Project Evaluation Report

Report title:	GEC-T STAGES Baseline Report
Evaluator:	School-to-School
GEC Project:	Supporting Transition of Adolescent Girls through Enhancing Systems (STAGES)
Country	Ethiopia
GEC window	GEC-Transition
Evaluation point:	Baseline
Report date:	May 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

Girls' Education Challenge

UKaid from the British people

GEC-T STAGES Baseline Report

May 2019

Project Number: 6473

Project Name: Supporting Transition of Adolescent Girls through Enhancing Systems

Implementing Agency: Link Community Development

Author: Hetal Thukral, Casey McHugh, and Randy Tarnowski

External Evaluator Institution: School-to-School International

Version Number: Final

Date: May 31, 2019

Acknowledgments

The School-to-School International team would like to thank the PricewaterhouseCoopers LLP and Girls' Education Challenge-Transition team for their support and guidance for the Supporting Transition of Adolescent Girls through Enhancing Systems baseline study. We would also like to recognize the strong and indispensable engagement of the Ministry of Education's Wolaita Zone and woreda-level education officials for serving as true partners and leaders throughout the baseline—including the learning assessment adaption workshop and piloting through baseline data collection. In addition, woreda-level education supervisors and experts played a crucial role in the data collection process. We would also like to acknowledge our appreciation for the woreda-level gender officers and female teachers who played a vital role in supporting and participating in the qualitative data collection as facilitators and notetakers. The research would not have been possible without the participation of all the students, teachers, school directors, and parents who participated in the assessments, surveys, and focus group discussions. Finally, this endeavor would not have been possible without the dedication and support of the Link Community Development International and Ethiopia teams who lent their continuous, thoughtful support throughout the baseline design, implementation, and analysis.

The following persons played a key role in the process:

Sandra Graham	Link Community Development International
Fiona Greig	Link Community Development International
Samantha Ross	Link Community Development International
Chris Martin	Link Community Development International
Taddesse Gemmeda	Link Community Development Ethiopia
Dagnachew Melese	Link Community Development Ethiopia
Wegayehu Asha	Link Community Development Ethiopia
Bethelhem Dereje	Link Community Development Ethiopia
Getahun Tsegaye	Link Community Development Ethiopia
Teferi Womber	Link Community Development Ethiopia
Ashley Doria	School-to-School International
Zewdu Gebrekidan	School-to-School International
Casey McHugh	School-to-School International
Alemneh Tadele	School-to-School International
Hetal Thukral	School-to-School International
Laura Zasoski	School-to-School International

List of Acronyms

BOWCYA	Bureau of Women, Children, and Youth Affairs
ChSA	Charities and Societies Agency
CP	Child Protection
CPP	Child Protection Policy
DFID	United Kingdom's Department for International Development
E.C.	Ethiopian Calendar
EGMA	Early Grade Mathematics Assessment
EGRA	Early Grade Reading Assessment
EMIS	Education Management Information System
FGD	Focus Group Discussion
FOI	Fidelity of Implementation
GAP	Gender Audit and Action Planning
GEAC	Girls' Education and Advisory Committees
GEC	Girls' Education Challenge
GEC1	Girls' Education Challenge 1
GEC-T	Girls' Education Challenge-Transition
GESI	Gender and Social Inclusion
GRP	Gender-responsive Pedagogy
IO	Intermediate Outcome
KETB	Kebele Education Training Boards
KII	Key Informant Interview
Link	Link Community Development
M&E	Monitoring and Evaluation
MEL	Monitoring, Evaluation, and Learning
MOE	Ministry of Education
MOI	Medium of Instruction
PTSA	Parent-teacher Student Association
RAM	Review and Adaptation Meeting
REB	Regional Education Bureau
SeGMA	Secondary Grade Mathematics Assessment
SeGRA	Secondary Grade Reading Assessment
SEL	Social-emotional Learning
SIC	Schools Improvement Committee
SLSP	School-level Survey Package
SNNPR	Southern Nations Nationalities and Peoples Region
SPAM	School Performance Appraisal Meetings
SRGBV	School-related Gender-based Violence
STAGES	Supporting Transition of Adolescent Girls Through Enhancing Systems
STS	School-to-School International
SVF	School Visit Form
TVET	Technical and Vocational Education and Training
USAID	United States Agency for International Development
WEO	Woreda Education Official
WG	Washington Group
WGQ	Washington Group Question
WLSP	Woreda-level Survey Package
WPM	Words Per Minute

Table of Contents

List of Figures

List of Tables

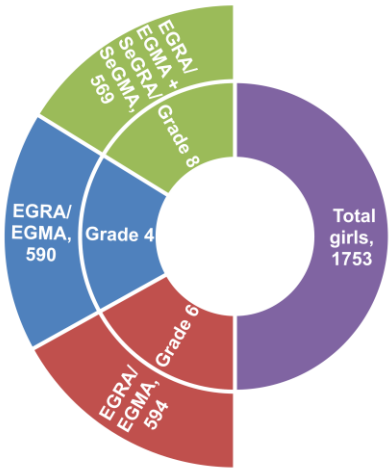
List of Supplementary Tables

Executive Summary

Background

The Supporting Transition of Adolescent Girls through Enhancing Systems (STAGES) project is part of the United Kingdom's Department for International Development's (DFID) Girls' Education Challenge-Transition (GEC-T) and is being implemented by Link Community Development (Link) in Ethiopia's Wolaita Zone of the Southern Nations Nationalities and Peoples Region (SNNPR). The project's interventions assume that improved attendance, quality in teaching, school management and governance, embedded positive community support for girls' education, and support for the well-being and self-esteem of marginalized girls are prerequisites for better learning, transition, and sustainability outcomes for these students. Activities in support of these goals are being implemented in primary schools and secondary schools in four districts in the Wolaita Zone.

Figure 1. Baseline Sample



Research Design

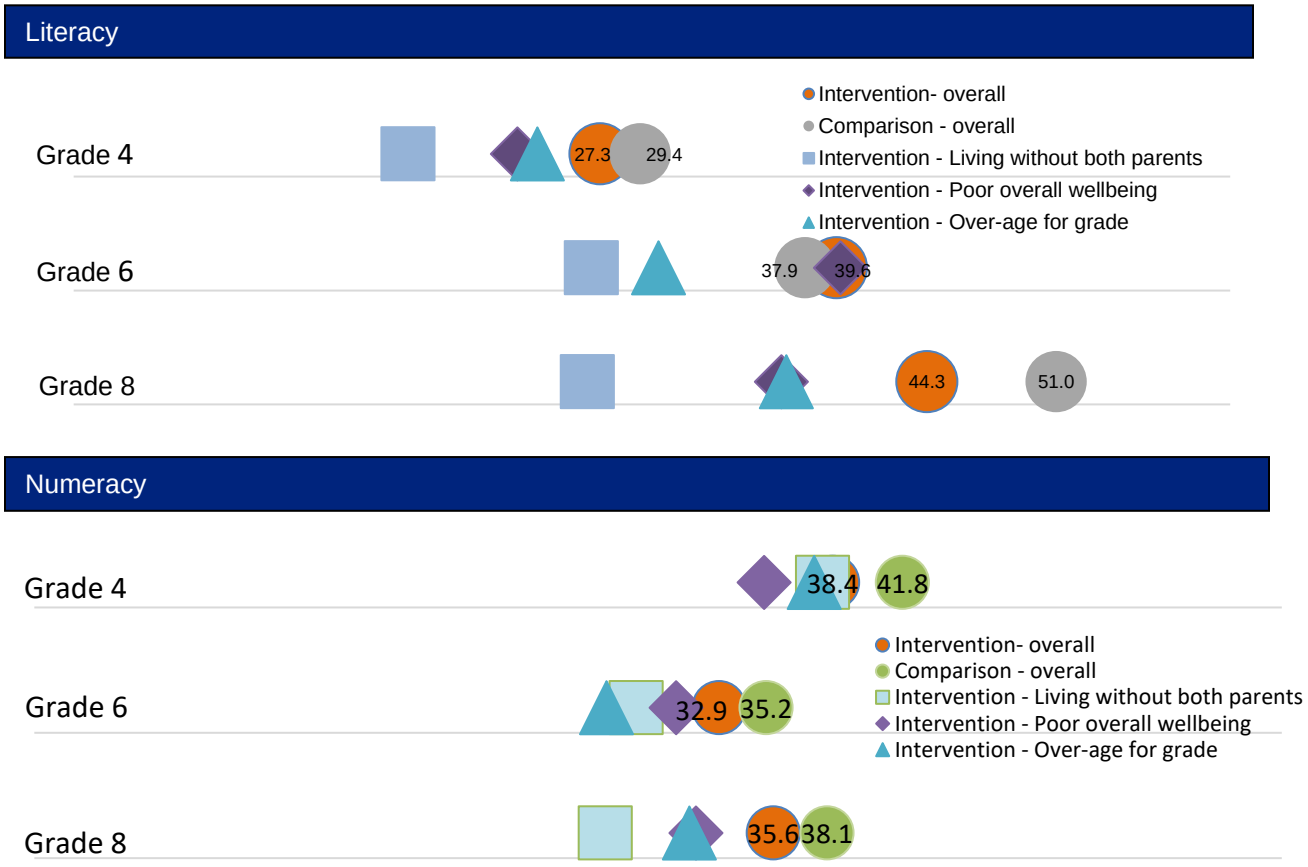
The evaluation for the STAGES project is a mixed-methods design that examines difference-in-differences between girls in treatment schools in four woredas compared to girls in comparison schools in one woreda.[1] It follows three cohorts of girls across six years of the project—from 2018 to 2024—utilizing existing government personnel at every stage of data collection and is conducted in a two-phase approach.[2] The baseline study results reported here include student, teacher and school director's data collected during phase one during the end of the school year in grades 4, 6, and 8 in primary schools only; as well as transition, parent and secondary teacher's data collected during phase one at the beginning of the school year in grades 5, 7, and 9.

Learning Outcome Findings

To assess literacy and numeracy ability, results from multiple assessments were compiled to create aggregate scores. Overall female students performed comparable across the treatment and comparison groups. However, grade 8 girls in the comparison group outperformed girls in the treatment group in literacy, and grade 4 girls in the comparison group outperformed girls in the treatment group on numeracy. Additionally, living without both parents, having a poor overall well-being, and being overage for their grade, were all identified as barriers that correlated with lower scores. This impact was most pronounced on literacy assessments and among girls in higher grades.

[1] The true comparability of the treatment and comparison woredas is unclear based on an examination of characteristics and barriers faced by girls in these two groups.
[2] The first phase of data collection, analysis, and reporting– focused on Outcome 1: Learning and Outcome 3: Sustainability– was completed in spring 2018. Data for Outcome 2: Transition, household survey data, and initial results for secondary schools were collected in December 2018 and results reported in May 2019.

Figure 2: Aggregate Scores for Intervention and Comparison Girls and Three Intervention Subgroups



Note: Aggregate scores are only shown for treatment group (overall) and comparison group (overall).

In the assessments of girls' literacy levels in Wolayttatto, girls in grades 4 and 6 in treatment schools struggled the most with fluency and reading comprehension. When considering their literacy in English, girls in grades 4, 6, and 8 struggled with all foundational skills, and many struggled with reading comprehension.[3] Girls in grade 8, also struggled with written tasks. In numeracy assessments, girls in grades 4 and 6 at treatment schools struggled with foundational skills.[4] Girls in grade 8 struggled with subtraction and written tasks.

Figure 3: Proficiency by Grade and Language[5]

	Literacy							
	SeGRA							
	Letter sound identification	Familiar word	Invented word	Oral reading fluency	Reading comprehension	Reading passage	Fill in the blank	Revising sentences
Grade 4—Wolayttatto	41.52%	17.65%	17.65%	1.04%	4.84%			
Grade 6—Wolayttatto	40.43%	23.47%	17.33%	7.22%	13.36%			

[3] The early grade reading assessment (EGRA) subtasks that capture foundational skills in literacy include letter sound identification, familiar word, invented word, oral reading fluency and reading comprehension.
[4] The early grade mathematics assessment (EGMA) subtasks that capture foundational skills in numeracy include number identification, quantity discrimination, missing number, addition, subtraction and word problems.
[5] Colors in the graphic show relative proportions of girls with proficiency in that subtask and assessments. Red colors indicate relatively lower proportions of girls with proficiency; green indicates relatively higher proportions with proficiency.

Grade 4—English	15.57%	4.15%	21.45%	5.54%	0.35%			
Grade 6—English	22.74%	14.08%	29.24%	23.10%	2.53%			
Grade 8—English		28.41%	36.74	32.58%	7.95%	19.39%	0.38%	0.76%

	Numeracy								
	Number Identification	Quantity Discrimination	Missing Number	Addition	Subtraction	Word Problems	SeGMA		
							Geometry	Fractions	Multiplication
Grade 4	10.03%	7.27%	0.69%	3.11%	0.69%				
Grade 6	11.91%	9.39%	3.25%	3.25%	1.08%	5.78%			
Grade 8				7.95%	0.38%	11.36%	13.41%	3.83%	13.41%

Transition Outcome Findings

As noted previously, the external evaluators implemented a two-phased approach to collect baseline data for STAGES. The first phase of the baseline focused on Outcome 1 Learning and Outcome 3 Sustainability and was conducted in April 2018 with cohort girls in grades 4, 6, and 8. In order to gather the most accurate data for Outcome 2 Transition, the second phase was conducted in December 2018 when cohort girls were expected to have transitioned into grades 5, 7, and 9. In other words, a successful transition was defined as the movement into the next grade level in the following academic cycle. This may differ from other projects since the transition data were collected in the academic cycle following when the baseline was completed. Transition data was collected in interventions schools only; no data were collected in comparison schools during phase two.

Transition rates tended to be higher in lower grades. Overall, the transition rate across grades in the sample was almost seven in ten girls. The attrition rate of over 30 percent exceeds the attrition rate assumed in the sample—attrition of 30 percent was assumed across a two-year period.

Figure 4: Transition Rates at Key Transition Grades by Cohort

	Total Sample	Transition Rates at Key Transition Grades					
		In-school progression	Retained in Same Grade	Moves into Secondary	Drops out of school	Lost from sample	Overall successful transition rate
Grade 4 into Grade 5	297	218	33	n/a	28	18	73.40%
Grade 6 into Grade 7	300	204	15	n/a	34	47	68.00%
Grade 8 into Grade 9	294	n/a	13	203	10	68	69.05%
Overall	891	Average transition rate for all 3 key transition grades					66.44%

Since data were collected in two phases in the STAGES baseline, transition rates represent the actual transition of girls from the phase one grade level to the phase two grade level, a single grade level progression. Using these data, baseline transition rates were examined for girls in grades 4, 6, and 8

moving into grades 5, 7, and 9, respectively. Baseline transition rates were lower for girls in higher grades than for girls in lower grades. By subgroup, transition rates for the following groups were lower than the overall average for their grade:

- Grade 4 girls who reported low overall well-being;
- Grade 6 girls who reported they were living without both parents; and
- Grade 4 and grade 8 girls who reported their teacher was often absent.

Sustainability Outcome Findings

Activities and outcomes under the STAGES project are geared towards embedding respect and support for girls' education within the target communities and the decentralized education system. Moreover, localized activities like School Performance Appraisal Meetings (SPAMs) and Gender Action Plans (GAPs) will be unique to each school, offering a bottom-up approach to ensure support for girls' education and maximum buy-in from local stakeholders. Considering these points, indicators for sustainability were selected to cover a wide range of domains—attitudes, support, engagement, and pursuit—and across interventions. As with Link's Girls' Education Challenge 1 project, it is expected that different communities will utilize the interventions in subtly different ways according to their need. To adequately capture this, future iterations of the sustainability outcome will seek to highlight emergent indicators of sustainability. At baseline, the sustainability score assigned to STAGES activities was 1.7 on a four-point scale.

Marginalization Analysis and Gender Analysis

The following barriers were identified to be prevalent among the girls in the treatment schools.

Individual-level

Girls are faced with significant challenges at the individual level, including a high burden of household chores with the burden often increasing with a girl's age. Girls' participation in income-generating activity has a negative impact on access to education and school attendance, especially in families with high levels of poverty. Additionally, girls often face early marriage—and its association with high levels of school drop-out—as well as social and cultural norms rooted in gender inequality, including prioritization of boys' education, absenteeism related to menstruation, and lack of support for girls with disabilities. Distance to secondary schools and migration were also noted as impeding girls'—and boys'—ability to transition to and access secondary school. Areas requiring further examination include school-related, gender-based violence (SRGBV)—including corporal punishment—and abduction.

Teacher-level

Several barriers at the teacher-level also emerged. Teacher attendance appears to be a challenge, with slightly more than half of the girls in both the comparison and treatment groups reporting that their teachers are often absent from class. Teachers' reports of frequent disciplinary actions also appear to be high overall but lower in treatment schools than in comparison schools. Also of note, less positive attitudes towards girls' education appear to be a school-level issue rather than a teacher-level issue.

