								
Non-learner 0%	0.00%	0.00%	1.65%	0.83%	0.83%	0.83%	3.31%	5.79%
Emergent learner 1–40%	0.83%	0.00%	14.88%	2.48%	2.48%	2.48%	9.09%	36.36%
Established learner 41–80%	5.79%	23.97%	67.77%	11.57%	11.57%	23.14%	34.71%	28.93%
Proficient learner 81–100%	85.95%	68.60%	8.26%	77.69%	77.69%	66.12%	45.45%	20.66%
Source: N= 121	93%	93%	93%	93%	93%	93%	93%	92%
Grade 7								
Non-learner 0%	0.00%	0.00%	0.00%	0.00%	0.00%	0.74%	2.96%	3.70%
Emergent learner 1–40%	0.00%	0.00%	2.22%	0.74%	2.96%	2.22%	0.74%	17.04%

Categories	Subtask 1 Number Recognition	Subtask 2 Quantity Discrimination	Subtask 3 Missing Numbers	Subtask 4 Addition (1)	Subtask 5 Addition (2)	Subtask 6 Subtraction (1)	Subtask 7 Subtraction (2)	Subtask 6 Word problems
Established learner 41–80%	1.48%	1.48%	8.89%	2.96%	22.96%	8.15%	31.11%	21.48%
Proficient learner 81–100%	98.52%	98.52%	88.89%	96.30%	74.07%	88.89%	65.19%	57.78%
Source: N= 135	100%	100%	100%	100%	100%	100%	100%	100%

Table 38: Expansion from Table 17b: Foundational literacy skills, Grades 3, 5 and 7

Categories	Subtask 1 Letter Sound	Subtask 2 Familiar Word Reading	Subtask 3 Oral Reading Fluency – Short	Subtask 4 Oral Reading Comprehension – Short	Subtask 5 Oral Reading Fluency – Long	Subtask 6 Oral Reading Comprehension – Long	Subtask 7 Listening Comprehension
Grade 3							
Non-learner 0%	37.80%	3.66%	2.44%	31.71%	4.88%	50.00%	32.93%
Emergent learner 1–40%	26.83%	9.76%	28.05%	37.80%	45.12%	21.95%	34.15%
Established learner 41–80%	28.05%	14.63%	39.02%	14.63%	24.39%	12.20%	15.85%
Proficient learner 81–100%	6.10%	64.63%	19.51%	4.88%	14.63%	4.88%	6.10%
Source: N= 82	98.78%	92.68%	89.02%	89.02%	89.02%	89.02%	89.02%

Categories	Subtask 1 Letter Sound	Subtask 2 Familiar Word Reading	Subtask 3 Oral Reading Fluency – Short	Subtask 4 Oral Reading Comprehension – Short	Subtask 5 Oral Reading Fluency – Long	Subtask 6 Oral Reading Comprehension – Long	Subtask 7 Listening Comprehension
Grade 7							
Non-learner 0%	36.30%	0.74%	0.74%	11.85%	1.48%	18.52%	17.78%
Emergent learner 1–40%	19.26%	0.74%	3.70%	26.67%	7.41%	17.78%	32.59%
Established learner 41–80%	22.96%	5.93%	23.70%	37.04%	18.52%	32.59%	30.37%
Proficient learner 81–100%	21.48%	92.59%	71.85%	24.44%	72.59%	31.11%	19.26%
Source: N= 135	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

Annex 14: Cronbach's Alphas

Annex 14_Cronbach's Alphas_Barriers and IC

Annex 15: Intermediate Outcome Frequencies

Annex 15_IO Frequencies.pdf

Annex 16: External evaluator declaration

Name of project: SAGE

Name of External evaluator and contact information: School-to-School International, 1005 Terra Nova Boulevard, Suite 1, Pacifica, CA 94044

Names of all members of the evaluation team: Laura Conrad, Hetal Thukral, Aimee Reeves, Anne Laesecke, Melyssa Sibal

Laura Conrad certifies that the independent evaluation has been conducted in line with the Terms of Reference and other requirements received.

Specifically:

- All of the quantitative data was collected independently (Initials:).
- All data analysis was conducted independently and provides a fair and consistent representation of progress (Initials:).
- Data quality assurance and verification mechanisms agreed in the terms of reference with the project have been soundly followed (Initials:).
- The recipient has not fundamentally altered or misrepresented the nature of the analysis originally provided by School-to-School International (Initials:).
- All child protection protocols and guidance have been followed (initials:).
- Data has been anonymised, treated confidentially and stored safely, in line with the GEC data protection and ethics protocols (Initials:).

Laura Conrad
(Name)

School-to-School International
(Company)

13 December 2019
(Date)

Annex 17: Useful resources

Evaluation, analysis and reporting

- World Bank, 2016, *Impact Evaluation in Practice – 2nd Edition* - https://www.worldbank.org/en/programs/sief-trust-fund/publication/impact-evaluation-in-practice
- HM Treasury, 'The Green Book: Appraisal and Evaluation in Central Government'. 2018 - https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/685903/The_Green_Book.pdf

- J-PAL, Introduction to Evaluations - https://www.povertyactionlab.org/sites/default/files/resources/Introduction%20to%20Evaluations%20%281%29.pdf
- Better Evaluation - https://www.betterevaluation.org/

Gender and power analysis

- Sida, 2013, *Power Analysis: Experiences and challenges* (Concept Note). Stockholm: Swedish International Development Cooperation Agency (Sida) - https://www.sida.se/contentassets/83f0232c5404440082c9762ba3107d55/power-analysis-a-practical-guide_3704.pdf
- DFID, 2009, 'Gender and Social Exclusion Analysis How To Note', A Practice Paper, Department for International Development, London, UK - http://www.gsdrc.org/docs/open/se9.pdf
- European Bank for Reconstruction and Development, Gender Tools and Publications - https://www.ebrd.com/gender-tools-publications.html

Annex 18: Project management response

Project to complete

- **What is the project's response to the key findings in the report? Make sure to refer to main conclusions**

This is an opportunity to describe where the project feels the evaluation findings have confirmed or challenged existing understanding and/or added nuance to what was already known. For instance, have findings shed new light on relationships between outputs, intermediate outcomes, and outcomes and the significance of barriers for certain groups of girls – and how these can be overcome? This should include critical analysis and reflection on the project theory of change and the assumptions that underpin it.

- **What is the project's response to the conclusions and recommendations in the report?**

The management response should respond to the each of the external evaluator's recommendations that are relevant to the grantee organisation. The response should make clear what changes and adaptations to implementation will be proposed as a result of the recommendations and which ones are not considered appropriate, providing a clear explanation why.

- **Does the external evaluator's conclusion of the projects' approach to addressing gender inequalities** across activities correspond to the projects' ambitions and objectives?
- **What is the project's response to any GESI risks identified by the evaluator?**
- **What changes to the logframe will be proposed to DFID and the fund manager?**

The management response should outline any changes that the project is proposing to do following any emergent findings from the baseline evaluation. This exercise is not limited to outcomes and intermediate outcomes but extends also to outputs.

- **What are the project's reflections on the ambition of the project?**

Given the learning base levels and characteristics of beneficiaries presented, does the project propose to change its learning and/or transition pathways and targets originally articulated?

Overall, the project acknowledges that the external evaluation team have conducted the evaluation and compiled results in a thorough, diligent and reliable manner and have been able to confirm and challenge existing understanding in a variety of results. Key findings of the evaluation report corroborate assumptions held at design stage and featured within the Theory of Change and confirm what the project is currently learning from the implementation of activities and ongoing field feedback.

However, the project does believe that there are significant challenges in terms of added nuance to what is already known, given the lack of disaggregation by demographic sub-groups and scarcity of qualitative data.

The lack of disaggregation by demographic sub-groups was mainly due to the high replacement rate of girls in the baseline sample and the challenges of enrolling and accurately[70] obtaining the basic demographic data of these girls in time for the baseline data analysis. However, the project is particularly eager to understand these barriers to and experiences of learning for marginalised girls, recognising their heterogeneity and how factors such as being a young mother, Apostolic girl and dropout grade impact learning and intersect one another. The project reports this as a key learning emerging from working with out-of-school girls, whereby their existing life demands can conflict with the project's evaluation and data collection efforts. With the remaining demographic data expected to be compiled by the end of 2019, SAGE will carry out the disaggregation and data analysis either internally or externally, depending on available funds and staff capacity.

As noted in the Executive Summary, the evaluation recommends that future evaluations "place greater emphasis on qualitative data at future evaluation points, particularly as the baseline was limited in exploring 'why' and 'how' to better understand the reasons behind the quantitative results observed". Qualitative data was somewhat limited in the main study as it was initially considered and confirmed by the Fund Manager that the gender analysis represented a sufficient amount of qualitative data for the baseline.

In terms of findings across each outcome:

Learning outcomes

Findings on learning outcomes of girls with wide-ranging and varying learning levels have confirmed the results of screening assessments conducted by the project to determine eligibility of girls in the programme, especially for girls who dropped out of school post grade 5 but have been found to be performing below expected Grade 5 levels. This finding has already created implementation challenges regarding the delivery of learning sessions, confirming that Community Educators (CEs) require longer-term support on implementing differentiated learning. In response to this finding and challenge of working with composite classes, the project has started the process of building educators' capacity around differentiated learning. Although the first module in the ALP (1a) did not bring to prominence the issues of differentiated learning (as the assumption was that girls would be operating at below grade 3 level); the project is pleased to report that differentiated learning has begun to be factored in in Module 1b and 1c following reflection sessions with CEs before the development of each module. Additionally, the framework for Module 2/Year 2 has a deliberate focus on differentiated learning including drawing lessons from formal school interventions such as Performance Lag Address programme (PLAP) and Early Reading Initiative (ERI). A recent co-design workshop with relevant MoPSE departments in materials development, the project captured "what success

[70] The project noted that a small number of girls aged 9 and 20 have therefore been included in the sample. While it is unclear whether they are enrolled or not, this highlights a challenge with certifying the age of beneficiaries, as fewer than 15% of the girls identified so far could present IDs upon enrolment or contact with SAGE staff. The project is working with communities to find the best ways to ensure only eligible girls are ever mobilised / enrolled by the project and avoid having to turn any marginalised girls away.

would look like" for different girls with different learning levels. This understanding will be factored in the subsequent CE trainings.

However, the programme does believe that there are other factors that need to be considered alongside this in terms of the sample and the wider context. Higher proportions of 15-19 year old girls are represented in the treatment sample versus the general Cohort 1, for example in Cohort 1 they represent 24.98%, yet in the sample they represent 63%. Based on our field observations, we can assume that older girls are more likely to have had secondary school experience, but further analysis of treatment and comparison cohort versus school experience and point of dropout data is required to explore the possibility that school experience significantly impacts on learning results.

This is reinforced by the finding that there was a weak but significant correlation between age and average aggregate literacy and numeracy scores (correlation between age and the overall EGRA score was 0.34 and the correlation between age and the overall EGMA score was 0.32) suggesting that although older girls perform better, there was high variability in performance despite age. Furthermore, girls who are 15-19 years old had higher literacy and numeracy aggregate scores than girls who are 10-14 years old. As explained earlier, further disaggregation by sub-groups would aid the project to understand more about the non-learners versus highly proficient girls and hence, how to tailor teaching and learning strategies for SAGE's wider ranging student cohort.

The programme also has some reservations in benchmarking girls against those in the formal system given that this assumes that girls in the formal system are effectively learning and performing at the requisite grade, in line with curriculum expectations. Given that only 7% of the Grade 5 girls are proficient in in listening and reading comprehension (page 59) this does raise questions as to whether they are fully demonstrating what is required by the Grade 5 curriculum, which could explain the limited differences in some sub-tasks to the treatment girls. Open University (who also work on the GECT IGATE programme in Zimbabwe) have flagged that they have observed low-attaining girls performing under grade-level expectations in that programme, so referring to other GEC programmes may also add to the project and Fund Manager's understanding on this issue.

It would also be interesting to understand more about the differences between achievement between literacy and numeracy, particularly as to why girls score lower in numeracy. For example, the finding on page 149, that "the perception of safety mattered among the treatment cohort, where girls who had low levels of perceived safety had lower EGMA aggregate scores, but comparable EGRA scores" indicates that factors outside of the classroom could have a bearing on numeracy acquisition. SAGE will look at carrying out additional analysis on the relationship between literacy and numeracy achievement once the complete dataset is available, such as regression or a multi-level model looking at the effect of proficiency bands in EGRA subtasks on EGMA scores.

Transition

The findings on preferred transition pathway have greatly challenged assumptions at design stage but do corroborate with programme feedback in the implementation phase. The programme had been reporting that high proportions of girls were stating a preference to

transition into vocational training skills and employment, as opposed to returning to formal school. Whilst this challenges the assumptions at design stage and existing logframe targets, whereby 70% beneficiaries were expected to wish to transition into formal school, it is possible that current developments within the Zimbabwean economic climate may be attributing to this shift of choice. For example, girls are more likely to choose immediate livelihood interventions rather than the longer-term benefits that education affords when faced with low and drastically reducing household incomes due to inflationary conditions. Additionally, since 2017, increasing number of formal teachers have either been going on strike or issued threats of going on strike, citing poor working conditions which may have impacted on the perceived value and stability of the formal education system. Based on the above, this change in preference will need to be reflected in logframe targets.

In response to this shift, the programme has brought forward the component of Integrated Skills Outreach Programme (ISOP), a vocational skills training programme, into year 2 to meet girls' expectations and to encourage girls' longer-term attendance. Although many girls expressed an interest in receiving skills training, limited budget and spaces will lead to participation of girls to be guided by strict eligibility criteria.

Sustainability

The lowered sustainability scores (1 for community and 0 for learning space) could be anticipated given the infancy of the programme and that the SAGE approach is not commonly understood by many as it is a non-formal education component, which often is overshadowed by the formal education system both in terms of funding and receiving support. However, the programme thinks it is positive that participation by school heads and community leaders has been reported. The project shares the concerns of misalignment in programme goals and community expectations but would flag that given the small sample number, it would be advisable to replicate these surveys to a wider sample to consider whether these opinions are truly representative. The programme is also intending to roll out inter-generational dialogues in Year 2 to launch awareness raising on gendered issues may also provide an avenue to clarify these misconceptions about SAGE. However, these results will be useful for adjusting future community messaging. Currently, to increase awareness levels on the benefits of community-based learning mechanisms, the programme is learning from current hub experiences of engaging communities. During CE trainings, CEs shared some of their emerging best practices such as conducting end-of-term hub closing ceremonies in which girls were given an opportunity to showcase what they are learning in hubs to the broader community. These innovative community engagement approaches are expected to yield positive results in terms of having communities appreciate SAGE activities and becoming motivated to support. The project will move towards encouraging other hubs to conduct such activities.

The project had existing concerns that MoPSE will not have funding available to support and sustain SAGE after the end of the programme. Furthermore, the project has identified existing policies which are less supportive of NFE programming for girls. For example, although MoPSE recognises non-formal education (NFE), it does not specify NFE learners as beneficiaries of some learning polices such as support on birth registration and some school heads seem to focus support on learners who are formally enrolled in schools, with those registered under NFE not receiving the same level of support. These observations may be attributable to policies which are unclear in terms of giving direction to school heads to include NFE learners. The

project recognises the need to enhance MoPSE engagement in key policy issues to address these identified gaps. The Fund Manager's recent guidance on viewing sustainability as "Long lasting girls' empowerment, for current and future generations" opens up further avenues to explore these findings with MoPSE, in the next review of the project's Sustainability Plan in 2020.

In terms of response to intermediate outcomes, findings indicate that overall, girls in both cohorts who had high levels of self-efficacy, high levels of positive gender attitudes, and high levels of SRHR knowledge had higher literacy and numeracy scores than girls who had low levels of these intermediate outcomes, was an interesting finding and supports the project's Theory of Change that increased self-efficacy and life skills can positively impact on learning outcomes. However, as mentioned earlier, further disaggregation of data is needed to understand the level of skills and knowledge across sub-groups and hence how to tailor approaches.

The results for high scores in self-efficacy are particularly surprising, with the project curious as to whether the length of intervention exposure ahead of data collection could possibly have influenced this, based on the highly positive field-based feedback from girls at the end of Term 1 on their experiences in the Champions of Girls Education (CoGE) component, which explored aspects such as confidence and assertiveness. For example, the hubs were officially opened in the first week of June, with lessons starting two weeks later, so there will have been roughly seven weeks between the start of lessons and data collection. The project is keen to explore these findings further through qualitative methods.

The project noted boys in the community received quite low scores in terms of their gender attitudes and perceptions (9.26 mean score out of a maximum of 18) which confirms the need for the CoGE component to be continued alongside ALP sessions. However, results exploring positive community gender attitudes are relatively limited due to the nature of the report template and that only a limited number of head of households and caretakers could be interviewed due to budget issues. Furthermore, the project would suggest that although negative perceptions are listed and a misalignment in terms of the community's expectations of the project, the misconceptions are of a positive nature in terms of how education can aid a girl's transition to a more positive future e.g. expected outcomes of the programme as explained by community leaders included employment creation for young women, reduction in adolescent pregnancy, increased literacy and reduction in child marriage. This suggests that positive perceptions of the project are evident as a result of previous community messaging efforts. To better understand and monitor this aspect, the project will include community perceptions in the development of regular monitoring tools and midline evaluation tools.

The below table maps where baseline evaluation findings support (or otherwise) in terms of the barriers identified in the project Theory of Change:

Barrier	Supported by main study	Supported by Gender Analysis	New
Time poverty	☑	☑	

Poor quality school infrastructure	☑ (for GWDs)		
Distance	☑	☑	
Quality of instruction	☑	☑	
Low perceived value of education	☑	☑	
Stigma around young mothers and girls with disabilities		☑	
Gender based violence and harmful practices	☑		
Discriminatory religious, social and gender norms	☑	☑	
Limited opportunities for adolescents to learn about and discuss SRHR	☑		
Economic barriers		☑	
Safety and security at and on the way to/from the learning hubs	☑	☑	☑

As yet, there are no findings in support of the following barriers: Limited skills development opportunities; limited opportunities for community-level dialogue and discussion on gender issues; weak community-level protection mechanisms; non-gender responsive education policies; lack of coordination among civil society actors interested in supporting girls' education. The project proposes to use monitoring and midline tools to gain more information on these barriers.

- **What is the project's response to the conclusions and recommendations in the report?**

Thematic focus	Implication and Recommendation	Response
Monitoring, evaluation and learning of the project	Considering the high scores at baseline on self-efficacy questions, the EE recommends adding in additional questions to the girls' survey to provide more detail on girls' experiences as they relate to self-efficacy. SAGE should consider adding practical scenario questions to gather more nuanced data on girls' perceptions and experiences related to self-efficacy.	Agreed for actioning at midline evaluation stage.
Monitoring, evaluation and learning of the project	The programme should develop additional equated learning assessment forms to be used at future evaluation timepoints. These equated forms also need to accommodate the ceiling effects described in this report for both the EGRA and EGMA assessments.	Agreed for actioning ahead of midline evaluation stage. See also below response.
Monitoring, evaluation and learning of the project	Given the baseline data, SAGE should consider developing additional subtasks to capture growth in literacy and in mathematics between timepoints before the next evaluation point. For EGRA, familiar word reading and oral reading fluency subtasks, and on EGMA, all but missing number identification subtasks, appear to have ceiling effects. SAGE should consider these ceiling effects and develop more complex items for each of these subtasks for subsequent evaluation points. The addition of items will also have implications for piloting and comparability to baseline results. Given the criteria for selection of girls as well, higher difficulty subtasks and items are recommended	Agreed for actioning ahead of midline evaluation stage. The project confirms that assessment tools should include more complex subtask corresponding to learning levels. At baseline some tasks included infant level tasks as expected learning levels were to vary from Grade 1 to Grade 5 but shifting to grade 3 to 5 syllabus level expectations may be necessary. As mentioned above, we believe this is due to the identification of a significant proportion of older girls (aged 15-19) who will

		have potentially undergone more schooling compared to younger girls.
Monitoring, evaluation and learning of the project	At baseline, data on whether respondents are young mothers, members of the apostolic community or engaged in labour was not collected in the girls' survey. Given the large number of replacement girls, this data was also unavailable in the enrolment database. To triangulate this information and ensure the data can be disaggregated, STS recommends adding additional items to the girls' survey to capture this information for future cohorts, particularly if replacement is likely.	As discussed above, the programme will seek to collate the outstanding replacement girls' data and explore whether internal or external analysis is possible, recognising the more immediate need for this data. The recommendation to include demographic data in future girls' surveys is also agreed to prevent this issue from re-occurring.
Monitoring, evaluation and learning of the project	Ensure that monitoring data continues to capture changes in enrolment as the high rate of replacement suggests that girls who were enrolled were not in attendance while other girls who were not enrolled were present at the time of baseline, which occurred 2 months after the start of CBLH activities.	The project was still in the process of finalising its attendance tracking and enrolment processes and tools during the baseline data collection. Meanwhile, there has been progress on this, and it is expected that the relevant tools, systems and capacity building will be rolled out in early 2020.
Monitoring, evaluation and learning of the project	Baseline data suggests regional disparities around GBV and girls' access to a safety net for GBV related issues. Additional items may be added at midline to explore these regional differences and to better understand how to support girls in these areas.	Agreed for actioning at midline evaluation stage.
Monitoring, evaluation and learning of the project	Place greater emphasis on qualitative data at future evaluation points, particularly as the baseline was limited in exploring 'why' and 'how' to better understand the reasons behind the quantitative results observed.	Agreed for actioning ahead of midline evaluation stage. See also related points in the headline paragraph above.

Programme Design	Given the range of scores on the EGMA at baseline, SAGE should consider incorporating differentiated mathematics instruction to support all beneficiaries, including the high proportion of girls who are scoring proficient at baseline and may benefit from more complex mathematics instruction than grade 3 equivalent content.	The project has started the process of building educator capacity on learning differentiation to enable them to support different learning needs of all girls. For example, the Module 2 framework developed has elaborated plans to draw lessons from the PLAP and ERI (formal school interventions) to ensure that struggling learners are supported to catch up with high performing learners. ALP materials to be developed will also include "facilitator tips" to differentiate learning tasks to meet learners' operating levels.
Programme Design	SAGE should consider reviewing the mathematics curriculum to ensure the modules provide opportunities for girls to continue to build on their mathematics skills and knowledge. Since many girls are already performing at a grade 3 level, which is the target for one year of CBLH instruction, the current modules may not have enough new and complex content to ensure the girls continue to build on their mathematics understanding and skills.	Module 2 framework currently being developed is incorporating these points by ensuring that the curriculum covers up to Grade 5 level. Whereas as the current modules typically reflect one grade per unit e.g. 2a= Grade 3, the project will explore whether Module 2 grades should cover an array of activities covering Grades 3-5 in each module to accommodate all learning levels.
Programme Design	In mathematics, 7.63% of girls scored proficient on the missing number subtask. SAGE should investigate whether this is a misalignment in how girls were previously taught early multiplication skills, and whether the SAGE modules incorporate	This will be considered within the framework for Module 2 / Year 2 curriculum.

	skip counting and repeated addition to build these foundational multiplication skills.	
Programme Design	In general, respondents appeared to struggle on the reading and listening comprehension subtasks as well as in decoding. SAGE should consider reviewing the curriculum to ensure classroom instruction provides enough opportunities for girls to build and practice these skills and whether girls' higher proficiency in familiar word reading is a consequence of their prior instruction and if and how decoding skills may have been previously taught.	This is being considered within the framework for Module 2 / Year 2 curriculum.
Programme Design	Given the programme aims to provide the equivalence of a Grade 3 education at the end of Year 1, the programme may consider focusing recruitment and enrolment on girls with less than Grade 3 schooling.	The programme has developed a three-phase recruitment approach recognising that girls' willingness to enrol and attend is variable given their level of marginalisation, age and previous school experience. The programme will continue to pursue enrolment strategies that aid the enrolment of girls with lower / no school experiences, for example through peer-based activities.
Programme Design	There appears to be a mismatch between the programme's transition pathways for beneficiaries and the intended transition pathways girls reported. At baseline, transition pathways are estimated based on girls' stated intentions to transition following CBLH. 98.04 percent of beneficiaries believe they will finish CBLH with 2.76 percent indicating they hope to enrol in formal education, 47.62 percent hoping to enrol in vocational training, 47.12 percent planning for employment/self-employment and	The programme appreciates and agrees with this finding and will look to understand further these existing bias or perceptions. Factors such as age need further exploration, with results on page 72 suggesting that for in-school girls, as they get older, they see staying in formal education as a more interesting transition route conversely, in the out of

	2.51% plan to get married or did not know what they plan to do following CBLH. SAGE should ensure the program is providing adequate support for girls to understand all pathways available to them and provide support to identify an intended transition pathway. The programme should work to deconstruct existing biases or perceptions of access to formal education at the girl and community level, to ensure beneficiaries do not assume vocational training and jobs are their only option.	school population it is the younger girls are more interested in school.
Programme Design	At baseline, girls identified as having a functional disability on the Washington Group Child Functioning Questions, had significantly lower literacy and mathematics scores on the learning assessments. More than one-quarter of the baseline sample was girls with one or more functional difficulties. To meet the learning needs of this population, SAGE should provide training to CBLH facilitators on differentiated instruction and inclusive education strategies to meet the needs of all learners.	Training on disability inclusive approaches has been a part of existing CE induction and refresher training. However, the original project design to provide longer-term, community-based specialised support has been challenged, as the intended support of Special Needs Education teachers from the MoPSE is not possible as these specialised teachers are not readily available. The project has started a process of engaging teacher training colleges for possible partnership and training of CEs to fill this gap.
Sustainability	At future evaluation timepoints, SAGE should consider increasing the number of community leaders to participate in KIIs. By incorporating additional perspectives, SAGE can gain a better understanding of the enabling environment for sustainability at the community level.	Agreed for actioning at midline evaluation stage.

Sustainability	Based on findings in the KIIs with MoPSE officials, SAGE should evaluate their strategy for strengthening government support and capacity to lead SAGE-like programmes in the future.	This will be reviewed as part of the project's Sustainability Plan review in 2020. As above, the project recognises the need to enhance MoPSE engagement in key policy issues to address identified gaps.

In addition, the project suggests other factors that need to be considered for future evaluations:

- Given a small number of comparison girls who reported as being enrolled in the programme and receiving the intervention (5.2%), we would recommend for these girls to be included in the treatment group, in order to avoid their results skewing comparison scores. The project assumes for this time, that comparison scores are unaffected given the relatively limited intervention exposure time.

- **Does the external evaluator's conclusion of the projects' approach to addressing gender inequalities** across activities correspond to the projects' ambitions and objectives?

The project considers itself gender-accommodating in its programming, as evidenced by the gender analysis study having informed the development of a project gender strategy and a review of the GESI tool. The project's key focus following the baseline results will be on ensuring that community members and learners themselves are more gender-aware, while a key action going forward will be to support CoGE facilitators to ensure they engage in delivering CoGE sessions in a transformative manner.

- **What is the project's response to any GESI risks identified by the evaluator?**

The evaluator has highlighted three elements of risks to which the project has responded to. The risks are all previously known to the programme, with further evaluations and data collection and analysis to understand what mitigating measures would be most effective.

Risk flagged	Project Response
Given the high level of sensitivity of SAGE Zimbabwe beneficiary girls, the project should be aware of any heightened stigma or security threats that arise for the girls who are attending CBLHs.	The project is aware of this. Beyond ensuring a clear understanding of safeguarding and reporting mechanisms, the programme looks to minimise negative community perceptions through the establishment of community-led Hub Development Committees and the participation of boys and husbands through the Champions of Girls Education component.
Girls and their caregivers noted safety and security at and on the way to school as barriers, so the project should closely monitor any threats faced by participants as a result of their attendance.	The project agrees that this is an ongoing issue and looks to mitigate this through a range of actions – for instance, girls are encouraged to group up or buddy with one another so they can travel to hubs together; the project is trialling satellite hubs to minimise the distance travelled; and the Champions of Girls Education Module 10 will work with girls to explore issues around movement and safety.
Given mentions of physical and sexual violence against girls, the project should also ensure proper safeguarding training, particularly of staff, to be aware of signs and reporting mechanisms.	Safeguarding training is already part of the programme design for all staff and hub-level volunteers.