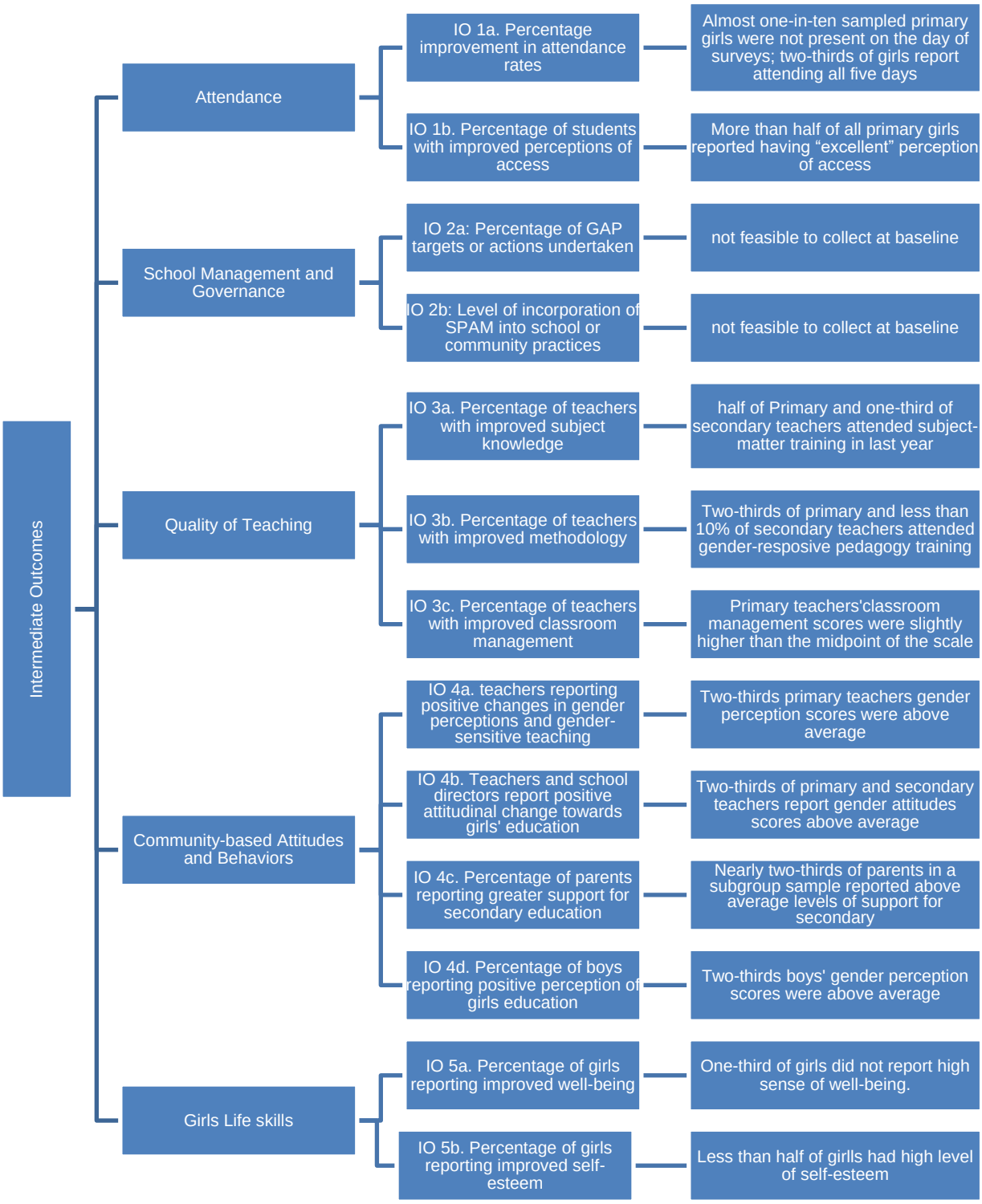
Home- and community-level

When looking at the home level, the baseline analysis indicated that when girls do not have high levels of household support, they have lower scores on student- and school-level outcomes of interest. Overall, gender perceptions were inconsistent across groups; one in three woreda staff members demonstrated low attitudes towards girls' education, which indicates that there continues to be room for growth in this area.

Intermediate Outcomes Findings

Intermediate outcomes were gathered at baseline to track changes in these areas over time. Their relationship to learning, transition, and sustainability outcomes will provide a complete picture of shifts in the landscape for girls in the Wolaita Zone.

Figure 5: Intermediate Outcomes Summary Findings



1. Background to project

1.1 Project context

The Supporting Transition of Adolescent Girls through Enhancing Systems (STAGES) project is part of the United Kingdom's Department for International Development's (DFID) Girls' Education Challenge-Transition (GEC-T) and is being implemented by Link Community Development (Link) in Ethiopia's Wolaita Zone of the Southern Nations Nationalities and Peoples Region (SNNPR). The purpose of the STAGES project is to transform access to secondary education in Ethiopia's Wolaita Zone, establish locally owned systems to improve the quality of schooling over eight years, and create sustainable support for girls long-term. By concentrating the project's implementation in a single zone, the overall geographic context the project works within does not vary. Within the Wolaita Zone, the project has identified four districts, or *woredas*, all of which are remote, densely populated, subsistence farming communities of the Wolaita ethnic group with high absolute poverty levels.

Moreover, some variations in gender equity indicators between the target woredas should also be noted. For example, in most indicators, Damot Pulasa consistently showed lower gender equity measures.

Supplementary Table 1. Woreda Gender Equity Measures, 2008 E.C. (2015–16)

Woreda	Gender-parity index		Percentage of female students		Percentage of female teachers	
	Primary (1–8)	Secondary (9–10)	Primary (1–8)	Secondary (9–10)	Primary (1–8)	Secondary (9–10)
Damot Pulasa	0.84	0.64	46.1	40.2	21.9	2.4
Damot Sore	0.93	0.83	48.2	46.1	31.4	19.4
Damot Woide	0.91	0.72	47.4	46.7	24.8	17.0
Kindo Koisha	0.89	0.89	47.0	46.5	21.5	24.5
Ofa (Comparison)[6]	0.93	0.89	47.8	49.2	24.9	19.6

Source: Wolaita Education Sector, Education Management Information System (EMIS), Education Statistics Annual Abstract 2008 E.C. (2015/2016), (2017)

The STAGES project's design documents indicate the main contextual factors influencing the project relate to multiple complex and intersecting barriers that create obstacles to girls' education in the Wolaita Zone.[7] STAGES' monitoring, evaluation, and learning (MEL) framework notes the project operates within an environment with unsupportive cultural and societal norms. There, girls are marginalized economically, socially, and culturally; they face high drop-out rates related to poverty, school distance, and early marriage; low levels of adult education and literacy are also present.[8] Girls primary completion rates in the

[6] While four target woredas of the STAGES project receive interventions at the school level, a neighboring woreda—Ofa—is also be examined for comparison as part of the quasi-experimental design of the evaluation and included in this table as background context. More in-depth detail about the design methodology, including the use of Ofa as a comparison woreda, will be presented later in this report.

[7] Design documents include the STAGES theory of change; the STAGES proposal submitted to the GEC in September 2016; and the STAGES MEL framework submitted in October 2017.

[8] Tesfaye Semela et al., eds., *Impacts of women development and change packages on the socio- economic and political status of women in SNNPR: Promise, Success and Challenges* (Hawassa: Center for Policy and Development Research, Hawassa University, 2015).

Wolaita Zone also continue to lag behind boys—with a grade 5 completion rate of 65.2 percent for female students as compared to 75.4 percent for males, and grade 8 completion rate of 45.9 percent for female students versus 54.5 percent for their male counterparts.[9] Research suggests the Wolaita school system has limited capacity to support girls in their learning given the poor quality of education delivery, gender disparities in enrollment and performance outcomes, an acute lack of secondary schools in rural areas, and few female role models and teachers.[10] In addition, Link has documented how schools struggle to provide conducive learning environments with overcrowded classrooms and limited school infrastructure—especially at the secondary level—as well as challenges related to the medium of instruction (MOI) and teachers' poor English-language competencies.[11] The STAGES project's design documents also note that girls and families often struggle or are unable to meet the hidden costs of education. Moreover, supporting girls through upper primary and secondary school is viewed as financially impossible for most families and perceived as a "wasted investment." These barriers are further exacerbated as girls move through the education cycle to upper primary and secondary levels—and is compounded as girls increasingly face the challenges of adolescence such as low aspirations, limited female role models, and harmful traditional practices related to early marriage and childbearing.

However, it should be noted that despite these challenges, the overall policy framework the STAGES project is working within in Ethiopia is strong. There is a solid foundation of government policies and frameworks clearly identifying challenges in attaining gender equality and parity in the education sector at the primary and secondary level as well as strategies and mechanisms identified by the Ethiopian Ministry of Education (MOE) to combat these issues. This includes the MOE's 2010 *National Girls' Education Strategy* which "focuses on the current status of girls' participation in education, obstacles to the education of girls, real change processes, analyzing lessons of what constitutes good practice for girls' education, and initiates strategic directions to ensure girls enrollment and achievement in education."[12] In addition, national policies are contextualized for the SNNPR by regional education bureaus (REBs). The MOE also produced the *Gender Strategy for the Education and Training Sector* in October 2014, which provides a "working roadmap for stakeholders in a bid to ensure gender equality at all levels of the education and training sector."[13]

1.2 Project Theory of Change and assumptions

The STAGES project's theory of change links project activities, outputs, and outcomes needed to address the multiple and intersecting barriers that currently prevent girls in the Wolaita Zone from progressing through and completing their education. Strong project performance in Link's Girls' Education Challenge 1 (GEC1) intervention and additional GEC-T research revealed five different domains as prerequisites for regular school attendance, retention, and learning of adolescent girls:[14]

- Improved leadership for girls' learning at all levels
- Improved quality of learning for students
- Improved access to secondary schools, specifically in extreme and remote areas

[9] Wolaita Education Sector, EMIS, *Education Statistics Annual Abstract 2008 E.C. (2015/2016),* (2017).
[10] The average proportion of female teachers in the Wolaita Zone is 28.2 percent at the primary level and drops down to 17.2 percent for secondary level grade 9 and 10. Wolaita Education Sector, EMIS, *Education Statistics Annual Abstract 2008 E.C. (2015/2016),* (2017)
[11] The official medium, or language, of instruction is Wolayttatto in grades 1 through 4; English is taught as a separate subject. The official transition of MOI from the mother tongue language (Wolayttatto) to English takes place in grade 5. However, as noted, in the Language and Literacy Baseline Report 2015—prepared for Link during GEC1 by their external M&E team—many teachers lack the English language competency skills to provide quality instruction in English in the upper primary level.
[12] The Federal Democratic Republic of Ethiopia, MOE, *National Girls' Education Strategy* (Addis Ababa: August 2010), http://info.moe.gov.et/gendocs/MOEGE.pdf.
[13] The Federal Democratic Republic of Ethiopia, MOE, *Gender Strategy for the Education and Training Sector*, (s.l.: October 2014), http://www.moe.gov.et/documents/20182/36315/GENDER+STRATEGY.pdf/b9e68a15-bc9e-4930-a5d2-1c1981ca264c.
[14] Also known as "Improved Girls' Learning in Rural Wolaita," Link most commonly refers to the previous iteration of this project in the Wolaita Zone as GEC1. The DFID supported intervention was implemented by Link from March 2013 to March 2017. STAGES is a follow-on to the GEC1 project.

- Direct inputs to provide an environment where girls can be “ready to learn”
- More support from girls’ parents, boys, and communities

The project’s interventions assume that improved attendance, quality in teaching, teacher supervision and support, school management and governance, embedded positive community support for girls’ education, and support for the well-being and self-esteem of marginalized girls are prerequisites for better learning, transition, and sustainability outcomes for these girls. Activities in support of these goals for marginalized girls will be implemented at the following levels: both primary- **and** secondary-school level, secondary-school level only, or system level. These activities are described further in the subsequent sections.

Since the original submission of the baseline report in the summer of 2018, Link has gone through an intensive review of the STAGES Theory of Change and associated interventions with STAGES new project leadership and staff in Ethiopia.[15] In turn, based on this review—as well as learnings from the baseline—in April 2019 Link updated the STAGES the Theory of Change and associated interventions to more accurately capture the project’s current approach, planned interventions, and work-plan.[16] The revised and latest Theory of Change is provided in Annex 22.

Primary- and Secondary-School Level Interventions

This section has been revised to reflect STAGES planned intervention and approach as of April 2019. At both the primary- and secondary-school level, direct inputs will seek to establish an environment where **girls are ready to learn.** These include cost-effective tutorial programs, which will provide improved opportunities for girls to catch-up on missed learning. Tutorial sessions will focus on literacy, numeracy, and science classes during critical years of grade 4 and upwards. Other inputs include provisions of sanitary packs, social-emotional learning (SEL) support via guidance and counseling, awards for high-performing girls, mentoring by female role models, life-skills, financial literacy, and careers advice for secondary girls. Both particularly marginalized primary and secondary girls will benefit from support to their basic needs to attend school, including the provision of scholastic materials, bursaries, and school uniform.

In order to **foster mobilized, gender- and inclusion-aware communities**, STAGES will support Girls’ Education and Advisory Committees (GEACs) based out of all project schools. GEACs are actively engaged in the promotion of gender equality and girls’ education at the school level, including gender club guidelines and operational plans. GEACs will also participate in awareness-raising campaigns, sharing materials surrounding gender disparities in schools, promoting female role models, and addressing negative social and cultural practices that mitigate female students’ attendance and performance. STAGES will also build the capacity of the formal existing structures in Ethiopia which link community and school, the parent-teacher associations (PTSAs) and School Improvement Committees (SICs), to support girls’ attendance, transition and learning, work with traditional and religious leaders, and strengthen the relatively new Mothers’ Group structure in all project schools. Fathers’ Groups, Gender Clubs, Good Brother Awards, community campaigns, and local radio advocacy are also integral to STAGES community engagement strategy. The participation of communities in School Performance Appraisal Meetings (SPAMs) will improve accountability for girls’ learning and transition.

To **improve the quality of learning** from which both girls and boys at primary and secondary level will benefit, Link will work in partnership with the woreda, Zone and Region to strengthen the capacity of teachers in gender and inclusion responsive pedagogy, teaching methodologies, safeguarding and protection, and positive classroom management. With more gender and inclusion responsive teaching in the classroom, it is more likely that the needs of all of the children in the class will be met, and that girls as

[15] A new Programme Director, Programme Manager, and several other key positions on the Link-E team have only recently come on board as of January 2019. While shifts in key personnel, including ensuring meaningful and in-depth engagement and understanding of the baseline report as part of new staff’s onboarding to STAGES contributed to delays in the submission of the revised baseline report—this was critical to laying the foundation for a solid understanding and commitment to the interventions and MEL approach for the remainder of the STAGES program.

[16] Link’s updated Theory of Change can be found in Annex 22.

well as boys will be present in the classroom, participate in lessons, and achieve better learning outcomes. Strengthening training on differentiated teaching methodologies will help teachers to understand and identify children who may be struggling and find ways to respond to their individual needs. Adaptions that teachers make to help children who are struggling in class, often also help all of the children in the class (e.g. using a variety of active teaching methods to reach all children; letting children 'buddy' with each other for peer support; using locally made teaching aids and resources in a fun way that all children can access; considering seating arrangements in the classroom; using the blackboard effectively).

Teachers will also receive support to teach literacy and numeracy, as the subjects that enable children to access all other core subjects, and as core Outcome 1 of STAGES "improved learning outcomes in literacy and numeracy." As children's foundational lessons (grade 1 to 4) are taught in Wolayttatto with the switch to English currently in grade 5, both Wolayttatto and English teachers will receive training in language competency and related gender and inclusion responsive teaching methodology.

Link will also work with the relevant government structure (quality assurance) to provide capacity development for Cluster Supervisors and Woreda Experts at woreda level. These officers are responsible for teacher supervision, monitoring, mentoring and coaching, and their support to teachers is critical to embed improved and gender-responsive teaching practice in schools (see also below under improved leadership). Support will include a review of and strengthening of classroom observation monitoring instruments for gender, inclusion, and safeguarding.

Also, at the primary and secondary school level, the STAGES project will work to **improve leadership for girls' learning at the school, woreda, zone, and regional level**. Interventions to improve leadership for girls' learning includes school leadership training (focus on instructional leadership) for both School Directors and Deputy Directors, as well as capacity development of Cluster Supervisors and Woreda experts to supervise and support teachers as highlighted above.

Whilst the training of Kebele Education and Training Boards, PTSAs and School Improvement Committees feature under Output 5 (community), they could equally appear under Output 1 as they contribute to the overall management and governance of the school including through their role in developing school improvement plans and making decisions on how school/community resources should be best utilized. More gender, inclusion, and safeguarding responsive leadership at woreda and school level will lead to more relevant and responsive school improvement planning and resourcing.

Training on school-related gender-based violence at all levels, as well as strengthening mechanisms for reporting of abuse, bullying, and harassment in schools, and improved case-management will create a safer school environment for girls (and boys) in which efforts are made to prevent abuse, and respond appropriately when it does happen. Preventative action, reporting, and response mechanisms for violence and abuse that happens in and around schools cut across all STAGES interventions. Training at all levels on gender mainstreaming will help to embed the cross-cutting nature of gender, inclusion, and safeguarding.

STAGES provides an amount of resources towards regional sustainability activities, including 'rollout' of selected trainings/activities beyond the 4 operational woredas of Wolaita Zone. The amount provided will support Training of Trainers or Orientation interventions at the level of the Zonal/Regional Bureau, who can then cascade the training as they see relevant to other zones and woredas. While the current design of 'rollout' activities is selective in the trainings to be rolled out, flexibility will be provided for regional and zonal partners to select the trainings that they identify with being the most valuable and impactful towards achieving their 'education for all' agenda.

Secondary-School Level Interventions

The barriers girls face in rural Wolaita are heightened during adolescence—the developmental phase during which social, physical, and educational changes accelerate and intensify. The STAGES project offers a strong opportunity to ensure that girls' learning opportunities are not cut short, and that

adolescent girls are given the investment they deserve. To address adolescence challenges, certain STAGES activities will occur exclusively at the secondary school level. These include the provision of life skills, financial literacy, and career advice to secondary girls and the construction of four new low-cost secondary schools in 'black hole' rural areas to allow for better access to secondary education in extreme and remote areas.

Bursary provision to ensure that all **girls are ready to learn**, for extremely vulnerable girls, initially only for secondary school girls will be extended to primary school girls, also identified to be extremely vulnerable and otherwise likely not to attend (girls with disabilities, girls who are orphaned, girls who are young mothers). This will not represent a huge shift in resources towards direct support but respond to the same need at the primary school level if primary school girls are to make the transition. Bursaries are expected to provide items such as uniform, school registration fees, textbooks, and in some cases, house rent where children have to travel long distances to school and cannot return daily to their homes.

The new secondary schools constructed will be closer to primary schools and will be staffed, maintained, managed, and monitored by the MOE. Secondary school construction will also aid in reducing the ratio of secondary to primary schools down to one secondary school per 7.7 primary schools in the four target woredas—as compared to the current rate of one secondary school to 10.8 primary schools.[17]

System-Level Interventions

The STAGES project's inputs at the school and community level will benefit from and inform inputs occurring on the system level. By partnering with woreda education officers and the REB to co-implement project activities, Link seeks to embed respect and support for girls' education sustainably within target communities and throughout the decentralized education system. This engagement with system level actors will take two forms: first, as a capacity building exercise using woreda staff in data collection and as a thought partner in implementation; second through capacity development on monitoring and supervision of teachers as described above under quality of learning, and third by extending school and community level inputs to the system level. For instance, gender mainstreaming training will support **improved leadership at the school, woreda, zone, and regional level** and involve woreda education officers as well as school directors and department heads in the training.