- **What changes to the logframe will be proposed to DFID and the fund manager?**

No major changes are expected, but adjustment to indicator targets are noted for those with existing ceiling effects, while adjustments to other indicator targets will also be considered. Logframe changes are listed below and will be explored further with the Fund Manager and wider consortium members:

Outcome:
Transition targets: reduce expectation of number of girls to pursue formal schooling route (based on current findings, the project may look to proceed with 5% for formal school, then 45% for skills and 45% for employment)
Revisions to be considered for IO.2 and IO.4.3 targets
Targets to be set for IO.3, IO.4.2 and IO.5
Outputs
Revisions to be considered for all output targets

- **What are the project's reflections on the ambition of the project?**

The project believes that the evaluation report findings indicate that no major changes are required in terms of shifts within the learning and transition-focused outcomes. However, the programme will make specific adjustments within the overall project ambition focusing on the following areas:

- The curriculum of the accelerated learning programme will be adjusted to ensure that it is relevant to the current learning levels, that it is addressing a wide range of learning levels as well as particular gaps, for example in reading and listening comprehension skills.
- Transition pathway targets will be revised with regards to those expected to transition into formal school versus skills training and (self-) employment.

STS conducted a pilot of one version of the existing EGRA and EGMA used previously in Zimbabwe with extended sets of questions in the two oral reading fluency passages, addition level 2 and subtraction level 2 and the word problems per the recommendation from the FM.

The pilot sample was determined in collaboration with SAGE Zimbabwe to represent the types of communities and girls found in the rest of the project. To prepare for the pilot, STS trained 14 female enumerators from Select Research on standard EGRA and EGMA protocols, as well as the types of accommodations provided for students with disabilities. The pilot was conducted on 29 July, 2019 with 167 beneficiaries identified by the SAGE team.

The EGRA and EGMA subtasks used during the piloting had been reviewed for fit and quality prior to the pilot by Sage and Select Research.

The EGRA contained the following subtasks:

1. Letter Name identification
2. Familiar word reading
3. Oral reading fluency (short)
4. Reading comprehension (for short passage)
5. Oral reading fluency (long)
6. Reading comprehension (for long passage)
7. Listening comprehension

The EGMA contained the following subtasks:

1. Number recognition
2. Quantity discrimination
3. Missing numbers
4. Addition levels 1 & 2
5. Subtraction levels 1 & 2
6. Word problems

71 girls took the pilot version of the EGRA and EGMA with the following protocols:

1. 2 minutes for timed subtasks
2. 3 seconds on the EGRA/5 second on EGMA before prompt to move to the next item
3. 18 point font on the stimuli.

Two students identified as possibly having low-vision used the following with adapted protocols:

1. 2 minutes for timed subtasks
2. 5 seconds on EGRA/10 seconds on EGMA before prompting no-response
3. 24 point font stimuli

Two students identified as possibility having difficulty concentrating used the adapted protocols:

1. 2 minutes for timed subtasks
2. 5 seconds on EGRA/10 seconds on EGMA before prompting no-response
3. 24 point font stimuli
4. Allowed to take breaks.

Following the pilot, STS met with the enumerators to debrief the pilot experience by getting both their impressions about the girls' responses to the learning assessments. Given the small sample for the pilot, it is not possible to generalize to the larger beneficiary population with confidence so the results are provided to give an illustrative example of what might occur during the baseline data collection. The results show that the 75 girls who took the assessment have some ability in the early literacy skills, sound identification and familiar word reading, but begin to struggle when asked to read a story and answer comprehension questions. This same trajectory was found in the EGMA results, with stronger scores on number recognition, quantity

discrimination and number patterns. The results show that the girls started to struggle with more complex addition and subtraction problems, as well as word problems.

The following table presents the key lessons learned and proposed next steps for each issue raised during the pilot exercise.

Lessons learned and next steps:

Topic	Pilot issue	Next steps
EGRA – Reading Comprehension (short passage)	The EGRA included three subtasks with extended items. The pilot included 10 reading comprehension questions immediately after the short reading passage. Best practice recommends 5 reading comprehension, including inferential questions.	STS recommends keeping the following questions: 1a. What does Mary's father send her to buy? • This item had a higher percent correct. 2b. Where does Mary drop the money? • This item has higher total item correlation than item 2a 3a. How does Mary feel after losing the money? • This item is correlated with total score and it is a more challenging inferential question, providing room for growth. 4a. Who does Mary ask to help her? • This item has a stronger overall correlation with total score. 5b. Why does her friend give Mary a bag of fruit? • This item has a stronger overall correlation with total score.
EGRA – Reading comprehension (long passage)	The EGRA included three subtasks with extended items. The pilot included 10 reading comprehension questions immediately after the long reading passage. Best practice recommends 5 reading comprehension, including inferential questions.	STS recommends keeping the following five reading comprehension questions for the long passage: 1a. Where was Anna working? • This question has higher percent correct scores and is correlated with the total score. 2b. Where was the bee flying? • This item has a stronger correlation with the total score. 3a. Why did Anna drop her bucket? • This item has a stronger correlation with oral reading fluency. 4b. What did Anna's sister tell her? • This item has a higher correlation with oral reading

		fluency. 5a. Why was Anna happy she didn't kill the bee? • This item has a stronger correlation with total score.
EGRA – Listening Comprehension	The EGRA included three subtasks with extended items. The pilot included 8 listening comprehension questions. Best practice recommends 5 listening comprehension questions.	STS recommends keeping the following items, as they are all correlated with the total score: 4. Why couldn't the dog stand up? 5. Where did Benjamin go? 6. What did Benjamin find when he came home? 7. Why was the dog fat? 8. How many puppies did the dog have?
EGMA – Addition Level 2	The EGMA included three subtasks with extended items. The pilot included 10 items in the Addition Level 2 subtask. Best practice recommends 5 addition level 2 items.	STS recommends keeping the following addition level 2 items. These items are all correlated with the total score and leave room for students to demonstrate gains. 18+2 14+16 25+30 41+15 37+29
EGMA-Subtraction Level 2	The EGMA included three subtasks with extended items. The pilot included 10 items in the Subtraction Level 2 subtask. Best practice recommends 5 addition level 2 items.	STS recommends keeping the following subtraction level 2 items. These items are all correlated with the total score and leave room for students to demonstrate gains. 18-2 25-13 45-15 56-41 67-29
EGMA – Word Problems	The EGMA included three subtasks with extended items. The pilot included 12 items in the word problems subtask. Best practice recommends 6 items in this subtask.	STS recommends keeping the following word problems. These items are correlated with the total score. • 1b. The boy had 16 candies. His mother gave him 13 more candies. How many candies did the boy have? • 2b. The sports team had 18 students. 13 students were girls and the rest were boys. How many boys were on the sports team? • 3b. Phiri had 26 grandchildren Banda had only 11. How many

		more grandchildren does Banda need to have the same number as Phiri? • 4b. Pempho added 25 apples to a basket. Now Pempho has 38 apples in the basket. How many apples did Pempho have to begin with? • 5b. There are 15 people and 45 oranges in a room. If each person has to receive the same number of oranges, how many oranges does each sister receive? • 6a. Each student in the classroom has 2 pencils each. If there are 7 students in the classroom, how many pencils are there?

I. Quantitative surveys

During the pilot data collection day, 28 girls, 17 Caregivers, 14 Head of Households, 11 boys, and 22 girls who fit the benchmarking profile took their respective surveys. Prior to the pilot the tools were reviewed and response options for several items were updated, as were some translations. Specific enumerator instructions – particularly around when response options should be read and when they should not – were determined and programmed.

A summary of key item updates to the girls' survey are listed in Table 2.

Table 2: Recommendations for girls' survey:

Section	Original Item	Decision
GEM	82. Girls provoke boys with short dresses.	82. Girls wearing short dresses provoke boys.
SRH	35. Apart from your family, is there a woman in the community you trust that you could go to if you had a problem or felt unsafe? 36. If girls in this community experience violence, do you know where they can go for support?	Based on observations, these questions yielded the same results. Question 35 was dropped and Question 36 was retained.
Self-efficacy	61. I will be able to achieve most of the goals that I have set for myself. 62. When facing difficult tasks, I am certain that I will accomplish them. 65. I will be able to successfully overcome many challenges	Based on observations, these questions yielded the same results. Question 65 was retained and the others were dropped.

Section	Original Item	Decision
SRH	93a. What material do you use to help you manage your monthly period? 93b. Do you miss school when you are on your period?	Add this question immediately before: “Have you started your period?” -yes -no -don’t know -refused Skip 93a and 93b if the answer is no/don’t know/refused.
SRH	97. A girl can get pregnant if the boy withdraws before ejaculation.	Question was removed.
Practices	109. I feel able to work together with others to make things better for girls in my community.	Question was re-worded to: Do you feel confident to work with others to make sure other girls can access education too?

The item update to the caregiver and head of household surveys is listed in Table 3.

Table 3: Recommendations for caregiver survey and head of household survey:

Section	Original Item	Decision
Caregiver details	4. What type of employment are you in? -Employed -Self-employed -Not employed 5. What is the main type of work that you do?	Both questions were kept in the survey.

The item update to the transition benchmarking survey are listed in Table 4.

Table 4: Recommendations for Transition benchmarking survey:

Section	Original Item	Decision
Benchmarking girl details	8. What is the highest level of education you have completed?	Question re-worded to: 8. What grade are you in?

I. Annex 1.

EGRA pilot results

			DISABILITY				STIMULI			
	Total		NO		YES		STANDARD		LARGE PRINT	
	Count	Mean	Count	Mean	Count	Mean	Count	Mean	Count	Mean
Letter Sound Identification	75	24.63	71	24.73	4	22.75	73	24.42	2	32.00
Familiar Word Reading	75	27.75	71	28.62	4	12.25	73	27.85	2	24.00
Oral Reading Fluency-short	75	37.91	71	39.14	4	16.00	73	38.08	2	31.50
Oral Reading Fluency-long	75	52.89	71	53.30	4	45.75	73	53.11	2	45.00
Reading Comprehension-short	75	1.75	71	1.85	4	0.00	73	1.79	2	0.00
Reading Comprehension-long	75	1.28	71	1.32	4	0.50	73	1.32	2	0.00
Listening Comprehension	75	2.49	71	2.52	4	2.00	73	2.49	2	2.50

EGMA pilot results

			DISABILITY				STIMULI			
	Total		NO		YES		STANDARD		LARGE PRINT	
	Count	Mean	Count	Mean	Count	Mean	Count	Mean	Count	Mean
Number Identification	75	14.43	71	14.59	4	11.50	73	14.55	2	10.00
Quantity Discrimination	75	6.75	71	6.87	4	4.50	73	6.73	2	7.50
Missing Number	75	4.13	71	4.27	4	1.75	73	4.19	2	2.00
Addition Level 1	75	13.08	71	13.04	4	13.75	73	13.21	2	8.50
Addition Level 2	75	8.60	71	8.85	4	4.25	73	8.66	2	6.50
Subtraction Level 1	75	11.40	71	11.68	4	6.50	73	11.56	2	5.50
Subtraction level 2	75	7.19	71	7.32	4	4.75	73	7.18	2	7.50
Word Problems	75	4.76	71	4.93	4	1.75	73	4.81	2	3.00

Barrier	Number of Items	Survey Questions Used	Barrier description	Index type	Index Scale	Reliability Coefficient (alpha)
Long distance to CBLH	1	q4 girls survey	girls who reported traveling more than 30 minutes to CBLH	single item	0 , 1	n/a
lack of safety net for GBV	3	q33,34,36	girls report not having a safe place in community, somewhere safe to go outside the home, and knowing where to go for support if they experience violence	sum	0-3	not computed, will be low because of 3 items
lack of right to education	5	q20- 24	girls perceive that there is no right to go to school and CBLH for children, girls, boys and children with disabilities	sum	0-4	0.119
Lack of quality education opportunities	8	q5,q6,q7,q8,q10,q16_r,q17_r,q18_r	school has books, computers, drinking water facilities, seats, toilet to use, and CE makes students feel welcome, treats boys and girls differently and often absent for class	sum	0 to 8	0.166
	3	q93a, q93b, q93c,qadded1	barrier includes items about whether they have materials, miss school, had questions/who ask. not about their knowledge (that's the IO)	1 if girl faces any of the 3 barriers (no materials, miss school, no one to talk to) about period	0 to 1	not computed, will be low because of 4 items
Relevance of school/CBLH to livelihood plans	1	q19	girls percieve that going to school or cblh is not important for what they want to do when they grow up	one item	0, 1	n/a
lack of ability to speak up	3	q107 to 109	girl does not feel able to talk to parent/cg/spouse about issues that are important to them, to speak up for girls rights in community or confident to work with others to make sure other girls can access education	1 if girl faces any of the 3 barriers	0,1	not computed, will be low because of 3 items

Indicator	Name of Index	Number of Items	Survey Questions Used	Indicator description	Index type	Index Scale	Reliability Coefficie
2.1: % of marginalised girls demonstrating improved self-efficacy	Self-Efficac	6	q63 q64 q65 q66 q67 q68	A higher score indicates higher self-efficacy	Mean	0-3	0.83
2.2: % of marginalised girls demonstrating improved knowledge, attitudes and practices on gender and SRHR	Gender	17	q69 q70 q71 q72 q73 q76 q79 q80 q81 q82 q83 q84 q85 q86 q87 q90 q91	A higher score indicates more positive attitudes towards women and opposition to SGBV	Mean	0-2	0.63
2.2: % of marginalised girls demonstrating improved knowledge, attitudes and practices on gender and SRHR	SRHR	30	q92_1 q92_2 q92_3 q92_4 q92_5 q92_6 q94 q95 q96 q98 q99 q100_1_1 q100_1_2 q100_1_3 q100_1_4 q100_1_5 q100_1_6 q100_1_7 q100_1_8 q101 q102 q103 q104 qadded2 qadded4 qadded5 qadded6 q105 qadded7 q106	A higher score indicates greater knowledge and practice of safe SRHR	Sum	0-30	0.82
4.2: Perception of safety and security amongst girls in the community	rceived Saf	5	q12 q14 q33 q34 q36	A higher score indicates higher perceived safety	Sum	0-5	0.41
4.3: % of marginalised girls who feel they are given appropriate support to stay in school / learning environment	-BH Faciliti	5	q5 q6 q7 q8 q10	A higher score indicates more supportive CLBH facilities	Sum	0-5	0.39
4.3: % of marginalised girls who feel they are given appropriate support to stay in school / learning environment	Caregiver Support for Education	5	q27 q28 q29 q30 q31	A higher score indicates more suportive caregivers	Sum	0-5	0.63
4.3: % of marginalised girls who feel they are given appropriate support to stay in school / learning environment	ducation Su	10	Questions for CLBH Facilities and Caregiver Support for Education Indices	A higher score indicates the girl has a more supportive learning environment and	Sum	0-10	0.58
Lifeskills							
4.1: % of community members demonstrating improved gender attitudes	Gender Att	18	bq10 bq11 bq12 bq13 bq14 bq16 bq17 bq20 bq21 bq22 bq23 bq24 bq25 bq26 bq27 bq28 bq31 bq32		Mean	0-2	0.65
4.1: % of community members demonstrating improved gender attitudes	er Gender A	12	cgq82 cgq83 cgq84 cgq85 cgq86 cgq87 cgq88 cgq89 cgq90 cgq91 cgq92 cgq93		Sum	0-12	0.71
4.1: % of community members demonstrating improved gender attitudes	sehold Gen	12	hhQ19 hhQ20 hhQ21 hhQ22 hhQ23 hhQ24 hhQ25 hhQ26 hhQ27 hhQ28 hhQ29 hhQ30		Sum	0-12	0.71
4.1: % of community members demonstrating improved gender attitudes	ity Gender	3	Boys, Caregiver, and HH Gender Attitudes Indices	A higher score indicates communities have more inclusive gender attitudes	Sum	0-26	0.02

O1 Composite items by indicator

Indicator	Subscale	Survey	Item	Response options	Freq	%
O1.3 Number of highly marginalised girls supported by GEC with improved life skills outcomes	Self-esteem	Girls	Q21. How much would you agree with the following statement: I cannot choose whether to attend or stay in education. I just have to accept what happens.	[_] 3 Agree a lot	190	52.63%
				[_] 2 Agree a little	34	9.42%
				[_] 1 Disagree a little	20	5.54%
				[_] 0 Disagree a lot	110	30.47%
				[_] 777 Don't know	7	1.94%
		Girls	Q52. Generally, I am satisfied with myself.	[_] 3 Completely true	146	40.44%
				[_] 2 Mostly true	59	16.34%
				[_] 1 Slightly true	51	14.13%
				[_] 0 Not true	96	26.59%
				[_] 777 Don't know	9	2.49%
		Girls	Q53. At times, I think I am no good at all.	[_] 3 Completely true	99	27.42%
				[_] 2 Mostly true	48	13.30%
				[_] 1 Slightly true	39	10.80%
				[_] 0 Not true	164	45.43%
				[_] 777 Don't know	11	3.05%
		Girls	Q54. I feel that I have a lot of good qualities.	[_] 3 Completely true	61	16.90%
				[_] 2 Mostly true	20	5.54%
				[_] 1 Slightly true	36	9.97%
				[_] 0 Not true	236	65.37%
				[_] 777 Don't know	8	2.22%
		Girls	Q55. I can do things as well as most other girls my age.	[_] 3 Completely true	124	34.35%
				[_] 2 Mostly true	67	18.56%
				[_] 1 Slightly true	44	12.19%
				[_] 0 Not true	119	32.96%
				[_] 777 Don't know	7	1.94%
		Girls	Q56. I feel I do not have much to be proud of	[_] 3 Completely true	128	35.46%
				[_] 2 Mostly true	83	22.99%
				[_] 1 Slightly true	28	7.76%
				[_] 0 Not true	117	32.41%
				[_] 777 Don't know	5	1.39%
		Girls	Q57. I certainly feel useless at times.	[_] 3 Completely true	84	23.27%
				[_] 2 Mostly true	49	13.57%
				[_] 1 Slightly true	46	12.74%
				[_] 0 Not true	173	47.92%
				[_] 777 Don't know	9	2.49%
		Girls	Q58. I feel that I am just as important as anybody else.	[_] 3 Completely true	151	41.83%
				[_] 2 Mostly true	103	28.53%
				[_] 1 Slightly true	41	11.36%
				[_] 0 Not true	60	16.62%
				[_] 777 Don't know	6	1.66%
		Girls	Q59. I wish I could have more respect for myself.	[_] 3 Completely true	174	48.20%
				[_] 2 Mostly true	97	26.87%
				[_] 1 Slightly true	44	12.19%
				[_] 0 Not true	40	11.08%
				[_] 777 Don't know	6	1.66%
		Girls	Q60. I am afraid that I will fail.	[_] 3 Completely true	109	30.19%
				[_] 2 Mostly true	93	25.76%
				[_] 1 Slightly true	34	9.42%
				[_] 0 Not true	112	31.02%
				[_] 777 Don't know	13	3.60%
		Girls	Q61. I feel positively about myself.	[_] 3 Completely true	137	37.95%
				[_] 2 Mostly true	72	19.94%
				[_] 1 Slightly true	65	18.01%
				[_] 0 Not true	80	22.16%
				[_] 777 Don't know	7	1.94%
	Self-confidence	Girls	Q62. I can make decisions that will help me in my life.	[_] 3 Completely true	146	40.44%
				[_] 2 Mostly true	80	22.16%
				[_] 1 Slightly true	60	16.62%
				[_] 0 Not true	66	18.28%
				[_] 777 Don't know	9	2.49%
		Girls	Q63. I feel confident answering questions when in a group.	[_] 3 Agree a lot	41	11.36%
				[_] 2 Agree a little	30	8.31%
				[_] 1 Disagree a little	43	11.91%
				[_] 0 Disagree a lot	240	66.48%
				[_] 777 Don't know	7	1.94%
		Girls	Q64. I can describe my thoughts to others in the group when I speak.	[_] 3 Agree a lot	221	61.22%
				[_] 2 Agree a little	34	9.42%
				[_] 1 Disagree a little	36	9.97%
				[_] 0 Disagree a lot	64	17.73%
				[_] 777 Don't know	6	1.66%
		Girls	Q65. Can you remember the last time you had a problem? Please tell me the problem. How much would you	[_] 3 Agree a lot	109	30.19%
				[_] 2 Agree a little	41	11.36%
				[_] 1 Disagree a little	39	10.80%
				[_] 0 Disagree a lot	105	29.09%
				[_] 777 Don't know	67	18.56%
		Girls	Q70. I have access to sanitary products if I need them	[_] 3 Agree a lot	142	39.34%
				[_] 2 Agree a little	28	7.76%
				[_] 1 Disagree a little	29	8.03%
				[_] 0 Disagree a lot	44	12.19%
				[_] 888 Refused	2	0.55%
				[_] 777 Don't know	2	0.55%

		Girls	Q71. When I'm on my period, I believe that I can do evertyhing that I normally can do.	[_] 3 Agree a lot	136	37.67%
				[_] 2 Agree a little	17	4.71%
				[_] 1 Disagree a little	16	4.43%
				[_] 0 Disagree a lot	74	20.50%
				[_] 888 Refused	3	0.83%
				[_] 777 Don't know	1	0.28%
		Girls	Q77. I believe that I have the right to say no to unwanted sex.	[_] 3 Agree a lot	240	66.48%
				[_] 2 Agree a little	19	5.26%
				[_] 1 Disagree a little	5	1.39%
				[_] 0 Disagree a lot	62	17.17%
				[_] 888 Refused	4	1.11%
				[_] 777 Don't know	31	8.59%
		Girls	Q78. I believe that I can decide when I want to get married.	[_] 3 Agree a lot	210	58.17%
				[_] 2 Agree a little	39	10.80%
				[_] 1 Disagree a little	9	2.49%
				[_] 0 Disagree a lot	70	19.39%
				[_] 888 Refused	6	1.66%
				[_] 777 Don't know	27	7.48%
		Girls	Q79. If I were pregnant, I would feel comfortable going to school during the pregnancy.	[_] 3 Agree a lot	150	41.55%
				[_] 2 Agree a little	25	6.93%
				[_] 1 Disagree a little	7	1.94%
				[_] 0 Disagree a lot	137	37.95%
				[_] 888 Refused	13	3.60%
				[_] 777 Don't know	25	6.93%
		Girls	Q80. If I were pregnant, I would know where to go to get support and information about the pregnancy.	[_] 3 Agree a lot	194	53.74%
				[_] 2 Agree a little	27	7.48%
				[_] 1 Disagree a little	9	2.49%
				[_] 0 Disagree a lot	72	19.94%
				[_] 888 Refused	6	1.66%
				[_] 777 Don't know	47	13.02%
		Girls	Q81. Boys need sex more frequently than do girls.	[_] 3 Agree a lot	120	33.24%
				[_] 2 Agree a little	15	4.16%
				[_] 1 Disagree a little	9	2.49%
				[_] 0 Disagree a lot	135	37.40%
				[_] 888 Refused	7	1.94%
				[_] 777 Don't know	66	18.28%
		Girls	Q82. A girl can suggest to her boyfriend that they use a condom	[_] 3 Agree a lot	155	42.94%
				[_] 2 Agree a little	15	4.16%
				[_] 1 Disagree a little	4	1.11%
				[_] 0 Disagree a lot	106	29.36%
				[_] 888 Refused	9	2.49%
				[_] 777 Don't know	0	0.00%
		Girls	Q89. I believe that girls have the right to be treated with the same respect as boys.	[_] 3 Agree a lot	258	71.47%
				[_] 2 Agree a little	18	4.99%
				[_] 1 Disagree a little	4	1.11%
				[_] 0 Disagree a lot	47	13.02%
				[_] 888 Refused	4	1.11%
				[_] 777 Don't know	30	8.31%
		Girls	Q90. If I saw abuse, I would report it.	[_] 3 Agree a lot	272	75.35%
				[_] 2 Agree a little	10	2.77%
				[_] 1 Disagree a little	7	1.94%
				[_] 0 Disagree a lot	48	13.30%
				[_] 888 Refused	5	1.39%
				[_] 777 Don't know	15	4.16%
		Girls	Q91. If I experienced abuse, I would report it.	[_] 3 Agree a lot	293	81.16%
				[_] 2 Agree a little	19	5.26%
				[_] 1 Disagree a little	1	0.28%
				[_] 0 Disagree a lot	30	8.31%
				[_] 888 Refused	2	0.55%
				[_] 777 Don't know	16	4.43%
		Girls	Q92. I know to whom or where to report abuse.	[_] 3 Agree a lot	295	81.72%
				[_] 2 Agree a little	6	1.66%
				[_] 1 Disagree a little	2	0.55%
				[_] 0 Disagree a lot	32	8.86%
				[_] 888 Refused	2	0.55%
				[_] 777 Don't know	24	6.65%

IO2 Composite items by indicator

Indicator	Subscale	Survey	Item	Response options	Freq	%
		Girls	Q68. I know where to go if I need support or information about menstrual periods	[_] 1 Yes	175	48.48%
				[_] 0 No	155	42.94%
				[_] 888 Refused	0	0.00%
				[_] 777 Don't know	31	8.59%
		Girls	Q75_x. Other than condoms, what are other types of contraception that you know about?	[_] 0 None/Don't know any others	265	73.41%
				[_] 1 Pill: Women can take a pill every day to avoid becoming pregnant	37	10.25%
				[_] 2 Injectables: Women can have an injection by a health provider that stops them from becoming pregnant for one or more months	75	20.78%
				[_] 3 Emergency contraception: Women can take pills up to 72 hours after sexual intercourse to avoid becoming pregnant	2	0.55%
				[_] 4 Implants: Women can have an implant under the skin on arm which can last for up to three years, or within their womb which can last for three to ten years	61	16.90%
				[_] 555 Other	8	2.22%