The MEL framework outlines the MEL approach for the STAGES project to examine the extent that proposed activities support the intended outcomes. The MEL framework recognizes and utilizes the collaborative relationship Link has built with woreda government staff; this relationship has been critical to both GEC1's success and sustainability of results. Furthermore, the MEL framework was designed with the aim that all data collected is purposeful and will provide relevant and useful information; data collection methods and administration are feasible and do not create an undue burden on the project; and findings and results can provide meaningful insights and learnings for the STAGES project, woredas, and Wolaita Zone MOE officials, GEC-T, and DFID.

Project to complete

Outline the project design and interventions. Complete the following table.

This table has been revised from the original baseline submission to reflect updates made to STAGES planned intervention and approach as of April 2019.

[17] Within the Wolaita Zone, Link's noted a secondary school to primary school ratio of 1:8 and 1:13 nationally.

Table 1. Project Design and Intervention

Intervention types	What is the intervention?	What Intermediate Outcome (IO) will the intervention will contribute to and how?	How will the intervention contribute to achieving the learning, transition, and sustainability outcomes?
Output 1: Improved leadership for girls' learning at school, woreda, zone, and regional level			
Capacity Building	School Leadership Training	IO 3: School Management and Governance	Improved capacity of School Directors and Deputy Directors on instructional leadership will greatly strengthen the quality of teaching and learning in the school. It will also strengthen the contribution that the PTSA and SIC make in strengthening the links between the school and community on gender, inclusion, and safeguarding, and school improvement planning and resourcing.
Capacity Building	Training for Cluster Supervisors and Woreda Experts on monitoring, supervision and mentoring of teachers in schools	IO 3: School management and governance	Improved capacity of Cluster Supervisors and Woreda Experts to supervise, monitor and mentor teaching in schools following training will greatly enhance the quality of teaching in schools and add to the sustainability of school leadership and teaching quality interventions.
Capacity Building	Training in gender-mainstreaming	IO 3: School management and governance	Improved capacity at regional, zonal and woreda level to look at education programs, interventions, and activities through a gender lens and address the specific barriers that girls face will help improve girls learning and transition, and ultimately contribute to the sustainability of the STAGES GEC model. Gender mainstreaming is a policy requirement in Ethiopia. Mainstreaming the use of a gender lens at all stages of education service delivery (planning, implementation, monitoring, and evaluation) will ensure that gender is considered holistically at all stages of programming and that it is more likely to be sustained beyond STAGES.
Capacity Building; Community initiatives	Gender Audit and action planning (GAP)	IO 3: School management and governance	Schools to identify how schools are doing on the mainstreaming of gender—challenges, and strengths. Used for school improvement planning and compare schools at woreda and zone. The gender audit process will help to improve the capacity to manage girls' education support systems and therefore improve outcomes. It will improve the planning and monitoring of girls' education, including transition, supporting the REB who are motivated by the need for accurate data to develop a regional gender database and analysis protocol. As systems are strengthened and embedded, the outcomes will be more sustainable in the long term.
Capacity building; Community initiatives	Support Girls' Education Advisory Committees (GEACs), Schools	IO 1: Attendance	Capacity development of the GEACs, SICs, KETBs, and PTSAs (all existing structures) to support girls' education and sensitize others to

Intervention types	What is the intervention?	What Intermediate Outcome (IO) will the intervention will contribute to and how?	How will the intervention contribute to achieving the learning, transition, and sustainability outcomes?
	Improvement Committees (SICs), PTSAs and Kebele Education Training Boards (KETB)	IO 2: Quality in teaching IO 3: School management and governance IO 4: Positive community attitudinal change	support it will bring about positive change in attendance through attitudinal change for girls education. PTSAs and SICs are involved in school improvement and action planning and have strong potential to mobilize community resources for girls' education. Community engagement in School Performance Appraisal Meetings (SPAMs) where data about the impact of teaching quality on children's learning and school improvement is presented, will lead to community contribution to action/gender action plans based on evidence. The capacity development and participation of community school structures, including the PTSAs and SICs, will help to improve school management and governance through improved accountability.
Capacity building	School Management Simulation Tool **This activity merged into training for KETB/PTSA and SIC—no longer a stand-alone activity.**	IO 3: School management and governance	Using the instrument will improve the capacity of PTSAs and SICs to play their role in school management and governance with a focus on girls, based on simulation of challenges and situations they may face. Improved capacity of school management and governance structures to consider the needs of both boys and girls will contribute to better learning and improved transition.
Capacity building; Safe spaces	Training in SRGBV	IO 3: School management and governance	Girls suffer disproportionately from school-related gender-based violence (SRGBV)—although boys also suffer, and it is less reported. As schools become safer places in which girls—and boys—feel more comfortable to learn and remain, there will be an impact on attendance, retention, and learning. As this training focuses on government structures at all levels, it is likely to be more sustainable.
Capacity building; Safe spaces	Embed mechanisms to report abuse	IO 3: School management and governance	Embedding mechanisms which enable girls to report abuse anonymously which Child Protection Committee will respond to (Coordinator, local police, school directors, women and children's affairs at *Kebele* level).[18] Providing the response to the report is appropriate—will support girls to attend and remain in school and learn better and support the system to respond.
Female voice and Governance	Accreditation of "girl-friendly" woredas or schools	IO 3: School management and governance	More girl-friendly schools and woreda offices will contribute overall to the outcome and to girls' learning and transition. The accreditation will motivate schools and officials to continue their efforts to support girls.
Output 2: Improved quality of learning			

[18] *Kebele* is the smallest administrative unit of Ethiopia,

Intervention types	What is the intervention?	What Intermediate Outcome (IO) will the intervention will contribute to and how?	How will the intervention contribute to achieving the learning, transition, and sustainability outcomes?
Teaching input	Intensive training and mentoring for all teachers in mathematics, English, knowledge, classroom management, and Gender and Inclusion Responsive Pedagogy	IO 2: Quality in teaching	The training of teachers in the areas mentioned should lead to improved lesson planning and delivery, and more conducive classrooms in which girls can learn. Better quality teaching will have a positive impact on girls' learning and transition.
Teaching input	Improved monitoring of teaching quality by school managers.	IO 2: Quality in teaching IO 3: School management and governance	With the right kind of monitoring and support for teachers by school managers, the quality of teaching should improve, impacting learning, and transition. Embedding monitoring and support for teachers in school management structures will lead to the same and contribute towards sustainability
Teaching input	Teacher language competency training (English and Wolayttatto)	IO 2: Quality in teaching	This should have a direct impact on girls' learning—if teachers are better able to articulate and teach their lessons in either Wolaita or English as is relevant for the grade. Better teaching should also lead towards girls dropping out less at key transition points.
Output 3: Better access to secondary schools in extreme and remote areas			
Access	Construction of 4 low-cost new inclusive schools in "black hole" rural areas	IO 1: Attendance IO 5: Greater well-being and self-esteem of girls	This will improve access to girls at secondary level in areas where they would otherwise have to walk a very long distance to the nearest school. It will improve attendance, and therefore impact on learning and the transition of girls from primary to secondary
Access and safe spaces	Upgrading inclusive female toilets and sanitation rooms in 13 existing secondary schools.	IO 1: Attendance IO 5: Greater well-being and self-esteem of girls	Poor sanitation and toilets are barriers to girls' attendance, especially during menstruation. If this is improved, attendance will improve, which will also impact learning and transition.
Output 4: Girls "Ready to Learn"			
Access	Cost-effective tutorial program	IO 1: Attendance IO 5: Greater well-being and self-esteem	Tutorial classes focus on improving literacy and numeracy outcomes for girls who are struggling in class, and who may be at risk of drop-out. The additional academic support will help to improve their learning, their self-esteem/ confidence and their attendance in school
Female Voice	Life skills (all grades)	IO 1: Attendance	With the relevant life-skills, girls will be able to better manage/avoid SRGBV, unwanted

Intervention types	What is the intervention?	What Intermediate Outcome (IO) will the intervention will contribute to and how?	How will the intervention contribute to achieving the learning, transition, and sustainability outcomes?
	Financial literacy Career advice (girls who may fail grade 10 to plan for the future, and girls in grade 11–12)	IO 5: Greater well-being and self-esteem	advances, and the changes that come with adolescence, allowing them to attend regularly. Financial literacy and career advice will help them to make the transition from school to work with basic entrepreneurial skills—linked to transition.
Material Support	Provision of uniform, textbooks, stationery, and bursaries for vulnerable secondary and some primary girls	IO 1: Attendance IO 5: Greater well-being and self-esteem	When vulnerable girls do not have access to these items or to any resources to support their education, they are unlikely to attend school, and if they do attend, more likely to drop out. The provision of these items and bursaries (now for both primary and secondary girls) will support vulnerable girls to attend, which will contribute to better learning, and transition to the next level.
Material Support; Safe spaces	Sanitary packs, Sexual reproductive health advice	IO 1: Attendance IO 5: Greater well-being and self-esteem	Sexual reproductive health will help girls to better manage menstruation, improve self-esteem and confidence, and manage adolescence generally. Better self-esteem will lead to girls believing they can learn and make the transition to the next level.
Female Voice; Safe spaces	Social Emotional Learning (SEL): Guidance and Counselling	IO 1 Attendance IO 5: Greater well-being and self-esteem	Girls who believe they can learn and have increased confidence will attend school and be ready to learn. SEL will help girls to resolve some of the barriers and issues they face in attending school.
Female Voice; Safe spaces	Gender Clubs and Girls Clubs	IO 1: Attendance IO 5: Greater well-being and self-esteem of marginalized girls	Gender clubs will result in boys and men understanding the barriers that girls face in attending, learning, and transitioning, and supporting their education. They will help girls to be better able to 'safeguard' and protect themselves from abuse/harassment. Girls Clubs provide an opportunity for girls to share and address common experiences and challenges linked to their education. Also, address traditional social norms associated with girls' education. They will contribute to girls' attendance and greater well-being and self-esteem
Female voice	Awards: (outstanding girls, Head Teachers, teachers, GEAC members, Cluster Supervisors, supportive	IO 1: Attendance IO 2: Improved	Receiving awards at all levels will motivate stakeholders to continue to work towards schools which are responsive to boys and girls, and to children who face additional and/or multiple barriers to attending, participating and learning in school.

Intervention types	What is the intervention?	What Intermediate Outcome (IO) will the intervention will contribute to and how?	How will the intervention contribute to achieving the learning, transition, and sustainability outcomes?
	mothers and fathers, and good brothers)	teaching quality IO 3: Improved management and school governance IO 4: Positive community attitudinal change IO 4: Positive community attitudinal change IO 5: Greater well-being and self-esteem	
Output 5: Mobilized, gender-aware communities demanding high-quality education			
Community initiatives	Role-modeling and awareness-raising campaigns	IO 4: Positive community attitudinal change IO 5: Greater well-being and self-esteem	Local role-models (male and female) will be selected to champion girls' education in the community and at community campaigns which will take place in market areas close to school communities. Role models might be females who have completed their education and as a result been able to contribute positively to their family and to the wider community; they might be boys or men who believe and are willing to express that girls and boys should have the same opportunities to education, and that education is valuable for girls/women as well for boys and men. Campaigns will target a wide and diverse range of education stakeholders, contributing to a change in attitude towards girls' education, The campaigns will help to raise girls' self-esteem and confidence in their education and learning. They will also address negative attitudes and perceptions of girls and boys with disability and education, girls who are young mothers, and for example, girls who are overage for their grade. They are also a forum in which issues of migration or abduction could be expressed (found to be an issue in the Baseline 1 survey).

Intervention types	What is the intervention?	What Intermediate Outcome (IO) will the intervention will contribute to and how?	How will the intervention contribute to achieving the learning, transition, and sustainability outcomes?
Community initiatives	School Performance Appraisal Meetings (SPAMS)	IO 3: School management and governance IO 4: Positive community attitudinal change	School Performance Appraisal Meetings, summarized at the cluster, woreda, and zonal level ensure that school planning is based on data: data-driven school improvement. SIP and performance data presented to school communities enables stakeholders to discuss and identify gaps in gender and safeguarding (and inclusion) responsive education service delivery, and develop action plans to address the gaps. School communities who are informed on what constitutes quality, inclusive and safe education, a good school, a good teacher, and on the gaps related to school improvement domains (teaching and learning, school leadership, community participation and school environment for girls), can play a major role in attitudinal change and in mobilizing resources for girls' attendance, participation, retention, safeguarding and learning.
Community initiatives	Mothers and fathers groups	IO 1: Attendance IO 4: Positive community attitudinal change	The support of mother and father groups will provide another community level support for girls' education. Mother groups are already established in all schools, and cultural norms, girls' performance, attendance, transition, and other items are discussed by the Mothers to support girls. The role of Fathers will be different. They are mostly the decision-makers on whether children attend school or not, and on whether and how family resources are utilized. The support to Father Groups is designed with their specific role in girls' education in mind. Mothers and fathers are part of the community and have much potential to create further attitudinal change within their own communities for girls' education.
Community Initiatives	Training of Kebele Education and Training Boards (KETBs), PTSAs and School Improvement Committees	IO 1: Attendance IO 3: School management and governance IO 4: Positive community	Community School Structures including KETBs, PTSAs, and SICs know their school communities, link the community and school, and play a major role in contributing to school improvement planning. They also play a role in mobilizing resources for school improvement, and in deciding how school/community resources are utilized. These structures, provided with information on what constitutes quality, inclusive and safe education for girls and boys, and on education rights and entitlements, can also play a strong

Intervention types	What is the intervention?	What Intermediate Outcome (IO) will the intervention will contribute to and how?	How will the intervention contribute to achieving the learning, transition, and sustainability outcomes?
		attitudinal change	role in holding schools accountable for what they are supposed to deliver. Providing capacity development, and strengthening roles and responsibilities around gender, inclusion, and safeguarding, with follow up mentoring and monitoring support to these structures can lead to transformational changes in attitude around the social and cultural barriers which often keep girls and children with disabilities out of school. As these are long-standing, well established and existing structures that link the community and school in Ethiopia, they are critical to sustainability beyond the project.
Output 6/7: **Rollout Activities (exact details to be agreed upon and finalized through discussion with Regional partners.**			
Governance	Regional partnership rolls-out key training	IO 3: School management and governance IO 2: Improved teaching quality IO 4: Positive attitudinal change	With the Regional Education partners providing and rolling out training which they deem to be valuable will contribute to sustainability. Possible areas include Gender Mainstreaming Training, SRGBV Training, Training on Inclusive Education, Gender and Inclusion Responsive Training, Language Competency Training, Woreda Cluster Supervisor and Expert Training; PTSA/SIC training—or a mixture of some/all of these. Link has limited resources and can provide support at TOT level only.
Governance	Ratification of GEC models and modules	IO 3: School management and governance	Sustainability—government taking ownership of adapted GEC models and modules
Governance	Dissemination with policymakers	IO 3: School management and governance	Supporting more sustainable outcomes as policymakers understand and buy-in to program interventions.

1.3 Target beneficiary groups and beneficiary numbers

Box 1: Project's contribution [Link Community Development]

This section has been revised to reflect STAGES beneficiary numbers as of April 2019.

Primary target group. The project's primary target group are girls in grades 1– 10 of STAGES supported primary and secondary schools across 4 woredas of Wolaita Zone. These are the girls who stand to significantly benefit from the range of STAGES interventions. Girls from **127 primary schools** in grades 1–8 and **13 secondary schools** in grades 9 and 10 will benefit.

Target number of girls' beneficiaries. The total beneficiary number is **61,345 girls** of which 2,061 are secondary grade 9 students, and 52,678 are primary girls from grades 1 to 8. A total of 2,227 grade 10 girls are not included in this beneficiary number as at the time of baseline they were already leaving school, but from year 2 onwards, grade 10 girls will be included as direct beneficiaries, as will an estimated number of girls who will join STAGES supported schools from other woredas/zones (estimated 6,606) over the life of the project. This will bring the total direct beneficiary number to **61,345** girls.

This data is more up-to-date than the EMIS data used to estimate beneficiary numbers at the time of writing the proposal.

Differences between GEC1, GEC-T proposal, and MEL Framework. The difference between GEC1 and GEC-T is the focus on transition, and this has meant extending the project focus to secondary level (to grade 10), including an infrastructural intervention for the first time. GEC-T addresses transition for the upper cohort of GEC1 beneficiaries. The justification for this change was to more specifically address transition.

GEC External Evaluator Response to Box 1

This section has been revised in response to STAGES updated beneficiary numbers as of April 2019.

Link's proposed methodology and calculations indicate an increased level of accuracy as compared to previous beneficiary number estimates in the GEC-T proposal and STAGES MEL framework. More specifically, Link's previous total direct beneficiary number estimates drew on GEC1 enrollment figures and EMIS data from several years ago, whereas the current estimates utilize enrollment data from 2017/2018 academic year—the first year of the STAGES project. This approach, in turn, should provide more reliable figures as it is the most up-to-date data practically available. Moreover, during phase two of the baseline analysis, more detailed discussions and review with Link of the target beneficiary groups and numbers also took place between the external evaluators and Link. This included examining the counting methodology and assumptions that produced the final numbers and, based on the available evidence, the proposed beneficiary numbers look reliable.

2. Baseline Evaluation Approach and Methodology

This section outlines the external evaluator's approach to the baseline evaluation and methodology. A more comprehensive presentation of the approach can be found in Annexes 5 and 6.

2.1 Key evaluation questions and the role of the baseline

The evaluation design for the STAGES project is intended to facilitate the measurement of primary and IOs needed to answer three primary research questions:

> *Research Question 1: Was the STAGES project successfully designed and implemented? Was the STAGES project good value for money?*
>
> *Research Question 2: What impact did the STAGES project have on the transition of marginalized girls through education stages and on girls learning?*
>
> *Research Question 3: What parts of the intervention work to facilitate the transition of marginalized girls through education stages and increase their learning?*

The baseline data collection will lay the groundwork for addressing research questions two and three during future evaluation points; research question one will be answered through the internal monitoring of STAGES activities. External project evaluations will take place at four time points—baseline in 2018, midline 1 in 2020, midline 2 in 2022, and endline in 2024.