				[ ] 888 Refused	2	0.55%
		Girls	Q76_x. What are examples of sexual and reproductive health rights that you know about?	[ ] 0 None/Don't know any	309	85.60%
				[ ] 1 The right to correct sexual and reproductive health information and education	13	3.60%
				[ ] 2 The right to equal treatment and access to services	21	5.82%
				[ ] 3 The right to physical integrity and safety	24	6.65%
				[ ] 4 The right to choose when to marry and have a family	15	4.16%
				[ ] 555 Other	4	1.11%
				[ ] 888 Refused	6	1.66%
		Girls	Q80. If I were pregnant, I would know where to go to get support and information about the pregnancy.	[ ] 3 Agree a lot	194	53.74%
				[ ] 2 Agree a little	27	7.48%
				[ ] 1 Disagree a little	9	2.49%
IO2.1 Number of girls with improved understanding of SRHR	NA			[ ] 0 Disagree a lot	72	19.94%
				[ ] 888 Refused	6	1.66%
				[ ] 777 Don't know	47	13.02%
		Girls	Q81. Boys need sex more frequently than do girls.	[ ] 3 Agree a lot	120	33.24%
				[ ] 2 Agree a little	15	4.16%
				[ ] 1 Disagree a little	9	2.49%
				[ ] 0 Disagree a lot	135	37.40%
				[ ] 888 Refused	7	1.94%
				[ ] 777 Don't know	66	18.28%
		Girls	Q82. A girl can suggest to her boyfriend that they use a condom	[ ] 3 Agree a lot	155	42.94%
				[ ] 2 Agree a little	15	4.16%
				[ ] 1 Disagree a little	4	1.11%
				[ ] 0 Disagree a lot	106	29.36%
				[ ] 888 Refused	9	2.49%
				[ ] 777 Don't know	58	16.07%
		Girls	Q83_x. Other than HIV, what are other sexually transmitted diseases that you know about?	[ ] 0 None/Don't know any	224	62.05%
				[ ] 1 Genital herpes	10	2.77%
				[ ] 2 Chlamydia	11	3.05%
				[ ] 3 Gonorrhea	94	26.04%
				[ ] 4 Syphilis	84	23.27%
				[ ] 555 Other	19	5.26%
				[ ] 777 Don't know	0	0.00%
				[ ] 888 Refused	1	0.28%
		Girls	Q84. Do you believe this statement is true or false: Some medical drugs can prevent the transmission of HIV from mother	[ ] 1 True	192	53.19%
				[ ] 0 False	67	18.56%
				[ ] 888 Refused	3	0.83%
				[ ] 777 Don't know	99	27.42%
		Girls	Q85. Do you believe this statement is true or false: A person with HIV always looks emaciated or unhealthy in	[ ] 1 True	261	72.30%
				[ ] 0 False	60	16.62%
				[ ] 888 Refused	1	0.28%
				[ ] 777 Don't know	38	10.53%
	Self-esteem	Girls	Q52. Generally, I am satisfied with myself.	[ ] 3 Completely true	146	40.44%
				[ ] 2 Mostly true	59	16.34%
				[ ] 1 Slightly true	51	14.13%
				[ ] 0 Not true	96	26.59%
				[ ] 777 Don't know	9	2.49%
	Self-esteem	Girls	Q53. At times, I think I am no good at all.	[ ] 3 Completely true	99	27.42%
				[ ] 2 Mostly true	48	13.30%
				[ ] 1 Slightly true	39	10.80%
				[ ] 0 Not true	164	45.43%
				[ ] 777 Don't know	11	3.05%
	Self-esteem	Girls	Q54. I feel that I have a lot of good qualities.	[ ] 3 Completely true	61	16.90%
				[ ] 2 Mostly true	20	5.54%
				[ ] 1 Slightly true	36	9.97%
				[ ] 0 Not true	236	65.37%
				[ ] 777 Don't know	8	2.22%
	Self-esteem	Girls	Q55. I can do things as well as most other girls my age.	[ ] 3 Completely true	124	34.35%
				[ ] 2 Mostly true	67	18.56%
				[ ] 1 Slightly true	44	12.19%
				[ ] 0 Not true	119	32.96%
				[ ] 777 Don't know	7	1.94%
	Self-esteem	Girls	Q56. I feel I do not have much to be proud of.	[ ] 3 Completely true	128	35.46%
				[ ] 2 Mostly true	83	22.99%
				[ ] 1 Slightly true	28	7.76%
				[ ] 0 Not true	117	32.41%
				[ ] 777 Don't know	5	1.39%
	Self-esteem	Girls	Q57. I certainly feel useless at times.	[ ] 3 Completely true	84	23.27%
				[ ] 2 Mostly true	49	13.57%
				[ ] 1 Slightly true	46	12.74%
				[ ] 0 Not true	173	47.92%
				[ ] 777 Don't know	9	2.49%
	Self-esteem	Girls	Q58. I feel that I am just as important as anybody else.	[ ] 3 Completely true	151	41.83%
				[ ] 2 Mostly true	103	28.53%
				[ ] 1 Slightly true	41	11.36%
				[ ] 0 Not true	60	16.62%
				[ ] 777 Don't know	6	1.66%
	Self-esteem	Girls	Q59. I wish I could have more respect for myself.	[ ] 3 Completely true	174	48.20%
				[ ] 2 Mostly true	97	26.87%
				[ ] 1 Slightly true	44	12.19%
				[ ] 0 Not true	40	11.08%
				[ ] 777 Don't know	6	1.66%
IO2.2 Number of girls with improved self-esteem, self-confidence and well-being	Self-esteem	Girls	Q60. I am afraid that I will fail.	[ ] 3 Completely true	109	30.19%
				[ ] 2 Mostly true	93	25.76%
				[ ] 1 Slightly true	34	9.42%
				[ ] 0 Not true	112	31.02%

				[_] 777 Don't know	13	3.60%
	Self-esteem	Girls	Q61. I feel positively about myself.	[_] 3 Completely true	137	37.95%
				[_] 2 Mostly true	72	19.94%
				[_] 1 Slightly true	65	18.01%
				[_] 0 Not true	80	22.16%
				[_] 777 Don't know	7	1.94%
	Self-confidence	Girls	Q62. I can make decisions that will help me in my life.	[_] 3 Completely true	146	40.44%
				[_] 2 Mostly true	80	22.16%
				[_] 1 Slightly true	60	16.62%
				[_] 0 Not true	66	18.28%
				[_] 777 Don't know	9	2.49%
	Self-confidence	Girls	Q63. I feel confident answering questions when in a group.	[_] 3 Agree a lot	41	11.36%
				[_] 2 Agree a little	30	8.31%
				[_] 1 Disagree a little	43	11.91%
				[_] 0 Disagree a lot	240	66.48%
				[_] 777 Don't know	7	1.94%
	Self-confidence	Girls	Q64. I can describe my thoughts to others in the group when I speak.	[_] 3 Agree a lot	221	61.22%
				[_] 2 Agree a little	34	9.42%
				[_] 1 Disagree a little	36	9.97%
				[_] 0 Disagree a lot	64	17.73%
				[_] 777 Don't know	6	1.66%
	Self-confidence	Girls	Q65. Can you remember the last time you had a problem? Please tell me the problem.	[_] 3 Agree a lot	109	30.19%
				[_] 2 Agree a little	41	11.36%
			How much would you agree with the following statement: I was able to find several solutions to the problem.	[_] 1 Disagree a little	39	10.80%
				[_] 0 Disagree a lot	105	29.09%
				[_] 777 Don't know	67	18.56%

IO4 Composite items by indicator

Indicator	Subscale	Survey	Item	Response options	Freq	%
IO4.2 Improved community support for SRHR and child protection	Child protection	Girls	Q90. If I saw abuse, I would report it.	[_] 3 Agree a lot	272	75.35%
				[_] 2 Agree a little	10	2.77%
				[_] 1 Disagree a little	7	1.94%
				[_] 0 Disagree a lot	48	13.30%
				[_] 888 Refused	5	1.39%
				[_] 777 Don't know	19	5.26%
	Child protection	Girls	Q91. If I experienced abuse, I would report it.	[_] 3 Agree a lot	293	81.16%
				[_] 2 Agree a little	19	5.26%
				[_] 1 Disagree a little	1	0.28%
				[_] 0 Disagree a lot	30	8.31%
				[_] 888 Refused	2	0.55%
				[_] 777 Don't know	16	4.43%
	Child protection	Household	Q101. If I saw or learned about abuse against a child, I would report it.	[_] 1 Agree a lot	308	88.51%
				[_] 2 Agree a little	9	2.59%
				[_] 3 Disagree a little	2	0.57%
				[_] 4 Disagree a lot	25	7.18%
				[_] 888 Refused	0	0.00%
				[_] 777 Don't know	4	1.15%
	Child protection	Household	Q102. If I saw or learned about abuse against a child, I would know to whom or where to report it.	[_] 1 Agree a lot	320	91.95%
				[_] 2 Agree a little	6	1.72%
				[_] 3 Disagree a little	0	0.00%
				[_] 4 Disagree a lot	14	4.02%
				[_] 888 Refused	1	0.29%
				[_] 777 Don't know	7	2.01%
	SRHR	Household	Q104. I believe that girls have the right to go to school while pregnant.	[_] 3 Completely true	154	44.25%
				[_] 2 Mostly true	24	6.90%
				[_] 1 Slightly true	8	2.30%
				[_] 0 Not true	156	44.83%
				[_] 888 Refused	2	0.57%
				[_] 777 Don't know	4	1.15%
	SRHR	Household	Q105. I believe that girls have the right to go back to school after they have children.	[_] 3 Completely true	329	94.54%
				[_] 2 Mostly true	3	0.86%
				[_] 1 Slightly true	0	0.00%
				[_] 0 Not true	13	3.74%
				[_] 888 Refused	0	0.00%
				[_] 777 Don't know	3	0.86%
	SRHR	Household	Q106. It is a woman's repsonsibility to avoid getting pregnant.	[_] 3 Completely true	319	91.67%
				[_] 2 Mostly true	6	1.72%
				[_] 1 Slightly true	3	0.86%
				[_] 0 Not true	20	5.75%
				[_] 888 Refused	0	0.00%
				[_] 777 Don't know	0	0.00%
	SRHR	Household	Q107. I believe that girls and women have the right to say no to unwanted sex.	[_] 3 Completely true	316	90.80%
				[_] 2 Mostly true	7	2.01%
				[_] 1 Slightly true	4	1.15%
				[_] 0 Not true	19	5.46%
				[_] 888 Refused	2	0.57%
				[_] 777 Don't know	0	0.00%
	SRHR	Household	Q108. I believe that girls have the right to say no to getting married before they are 18.	[_] 3 Completely true	327	93.97%
				[_] 2 Mostly true	4	1.15%
				[_] 1 Slightly true	6	1.72%
				[_] 0 Not true	11	3.16%
				[_] 888 Refused	0	0.00%
				[_] 777 Don't know	0	0.00%

IO4.3 Improved community support for girls education through CBEs and primary school	NA	Girls	Q16. Do you think going to school is important for what you want to do when you grow up?	[_] 1 Yes	349	96.68%
				[_] 0 No	11	3.05%
				[_] 888 Refused	1	0.28%
				[_] 777 Don't know	0	0.00%
		Girls	Q17. Do you think that it is important for children to go to school?	[_] 1 Yes	355	98.34%
				[_] 0 No	4	1.11%
				[_] 888 Refused	1	0.28%
				[_] 777 Don't know	1	0.28%
		Girls	Q18. Do you think girls have a right to go to school?	[_] 1 Yes	348	96.40%
				[_] 0 No	9	2.49%
				[_] 888 Refused	2	0.55%
				[_] 777 Don't know	2	0.55%
		Girls	Q19. Do you think boys have a right to go to school?	[_] 1 Yes	344	95.29%
				[_] 0 No	12	3.32%
				[_] 888 Refused	0	0.00%
				[_] 777 Don't know	5	1.39%
		Girls	Q20. Do you think children with disabilities have a right to go to school?	[_] 1 Yes	65	18.01%
				[_] 0 No	291	80.61%
				[_] 888 Refused	0	0.00%
				[_] 777 Don't know	5	1.39%
		Household	Q62. After finishing CBE, what do you hope [GIRL] will do?	[_] 1 Go to primary school	63	18.10%
				[_] 2 Go to vocational training	171	49.14%
				[_] 3 Work in a safe, fairly paid job	161	46.26%
				[_] 4 Become self-employed	136	39.08%
				[_] 888 Refusal	0	0.00%
				[_] 777 Don't know	39	11.21%
		Household	Q63. What level of schooling would you like [GIRL] to achieve?	[_] 1 None	1	0.29%
				[_] 2 Primary	64	18.39%
				[_] 3 Lower secondary	33	9.48%
				[_] 4 Upper secondary	170	48.85%
				[_] 5 College or university	50	14.37%
				[_] 777 Don't know	30	8.62%
		Household	Q64. To what extent do you agree that "even when funds are limited it is worth investing in [GIRL]'s education"	[_] 4 Strongly agree	303	87.07%
				[_] 3 Agree	40	11.49%
				[_] 2 Neither agree or disagree	1	0.29%
				[_] 1 Disagree	4	1.15%
				[_] 0 Strongly disagree	0	0.00%
		Household	Q65. To what extent do you agree "a girl is just as likely to use her education as a boy"	[_] 4 Strongly agree	265	76.15%
				[_] 3 Agree	59	16.95%
				[_] 2 Neither agree or disagree	7	2.01%
				[_] 1 Disagree	14	4.02%
				[_] 0 Strongly disagree	3	0.86%

Menstruation Barriers Composite items by indicator

Survey	Item	Response options	Freq
Girls	Q68. Do you know where to go if you need support or information about menstrual	[_] 0 No	155
		[_] 1 Yes	175
		[_] 777 Don't know	7
	Q69. Where would you go if you need support and information about menstrual periods?	[_] 1 Shop	0
		[_] 2 Pharmacy	0
		[_] 3 Government hospital, health centre, or clinic	38
		[_] 4 Private doctor, nurse, or clinic	1
		[_] 5 Mother or female family member	127
		[_] 6 Father or male family member	7
		[_] 7 Church	2
		[_] 8 Community member	21
		[_] 9 NGO or CBO	3
		[_] 555 Other	16
		[_] 888 Refused	0
		[_] 777 Don't know	4
	Q.70 I have access to sanitary products if I need them	[_] 3 Agree a lot	142
		[_] 2 Agree a little	28
		[_] 1 Disagree a little	29
		[_] 0 Disagree a lot	44
		[_] 888 Refused	2
		[_] 777 Don't know	2
	Q71. When I'm on my period, I believe that I can do everyhing that I normally can do.	[_] 3 Agree a lot	136
		[_] 2 Agree a little	17
		[_] 1 Disagree a little	16
		[_] 0 Disagree a lot	74
		[_] 888 Refused	3
		[_] 777 Don't know	1
	Q72. I feel ashamed of my body when I have my period	[_] 3 Agree a lot	114
		[_] 2 Agree a little	25
		[_] 1 Disagree a little	9
		[_] 0 Disagree a lot	93
		[_] 888 Refused	3
		[_] 777 Don't know	0

Annex 19. Additional sub-group analysis

As noted in the Management Response (Annex 18) due to challenges with accurately [(1)] obtaining the required demographic data of replacement girls in the baseline sample, an additional sub-group analysis was undertaken in February-March 2020 to accompany baseline findings. Data from treatment areas was re-analysed by the external evaluator (EE) with the below narrative developed by the project. This document accompanies the dataset and explains key findings, as well as the project response to challenges or aspects hindering full data-analysis.

This annex details the following:

1) Confirmation of data available for analysis
2) Definition of key sub-groups
3) Sample breakdown by main demographics
4) Learning outcome data by sub-group
5) Analysis of dropout by age
6) Intermediate outcome (IO) scores by sub-group
7) Barriers reported by girls by sub-group
8) Specific findings by sub-group
9) Transition outcomes by sub-group
10) Further analysis requested by the project
11) Further programme actions based on sub-group analysis

The sub-group analysis dataset is the following:

Tables_by_subgroups_May_072020.xlsx

(1) The project noted that a small number of girls aged 9 and 20 have been included in the sample. While it is unclear whether they are enrolled or not, this highlights a challenge with certifying the age of beneficiaries, as fewer than 15% of the girls identified so far could present IDs upon enrolment or contact with SAGE staff. The project is working with communities to find the best ways to ensure only eligible girls are ever mobilised / enrolled by the project and avoid having to turn any marginalised girls away.

1) Confirmation of data available for analysis

The project was only able to obtain demographic data from 66.8% of the girls sampled from the treatment areas and 35.1% of girls sampled from the comparison areas, as per the below table:

Table I. Percentage of total sample vulnerability data available

<table>
<tr><td colspan="4">Total sample:
720 girls</td></tr>
<tr><td colspan="2">Treatment sample:
458 girls</td><td colspan="2">Comparison sample:
262 girls</td></tr>
<tr><td>Original treatment sample:
83 girls</td><td>Replacement treatment sample:
375 girls</td><td>Original comparison sample:
82 girls</td><td>Replacement comparison sample:
180 girls</td></tr>
<tr><td>Of which % demographic data available:
89% (74 girls)</td><td>Of which % demographic data obtained:
62%% (232 girls)</td><td>Of which % demographic data available:
100% (82 girls)</td><td>Of which % demographic data obtained:
5% (10 girls)</td></tr>
<tr><td colspan="2">Treatment sample % demographic data available:
66.8% (306 girls)</td><td colspan="2">Comparison sample % demographic data available:
35.1% (92 girls)</td></tr>
<tr><td colspan="4">Total sample % demographic data available:
55.2%</td></tr>
</table>

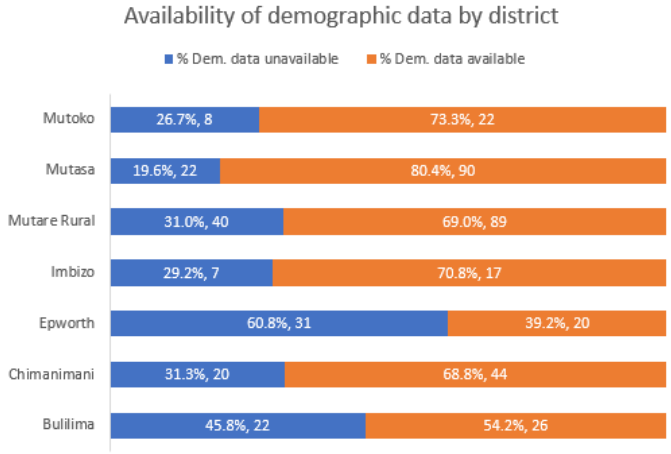
It is important to clarify that all treatment area replacement girls were actual beneficiaries of the programme. However, the rolling enrolment approach meant that some of them joined the hubs after the sampling had been done, and there was not enough time to collect their identification information before the follow-up was carried out. During that time, some of them had already left the hub (either migrating from their areas or due to marriage). The missing information for those still in the hub but who were unavailable

during the follow-up was not collected in time as per the established learner registration process, a challenge that the programme is aware of and currently working to resolve.

In light of this, as well as the likelihood of significant changes to the comparison cohorts envisaged by the project redesign (ongoing at the time of writing) we have decided to focus the sub-group analysis on the treatment group, where we have 66.8% of the demographic data – as adding in the comparison group would lower that percentage to 55.2%.

The above challenges present an issue with the generalisability of the results since demographic data are available to varying degrees among sub-groups of girls. The impact of this is reflected across the sub-group analysis results, which makes it difficult for the programme to make broad generalisations. Throughout the analysis, caveats are provided to flag where the percentage of complete data is particularly low or disproportionate. For instance, the data available from Mutasa and Mutoko was more complete than that available from Epworth and Bulilima districts. This means some of the conclusions are biased towards districts where the data was more complete.

Figure I



2) Definition of key sub-groups

As per the SAGE programme MEL Framework and at proposal phase, the programme proposed to work with girls from the following sub-groups:

- Adolescent mothers
- Adolescent married girls
- Girls from single-parent households
- Girls with disabilities
- Girls living with extended family
- Girls at risk of early marriage
- Girls engaged in labour (2)
- Girls who have been internally displaced
- Girls from Apostolic communities
- Girls from ethnic minorities (defined as Kalanga, Ndebele and San)
- Girls from migrant communities
- Girls whose parents cannot afford fees

(2) Please note the programme has adapted this category name from its previous term of 'worst forms of child labour, including transactional sex' given the difficulty to ascertain this at identification stage and the sensitivity of disclosure.

3) Sample breakdown by main demographics

When reviewing the dataset, the tables should be read with the following in mind:

- 0 – Meets criterion / 1 – Does not meet criterion / 999 – Missing relevant data. Where 1 is missing it means no girls met the criterion.
- N% is the percentage meeting the relevant criterion.
- Where A, B or C is found alongside a value in the dataset, this denotes a significant statistical difference between that sub-group and rest in that category. Results are based on two-sided tests

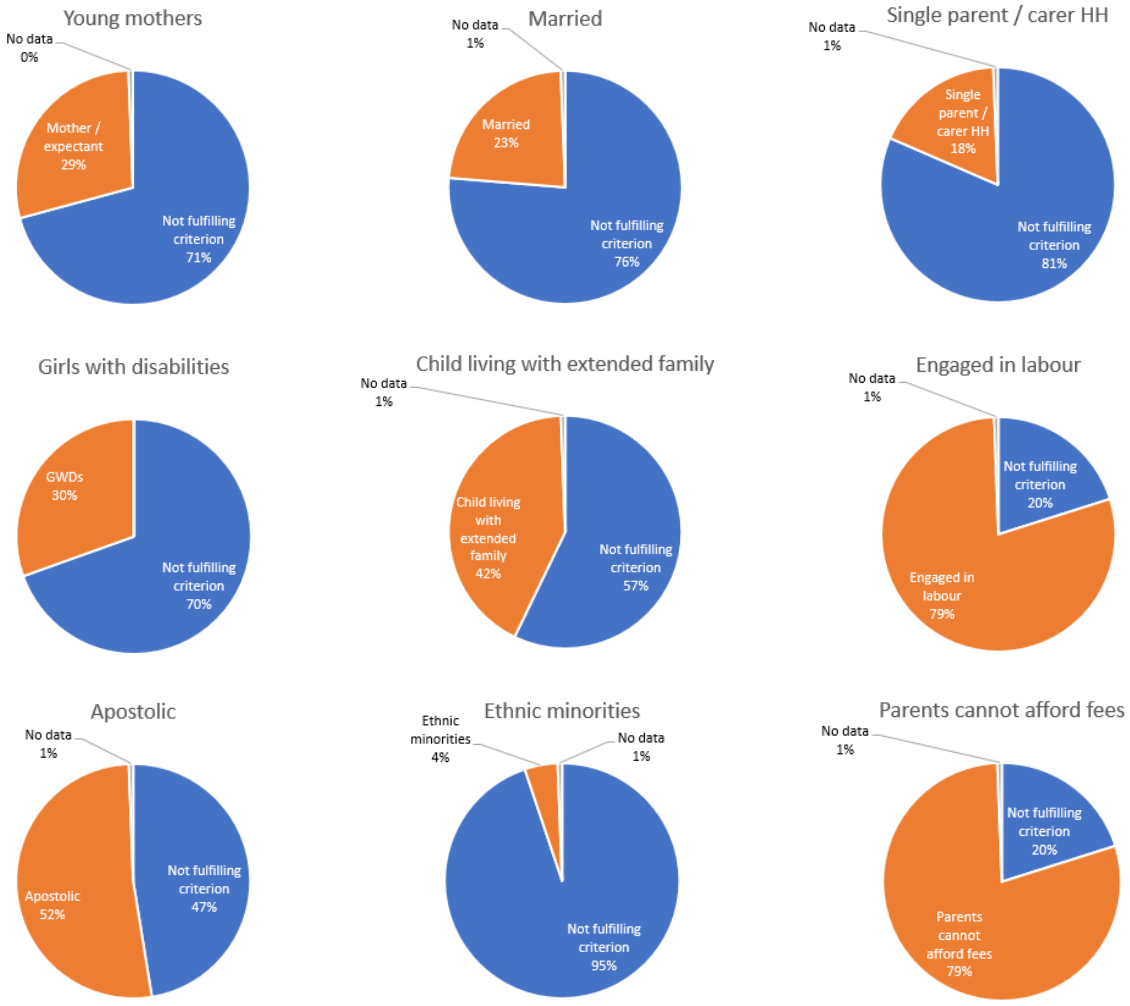
assuming equal variances. For each significant pair, the key of the smaller category appears in the category with the larger mean. The significance level for upper case letters (A, B, C) is .05. Categories were not used in comparisons if the sum of case weights is less than ten. Tests are adjusted for all pairwise comparisons within a row of each innermost sub-table using the Bonferroni correction.

There were no girls in the sample who met the 'At risk of early marriage' and 'Refugee / migrant / IDP' criteria [3] therefore these subgroups are not analysed further.

(3) This information came from our identification survey – if a girl had a married sibling of a similar age she was considered 'at risk' and if a girl said she had moved from a different part of the country for reasons of natural disaster or conflict. In addition, two more vulnerability criteria appear in the data tables which were not included in the above sub-groups (which correspond to the proposal stage) namely 'Chronic illness' and 'Distance to school over govt. limit' both of which come from direct questions in the identification survey. Our sampling however was not purposive to this extent and thus no girls from these categories were found in the baseline sample.

For the treatment girls whose demographic data we have, the proportion of each sub-group vs the rest of the sample with available demographic data is as follows:

Figure II. Sub-groups by vulnerability criteria



4) Learning outcome data by sub-group

Within each sub-group, reading and math scores for girls who met the criteria were compared to those who did not meet the criteria. Table II summarizes the results of the pairwise significance tests comparing the mean scores of the two groups, by sub-group. Note that the analyses did not include additional control variables and the results presented below exclude girls who were missing the data (i.e., comparison is only between girls who were '0' on the criteria and those who were '1' on the criteria).

Figure III

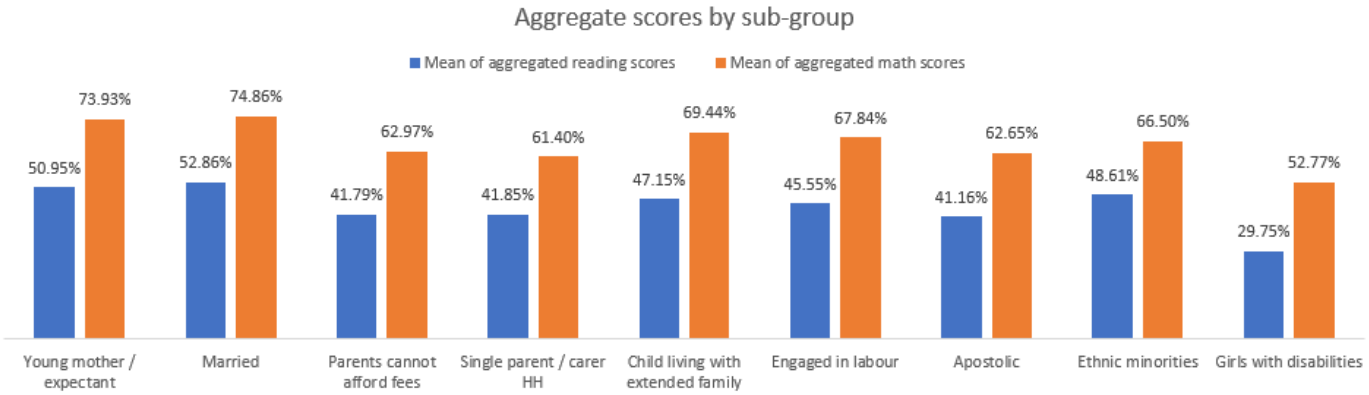
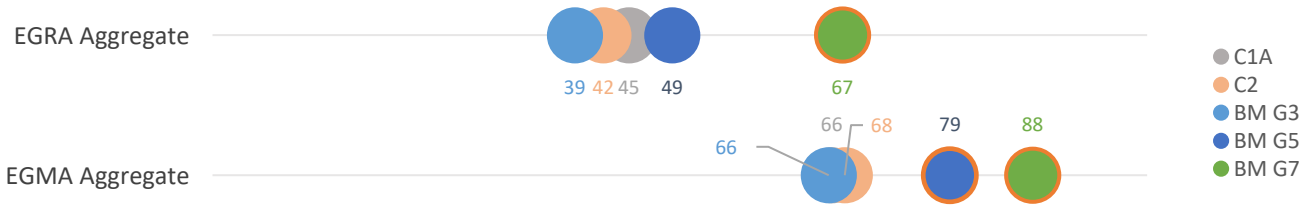


Table II. Significance of differences by sub-group on mean math and reading aggregate scores

Sub-group	**Reading scores of girls who meet the criteria vs girls who do not meet criteria**	**Math scores of girls who meet the criteria vs girls who do not meet criteria**	**Additional details on sample with available data**
Young mother / expectant	Significantly higher	Significantly higher	Does not control for age of girls, which is highly correlated with this criterion. Highest number of young mothers / expectant were in Mutare rural (33) and Mutasa (33) with fewest in Bulilima (2) and Epworth (1).
Married	Significantly higher	Significantly higher	Does not control for age of girls, which is highly correlated with this criterion. Highest number of girls who reported being married were in Mutasa (31) and Mutare rural (22) with none in Bulilima, Epworth and Imbizo.
Parents cannot afford fees	No difference	Significantly lower	Majority of girls whose parents cannot afford fees were in Mutare rural (80) and Mutasa (69) and the fewest were in Imbizo and Bulilima (14 each).
Single parent / carer HH	No difference	No difference	Highest number of single parent/carer households were in Bulilima (15) and fewest in Mutoko (2).
Child living with extended family	No difference	Significantly higher	Highest number of children living with extended family were in Mutasa (57) and fewest were in Bulilima and Chimanimani (4 each).
Engaged in labour	Significantly higher	Significantly higher	Correlated with age. Majority of girls engaged in labour were in Mutasa (88) and Mutare rural (72) and the fewest were in Epworth (11) and Imbizo (3).
Apostolic	No difference	No difference	Highest proportion of Apostolic girls were in Mutare rural (63) and Mutasa (50) and fewest were in Bulilima (2).
Ethnic minorities	No difference	No difference	Overall 14 girls identified as ethnic minorities. 6 were in Bulilima, 7 in Imbizo and 1 in Mutasa.
Girls with disabilities	Significantly lower	Significantly lower	The greatest proportion of girls with disabilities were in Mutare rural (37) and Mutasa (29) and fewest were in Mutoko (7).