The baseline evaluation for the STAGES project will provide evidence and analysis that reflect on the validity and relevance of the project's theory of change; identify barriers to education that girls face in the Wolaita Zone; set baseline measures and targets for improvements to the project's outcomes, IOs, and outputs; and provide the GEC fund manager, DFID, and external stakeholders with quality analysis and data for aggregation. The baseline evaluation took a two-phase data collection approach; baseline phase one included **Outcome 1: Learning** and **Outcome 3: Sustainability** and baseline phase two included **Outcome 2: Transition**, as well as a subsample of household level surveys and secondary teachers and secondary school audits.

The first submitted version of this report in April 2019 included phase one results; this version now includes phase two results as well and represents the final baseline report for STAGES.

2.2 Outcomes and Intermediate Outcomes

This section outlines the indicators for each of STAGES' outcomes and IOs.

Table 2: Outcomes for Measurement

Outcome	Level at which measurement will take place	Tool and mode of data collection	Rationale	Frequency of data collection
OUTCOME 1: Learning. Number of marginalized girls supported by GEC with improved literacy and numeracy outcomes	School	Early Grade Reading Assessment (EGRA), Early Grade Mathematics Assessment (EGMA), Secondary Grade Reading Assessment (SeGRA), Secondary Grade Mathematics Assessment (SeGMA)[19]	Measure growth from grade 4 to grade 6, grade 6 to grade 8, and grade 8 to grade 10	Per evaluation point
OUTCOME 2: Transition. Number of marginalized girls who have transitioned through key stages of education or vocational training to safe employment	School and household	Girls' transition intentions survey; transition follow-up and household survey, key informant interviews (KIIs), focus group discussions	Measure transition rates for girls at key STAGES transitions; track overall enrollment in woredas	Per evaluation point, with follow-ups in subsequent years
IO 1: Attendance	School	School register, spot checks; girls survey	Track attendance for all girls using school registers; girls attendance behaviors surveyed	Per evaluation point (girls survey) and quarterly (attendance monitoring)
IO 2: Quality in teaching	School	Teacher surveys, monitoring visits/classroom observation instrument	Develop descriptions of classrooms with high- and low-quality teaching	Per evaluation point (teacher surveys) and quarterly (monitoring visits—

[19] The EGRA and EGMA are administered orally one-to-one whereas the SeGRA and SeGMA are written assessments.

Outcome	Level at which measurement will take place	Tool and mode of data collection	Rationale	Frequency of data collection
				classroom observations)
IO 3: School management and governance	Woreda and Community	Woreda-official survey, School audits; KIIs	Characterize the type, nature, and quality of woreda-level support provided to schools and girls	Per evaluation point
IO 4: Positive community attitudinal change	Woreda and Community	Woreda-official survey, School audit, Teacher Survey, Boys survey, Parent survey	Characterize the type, nature, and extent of attitudinal changes among community stakeholders	Per evaluation point
IO 5: Greater well-being and self-esteem of marginalized girls	School	Girls Survey	Develop descriptions of well-being and self-esteem scales among subgroups of girls	Per evaluation point

Sustainability. To determine the extent learning and transition outcomes for girls participating in the STAGES project are sustainable, evaluators will use data collected at the key evaluation stages as well as monitoring data to gauge sustainability outcomes using the sustainability scorecard. Major considerations with sustainability are the holistic nature of the STAGES interventions, and the close relationship Link has fostered with woreda education officials.

Table 3: Sustainability Outcome for Measurement

Sustainability level		Where will measurement take place?	What source of measurement and verification will be used?	Rationale	Frequency of data collection
School	Level of incorporation of SPAM into practices of school and community	School and community	Case studies and KIIs	KIIs and case studies provide context on how schools are deploying SPAMs. They allow evaluators to distinguish between highs and emerging sustainability within a given school-community.	Per evaluation point
	Proportion of community-stakeholder members who demonstrate high levels of knowledge of the STAGES interventions	School and community	School- and woreda-level survey packages	Questions regarding knowledge of and support for STAGES activities capture the level of knowledge surrounding interventions at the school, community, and systems level.	Per evaluation point
	Proportion of community-stakeholder members who demonstrate increased levels of support for the STAGES interventions	School and community	School- and woreda-level survey packages	Questions regarding knowledge of and support for STAGES activities capture the level of knowledge surrounding interventions at the school, community, and systems level.	Per evaluation point
Community	Percentage of GAP targets and actions that have been undertaken	School and community	School improvement plans, gender action plans, and KIIs	The use of GAPs indicates the successful community-based generation of school targets toward girls' education. Interviews highlighting successful and unsuccessful actions, as well as adaptive changes made, offer evaluators useful data to determine the level of sustainability.	Per evaluation point
	Proportion of school staff who demonstrate high levels of knowledge of the STAGES interventions	School and community	School- and woreda-level survey packages	Questions regarding the knowledge of and support for STAGES activities capture the level of knowledge surrounding interventions at the school, community, and systems level.	Per evaluation point
	Proportion of school staff who demonstrate increased levels of support for the STAGES interventions	School and community	School- and woreda-level survey packages	Questions regarding the knowledge of and support for STAGES activities capture the level of knowledge surrounding interventions at the school, community, and systems level.	Per evaluation point
Syste			Cost analysis	The cost analysis instrument provides a system-level assessment of financial sustainability. This	Per evaluation point

Sustainability level		Where will measurement take place?	What source of measurement and verification will be used?	Rationale	Frequency of data collection
	Cost analysis[20]	Government, school, and Community		captures the sustainability of Link's unique partnership with the woreda government, measuring how much the government is picking up of costs and services related to the project although in-kind support.	
	Proportion of government officials who demonstrate high levels of knowledge of the STAGES interventions	Government	Woreda-level survey package	Questions regarding the knowledge of and support for STAGES activities capture the level of knowledge surrounding interventions at the school, community, and systems level.	Per evaluation point
	Proportion of government officials who demonstrate increased levels of support for the STAGES interventions	Government	Woreda-level survey package	Questions regarding the knowledge of and support for STAGES activities capture the level of knowledge surrounding interventions at the school, community, and systems level.	Per evaluation point

[20] Cost analysis was not completed as part of the baseline. This may be done as more of the STAGES activities are underway later in the calendar year (2018). Proxy estimates based on data collectors costs are reported in the sustainability section of this report.

2.3 Evaluation methodology

Evaluation design. The evaluation uses a mixed-method, repeated measure, quasi-experimental design using comparison schools within a woreda where no Link interventions have been conducted. This design is most appropriate because the STAGES interventions will reach all schools and girls within the selected grades in target woredas; therefore, the only possible comparison schools would need to come from a neighboring woreda. Furthermore, since the current project works with primary and secondary girls—as opposed to primary grade girls only in GEC1—the design follows three cohorts of girls through higher grades rather than adding new cohorts of girls. Monitoring data will complement the data collected at the four evaluation points, and throughout all data collections, an effort has been made to minimize the amount of data collection conducted outside of STAGES-focused schools.

Target beneficiary. The most directly impacted beneficiaries of the STAGES project are GEC1 girls enrolled in the primary and secondary schools of the four target woredas—Damot Pulasa, Damot Sore, Damot Woide, and Kindo Koisha—in the Wolaita Zone, SNNPR, Ethiopia. Subgroups within this beneficiary group include girls with disabilities, girls who are orphans, and girls who are pregnant or have children. Boys enrolled in the same schools will also be indirect beneficiaries of interventions aimed at the school level. Additional beneficiaries include new enrollees in grades being served by STAGES in each year, primary and secondary school teachers, community members and stakeholders participating in GEAC, PTSAs, mother and father groups, and woreda- and zone-level education officials and staff engaged in the STAGES interventions.

Evaluation cohorts. The design will follow three cohorts of girls in treatment and comparison schools. When the cohorts transition from a primary school to a secondary, the data collection will occur in the secondary school. To ensure that tracking girls into secondary will be feasible, the evaluator has excluded schools from the sampling frame where girls matriculate into a secondary school outside of the treatment woreda; therefore, tracking girls in the sample will be feasible through secondary.[21] Difference-in-difference comparisons of treatment and comparison school students will be made at each of the four evaluation time points on learning outcomes; transition outcomes will be compared only within the treatment group since no transition data will be collected for the comparison group by design. Comparisons across groups at each time point will allow the evaluators to determine the effect of multiple years of STAGES exposure on outcomes for girls in each cohort.[22] Data collection was conducted through a two-phased approach at baseline and will be consolidated in subsequent evaluation points. Supplementary Table 2 shows the tracking of cohorts across the life of the project.

Supplementary Table 2. Cohorts Tracked Across Each Year of STAGES

Cohort	Baseline 2018		Midline 1 2020		Midline 2 2022		Endline 2023
	Spring	Fall	Spring	Fall	Spring	Fall	Spring
1	Grade4	Grade5	Grade6	-	Grade8	-	Grade10
2	Grade6	Grade7	Grade8	-	Grade10	-	-
3	Grade8	Grade9	Grade10	-	-	-	-

For transition reporting, the learning sample of girls will be tracked every two years to compare the proportion of the sample returning to school in two years. The baseline two-phased approach allowed for an estimate of the one-year transition rate across the key transition points as well as estimates against which targets can be set. For those girls who return to school, parents will be surveyed through the profile

[21] Within each selected grade level, cohort girls were identified through a randomized sample.
[22] Furthermore, what is learned from each successive cohort will be applied to improving the process in subsequent cohorts. For example, cohort 3's learning and transition results from grade 8 as well as challenges in data collection and instrumentation will be applied to subsequent years and grades.

groups approach.[23] Throughout evaluation points, unique identifiers used by both the external evaluator and Link will be used to identify and confirm girls needing follow-ups. Therefore, transition tracking results for girls using profile groups will, following each evaluation point, be based on a subsample of girls.

Role of quantitative and qualitative data. The objective of the evaluation surveys is to provide a deeper understanding of the state of STAGES beneficiaries and their supporting environments. The survey instruments were designed to connect with learning assessment data, offering greater insight into the ways that STAGES IOs—attendance, quality in teaching, school management, and governance, community attitudes, and girls' self-esteem and well-being—are linked to the outcomes of learning and transition. Data from the survey instruments were also considered alongside qualitative data to understand better the socioeconomic dynamics surrounding girls' transition and the factors contributing to the program's sustainability.

The purpose of the qualitative instruments and data collection for the STAGES baseline evaluation was to provide a deeper understanding of the current state of beneficiaries and supporting environments—schools, homes and families, communities, and zone or woreda structures—as well as enrich the quantitative component of the baseline study.[24] Qualitative data collection was designed to help evaluators understand the context within which the relationship between the Link interventions and student outcomes by describing the conditions that facilitate and mitigate the intended outcomes. Qualitative data were only collected in phase one of the baseline. The GEC-T baseline instrument focus on the socioeconomic dynamics surrounding girls' transition and the factors contributing to the program's sustainability. The qualitative instruments capture data that will support identifying the marginalization status of beneficiaries targeted and participating in project activities, explore the prevalence and importance of barriers to girls' education, and establish the extent to which project interventions have affected them.

Incorporation of gender and social inclusion minimum standards. In the context of the STAGES project, the focus on girls, and more specifically on marginalized girls, places an importance on providing equitable programming and purposefully inclusion in the evaluation. A gender and social inclusion (GESI) analysis was conducted by School-to-School International (STS) in July 2017 and identified areas in design, planning, implementation, and monitoring and evaluation (M&E) where gender disparities may exist as well as the appropriateness of the proposed activities to remove or reduce these disparities. Following that analysis, a wide range of steps and approaches were implemented to reduce gender disparities in order to promote the inclusion of girls and especially marginalized girls during the baseline evaluation.[25]

Moreover, the wide and diverse range of stakeholders included in the baseline data collection provided multiple opportunities to ensure the inclusion of experiences and perspectives from respondents who may not otherwise be directly captured. For example, while the project's main beneficiaries are female students, male students were also included in the baseline research through their participation in the upper primary male student focus group discussions (FGDs) as well as in the boys student survey. In addition, given the likelihood that school audit and classroom teacher survey respondents were likely to skew male—less than a third of primary school teachers are female within the Wolaita Zone—perspectives and insights from female teachers were included through female teacher FGDs.[26] Where

[23] Profile groups are described in the MEL as groups of girls based on their baseline survey data. These groups will be based on likelihood to transition as well as characteristics that may inhibit their ability to transition. The groups will serve as a basis to identify a subgroup of girls to follow up with in the fall, after they have transitioned to the next grade.
[24] This includes learning assessments and evaluation surveys.
[25] For example, the final logframe was reviewed to ensure gender-sensitive and disability focused indicators were included. Do no harm, child protection policies, and risk analysis—corresponding mitigation and response strategies—were also incorporated into the baseline evaluation design and implementation. In addition, all assessment materials and instruments were reviewed for potential gender-bias by the external evaluation team members with gender expertise. Evaluation survey instruments and FGDs were also administered in Amharic or Wolayttatto to mitigate language barriers negatively impacting or reducing participation. Additionally, while it is to be expected that woreda officials and school directors will have the requisite Amharic language skills to understand questions and respond in Amharic; students, community members or parents may have a wider range (or no) Amharic language skills; therefore, we plan to administer surveys, interviews or FGDs with these respondents in Wolayttatto.
[26] The Federal Democratic Republic of Ethiopia, EMIS, and ICT Directorate and MOE, *Education Statistics Annual Abstract, 2008 E.C. (2015/16)* (Addis Ababa: June 2017).

appropriate and logistically feasible, single-sex FGDs were utilized, and female facilitators and notetakers led all female-only focus groups.[27]

Several types of accommodations were also incorporated into the baseline study learning assessment design and administration in recognition of some of the challenges students with disability may encounter. For example, given that some students may have low or limited vision, all learning assessment student stimuli for EGRA and EGMA subtasks were developed using large print (at least font 16 or above) and spacing between assessment items. In addition, while it is common practice for timed EGRA and EGMA subtasks to be limited to 60 seconds each, for the baseline study the timed EGRA subtasks for letter sound identification, invented word, and familiar word, as well as the EGMA subtasks for number identification, addition and subtraction) were extended to 120 seconds to enable extra time for students to complete the subtask.[28]

Lastly, M&E processes that include and differentiate girls form a variety of subgroups—such as girls who are pregnant, those who are orphans, and those who have disabilities—are in the process of being established with the project.[29] Their goal is to ensure that data are being collected to track girls' experiences and examine whether interventions are responding to girls' unique needs.[30] In order to support this process, data in this report are disaggregated by sex and age but not by disability. Results by disability are excluded since the phase one disability prevalence was deemed to be very high using the Washington Group short set and therefore, the Washington Group Child Functioning questions were used in phase two to verify the prevalence rate. As such, disability rates are based on the subgroup of girls who were surveyed in phase two and not all girls. To avoid confusion, disability results are excluded in the baseline report but will be included in subsequent evaluation points.

Benchmarking. Benchmarking was completed for learning outcomes by using available data from the baseline sample for grades 6 and 8. Benchmarking data are for students two grades **after** the grade-level of interest; this allows the evaluators to identify the learning levels that students are expected to move into at the next evaluation point. In other words, baseline data from students in grades 6 and 8 serve as benchmarks at evaluation point two for students currently in grades 4 and 6. Additional details on grade 6 and 8 transition benchmarking are discussed in section 4.5.

Benchmarks for students in grade 8 (when they move into grade 10) were established through a separate grade 10 Benchmarking exercise with data collected from grade 10 girls in October 2018. This shift in benchmarking from the approach outlined in the MEL framework (i.e., conducting benchmarking data collection in the fall of 2018 instead of spring 2018) was required because by fall 2018 the relevant population of female students—i.e., the first cohort girls from Link's GEC1 interventions—were now in grade 10. In addition, by this time, the project also had better-established working relationships with secondary schools, enabling a smoother process for data collection. For additional results on grade 10 benchmarking, including sampling and analysis, please see Annex 20.

[27] It should be noted however, one of the constraints in the evaluation encountered is a lack of gender-balance within the data collection teams as the external evaluator does not have control over the selection of enumerators or teaming. As part of the focus on sustainability, local ownership and capacity building, teams are determined internally at the woreda level and enumerators are drawn from existing woreda experts and supervisors who are predominantly male. While it would be preferable to pair all female respondents—especially girls—with female data collectors, this was not feasible. However, additional space for women to participate, contribute to and learn from the data collection was established through the inclusion of female teachers as the FGD facilitators and notetakers. Furthermore, the five women who serve as woreda-level gender officers also served in leadership roles supervising and coordinating woreda-level school-based teams—along with male woreda-level quality assurance officers—during data collection.

[28] Please note, however, the reading passages were limited to 60 seconds per GEC-T guidance.

[29] In addition, midline qualitative research will include a more targeted and purposeful sampling of girls with disabilities and their parents for key informant interviews and focus groups discussions to better understand the challenges, opportunities and unique circumstances these girls face in their learning and transition within the Wolaita Zone.

[30] This is an ongoing process with the hopes of a more targeted approach and understanding of these subgroups which will be established through the transition profile group data collection to be conducted at the start of the next school year.

2.4 Baseline data collection process

This section outlines the process used to collect quantitative and qualitative baseline phase one and quantitative phase two data. It provides details on the preparation, administration, and post-data collection processes for each baseline component: learning assessments, evaluation surveys, and qualitative research.

The baseline data collection was divided into **two phases**. The first phase of data collection was conducted in spring 2018 and included learning assessments, evaluation surveys, and qualitative research. Most of the baseline data collection, analysis, and reporting writing was conducted during phase one, with the associated baseline report submitted and approved in the summer of 2018. The second phase of baseline data collection was conducted in fall 2018 (December) which included an additional set of evaluation surveys in order to gather the remaining baseline data needed for updating, revising and finalizing the STAGES baseline report. More specifically, phase two data collection focused on determining the STAGES baseline transition rate and further characterizing challenges that girls overcome in order to return to school; gathering parental-level information as required by GEC and for baseline indicators;[31] and establishing baseline levels for girls at the secondary school level.[32] Phase two also included the collection of additional Washington Group Question data utilizing the Child Functioning set to further inform analysis and disaggregation on disability prevalence.