To aid comparison with the wider sample, the below diagram showing EGRA and EGMA aggregate mean scores for girls in C1A, C2 and benchmarked girls in Grades 3, 5 and 7 has been extracted from the main baseline report:

Figure IV



Aggregate reading scores

On the aggregate reading score, young mothers / expectant, married girls, girls engaged in labour and girls without disabilities have significantly higher scores than girls who are not mothers, those who are not married, those not engaged in labour and those with disabilities respectively.

Aggregate math scores

On the aggregate math score, young mothers / expectant, girls who are married, girls whose parents can afford fees, girls living with extended family, girls engaged in labour and girls without a disability have significantly higher aggregate math scores than girls who are not mothers, those who are not married, those not engaged in labour, those whose parents cannot afford fees, girls not living with extended family and those with a disability respectively. Differences were observed whether those missing vulnerability data were included or not.

- As the above overview indicates, the vast majority of reading aggregated scores and all math aggregated scores for all sub-groups rank as ‘established learners’ (41-80%).
- Only girls with disabilities would rank as ‘emergent learners’ (1-40%) for reading scores and none would rank as ‘non-learner’ (0%) or ‘proficient learners’ (80-100%).
- The highest three aggregate scores were all attained in math by married girls (74.86%); young mothers / expectant (73.93%) and girls living with extended family (69.44%). All are above the treatment group EGMA average aggregate score of 66.25%.
- The lowest three scores were all attained in reading by girls with disabilities (29.75%); Apostolic girls (41.16%); and girls whose parents cannot afford school fees (41.79%). All are below the EGRA average aggregate score of 44.55%.

Figure V

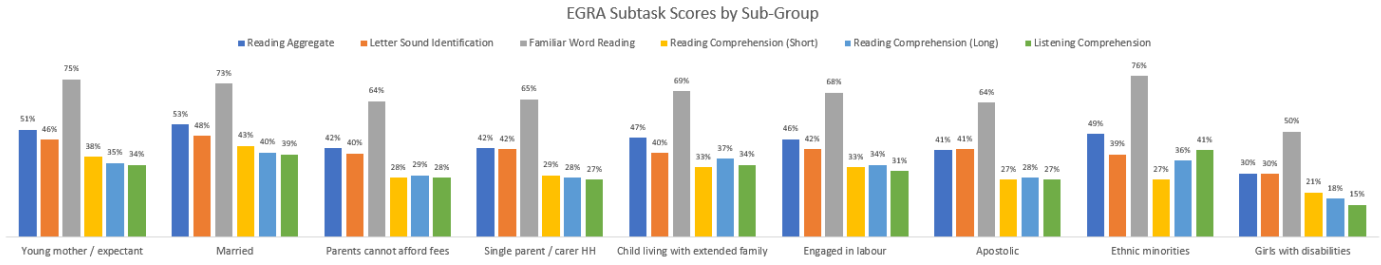
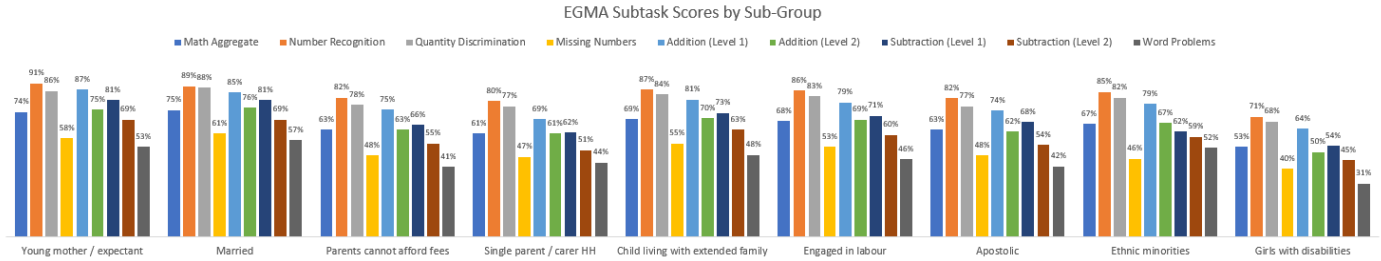


Figure VI



Sub-task scores

- All sub-groups performed best at the familiar word reading and number recognition tasks.
- Young mothers / expectant, married girls and girls who are engaged in labour have significantly higher scores than those who are not in all but one sub-task.
- The learning comprehension score of girls with disabilities was the lowest score obtained for any reading sub-task (15%). The familiar word reading score of girls from ethnic minorities was the highest score obtained for any reading sub-task (76%).
- The word problems score of girls with disabilities was the lowest score obtained for any math sub-task (31%). The number recognition score of young mothers / expectant was the highest score obtained for any math sub-task (91%).

Figure VII

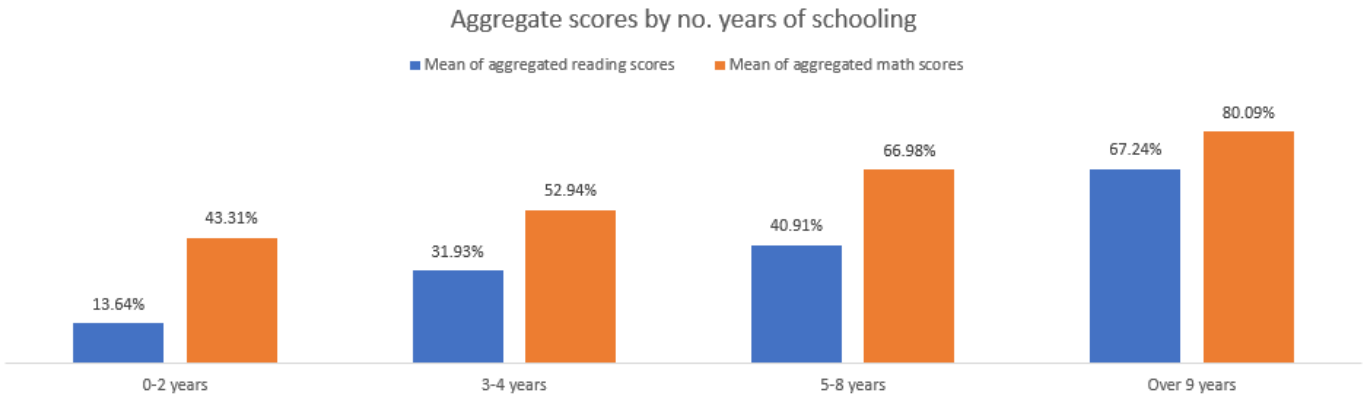
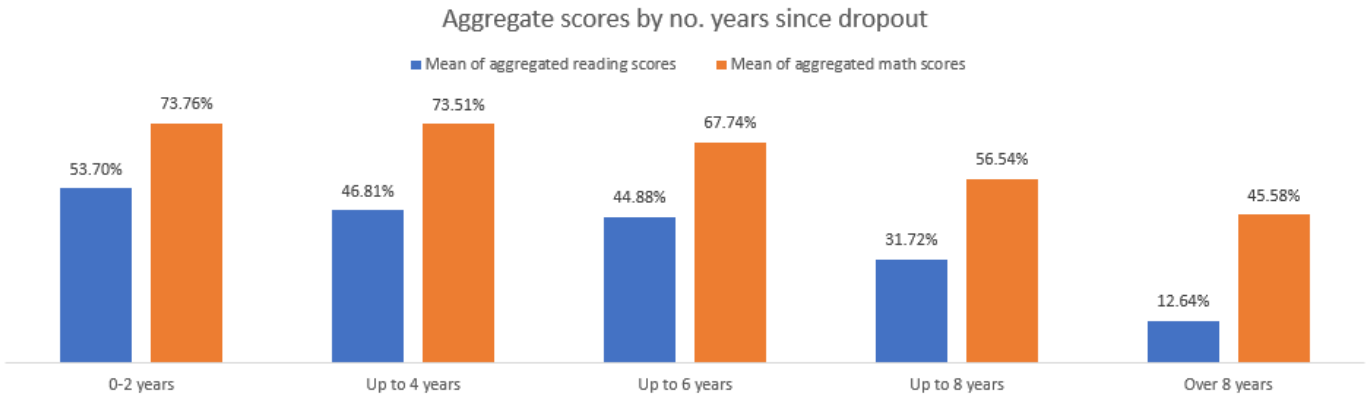


Figure VIII



School experience

- Girls were asked if they had ever been to school, then for those who responded yes, they were asked the number of years of since they've dropped out of school (see Figure VIII above) and the total number of years of schooling they completed (not shown). By age group, about one-in-ten girls (11.3%) ages 10-14 had never been to school and one-quarter (26.9%) of girls ages 15-19 had never been to school. These data must be interpreted with caution since schooling experience information was missing from two-thirds of girls ages 10-14 and one-half of girls ages 15-19.
- Comparing girls by their response on the first question, there is no significant difference on the reading or the math aggregate score between girls who have never been to school (n=121, mean reading score = 46.98%, mean math score = 66.55%) and those who have (n=103, mean reading score=44.50%, mean math score = 68.14%). Of note, girls who were missing information on prior schooling also had comparable scores to those with data (n=234, mean reading score=43.40%, mean math score = 43.40%).. This is an interesting finding which we will be investigating further at the midline stage.
- Among those who said they had been to school, more years of schooling completed was associated with higher reading and math scores. Girls who had completed 5-8 years of schooling had significantly higher reading and math aggregate scores than girls who had 0-2 years of schooling, but there were no significant differences between girls with 3-4 years schooling and those with 5-8 years schooling.
- Girls with 9 or more years of schooling significantly outperformed all other girls.
- Among those who said they had been to school, the trend suggests that the more years that have passed since a girl dropped out the lower the girls' math and reading scores (Figure VIII). However,

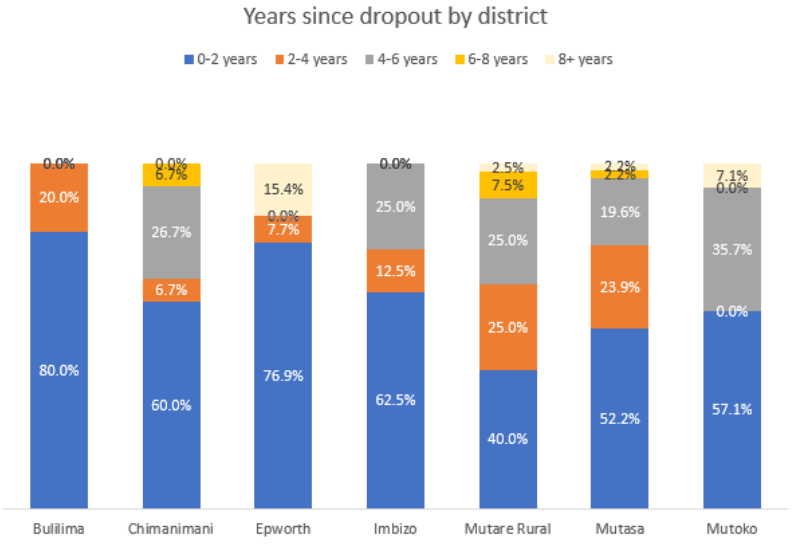
statistically significant differences between groups were only observed with those who most recently dropped out and those who dropped out 8 or more years ago, with girls who had dropped out in the last 2 years had significantly higher reading scores than girls who dropped out 8 years or more ago.

- There were no significant differences in reading aggregate scores between girls who had dropped out between 2-4 years ago, 4-6 years ago or 6-8 years ago.

5) Analysis of dropout by age

- For the purposes of this analysis, the number of years of schooling and number of years since dropout were grouped into 2-year intervals.
- Full data on the dropout stage was available for only 30% of the treatment sample.
- The average number of years since dropout is 2.74 years.
- The sampled girls were most likely to have between 5-8 years of schooling – half of girls aged 15-19 and two-thirds of girls aged 10-14.
- Most sampled girls also seem to have dropped out in the past 2 years – three quarters of girls 10-14 and half of those aged 15-19 are in this category – consistent with beneficiary identification findings. One quarter of girls aged 15-19 dropped out between 4-6 years ago.
- Bulilima and Epworth districts have the highest percentages of girls who have dropped out up to 2 years prior to the baseline, while Epworth and Mutoko have the highest number of girls who have dropped out over 8 years before that (with the caveat that the actual sample numbers involved are quite small).
- The highest actual numbers of girls who dropped out were in Mutasa and Mutare Rural districts.

Figure VIII



6) Barriers reported by girls by sub-group

- On average, significantly more girls who said their parents cannot afford fees (59%) girls engaged in labour (50%) and girls with disabilities (42%) have reported facing barriers to education.
- At the opposite ends of the spectrum, on average, significantly fewer girls from ethnic minorities (5%) married girls (10%) girls from single parent households (10%) and young mothers / expectant (15%) have reported facing barriers to education.
- Having at least one functional difficulty is the most common barrier to education (35% of girls have said they faced this).
- The least common barrier to education is the lack of voice and ability to speak up (reported by 22% of girls) though with great variation among the sub-groups – for instance, 67% of girls whose parents cannot afford fees and the same percentage of Apostolic girls said they face it.
- 62% of girls engaged in labour reported low self-esteem and 54% of them said menstruation is a barrier to education.

- Accessibility is a barrier reported by 28% of girls on average – highest for girls whose parents cannot afford fees (55%) and girls engaged in labour (53%).