Supplementary Table 3. Tools Administered at Baseline Phase 1 and Phase 2

Baseline Data Collection Phase	Student Learning Assessments	Girls Surveys	Boys Survey	Woreda Official Surveys	Teacher Surveys	Guidance Counselor Surveys	School Audit/ Gender Audit	Parent/ Caregiver Surveys	Qualitative Study
1 (Primary only)	X	X	X	X	X		X		X
2 (Transition and secondary)		X			X	X	X	X	

Pre-data collection

Learning Assessments. The sampling framework for reporting on Outcome 1: Learning and its associated learning assessments was developed as part of the MEL framework process. The finalization of selected sample primary schools—including the comparison woreda—took place during the inception phase prior to data collection. The baseline learning assessment sample framework includes 30 primary schools in the Wolaita Zone of SNNPR; the sample consisted of 15 treatment schools and 15 comparison schools.[33] The 15 treatment schools were distributed proportionally across the four target woredas. Schools that did not have any of the grades of interest were excluded from the sample, as were schools that could not be verified prior to the start of data collection—see Annex 10 for additional details of schools that were excluded from the sampling frame. Because the STAGES interventions take a full saturation approach—

[31] These surveys are called the Household survey in the GEC templates, however, to reduce confusion on STAGES, they are referred to as the Parent/Caregiver survey since they are conducted at the school location and not household-to-household. The Parent/Caregiver surveys are conducted with a purposive subsample of girls' families who are selected from the baseline learning sample. The sample of girls are selected from within profile groups

[32] The teacher and school audit use the same existing instruments used during the spring with primary schools, and the Guidance Counselor Survey will be a slightly adapted version of the existing school audit and Teacher Survey.

[33] Secondary schools will be incorporated at future evaluation points as the selected cohort of girls progresses in their education cycle.

that is, it covers all schools in the four target woredas—all 15 comparison schools were selected from a neighboring woreda that receives no direct intervention.[34]

Within the sampled primary schools, a list of 60 female students was generated from enrollment lists; 20 girls were randomly selected per cohort grade—grades 4, 6, and 8—to take part in the baseline learning assessment.[35] This list was first used for the evaluation survey sampling to establish the cohort sample. The same girls then participated in the learning assessment, with a small number of replacements students chosen for those individuals unavailable on the day of data collection. Students assessed during the learning assessment data collection will be tracked for future evaluations. Additional details on the sampling approach may be found in Annex 10.

The learning assessments consist of five instruments: grade 4 EGRA/EGMA, grade 6 EGRA/EGMA, grade 8 EGRA/EGMA, grade 8 SeGRA, and grade 8 SeGMA. Content for the learning assessment instruments was developed in Wolayttatto and English as appropriate.[36] The assessments drew on GEC-T guidance, previous literacy and numeracy assessments conducted in Ethiopia, as well as existing Ethiopia MOE minimum learning competencies, syllabi, and textbooks.[37] In addition, STS led a five-day learning assessment adaptation workshop with 15 Wolaita Zone and woreda-level education officials to adapt, review, and revise the instruments' content and ensure assessments were appropriately contextualized for the Wolaita Zone prior to piloting.

Learning assessments were piloted in February 2018 over three days in four primary schools—one school per each targeted woreda. Nearly 300 female primary school students—about 100 per sampled grade—participated in the pilot. The pilot enabled STS and Link to assess the quality of the instruments as well as the data collection process and procedures for the baseline study. This included observing first-hand 109 tablets being utilized for electronic data capture, troubleshoot issues that emerged with tablets and on the Tangerine electronic data collection platform, and identifying topics and practices requiring additional training prior to operational data collection.[38]

The quality of the pilot learning assessments was measured by examining their internal consistency, item difficulty, and item-total correlations. Based on those results, the most appropriate items and subtask were selected or revised to ensure high internal consistency, appropriate levels of difficulty, and adequate item discrimination. Analysis of test quality showed that the learning assessments were of high quality. Finalized instruments were reviewed and approved by the evaluation manager. Additional details on the learning test pilot and calibration can be found in Annex 9, and the finalized learning assessment instruments can be found in Annex 7.

The STAGES project utilizes existing government personnel at every stage; this is a key aspect of Link's core model of capacity-building within the system and generating sustainable models with the potential for up-scaling within MOE systems and budgets. Therefore, the baseline learning assessment data collection was carried out by governmental officials. Specifically, approximately 152 woreda experts and supervisors from the target and comparison woredas were trained and served as the learning

[34] One comparison woreda was selected due to geographic and logistical constraints. In addition, there were no other viable comparison woredas after excluding woredas that had received previous Link interventions, previously served as comparisons in GEC1, or with fewer than the necessary number of primary schools.
[35] A small number of primary schools had fewer than 20 girls enrolled in the targeted grades; therefore, all enrolled female students were included in the sample.
[36] The MOI and administration for the learning assessments was informed by the official MOI in the Wolaita Zone as well as by recommendations from zone and woreda officials. For grades 4 and 6 assessments, both a Wolayttatto and an English EGRA were developed and administered. The EGMA was conducted in Wolayttatto in grade 4 and English in grade 6. The grade 8 EGRA/EGMA, grade 8 SeGRA, and grade 8 SeGMA were conducted in English only.
[37] Previous literacy and numeracy assessments include, but are not limited to, Link EGRA/EGMA instruments from GEC1, Link's grade 4 and 7 core learning subject tests in English and mathematics, EGRAs produced under the USAID READ project in Ethiopia, and secondary-level assessments developed as part of the Young Lives study in Ethiopia.
[38] Except for the grade 8 SeGRA and grade 8 SeGMA, pilot learning assessment data were collected on tablets utilizing a customized Tangerine platform. Tangerine is an open-source electronic data collection software specifically designed for administering the EGRA and EGMA. The SeGRA and SeGMA are written exams, and therefore administered with grade 8 female students as a group with enumerators overseeing their completion as invigilators.

assessment enumerators.[39] Enumerators received two rounds of training led by STS to familiarize themselves with the instruments and content; strengthen their understanding and application of data collection protocols, procedures, and best practices; and build confidence and abilities in utilizing tablets and Tangerine.[40] These trainings also included a session on child protection led by Link's child protection officer.

Evaluation Surveys (Phase 1). The sampling framework for the evaluation surveys during phase one aligns with the learning assessment sampling framework—with surveys conducted in the same 15 target primary schools and 15 comparison primary schools.

Evaluation Surveys (Phase 2). The sampling framework for the phase two evaluation surveys aligns with the learning assessment sampling framework but with some slight variations. The sampling framework was driven mainly by reporting needs on Outcome 2: Transition and its associated evaluation surveys–girls transition survey and parent/caregiver survey. Surveys were conducted in the same 15 treatment primary schools as phase one, plus 8 treatment secondary schools.[41] The secondary schools were selected based on the feeder primary schools from the sample. The phase two evaluation surveys consisted of five instruments: girls transition survey, parent/caregiver survey, guidance counselor survey, classroom teacher survey,[42] and school audit instruments.[43]

Within the fifteen treatment primary schools, grade 4 and 6 cohort girls who previously participated in the learning assessments were identified to participate in the transition data collection as well as grade 8 cohort girls who were repeating eighth grade. Parent/caregivers of a subsample of grade 4 and 6 cohort girls were also identified to participate in data collection at the primary school. The parent/caregiver sample was purposively selected from within profile groups for girls in treatment woredas only.[44]

Within the eight treatment secondary schools, grade 8 cohort girls from phase one who successfully transitioned into grade 9 at the targeted secondary school were included in the girls' transition survey sample. In addition, within each sampled secondary school, two teachers—one English and one math—were selected for the classroom teacher survey, one school audit per school was completed with the school director and one school guidance counselor survey.

The phase two evaluation survey instrument design drew on the GEC-T guidance survey templates, as well as other validated survey instruments such as Rosenberg Self Esteem Scale,[45] General Self-Efficacy

[39] These figures also include woreda-level quality assurance officers and gender officers. The number of evaluation survey trainees by woreda included the following: 31 participants from Kindo Koisha, 26 from Damot Sore, 28 from Damot Pulasa, 31 from Damot Woide and 36 from Ofa—the comparison woreda.

[40] Enumerators from the four target woredas partook in a three-day pilot learning assessment enumerator training in preparation for their role as enumerators for the pilot in February 2018. However, the comparison woreda did not partake in the pilot. A separate two-day intensive training was held with Ofa enumerators on February 17 and 18 to ensure they received similar training and exposure to the learning assessment content. All enumerators participated in the three-day learning assessment refresher training on April 13 through 15, 2018, that immediately preceded baseline data collection. A number of strategies and training approaches were utilized, including providing an overview on the background, key concepts and content of the instruments—including a detailed overview of specific subtasks included in the instrument; sharing and utilizing two key EGRA resources—an adapted EGRA/EGMA Assessor Cheat Sheet and the EGRA/EGMA Supervisor Observation Checklist—to help participants keep track of and apply the key EGRA/EGMA administration procedures, rules, and best practices; and embedding multiple opportunities for participants to practice instrument administration and receive feedback—plenary role playing of the instrument administration, practice in pairs, and practice in groups of threes.

[41] In order to enable continued tracking and follow-up with sampled cohort girls between phase one and phase two of the baseline data collection, as well as later evaluation points, the feeder secondary schools were identified for each sampled primary school and added to the sampling framework.[41] The eight treatment secondary schools include one school in Damot Pulasa, two schools in Damot Sore, two schools in Damot Woide and 3 schools in Kindo Koisha. Phase two surveys were not conducted in comparison schools.

[42] This was the same instrument administered in phase one but was administered in secondary schools (grade 10) in order to establishing baseline figures at the secondary level.

[43] This was the same instrument administered in phase one but was administered in secondary schools only in order to establishing baseline figures at the secondary level.

[44] Profile groups were generated using phase one baseline survey data, more specifically, items regarding girls' intentions to transition, their demographics and barriers. Profile groups were generated from the existing girls' survey data by creating an index of 'risk' for girls; this enabled a relatively comparable representation in the profile groups of girls who are at higher and lower risk of dropout. Profile groups were developed at two primary grade-levels, with distinct samples from the Grade 4 cohort and the grade 6 cohort.

[45] Rosenberg, Morris. 1989. Society and the Adolescent Self-Image. Revised edition. Middletown, CT: Wesleyan University Press

Scale,[46] as well as an excerpt from the Room to Read Life Skills Assessment.[47] The Washington Group-Child Functioning set of questions were also included.

Draft instruments were shared with the fund manager for review and feedback. Detailed reviews of each instrument were conducted with Link's staff as well as with woreda experts and supervisors during the enumerator training. In addition, prior to enumerator training an informal field-test of the new instruments was completed with one woreda quality assurance officer and one cluster supervisor at a single primary school.[48] It included the administration of one girls' transition survey with one female primary school student and one parent/caregiver survey completed with the primary school girl's mother.[49] These reviews helped ensure the instruments and items were relevant and appropriate for the Wolaita Zone.

As with the phase one of baseline data collection, woreda experts and supervisors served as enumerators in phase two. STS led a two-day enumerator training in December directly prior to data collection to prepare enumerators to provide high quality, effective administration of the five instruments on tablets with the SurveyCTO data collection software.

The transition status of baseline girls was initially collected by woreda experts and supervisors for phase one grade 4 and 6 girls the month prior to (in mid to late November) through school site visits that included cross-checking the school enrollment lists as well as discussions with the relevant teachers and school director. The girls' status was then documented in the Girls Transition Verification Tracker which noted whether she successfully transitioned to the next grade level, repeated, transferred schools, dropped out, or if status was unknown. It was not possible to get grade 8/9 at the time as girls transitioned into new and separate schools.[50]

Qualitative Research. A purposive subsample of four primary schools was drawn from the wider baseline primary school sample for the qualitative component of the study; this was only conducted in phase one. To help ensure a range of school types and experience, the four primary schools—one per target woreda—were selected based on a number of criteria, including the grade 8 student pass rate, total primary student population, and location inside or outside a woreda capital town.[51] Qualitative data collection included a series of discussion groups and KIIs with different participant populations. To assist with these, the external evaluators created six guides: upper-primary girls FGD guide, upper-primary boys FGD, female teachers FGD guide, parents FGD guide, school management KII guide, and zone and woreda education officials KII guide. Four FGDs and one KII with school management were conducted at each of the sampled schools. In addition, eight KIIs were to be completed outside of schools; these included four interviews with Wolaita Zone education officials and four with the gender officers from each target woreda.

Discussion guides were based on Link's previous work and evaluation experience and drew from other girls' education evaluations; drafts were shared with the fund manager for review and feedback. A detailed review of each discussion guide was conducted with Link staff, woreda-level gender officers, and female teachers during the data collector training. Considerable revisions were undertaken during this period to simplify and streamline the guide and ensure that questions were relevant for respondents in the Wolaita Zone. This refinement was critical because logistical, administrative, and budget constraints made it infeasible to pilot the FGD guides prior to data collection.

[46] Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy scale. In J. Weinman, S. Wright, & M. Johnston, Measures in health psychology: A user's portfolio. Causal and control beliefs (pp. 35-37). Windsor, UK: NFER-NELSON.
[47] Room to Read (2019) Life Skills Assessment Scale, email communication.
[48] While the school was part of the STAGES intervention, it was not part of the evaluation sample.
[49] This was especially important as it was not feasible to pilot evaluation surveys prior to data collection due to logistical, administrative, and budget constraints.
[50] Preliminary transition tracking of phase one Grade 8 girls was later conducted in coordination and collaboration between primary-level and secondary-level school directors as part of the school mobilization meetings held during the weekend before data collection.
[51] The final qualitative sample included primary schools with grade 8 exam student pass rates between 16.67 and 100.00 percent, schools with student populations between 800 and 1,400 students, one school located within a woreda capital, and three schools located outside of capital towns. It was not feasible to include comparison schools during the baseline as the selection and initial orientation of the comparison woreda occurred concurrently with qualitative enumerator training and data collection.

In order to support the training on qualitative data collection, STS recruited an Ethiopian qualitative consultant with extensive training and qualitative data collection experience related to gender issues and girls' education in Ethiopia. In addition to leading the qualitative data collector training, the consultant conducted the KIIs with Wolaita Zone and woreda education officials as well as with school management personnel. Woreda officials recommended a total of 16 female teachers—four per target woreda—to participate in the training and serve as focus group facilitators and notetakers. Except for one secondary school teacher, all facilitators were female teachers at primary schools; approximately half had previous experience serving as focus group facilitators or notetakers with Link.

Twenty participants, including the 16 female teachers and four woreda-level gender officers, completed a two-day qualitative data collector training in late February 2018. The training focused on building participants' familiarity with the discussion guides and content; learning qualitative data collection protocols, procedures, and best practices; and building confidence and abilities to facilitate the FGDs in selected schools.

During collection

Learning Assessment. In phase one (April 2018), the baseline learning assessment data collection was successfully completed over five days in 30 primary schools approximately one month after the evaluation survey and qualitative data collection. To ensure protocols—such as ensuring ethical and child protection standards—were followed, targeted sessions on these topics were included in all data collection trainings and child protection policies established. External evaluators provided on-the-ground supervision of data collection in each woreda along with Link staff zone and woreda-level officials to address immediately if any issues that emerged.[52] This oversight mitigated the risk of self-censorship or bias from the enumerators. Additionally, the incorporation of electronic data capture via tablets contributed to data quality, consistency, and collection efficiency by streamlining field work and reducing measurement and data entry errors. Team leads completed school visit forms (SVF) in each school to document the number and type of assessments completed as well as note any issues or challenges in the field.[53] STS compiled this information into a single learning assessment data collection tracker.

Learning assessments were administered to girls on the electronic platform, Tangerine. SVFs were used to identify the total number of girls that were tested; 91.61 percent of this sample was available for analysis. Supplementary Table 4 provides the assessed sample and the final analytical sample for each of the learning assessments by grade level.

Supplementary Table 4. Final Sample Size—Learning Assessment by Grade Level

Grade level	Assessment type	Number of treatments (n)	Number of comparisons (n)	Total number of girls assessed per school visit forms	Total number of girls assessments available for analysis
Grade 4	*EGRA/EGMA*	298	292	590	563
Grade 6	*EGRA/EGMA*	300	282	594	531
Grade 8	*EGRA/EGMA* + *SeGRA/SeGMA*	294	272	569	512
Total		892	846	1,753	1,606[54]

[52] Three members of STS's home office team as well as an Ethiopian EGRA expert and four experienced Ethiopian EGRA/EGMA trainers comprised the external evaluation team. Woreda-level quality assurance officers and gender officers provided integral supervision and administrative support throughout data collection.
[53] Team leads were identified within each woreda; this role was usually assigned to cluster supervisors.
[54] The difference in number of girls assessed based on the School Visit Forms and those that were available for analysis is due to several factors- these are discussed in detail in the post-cleaning section below.

Evaluation Surveys (Phase 1). In late February 2018, evaluation survey data collection was successfully conducted in 30 primary schools—15 treatment and 15 comparison—over the course of five days. Surveys were conducted one week prior to qualitative data collection and approximately one month before the learning assessments. The same protocols and approaches were implemented as with the learning assessments to ensure quality, ethical data collection. The woreda staff survey was self-administered using tablets on the final day of the evaluation survey data collection training. SVFs were completed by each team lead to document the number and type of evaluation surveys completed in each school as well as any issues or challenges during data collection; the forms were later compiled by STS into a single evaluation survey data collection tracker.

Evaluation surveys were administered by enumerators using tablets and the SurveyCTO platform across the treatment and comparison school samples. Supplementary Table 5 provides the final sample size for each of the evaluation surveys conducted in phase one and in Supplementary Table 6 for Phase two.

Supplementary Table 5. Final Sample Size, Phase 1 Evaluation Surveys in Treatment and Comparison Primary Schools

Grade	Girls student survey	Boys student survey	Classroom teacher survey	Classroom observation	School audit survey[55]	Woreda official survey
Grade 4	595	154	T: 57	T: 34	T: 30 (1 per school)	T:116 M: 106 F: 10
			M: 37	M: 29		
			F: 20	F: 5		
Grade 6	594	155	T: 23	T: 36		
			M: 11	M: 25		
			F: 12	F: 11		
Grade 8	569	149	T: 39	T: 38		
			M: 29	M: 29		
			F: 10	F: 9		
Multi-grade			T: 61			
			M: 42			
			F: 19			
Total Sample	1,758	458	T: 180	T: 108	T: 30	T:116

Evaluation Surveys (Phase 2). In December 2018, phase two evaluation survey data collection was successfully conducted in 15 primary schools and eight secondary schools—over the course of four days. The same protocols and approaches were implemented as with phase one to ensure quality, ethical data collection. Girls' transition status—transitioned, repeated, transferred or dropped out—was also re-checked day of data collection by woreda officials when conducting the girls' transition survey and in consultation with the school director and relevant teacher for girls who were not present (absent, dropped, transferred, etc.).[56] Any discrepancies were documented in the School Visit Form (SVF). SVFs were completed by each team lead and documented the number and type of evaluation surveys completed in each school as well as any issues or challenges during data collection; the forms were later compiled by

[55] The school audit survey was administered at each school with the school director.
[56] The initial transition status collected in November was pre-loaded into the Girls Transition Verification form to facilitate cross-checking and/or confirmation of transition status.

STS into a single evaluation survey data collection tracker.[57] Supplementary Table 6 provides the final sample size for each of the evaluation surveys conducted.

Supplementary Table 6. Final Sample Size, Phase 2 Evaluation Surveys in Treatment Primary and Secondary Schools

Grade	Girls Transition Survey	Parent/ Caregiver Survey	Classroom teacher survey	School audit survey[58]	School Guidance Counselor
School Level	**Primary**		**Secondary**		
Grade 4/5	247	T: 107 M: 25 F: 82			
Grade 6/7	217	T: 79 M: 24 F: 55			
Grade 8/9	176	N/A			
Multi-grade			T: 16 M: 13 F: 3	T: 8 M:8 F:0	T: 7 M: 2 F: 5
Total Sample	640	T: 186 M: 49 F:137	T: 16 M: 13 F: 3	T: 8 M:8 F:0	T: 7 M: 2 F: 5

Qualitative Data. Qualitative data collection took place in late February during baseline phase one. It consisted of 12 KIIs completed by STS's qualitative consultant over the course of one week and 16 FGDs facilitated by female teachers at four schools over two days. To ensure protocols were followed when collecting the data, targeted sessions on child protection, research ethics, and corresponding protocols were included in the data collection training. As with the learning assessments, external evaluators provided on-the-ground supervision of data collection in each woreda along with Link's child protection officer and gender officer. Woreda-level gender and quality assurance officers provided supervision and coordination to immediately address any issues that emerged. Supplementary Table 7 provides the final sample size for each of the qualitative instruments.

[57] Phase two evaluation surveys were also administered by enumerators using tablets and the SurveyCTO platform across the treatment primary and secondary school samples.
[58] The school audit survey was administered at each school with the school director.

Supplementary Table 7. Final Qualitative Sample Size in Treatment Woredas (Phase 1)

Qualitative instrument	Number and type of participants targeted	Total number of participants	Total number of focus group discussions or key informant interviews
Upper-primary female student focus group discussions	Six participants each *(three grade 6 female students; three grade 8 female students)*	24 *(all female)*	4
Upper-primary male student focus group discussions	Six participants each *(three grade 6 male students; three grade 8 male students)*	24 *(all male)*	4
Female teacher focus group discussions	Up to six participants each *(preferably from targeted grade levels)*	15 *(all female)*	4
Parent focus group discussion	Six participants with children in primary school each *(three mothers; three fathers)*	24 *(14 male; 10 female)*	4
Total focus group discussions		**87 *(38 male, 49 female)***	**16**
School management KIIs	Four participants *(one school director per school)*	5 *(all male)*[59]	4
Zone and woreda official KIIs	Eight participants *(four zone level officials; four woreda-level gender officers)*	8 *(three male, five female)*[60]	8
Total KIIs		**13 *(eight male, five female)***	**12**
Total qualitative participants		**110 *(46 male, 54 female)***	

Post data collection

Learning Assessment. Upon completion of the data collection, all EGRA/EGMA data was uploaded from the tablets via Wi-Fi to Tangerine's server. Next, files were downloaded and securely stored on STS's password-protected server for cleaning and analysis using Excel and SPSS. Data from SeGRA and SeGMA forms were entered into Excel by STS staff members who also completed a 10 percent double data entry to ensure accurate data entry; hard copies of SeGRA and SeGMA forms are stored at STS's home office in Pacifica, California.

The final dataset for the learning assessments was reviewed by STS to ensure that it meets GEC standards. Three main criteria were used to guide data-quality assessments: completeness, accuracy, and internal consistency. The external evaluator utilized multistage data cleaning plans to ensure that all data values were within an allowable range and that reserve codes were used appropriately. The external evaluator followed the standard best practices for cleaning and finalizing data as outlined in the *Early Grade Reading Assessment (EGRA) Toolkit, Second Edition*, including developing and providing a master codebook and merging or appending data files where possible for easier use and manipulation.[61]

The external evaluator also maintained detailed documentation of all issues encountered during data collection, which was used to support the data cleaning process. Disposition codes were applied to

[59] Both the school director and deputy school director participated in one school.
[60] All the zone and woreda-level gender officials were women.
[61] RTI International, *Early Grade Reading Assessment (EGRA) Toolkit, Second Edition* (Washington, DC: United States Agency for International Development, 2015), 216-17.

categorize the various issues or problems that emerged in the data collection process as well as in the datasets. These disposition codes informed the cleaning rules that were incorporated into the database using syntax to clean the data accordingly. Disposition codes also enabled flagging any school level issues, such as sampling. These coding and flagging procedures helped to ensure that the various and nuanced context of the data collection at the school level were sufficiently cataloged and considered during the data cleaning, analysis, and reporting process. To anonymize the submitted data, the external evaluators followed GEC-T guidance and EGRA best practices to de-identify information before it was shared with the fund manager.[62]

Evaluation Survey. Upon completion of the evaluation survey data collection in both phase one and two, all data were uploaded from the tablets via Wi-Fi to the SurveyCTO server and then downloaded and stored securely on STS's password-protected server for cleaning, review, and analysis using Excel and SPSS. Using the evaluation survey data collector tracker and SVFs, data was cleaned based on pre-set criteria: time and date inconsistencies, consent checks, and survey sessions timing.

Quantitative Data Cleaning and Merging Process (Phase 1). As previously stated, a sample of 60 female students was selected from student enrollment lists—with 20 girls randomly selected per cohort grade—to take part in the baseline evaluation survey. The sample for the learning assessments, however, evolved over four stages (see Annex 14). Stage 1 represents the final list of survey respondents including originally sampled girls and those who served as replacements for absent girls; this list was also the sample for the learning assessments. When the enumerators visited each school to administer learning assessments in stage 2, they used SVFs to track absent or missing students, record replacement students assessed in their place, and confirm the completion of each assessment. During stage 3, SVF were reconciled with the actual learning data collected in the electronic files.[63, 64] In stage 4, the electronic records of grade 8 students were merged with data entered from scored, paper-based SeGRA and SeGMA. Finally, girls evaluation survey data were merged with the learning assessment data in stage 5 using unique student identifiers.

By examining relationships with survey data, any girls who could not be matched to survey data are excluded.[65] The results reported in Section 4 include all students who were administered the learning assessments and had valid data. Phase two data includes all girls from phase one who were also re-surveyed during phase two.

Quantitative Data Cleaning and Merging Process (Phase 2). Phase two data were merged into phase one data for girls. In other words, for each record from phase one, any available data from phase two were merged in. Additionally, parent/caregiver data were merged in so that each record consisted of a girl's phase one, phase two, and parent/caregiver data. Separate datasets were cleaned for teachers by merging in secondary teacher data from phase two into the primary teacher data from phase one. School audit and guidance counselor data were cleaned in a third dataset as these data were at the school-level (one per school, at most). The resulting data from the end of phase two will serve as the base files for analysis of midline 1 results.

Qualitative Data. In phase one, FGDs were facilitated using a mix of Wolayttatto and Amharic—with detailed field notes recorded by the notetakers by hand in Amharic. Trained teachers reviewed and finalized the field notes and then participated in a debrief discussion with STS's qualitative consultant to provide any additional clarifications, questions, or take-aways. The field notes were then typed into a Word document by Link's data entry officer before being translated into English. STS's qualitative consultant, as well as Link's child protection officer and gender officer, provided oversight and quality checks on the data entry and translation. KIIs were conducted in Amharic by the qualitative consultant

[62] Ibid.
[63] Electronic data captured in tangerine included EGRA and EGMA only with the Washington Group Questionnaire.
[64] Prior to merging, a three-step data cleaning process was used to detect incomplete or problematic cases. "Problematic" cases were identified as those that occurred outside of expected assessment periods, were duplicate records, fell outside the range of 2.58 standard deviations, or were flagged by enumerator comments.
[65] Annex 14 shows the total Completed Cases—cases that had full data on girls survey as well as all applicable learning assessments and were not replaced but persisted through all five stages.

who then produced field notes in English in a Word document. All field notes were shared and stored on STS's secured, password-protected server. Finalized field notes were imported into NVivo 12, a data analysis software package, in order to systematically code and analyze the data.

While it was not logistically or budgetarily feasible to complete full transcriptions and translation of the FGDs and KIIs, all were audio-recorded to serve as references when writing up the field notes. Once the field notes were entered, the audio files were uploaded to a secure, password protected server, and deleted from the audio-recording device.

The qualitative data analysis methodology incorporated an iterative approach and included content analysis and constant comparison of narrative data to identify and validate emerging themes. While a preliminary codebook was developed based on the MEL GEC-T Framework, Link's previous GEC1 evaluations in Ethiopia, and similar studies on girls' education in Africa, additional codes emerged during the data analysis and the codebook was updated as needed.[66] The qualitative data and emergent themes were also examined within the broader context of the quantitative results and indicators, with relevant findings woven into the report as appropriate to help provide additional insights and understanding into the STAGES baseline evaluation results, analyses, and external evaluator recommendations.

2.5 Challenges in baseline data collection and limitations of the evaluation design

The external evaluators were satisfied with the overall evaluation design, as well as the quality and content of the collected data. While no major issues emerged, some minor challenges and limitations were identified and are highlighted for additional background and context.

The baseline did not include any data for girls in grade 10. STAGES was still in the process of establishing working relationships with secondary schools, and therefore, data collection during phase one was not feasible in secondary schools. Girls in grade 9 were surveyed in phase two, as were teachers in grades 9 and 10 and school directors, but no girls in grade 10 completed the learning assessments during phase two for baseline purposes. Instead, midline 1 will include data for girls in grade 10 for the first time-when girls who were in grade 8 during baseline reach grade 10.

Secondary school sampling did not accurately account for one of the sampled primary schools feeding into a secondary school outside the target woredas. While Link provided a preliminary mapping of primary schools to their feeder secondary schools in the original sampling frame, it wasn't until phase two data collection planning that it became evident that students from one of the sampled Kindo Koisha primary schools feed into a secondary school outside the of four target woredas. Therefore, the evaluation will be unable to track students from this primary school beyond grade 8.

Due to the school-based approach for evaluation data collection, administration of the transition survey was limited to sampled cohort girls who were currently in school and present on the day of data collection. While the original design proposed in the MEL was to complete the transition survey with a subsample of girls, regardless of their transition status, this was not possible due to logistical and administrative constraints. Therefore, school directors and/or teachers were asked to provide additional information regarding transition status of girls who were absent or no longer enrolled at the relevant primary school as a proxy measure. Following up with girls who have left school will need to be a focus in the attendance monitoring as well as enrollment verification process associated with project intervention tracking. However, an effort was made to ensure that the profile groups used to reach out to parents and complete the parent survey included parents of girls who transition as well as those who did not.

The parent/caregiver survey was limited to a subsample of grade 4/5 and 6/7 cohort girls' parents selected from within profile groups. [67] As data collection was school-based, it was not feasible to include parents/caregivers of girls in secondary school, i.e., transitioned from grade 8 to grade 9, in the sample. This was deemed unfeasible when planning the data collection with Link staff since the distance

[66] The finalized, qualitative codebook is included in Annex 8.
[67] Profile groups were generated using phase one baseline survey data, more specifically, items regarding girls' intentions to transition, their demographics and barriers.

between a parents residence and the secondary school can be long. Instead, the survey was redesigned to capture general parent/caregiver perceptions of girls' experience in their communities and not just the daughters of parents surveyed.

Data on the prevalence of disabilities, and conclusions based on these data are limited because of the shift from the short set in phase one to the Child Function Questions in phase two. Specifically, the short set was used with all girls in the sample in phase one, but the unusually high prevalence rate warranted a reexamination of the results. In phase two, the Child Function questions were used instead. The results from the Child Function questions were deemed more reliable by the DFID Disability Experts and results from phase one, based on the Short Set, were discarded. Due to the difference in the denominator across phases, comparison of results for girls in the disability subgroup was not possible and are excluded in this final version of the report. However, this information will be provided starting at midline using the Washington Group Child Functioning questions. Additional details regarding the Washington Group questions administration, issues, and comparison between the Short Set findings and the Child Functioning data can be found in Annex 19.

The quasi-experimental design includes a "comparison" woreda as opposed to a true "control" group; the analysis, interpretations, and findings must take this into consideration when comparing results.[68] Full saturation of STAGES across all schools in the four woredas precluded the selection of a set of true comparison schools. This is especially true considering the woreda-level education management information system data where the comparison woreda—Ofa—demonstrated particularly strong gender-equity measures on indicators such as gender-parity index, the percentage of female teachers, and the proportion of female students.[69]

Sampling at the grade level within the selected schools required a small number of student replacements between evaluation surveys and learning assessments due to high levels of student absenteeism and drop-out. Student enrollment lists were in flux due to high levels of drop-out even within a short time frame such as a month. Some minor inaccuracies on student lists used for initial random sampling were also noted. See Annex 14 for complete details on the level of replacement between surveys and learning assessments.

The attrition rate after one-year exceeds the attrition rate assumed in the sample for a two-year period. The average rate of attrition in the sample was almost 35 percent in a period of six months. By comparison, the assumed attrition rate for each two-year period in the sampling strategy was 30 percent.

Delays in shipment and processing of tablets through local customs impacted data collection, analysis, and report writing timeline. In turn, adaptive contingency planning was utilized. Clear lines of communication were maintained regarding adjusted timelines among the evaluators, the project, and the fund manager. Once tablets were processed and received, the project and evaluators noted that electronic data capture's efficiency and the associated regain of time lost highly valuable.[70] However, as a result of this delay, changes in data collection had to be made. Particularly, baseline transition tracking surveys at the household level was not feasible, and surveys had to be shortened.

A state of emergency was declared in Ethiopia during the baseline evaluation that led to limited access to mobile internet for electronic data uploads and impacted some of the evaluation team's travel and concerns over the potential unrest. Solid contingency plans were established to support an adaptive approach to a potentially shifting political landscape, prevent delays in data collection, and ensure proper oversight and coverage of the various trainings and data collection. These proceeded despite needing to cancel one team member's travel to the field.[71]

[68] As noted in the previous methodology sections, a randomized control trial was not feasible or appropriate for this context.
[69] The Federal Democratic Republic of Ethiopia, EMIS, and ICT Directorate and MOE, *Education Statistics Annual Abstract, 2008 E.C. (2015/16)* (Addis Ababa: June 2017).
[70] For example, no quantitative data entry was required during data collection except for the SeGRA and SeGMA data.
[71] The external evaluation team also utilized domestic flights to avoid travel through areas of potential unrest. In addition, planned nightly upload of data was shifted all data being upload at the end of each phase of data collection via the office Wi-Fi.

Some enumerators had limited experience or exposure to tablets or mobile technology prior to the baseline. This shortcoming was addressed by repeatedly exposing enumerators to the tablets and promoting practice using tablets during trainings and data collection. These exposures occurred first during the learning assessment pilot, next during the evaluation survey, and lastly during the baseline learning assessment.[72]

Inconsistencies in enumerators' understanding, administration, and scoring of phonemic awareness-based learning assessments in Wolayttatto subtasks—letter sound identification and invented word reading—were observed both in the training and data collection. Inconsistencies or disagreements across enumerators on acceptable pronunciations were especially apparent on the letter sound identification subtask in Wolayttatto.[73] While this is not an insurmountable obstacle, the limited time available for the learning assessment enumerator training—in addition to the large number of subtasks that required training—made dedicating sufficient time and explicit training on Wolayttatto letter sounds unfeasible.

Communicating across three languages—English, Amharic, and Wolayttatto—during trainings and data collection could prove challenging. To reduce language gaps between trainers, enumerators, and respondents as well as provide stronger quality assurance, the evaluation team included Ethiopian nationals with fluent Amharic and English language skills and extensive training experience who co-led all trainings and supervised data collection.[74] While this was sufficient for the baseline—given that Amharic is the working language of the government—the inclusion of a fluent Wolayttatto-Amharic-English trainer would be ideal for future evaluation points.

The rich, nuanced, and textured discussions, reflections, and insights elicited from key stakeholders during focus groups discussion and KIIs may be limited due to a lack of full transcriptions as well as the multiple layers of translations required.[75] Due to time, budget, and logistical constraints, detailed field notes were utilized in place of fully translated transcriptions. Field notes produced by female teachers from FGD and by the evaluation team's qualitative consultant from KIIs enabled a quicker turnaround that was less labor intensive and fit within the current conditions and capacity of the project. [76, 77, 78]

The school-level qualitative data collection sample only included treatment primary schools in phase one due to logistical constraints. Given the evaluation's cohort approach and the focus on the baseline data collection on grades 4, 6, and 8, external evaluators determined it was most appropriate to

[72] Enumerators also demonstrated a wide range of experience and expertise on mobiles and peer-learning and support was encouraged. Enumerators with stronger demonstrated competency with tablets were also prioritized as data collectors.
[73] This is unsurprising as discussions with zone and woreda officials indicate that teaching letter sound identification in Wolayttatto is a recent practice within the Wolaita Zone, with a new approach set to be piloted imminently for grade 1. Moreover, previous versions of tis subtask in Link evaluations utilized letter name identification—not sound—and this subtask was also not included in the pilot.
[74] Amharic translations of data collection instruments and training materials were also reviewed by evaluation team members fluent in Amharic and English as well as Link staff as needed.
[75]Focus groups were conducted in Wolayttatto, or a mix of Amharic and Wolayttatto. KIIs were conducted in Amharic, with field notes produced in English by the interviewer.
[76] Insufficient time for full transcription of sixteen focus groups in Wolayttatto language, let alone translation into Amharic and English and back translations for quality control checks. Moreover, even with time, there is limited funds and capacity of individuals with required language expertise, computer skills, and understanding of local and project context to carry out full transcription and translations.
[77] Female teachers served as facilitators and notetakers for the focus groups. Focus groups were audio-recorded and teachers could use these as an additional reference when finalizing their field notes after the focus groups were completed. Upon submission of field notes the female teachers conducted a debrief session with STS's Ethiopian qualitative consultant and the notes were typed up in Amharic by a Link data entry officer, cross-checked by Link staff for quality and clarity, then translated from Amharic into English by an external translator. Upon completion of translations, a final cross-check and review of quality of the translation was conducted by STS's Ethiopian qualitative consultant. The engagement of the Ethiopia qualitative consultant on this component in creating a bridge between the Amharic and English—especially in terms of providing quality checks of translation—was essential in mitigating these challenges. In addition, interviews and focus groups were audio-recorded and the files password protected so it will be possible to complete transcripts and translations of this data if it were to become feasible financially or time-wise at a later date.
[78] Translation challenges were less of a concern on the KIIs, as they were conducted in Amharic with field notes produced in English by the same individual—STS's Ethiopia qualitative consultant—who has extensive technical experience in this area and strong fluency in Amharic and English.

focus baseline qualitative data collection within the primary school setting only. Secondary schools will be incorporated at later evaluation points as the cohorts move up the school cycle.[79]

Qualitative analysis captured limited demographics of participants in the focus group discussions. The qualitative data collection instruments and field notes did not adequately account for the demographics of focus group respondents.[80] In the future data collections, the evaluator will ensure all qualitative data collection instruments adequately capture key demographics for participants, including age, education level, marital status, and childbearing status; the evaluators will also emphasize the importance of capturing this information during training and data collection.