Figure IX

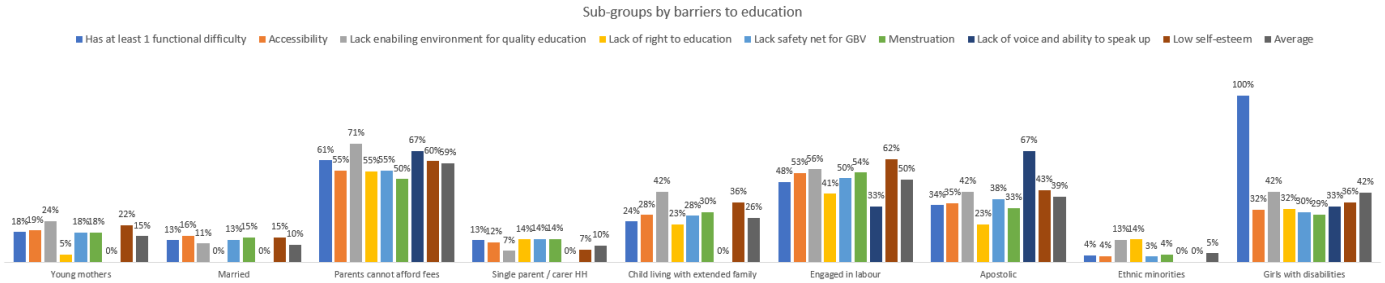
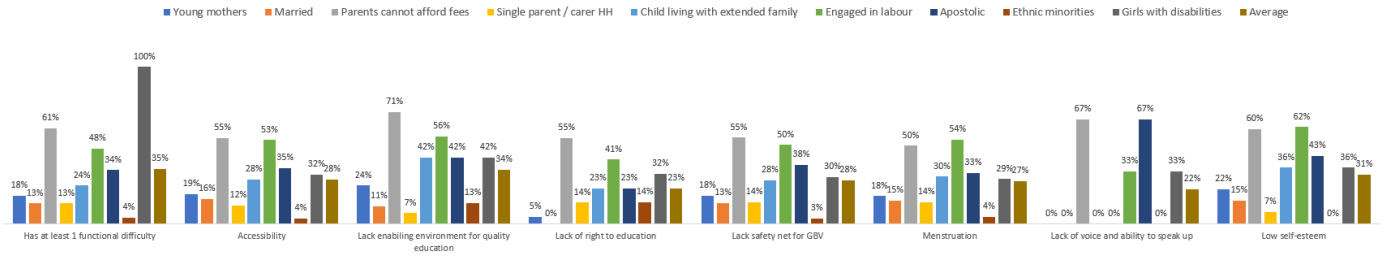


Figure X



7) Intermediate outcome (IO) scores by sub-group

IO2. Adolescent girls have increased self-efficacy and life skills

- Girls who were mothers / expectant, married girls and girls engaged in labour have significantly higher gender norm scores and SRHR knowledge than girls who were not mothers / expectant, unmarried and not engaged in labour respectively.
- Girls whose parents cannot afford fees scored significantly lower on SRHR knowledge than girls who did not indicate this. Girls living with their extended family on the other hand scored significantly higher than those who lived with their parents.
- Girls with disabilities scored significantly lower on self-efficacy, SRHR and gender norms than girls without disabilities.

IO4. Communities demonstrate more positive gender attitudes and actively support and protect girls

- Married girls and girls living with extended family scored significantly lower on the Complete Education Support Index than unmarried girls and those who lived with their parents.
- There were no significant differences between the sub-groups and the rest of the treatment sample in terms of the Perceived Safety Index.

8) Specific findings by sub-group

To understand girls learning journeys by sub-groups further, the specific findings by sub-groups are as follows:

Table III. Specific findings by sub-group

	Learning outcomes	**IO scores**	**Barriers**
Young mothers	• Significantly higher aggregate math scores (74%) and basic addition	• Significantly higher gender norm scores and SRHR knowledge than	• Only 15% reported facing barriers to education

	(87%) scores than girls who have no children (61% and 72% respectively) • The number recognition score of young mothers / expectant was the highest score obtained for any math sub-task (91%)	girls who have no children	
	• Average aggregate math score for full sample: 67% / Sub-group difference: +7% • Average basic addition score for full sample: 77% / Sub-group difference: +10% • Average number recognition score for full sample: 84% / Sub-group difference: +7%		
	Learning outcomes	**IO scores**	**Barriers**
Married girls	• Significantly higher aggregate math scores (75%) and missing numbers scores (61%) than unmarried girls (62% and 47% respectively) • Significantly higher aggregate reading scores (53%) than unmarried girls (41%)	• Higher gender norm scores and SRHR knowledge than unmarried girls	• Only 10% reported facing barriers to education
	• Average aggregate math score for full sample: 67% / Sub-group difference: +8% • Average missing numbers score for full sample: 51% / Sub-group difference: +10% • Average aggregate reading score for full sample: 43% / Sub-group difference: +10%		
Girls from single parent / carer households	N/A	N/A	• Significantly fewer girls (10%) have reported facing barriers to education compared to other sub-groups
Girls with disabilities	• Significantly lower aggregate reading scores (30%) and math scores (53%) than girls without disabilities (51% and 73% respectively) • The only sub-group to rank as 'emergent learners' (1-40%) for aggregate reading scores • The listening comprehension score of girls with disabilities was the lowest score obtained for any reading sub-task (15%) • The word problems score of girls with disabilities was the lowest score obtained for any math sub-task (31%)	• Scored significantly lower on self-efficacy, SRHR and gender norms than girls without disabilities	• Significantly more girls with disabilities (42%) have reported facing barriers to education compared to other sub-groups
	• Average aggregate reading score for full sample: 43% / Sub-group difference: -13%		

	• Average aggregate math score for full sample: 67% / Sub-group difference: -14% • Average listening comprehension score for full sample: 29% / Sub-group difference: -14% • Average word problems score for full sample: 48% / Sub-group difference: -17%		
	Learning outcomes	**IO scores**	**Barriers**
Girls living with extended family	• Obtained the third highest math aggregate score (69%)	• Scored significantly higher on SRHR knowledge and gender norms than those who lived with their parents • Scored significantly lower on the Complete Education Support Index than those who lived with their parents	N/A
	• Average aggregate math score for full sample: 67% / Sub-group difference: +2%		
Girls engaged in labour	• Obtained significantly higher aggregate math scores (68%) than girls who are not engaged in labour (54%) • Scored significantly higher than those who are not engaged in labour in all but one EGMA sub-task (advanced subtraction – 60% vs 47%)	N/A	• 62% reported low self-esteem • 54% and 53% of them said menstruation and accessibility respectively constitute barriers to education
	• Average aggregate math score for full sample: 67% / Sub-group difference: +1% • Average advanced subtraction score for full sample: 60% / Sub-group difference: 0%		
Apostolic girls	• Obtained the second lowest aggregate score for reading (41%) but overall had comparable scores to girls from non-Apostolic on reading, math and life-skills	N/A	• 40% reported facing barriers to education • 66% said they lack a voice and the ability to speak up
	• Average aggregate reading score for full sample: 43% / Sub-group difference: -2%		
Girls from ethnic minorities	• The familiar word reading score of girls from ethnic minorities was the highest score obtained for any reading sub-task (76%)	• There were no statistical differences between the reading, math and life-skills scores of girls from ethnic minorities than those who are not from ethnic minorities. These results may be skewed due to the small n size of these results (only 14 girls were identified in this sub-group) and findings should be interpreted with caution	• Significantly fewer girls from ethnic minorities (5%) have reported facing barriers to education compared to other sub-groups (again with the caveat that the actual sample numbers involved are quite small)
	• Average familiar word score for full sample: 67% / Sub-group difference: +9%		
Girls whose parents cannot afford fees	• Obtained the third lowest aggregate score for reading (42%) but overall had comparable	• Scored significantly lower on SRHR knowledge than girls who did not indicate this	• 60% reported facing barriers to their education

	reading and life-skills scores to girls who did not indicate this.		• 66% said they lack a voice and the ability to speak up • Accessibility is a barrier reported by 55% of girls in this category
	• Average aggregate reading score for full sample: 43% / Sub-group difference: -1%		

Table IV. Learning, life-skills and IO outcomes by sub-group

	No. girls	Aggregate EGRA	Aggregate EGMA	Life-skills mean index score	IO.2.1 (SE[1]) and IO.2.2 (GN, SRHR) mean index scores	IO.4.2 (CS) and IO.4.3 (CES) mean index scores
All girls C2 – Comparison cohort	264	41.82	67.65	27.46	SE: 2.63 GN:0.99 SRHR: 13.81	CS: 1.69 CES: 4.72
All girls C1A – Treatment cohort	459	44.55	66.25	29.22	SE: 2.67 GN:1.05 SRHR: 14.91	CS: 3.56 CES: 7.81
Young mothers	88	50.95**	73.93**	33.95**	SE: 2.66 GN:1.12** SRHR: 19.14**	CS: 3.71 CES: 7.53
Married girls	71	52.86**	74.86**	34.43**	SE: 2.71 GN:1.13** SRHR: 19.40**	CS: 3.75 CES: 7.34
Girls from single parent / carer household	55	41.85	61.40	27.96	SE: 2.68 GN:1.06 SRHR: 15.44	CS: 3.53 CES: 7.90
Girls with disabilities	12	29.75	52.77	23.63	SE: 2.57 GN:0.98 SRHR: 10.38	CS: 3.00 CES: 6.83
Girls living with extended families	130	47.15	69.44**	30.86**	SE: 2.67 GN:1.08 SRHR: 16.37**	CS: 3.58 CES: 7.48
Girls engaged in labour	244	45.55**	67.84**	30.10**	SE: 2.69 GN:1.08 SRHR: 15.58	CS: 3.65 CES: 7.60
Apostolic girls	160	41.16	62.65	29.10	SE: 2.67 GN:1.03 SRHR: 14.98	CS: 3.63 CES: 7.63
Girls from ethnic minorities	14	48.61	66.50	28.96	SE: 2.38 GN:0.99 SRHR: 16.00	CS: 3.50 CES: 8.21

[1] Acronyms used in this table: SE – Self-efficacy; GN – Gender norms; SRHR – Sexual and reproductive health rights; CS – Community safety; CES – Complete education support

Girls whose parents cannot afford fees	244	41.79	62.97**	28.87	SE: 2.66 GN:1.05 SRHR: 14.65	CS: 3.63 CES: 7.66

9) Transition outcome by sub-group

Transition outcomes by sub-group are presented in the updated Supplementary Table 12 below. This shows the percentage of girls from different sub-groups indicating their intentions after completing SAGE.

Supplementary Table 1: Percentage of girls' hopes after completing CBLH, C1A

Group (transition)	No. girls	Formal education	Vocational training	Employment or self-employment	Get married / Other / Don't know / Refused[2]
All girls	399	2.76%	47.62%	47.12%	2.51%
District					
Bulilima	48	8.51%	44.68%	44.68%	2.13%
Chimanimani	64	0.00%	71.11%	28.89%	0.00%
Epworth	51	0.00%	13.95%	86.05%	0.00%
Imbizo	24	0.00%	33.33%	52.38%	14.29%
Mutare Rural	130	4.39%	50.88%	40.35%	4.4%
Mutasa	112	1.98%	51.49%	45.54%	0.99%
Mutoko	30	0.00%	50.00%	50.00%	0.00%
Area					
Urban	3	0.00%	100%	0.00%	0.00%
Peri-urban	69	0.00%	27.54%	68.12%	4.35%
Rural	327	3.36%	51.38%	43.12%	2.14%
Language of instruction of assessment					
Shona	331	2.11%	48.94%	47.13%	1.81%
Ndebele	68	5.88%	41.18%	47.06%	5.88%
Age group					
Age 10–14	101	6.93%	33.66%	55.45%	3.97%
Age 15–19	272	1.10%	54.04%	43.38%	1.47%
Schooling history					
Never been to school	101	5.9%	55.1%	35.64%	9.90%
Dropped out less than 2 years ago	66	0.00%	54.55%	45.45%	0.00%
Dropped out 2.1-4 years ago	21	0.00%	47.62%	52.38%	0.00%
Dropped out 4.1-6 years ago	29	3.40%	44.83%	44.83%	3.40%
Dropped out 6.1-8 years ago	4	0.00%	50.00%	50.00%	0.00%
Dropped out more than 8 years ago	4	0.00%	50.00%	50.00%	0.00%
Sub-groups					
Young mothers	76	1.32%	53.95%	44.74%	0.00%
Married girls	58	1.72%	53.45%	44.83%	0.00%
Girls whose parents cannot afford fees	161	1.86%	50.31%	45.34%	2.48%

[2] The proportion of girls in the treatment cohort by sub-category are: 0.3% said 'Get married and care for my family', 0.5% said 'Other', and 1.8% said 'Don't know'.

Group (transition)	No. girls	Formal education	Vocational training	Employment or self-employment	Get married / Other / Don't know / Refused[2]
Girls from single parent / carer household	16	0.00%	62.50%	37.50%	0.00%
Girls living with extended families	115	0.87%	52.17%	46.09%	0.87%
Girls engaged in labour	174	1.72%	51.15%	46.40%	1.72%
Apostolic girls	114	1.80%	53.51%	42.11%	2.63%
Girls from ethnic minorities	13	7.7%	46.15%	46.15%	0.00%
Barriers					
Girls with at least 1 functional disability	117	2.56%	37.61%	56.41%	3.41%
Accessibility—long distances to school	273	2.20%	50.55%	44.69%	2.56%
Lack safety net for GBV	147	2.04%	49.66%	45.58%	2.72%
Lack of right to an education	22	9.09%	45.45%	40.91%	4.55%
Lack of enabling environment for quality education	43	4.65%	37.21%	51.16%	6.98%
Logistic barriers during menses	196	4.08%	47.96%	43.37%	4.59%
Lack of voice and ability to speak up	79	1.27%	45.57%	48.10%	5.06%

10) Further analysis requested by the project

The programme will be looking to complete the following by the time of the sign-off:

- Separating the girls who have never been to school as a sub-group for further analysis.

11) Further programme actions based on sub-group analysis

Areas of focus in the short and medium term will include:

- Wider consortium and staff consultation on baseline findings and how they will feed into our adaptive management approach.
- Recovering as much of the missing baseline sample data as possible and running further internal analysis to learn more about each sub-group – e.g. trends related to accessibility barriers – complemented by regular monitoring data.
- Supporting girls with disabilities, who obtained the lowest aggregate scores in both reading (24.92%); and math (52.77%).
- Supporting girls whose parents cannot afford fees and Apostolic girls, as two thirds of girls in those sub-groups said they lack a voice and the ability to speak up.