3. Key Characteristics of Baseline samples

3.1 Project Beneficiaries

The STAGES project targets female students enrolled in primary and secondary schools within the four woredas of the Wolaita Zone. The project defines all girls as "educationally marginalized" due to the remoteness of the areas and an associated shortage of schools—especially well equipped, quality schools. Further, girls are considered economically marginalized given the high poverty levels within the Wolaita Zone and socially marginalized in regards to harmful gender norms, stereotypes, and practices that permeate their families, communities, and broader school setting. Given these intersecting barriers and the full saturation approach for targeted grade levels, the STAGES project will continue to apply the definition of educational marginalization to all female students within the targeted grades; no additional criteria will be utilized for selection for general participation in the STAGES intervention.[81]

Within female student beneficiaries, three subgroups have been identified as having increased levels of marginalization. Female students within these subgroups will be prioritized in some STAGES activities and inputs.[82] These subgroups include girls with disabilities, girls who are orphans, and pregnant girls and young mothers. Based on the data from the Washington Group-Child Functioning short set questions administered phase two of the baseline, it is estimated that 5.32 percent of STAGES female student beneficiaries meet the definition of girls with disabilities.[83] Note this is a revised figure that supersedes the previous figures provided from phase one of the data collection that used the Washington Group Short set.

Individual-level identification of girls who are orphans, pregnant, or mothers was not done. Instead triangulation from multiple sources was used. All primary and secondary schools in the sample reported having female student beneficiaries who are orphans. On average, more than four in ten schools—across

[79] The evaluation team, enumerators, and Link staff were already stretched to maximum capacity due to the large-scale nature of the learning assessment and evaluation survey baseline data collection across thirty schools and three grade levels—which, in turn made the addition of more qualitative school sites (secondary or comparison) impractical at the time of data collection.
[80] One member of the evaluation team, who was to play a key role in the qualitative component oversight, canceled travel due to the declaration of the state of emergency in Ethiopia, and with the team unexpectedly stretched thin, this element within the qualitative instruments and data collection in the field was mistakenly overlooked.
[81] Male students within the target woredas and schools face similar educational marginalization related to geography and economical marginalization related to high levels of poverty. Boys will be reached through project interventions at the school level. For example, male students will benefit from the construction of new secondary schools as well as interventions aimed at improving the quality of teaching and learning through teacher trainings at the primary and secondary school level.
[82] These activities and inputs include, but are not necessarily limited to, prioritizing access to tutorials, sanitary packs, social and emotional learning support, mentoring, and bursaries at the secondary school level.
[83] This figure is based on the Washington Group—Child Functioning set that were administered during phase two of the baseline as part of the girls transition survey in December 2018 with the cohort sample of grade 6/7 and 8/9 female students who returned to school in the fall (i.e., it excludes girls who dropped out between phase one and phase two). This differs from the original 22.45 percent that was calculated during phase one of the baseline during spring 2018 which utilized the Washington Group Short Set question. Additional details on the rationale for replacing the original figures form the short set with the child functioning set can be found in Annex 19. For additional context, also note, estimates in the STAGES proposal from September 2016 identified approximately 1,387 girls with disabilities; 1,376 of the girls with disabilities are registered in the 123 primary GEC1 schools and 11 girls are registered in targeted secondary schools grade 9 and 10. The basis of determining an individuals' disability in the STAGES proposal was based not on the Washington Group—Child Functioning questions, and as such, the criteria for identifying individuals with disabilities differs from baseline. As such, these two data points are not comparable.

treatment and comparison primary schools—noted that they enrolled at least one female student who is a mother.[84] Within sampled treatment secondary schools, six out of eight secondary school directors indicated they are aware of girls that are pregnant or expecting in their school; when asked to approximate the number of pregnant girls, figures ranged from 0 to 10 girls within a secondary school. About half of secondary school teachers survey (53.85 percent) also indicated that they teach girls that are pregnant or expecting in their classes (that they are aware of).

3.2 Representativeness of the Learning and Transition Samples across Regions, Age Groups, Grades, Disability Status, and Sex of the Beneficiaries

Table 4, Table 5, and Table 6 describe the evaluation sample for the baseline study. The evaluation sample consists of girls from three grades—4, 6, and 8—from 15 treatment and 15 comparison primary schools. Treatment schools were drawn from four treatment woredas, and comparison schools were drawn from one neighboring woreda. Because the schools selected for comparison come from a geographically distinct area—all comparison schools come from a separate woreda rather than from within the same woredas as the treatment schools—the group is referred to as the **comparison** group and not the **control** group throughout this report.

Overall the sample is representative of the population of schools and students in the woredas from which they were selected. The number of schools selected from each woreda reflects the total number of schools in the woreda. For example, Kindo Koisha, the woreda with the largest relative number of schools, also has the most schools in the evaluation sample—five. By grade, the sample was equally split between grades 4, 6, and 8. By age, the sample consists of approximately three-quarters students who are on-age or at the expected age of that grade.[85] By disability status, the data reported below are based on the Child Functioning items administered in phase two and replaces the rates reported previously using phase one data and the WG Short Set. Furthermore, phase two did not include any comparison school data collection, and therefore we are extrapolating that the prevalence rates of disability in the comparison group are similar to those in the intervention group since the rates were comparable with the Short Set. However, actual prevalence figures are not available to report in the table below.

Table 4: Evaluation Sample Breakdown (by Woreda)

Woreda	Number of schools sampled	Intervention (baseline)	Comparison (baseline)[86]
Sample breakdown (girls)			
Damot Pulasa (percentage of overall sample)	3	179 (20.25%)	0
Damot Sore (percentage of overall sample)	3	171 (19.34%)	0
Damot Woide (percentage of overall sample)	4	237 (26.81%)	0
Kindo Koisha (percentage of overall sample)	5	297 (33.60%)	0
Ofa	15	0 (0.00%)	838 (100.0%)

[84] When girls were asked if there are any girls who are mothers in their current class this year, 17.41 percent of grade 4/5 girls, 31.34 percent of grade 6/7 girls and 30.34 percent of grade 8/9 girls indicated yes. Of those who reported yes, when asked how many girls in their current class were mothers, the minimum reported was one girl and maximum 5 girls in the grade 4/5 cohort, whereas in grade 6/7 the lowest was also 1 girl but the highest 8 girls; in the grade 8/9 cohort, half of the respondents reported one girl was a mother and two reported that five girls were mothers.

[85] The official entrance age for primary schools is seven years old in grade 1. The Global Education Monitoring Report (2018) reports that 22.2% of pupils in Ethiopian primary schools are overage for their grade based on data from the school year ending in 2015. The report defines overage "as those [students] whose age is at least two years higher than the official age of entry in a given grade."

[86] For all relevant tables, "control" has been replaced with "comparison" to more accurately reflect the sampling approach and framework.

Woreda	Number of schools sampled	Intervention (baseline)	Comparison (baseline)[86]
(percentage of overall sample)			
Girls		**884**	**838**
Sample breakdown (boys)			
Damot Pulasa (percentage of overall sample)	3	47 (20.52%)	0
Damot Sore (percentage of overall sample)	3	47 (20.52%)	0
Damot Woide (percentage of overall sample)	4	57 (24.89%)	0
Kindo Koisha (percentage of overall sample)	5	78 (34.06%)	0
Ofa (percentage of overall sample)	15	0 (0.00%)	229 (100.0%)
Boys		**229**	**229**

Table 5: Evaluation Sample Breakdown (by Grade)

Grade	Intervention (baseline)	Comparison (baseline)
Sample breakdown (girls)		
Grade 4 (percentage of overall sample)	299 (33.82%)	291 (34.73%)
Grade 6 (percentage of overall sample)	296 (33.48%)	283 (33.77%)
Grade 8 (percentage of overall sample)	289 (32.69%)	264 (31.50%)
Out-of-school girls (percentage of overall sample)	0 (0.00%)	0 (0.00%)
Girls	**884**	**838**
Sample breakdown (boys)		
Grade 4 (percentage of overall sample)	78 (34.06%)	76 (33.19%)
Grade 6 (percentage of overall sample)	76 (33.19%)	79 (34.50%)
Grade 8 (percentage of overall sample)	75 (32.75%)	74 (32.31%)
Out-of-school boys (percentage of overall sample)	0 (0.00%)	0 (0.00%)
Boys	**229**	**229**
Sample breakdown (parents/caregivers)		
Grade 4 (percentage of overall sample)	25 (23.36%) male, 82 (76.64%) female	n/a
Grade 6 (percentage of overall sample)	24 (30.38%) male, 55 (69.62%) female	n/a
Grade 8 (percentage of overall sample)	n/a	n/a
Parents/caregivers		

Table 6: Evaluation sample breakdown (by Age)

Age bands	Intervention (baseline)	Comparison (baseline)
Sample breakdown (girls)		
Aged 6–8 (percentage of overall sample)	3 (0.34%)	0 (0.00%)
Aged 9–11 (percentage of overall sample)	157 (17.76%)	187 (22.32%)

Age bands	Intervention (baseline)	Comparison (baseline)
Aged 12–13 (percentage of overall sample)	309 (34.95%)	307 (36.63%)
Aged 14–15 (percentage of overall sample)	344 (38.91%)	286 (34.13%)
Aged 16–17 (percentage of overall sample)	43 (4.86%)	47 (5.61%)
Aged 18–19 (percentage of overall sample)	14 (1.58%)	4 (0.48%)
Aged 20+ (percentage of overall sample)	14 (1.58%)	6 (0.72%)
Age missing	0 (0.00%)	1 (0.12%)
Girls	**884**	**838**
On-Age, Overage and Underage for the Grade Level[87]		
Underage for grade (percentage of overall sample	42 (4.75%)	58 (6.92%)
On-age for grade (percentage of overall sample)	673 (76.13%)	652 (77.80%)
Overage for grade (percentage of overall sample)	169 (19.12%)	128 (15.27%)
Sample breakdown (Boys)		
Aged 6–8 (percentage of overall sample)	0 (0.00%)	0 (0.00%)
Aged 9–11 (percentage of overall sample)	38 (16.59%)	45 (19.65%)
Aged 12–13 (percentage of overall sample)	87 (37.99%)	74 (32.31%)
Aged 14–15 (percentage of overall sample)	83 (36.24%)	86 (37.55%)
Aged 16–17 (percentage of overall sample)	13 (5.68%)	16 (6.99%)
Aged 18–19 (percentage of overall sample)	8 (3.49%)	3 (1.31%)
Aged 20+ (percentage of overall sample)	0 (0.00%)	5 (2.18%)
Boys	**229**	**229**
On-Age, Overage and Underage for the Grade Level		
Underage for grade (percentage of overall sample)	11 (4.80%)	18 (7.86%)
On-age for grade (percentage of overall sample)	174 (75.98%)	162 (70.74%)
Overage for grade (percentage of overall sample)	44 (19.21%)	49 (21.40%)

[87] On-age, underage, and overage ranges were determined based on a review of extant classifications of age bands. For the purposes of this report, ranges were determined as follows: for grade 4, students nine and younger were “under age,” 10 to 12 were “on age,” 13 and older were “overage;” for grade 6, students 11 and younger were “under age,” 12 to 14 were “on age,” and 15 to 17 were “overage;” for grade 8, students 13 and younger were “under age,” 14 and 15 were “on age,” and 16 to 18 were “overage.”

Table 7: Evaluation Sample Breakdown (by Disability)[88]

Sample breakdown (girls)	Intervention (baseline)	Comparison (baseline)[89]	Household survey and girls school survey—Washington Group Child Functioning questions
Girls with disability (percentage overall)	Grade 6/7: 13 (5.99%) Grade 8/9: 8 (4.49%)	Data not collected in phase two with CF questions	Washington group questions-Child Functioning Set[90]
Provide data per impairment			
Vision impairment (percentage overall)	Grade 6/7: 2 (0.92%) Grade 8/9: 2 (1.12%)	Data not collected in phase two with CF questions. Estimate similar prevalence rates as intervention schools since the rates with WG Short set were comparable across groups.	CS_D1s
Hearing impairment (percentage overall)	Grade 6/7: 0 (0.00%) Grade 8/9: 1 (0.56%)		CS_D2s
Mobility impairment (percentage overall)	Grade 6/7: 0 (0.00%) Grade 8/9: 1 (0.56%)		CS_D3s
Cognitive impairment (percentage overall)	Grade 6/7: 6 (2.76%) Grade 8/9: 3 (1.69%)		CS_D4s
Self-care impairment (percentage overall)	Grade 6/7: 2 (0.92%) Grade 8/9: 1 (0.56%)		CS_D5s
Communication impairment (percentage overall)	Grade 6/7: 4 (1.84%) Grade 8/9: 4 (2.25%)		CS_D6s

3.3 Educational Marginalization

The following section highlights findings from the girl's survey that describe key individual characteristics of sampled girls. It also details key barriers to girls' education as reported in the girls, boys, teachers, woreda officials, or school audit surveys. Household-level parent and caregiver surveys were administered as part of the second phase of the baseline evaluation data collection in December 2018, and those items can now be found in Table 8. However, please note only a targeted subgroup of parent/caregivers of cohort grade 4/5 and 6/7 girls in treatment primary schools were included. Given this may contribute to some gaps in household-level information, available data from comparable items in other surveys are also provided, as appropriate.

The analysis examines girls' characteristics and barriers to education that go beyond observable challenges, such as being from a poor household or having children. While these observable challenges are certainly present, the survey data show that unobservable challenges—such as levels of support for girls' education, perceptions of gender, gender-appropriateness, and responsiveness of the curriculum and its implementation—are equally present. As such, results from the surveys are summarized in a set of scales, which are intended to provide a score on the underlying construct of interest and do not rely on a single item alone. For example, the scale for one construct of interest—gender perceptions—is based on responses to seven statements about girls' and boys' education. Supplementary tables listing the details of the scales used to examine girls' characteristics and barriers to education can be found in

[88] Based on discussions and guidance from the GEC disability and inclusion advisors, this table has been revised since the original submission of the STAGES baseline report to reflect the most up to date and accurate figures associated with the Washington Group—Child Functioning questions.
[89] It was not possible to collect addition data utilizing the Child functioning questions with comparison schools and girls during phase two of the baseline. The Child Functioning set will be captured for both treatment and control girls in grades 6 and above for the remainder of the STAGES evaluation points moving forward.
[90] The Washington Group Child Functioning questions were administered to the cohort sample girls at the time of the girls' transition survey—December 2019—to cohort treatment girls in grades 6/7 and 8/9. Responses were analyzed in line with GEC-T guidance that the population identified as having a disability should include all those with difficulty *in at least one domain* recorded at *a lot of difficulty* or *cannot do at all*. This applies to both the Washington Group Child Functioning questions as this cut off point will provide the most accurate representation of the population that has an impairment which may interact with barriers leading to educational marginalization. Additional details may be found in Annex 19.

Annex 15. Higher scores on each scale are more desirable; for example, a higher gender-perception score corresponds to a more positive gender perception.[91] See Annex 15 for item-level and scale-level scores.

Quantitative and qualitative data were examined to identify characteristics and barriers. First, in order to provide context and framing, findings from KIIs and FGDs are presented, followed by quantitative findings from the evaluation surveys.

Girls' characteristics

Several characteristics emerged as critical for considering when discussing girls' educational marginalization and barriers within the Wolaita Zone during the qualitative analysis.[92] For example, most qualitative respondents cited **girls' high burden of household chores** when asked about barriers to girls' education. In addition, across FGDs, **girls from poor households** or families with limited financial resources were noted as having to **engage in income-generating activities** that negatively impacted their education; they are also more vulnerable to early marriage and subsequent early childbearing.[93] Generally, characteristics of **girls who are married** and **girls who have begun childbearing** were discussed as negatively impacting girls' continued education. KII respondents, teachers, and parents noted that some **girls who are menstruating** struggle with attendance; this was also echoed in the FGDs with upper primary school female students in grades 6 and 8.[94] Other characteristics for consideration that were mentioned in the qualitative data include **disabilities, orphans,** and **migration status**. A more in-depth discussion of these characteristics and their intersection with barriers to girls' education is presented later in this section.

Girls' characteristics of interest based on quantitative baseline data are summarized, followed by a table showing specific proportions of the populations with each characteristic.

1. **The MOI policy and the language girls report using with their family, friends, and teachers are disparate.**

   The baseline found that the majority of girls from grades 4, 6, and 8 speak Wolayttatto at home with their family and friends, and not English. While English is the medium of instruction from grade 5 onwards, 84.10 percent of grade 6 and 78.03 percent of grade 8 girls continue to speak Wolayttatto in the classroom with their teachers. Less than 1.00 percent of girls reported using Amharic at home, with friends, or with teachers. None of the girls surveyed in grades 6 and 8 reported using the MOI—English—at school with their teachers.

2. **Girl's self-esteem and well-being are associated with higher student- and school-level constructs of interest; yet, at least one-quarter of girls do not report high levels of well-being, and one-half do not report high levels of self-esteem.**

   More than one-third of comparison group girls and approximately one-quarter of treatment group girls responded "no" to at least one item asked about their well-being. Self-esteem items were asked during phase two, which were only administered to treatment group girls.

   Overall, higher well-being and higher self-esteem scores among girls were associated with higher scores for girls on three other scales: life skills, decision-making, and gender perceptions. This held true for treatment and comparison schools in the case of well-being.[95] Schools where girls

[91] To facilitate a higher score equating a more desirable score, some items are reverse coded; for instance, disagreeing with the statement may corresponded with a more positive gender perception.
[92] Key stakeholders included upper primary school female and male students, parents, female primary school teachers, school management personnel at the primary level, and zone and woreda-level education officials.
[93] This was discussed in the qualitative data both implicitly and explicitly by participants. Within the qualitative data and the broader project context early childbearing was mainly discussed or associated within the context of early marriage.
[94] A Population Council and United Nations Population Fund study noted the mean age of menarche is 14.8 years old for rural girls within their sample of seven regions—including SNNPR. Population Council and United Nations Population Fund. 2010. Ethiopia Young Adult Survey. A Study in Seven Regions. Addis Ababa: Population Council.
[95] To compute the life-skills scale mean score 12 items were used in grade 4 and 18 items were used in in grades 6 and 8. To compute the decision-making scale, grade 4 used six items and grades 6 and 8 used seven items. To compute the gender-perceptions scale, grades 4, 6, and 8, used seven items. See Annex 15 for individual items results and scale score means.

reported higher levels of well-being were also schools where boys reported higher levels of well-being. Where girls reported higher levels of well-being, schools also had higher scores on three scales of interest: curriculum design and implementation, girls support mechanisms, and community support.[96]

3. **Although the majority of girls report intentions to go to secondary school, their intentions and actual behavior differed when transition outcomes were tracked in the fall.**

   Based on current data, girls' high level of intention to enroll in secondary school is encouraging. Similarly, high aspirations were reported at the end of GEC1; however, the actual enrollment of grade 8 girls in grade 9 was lower than expected. At baseline, data show that the proportions of girls intending to enroll in the following year are comparable across all three grades. The high level of intention was mirrored within discussions of girls' educational aspirations in the upper primary girls FGD with grade 6 and 8 female students. The actual rates of transition from upper primary to lower secondary—i.e., grade 8 to grade 9—was lower than the associated level of intention.

Table 8: Girls' (and Boys') Characteristics

Construct of interest	Intervention (baseline)	Comparison (baseline)	Source (household and girls student survey)
Girls characteristics			
Orphans (percentage of total sample)[97] - Single orphans - Double orphans	To be collected at future evaluation points from beneficiaries 5 primary schools reported having 10 or fewer girls who are orphans, 10 schools reported having between 11–20 girls who are orphans	To be collected at future evaluation points from beneficiaries 3 primary school directors reported 10 or fewer girls who are orphans; 4 reported 11–20 girls who are orphans; 3 reported 21–30 girls who are orphans, and 5 reported more than 30 girls who are orphans	PCG_11g[98] PCG_13g Q70, Q73 school audit survey
Living without both parents (percentage of total sample)[99]	Girls: 50 (5.66%) Boys: 2 (0.87%) 6.74% of grade 4 and 8.06% of grade 6	Girls: 28 (3.34%) Boys: 7 (3.06%)	PCG_10g PCG_12g HH_number (girls student survey)

[96] To compute the mean scores for the gender-sensitive curriculum design and implementation scale, nine items were used; the girls support mechanisms used five items, and the community support scale used eight items. See Annex 15 for individual items results and scale score means.

[97] In the school audit survey, respondents were asked if there were girls who are orphaned in their school. At the primary level, all respondents—one school director per school—said "yes." Respondents were also asked to give an approximate number of orphaned girls in the school. These data are aggregated to provide an estimate of the total number of orphaned girls in the sampled primary schools; the minimum number of orphaned girls reported by a primary school director was 8 and the maximum 75. At the eight sampled treatment secondary school, nearly all respondents—school directors and guidance counselors—say "yes;" when asked if there were girls who were orphaned in their school. The one exception was a single guidance counselor who respond ended they did not know; however, the school director responded "yes" so it can be assumed there are girls who are orphans in all STAGES secondary schools. When asked to approximate the number of orphaned girls in their secondary schools, the minimum number of orphaned girls reported in a secondary school was 7, and the maximum was 35. These results cannot be aggregated up to the sample/beneficiary population because the data were not collected at the individual level. During the baseline though it was determined proxy measures were insufficient to inform estimates, and therefore, at future evaluation points, girls will be asked directly about their orphan hood status.

[98] The crossed-out font indicates were the source utilized diverges from the GEC-T template. For example, the GEC-T template recommends utilizing questions 11 and 13 from the parent/caregiver household survey to determine the number of girls who are orphans; however, as this survey was not administered during this time, an alternative source was used—questions 70 and 73 on the school audit survey.

[99] Girls and boys were asked to report who lives with them in their household. If the student did not select mother and did not select father, then they are reported here as living without both parents.

Construct of interest	Intervention (baseline)	Comparison (baseline)	Source (household and girls student survey)
	parent/caregivers reported father is no longer alive		
Living in female-headed household percentage of subsample of parents)-	Grade 4/5 parents: 49 (70.00%) Grade 6/7 parents: 31 (60.78%)	n/a; to be collected	HH_8
Married (percentage of total sample)[100]	n/a; proxy data to be collected	n/a; proxy data to be collected	~~PCG_22g~~
Mothers (percentage of total sample)[101] - Under 18 - Under 16	n/a; to be collected at future evaluation point	n/a; to be collected at future evaluation point	PCG_23g
Poor households (percentage of total sample)	351 (39.71%) * of girls report living in household unable to meet basic needs 159 (85.48%) parents report having at least one day they've gone without cash income 88 (47.31%) of parents report having access to clean water every day 116 (62.36%) parents report that it is difficult to afford their daughters tuition	201 (23.99%) girls report living in a household unable to meet basic needs	PCG_7enr PCG_10econ H_1 (girls student survey) PCG_8econ

[100] Girls were asked to report who lives with them in their household. If girls selected either husband, mother-in-law or father-in-law, they may be classified as married. Using this approach, the proportion of girls that were identified as being married was notably low—17 girls in treatment schools (1.92 percent) and 21 girls in comparison schools (2.51 percent). Since this may be a severe underestimate of married girls in the population, it is not reported in the table, but provide here as context instead. However, due to the sensitive nature of the topic of marital status, as well as concerns around unintended harms related to disclosures related to early marriage, the project and external evaluators will continue to collect proxy data on marital status through the item on who lives with them in their household as well as explore other sensitive and creative mechanisms for collecting data on marital status by continue working with existing school structures—including the GEAC and PTSAs—to further inform estimates on the number of direct beneficiaries who are married. Additional contextual data on prevalence of married girls within sampled primary school was also collected. When sample cohort treatment girls were asked if any girls in their current class were married, 14.57 percent of grade 4/5, 24.88 percent of the grade 6/7 and 20.22 percent of the grade 8/9 cohort said 'yes.' For those who said yes, when asked how many girls in their class were married, the minimum number reported was one girl across the all grade-level cohorts, and a maximum of three girls for the grade 4/5 and 6/7 cohorts, and a maximum of five for the grade 9/10 cohort.

[101] Due to the sensitive nature of this question, motherhood and pregnancy status were not directly asked of girls during the baseline. Instead, girls and other stakeholders were asked about the prevalence rates within their schools and communities. When girls were asked if there are any girls who are mothers in their current class this year, 17.41 percent of grade 4/5 girls, 31.34 percent of grade 6/7 girls and 30.34 percent of grade 8/9 girls indicated yes. Of those who reported yes, when asked how many girls in their current class were mothers, the minimum reported was one girl and maximum five girls in the grade 4/5 cohort, whereas in grade 6/7 the lowest was also one girl who was a mother but the highest reported eight girls; in the grade 8/9 cohort, half of the respondents reported they knew of one girl in their class who was a mother and two reported that five girls were mothers. In addition, within sampled treatment secondary schools, six out of eight secondary school directors indicated they are aware of girls that are pregnant or expecting in their school; when asked to approximate the number of pregnant girls, figures ranged from 0 to 10 girls within a secondary school according to school directors and guidance counselors surveyed. About half of secondary school teachers survey (53.85 percent) also indicated that they teach girls that are pregnant or expecting in their classes (that they are aware of). During the baseline though it was determined proxy measures were insufficient to inform estimates, and therefore, at future evaluation points, girls will be asked directly about their motherhood status, but not directly about pregnancy status.

Construct of interest	Intervention (baseline)	Comparison (baseline)	Source (household and girls student survey)
Language difficulties[102] - MOI different from mother tongue (percentage of total sample) - Girl does not speak MOI (percentage of total sample)	Grade 4: 283 (97.25%) speak Wolayttatto (MOI) with teachers, at home, and with friends Grade 6: 300 (100.00%) speak Wolayttatto (MOI) with teachers, at home, and with friends Grade 8: 294 (100.00%) speak Wolayttatto (MOI) with teachers, at home, and with friends	Grade 4: 279 (93.31%) speak Wolayttatto (MOI)with teachers, at home, and with friends Grade 6: 282 (100.00%) (MOI) speak Wolayttatto with teachers, at home, and with friends Grade 8: 272 (100.00%) (MOI) speak Wolayttatto with teachers, at home, and with friends	~~PCG_2enr~~ ~~PCG_3enr~~ Lan_1, Lan_2, Lan_3 (girls student survey)
Parental education—Primary caregiver has no education ((percentage of subsample of parents)	Grade 4/5 parents: 41 (38.32%) Grade 6/7 parents: 23 (29.11%)	n/a, not to be collected	HH_13 ~~PCG_6~~
Girl does not report a high level of well-being (percentage of total sample)[103]	289 (32.69%)	332 (39.62%)*	Three additional well-being items on girls student survey
Overage for grade	Grade 4: 62 (22.63%)* Grade 6: 49 (18.08%) Grade 8: 39 (14.94%)	Grade 4: 35 (14.88%) Grade 6: 36 (14.81%) Grade 8: 35 (15.35%)	Age
Intends to enroll in secondary school after completing grade 8 (percentage of total sample)	860 (97.29%)	812 (96.90%)	Two additional intention items on girls survey
Boys characteristics			
Living without both parents (percentage of total sample)	2 (0.87%)	7 (3.06%)	~~PCG_10g~~ ~~PCG_12g~~ HH_number (boys student survey)
Boys from poor households (percentage of total sample)[104]	81 (35.37%)	57 (24.89%)	H_1 (boys student survey)
Boy does not report a high level of well-being (percentage of total sample)[105]	58 (25.33%)	61 (26.64%)	Three additional well-being items on boys student survey

Note: Samples are reported from all students with data on evaluation surveys. Subsequent analyses with learning assessments are based on fewer students who had both learning and survey data. Significance tests were limited to those groups with sufficient n's presenting the characteristic. Significance test results are reported at the 0.05 level.

[102] Girls were asked which language they primarily use at home with their family, with friends at school, and with their teacher. The language or MOI differs by grade. In grade 4, the MOI is Wolayttatto; in grades 6 and 8, the MOI is English.
[103] This was indicated by girls' responding "no" to at least one of three well-being questions. Three items on well-being were asked on the girls student survey: Wb_1: Were you happy the last time you were at school; Wb_2: Did you learn or do something interesting the last time you were at school? And Wb_3: Did you have enough energy to get things done the last time you were at school? Response options were "yes," "no," and "I don't know." The well-being scale was computed based on the number of items the girl responded "yes" (range: 0–3); the "proportion of girls who do not report a high level of overall well-being" is based on the proportion of respondents who selected "no" for at least one of the three well-being items.
[104] This was defined as boys responding that their households are unable to meet basic needs on the boys student survey.
[105] This was indicated by boys' responding "no" to at least one of three well-being questions. Three items on well-being were asked on the boys student survey.

Barriers to education

As in the previous section, barriers to education are summarized first from the qualitative analysis, followed by the quantitative.

The majority of the barriers identified in the qualitative data were within the context of girls' families and communities and addressed issues of gender-related social norms, attitudes, and practices. Some barriers to the school and learning context were also noted. The key findings are summarized, and a more detailed analysis and discussion of the qualitative data can be found in Annex 16.

- While the degree to which a **high burden of household chores** impedes a girls' ability to attend and thrive in school may vary, it was clear across the FGDs and KIIs that it continues to be a major challenge for primary girls. The burden often increases along with a girls' age.[106] Woreda officials were asked about whether community mobilization was taking place to decrease girls domestic work burden. 40 percent (n=6) of woreda officials from treatment woredas said yes; 46.67 percent (n=7) said partially, and 13.33 percent (n=2) said no.
- KIIs highlighted **girls' participation in income-generating activities** as having a negative impact on girls' access to education and school attendance—especially during market days. Girls from families with high levels of poverty, as well as those in secondary school, were most impacted.
- **Early marriage**—and its association with high levels of school drop-out related to community social norms and pressure, husband's demands, or household responsibilities as a wife and potentially mother—was another major barrier to girls' education noted across qualitative respondents, with numerous key informants suggesting that the practice was a widespread phenomenon throughout the Wolaita Zone. However, some FGDs noted exceptions; in fact, some of the participants from upper primary school girls noted that they were married and that their husbands supported their education.[107]
- Girls' struggles to consistently engage in their education due to **menstruation** was discussed in varying degrees by respondents.
- The majority of respondent types cited rural to urban migration as well as international **migration**—especially at the end of primary school and the beginning of secondary—as a barrier to girls' education as well as a mechanism to improve students' financial status.
- Key informants readily identified **social and cultural norms rooted in gender inequality** as a major barrier to girls' education. Responses mentioned the prioritization of boys' education, lack of utility given to educating girls, and the devaluing of women or girls in society as known norms. While teachers and parents did not identify gender inequality by name, several respondents described difficulties girls face in navigating traditional cultural norms and accessing education. Parents mentioned how preference was given to male children to attend school because girls have the option to better themselves through marriage.
- While limited data and mention of **SRGBV** directly were noted in the qualitative data, this should not be interpreted as a lack of SRGBV issues within the context.[108] Individuals may be hesitant to discuss or disclose information regarding SRGBV due to stigma or concern of retaliation and/or punishment.

[106] Proportions of girls reporting chores as a burden in the girls student survey in treatment schools was 7.56 percent, 12.37 percent and 16.29 percent in grades 4, 6, and 8 respectively. Among comparison schools, however, the proportion of girls reporting a high chore burden was comparable across grades, at 8.36 percent, 10.14 percent, and 7.61 percent respectively. The level of chore burden for girls in secondary school was not collected at baseline and, therefore, cannot be triangulated with qualitative findings.
[107] While abduction and rape were not asked about extensively in surveys, school audit included a question about whether the school collaborates with community to conduct awareness raising to prevent harmful traditional practices such as early marriage, abduction and rape. two-thirds of respondents in treatment woredas (66.70 percent, n=10) and almost two-thirds in comparison woredas (60.00 percent, n=9) said yes; one-third in both groups said partially (33.33 percent, n=5) and one respondent in comparison and no respondents in treatment said no.
[108] In addition, data collection and research surrounding SRGBV is quite challenging given the sensitiveness of the topic; as such, it requires a strong, ethical methodological approach and protection protocols to ensure a do-no-harm approach is maintained. Therefore, the evaluators determined it would be best to post-pone more in-depth investigation of the topic within the project schools until a later point in the evaluation when a more purposeful, focused approach can be employed.

In addition, some forms of SRGBV, such as corporal punishment or harassment, may be considered as commonplace and internalized as acceptable.[109] Moreover, analysis from the evaluation surveys indicates one form of SRGBV—corporal punishment—persists within the school environments.

- **Abduction** was mentioned by several teachers as an issue when discussing barriers to girls' education.[110]
- While the **distance to secondary schools** was noted as a challenge by key informants, this was not discussed as a major concern at the primary level.
- Lack of resources, accommodation, or specialized training to support **girls with disabilities** was also noted as a barrier.
- **Girls who are orphans** were noted as less likely to attend secondary school and received limited, targeted interventions or support.
- **Lack of resources and poor school infrastructure** were commonly cited by parents and teachers as barriers to girls' education.

Potential barriers to learning and transition were explored within the evaluation survey data. As required in the guidance, the following barriers were examined at the home and community levels: safety, parental or caregiver support, and community support. Within these broad areas, specific factors affecting girls learning were explored—some factors align with those suggested by the guidance, while other factors draw on trends observed in the data. At the school level, barriers related to attendance, school facilities, and teachers support were considered. Where relevant, the proportion of girls—and boys—in the sample who face identified barriers is reported in Table 9. Note that proportions of girls facing each barrier are summarized across the three grade levels in the sample; disaggregated data are included in the IOs section, as appropriate.

Table 9: Potential Barriers to Learning and Transition

	Intervention (baseline)	Comparison (baseline)	Source
Sample breakdown (girls)			
Home and community level			
Safety			
Fairly unsafe or very unsafe travel to schools in the area (percentage of a subsample of parents)	G4/5 parents: 23 (21.50%) G6/7 parents: 15 (18.99%)	n/a	PCG_9
Doesn't feel safe traveling to or from school (percentage of total sample)	83 (9.39%)*	49 (5.88%)	CS_W13s
Parental or caregiver support			
Sufficient time to study: High chore burden - Girls whose parents **never or sometimes** decreased household chores to accommodate schoolwork (percentage of total sample) - Percentage of parents/caregivers who reported girls have high	 Grade 4: 205 (74.82%)* Grade 6: 181 (66.79%) Grade 8: 174 (66.67%) Grade 4/5 parents: 44 (41.90%) Grade 6/7 parents: 32 (41.03%)	Grade 4: 159 (67.09%) Grade 6: 165 (67.90%) Grade 8: 149 (63.35%) n/a, will not be collected	H_4 girls student survey PCG_26g

[109] Jenny Parkes et al., *Addressing School-Related Gender-Based Violence in Côte d'Ivoire, Togo, Zambia and Ethiopia: A Cross-country Report* (New York: UCL Institute of Education, 2017), http://discovery.ucl.ac.uk/10027909/7/Parkes_Addressing%20SRGBV%20Cross%20Country%20report_2017.pdf.
[110] In the context of the Wolaita Zone, abduction was noted by teachers as a form of escalating harassment of female students by male students. It may also be associated with 'arranged' or 'forced' marriage."

	Intervention (baseline)	Comparison (baseline)	Source
chore burden[111] (percentage of subsample of parents)			
Does not get support to stay in school and do well[112] - girls in primary grades (percentage of total sample)[113]	29 (3.28%)	38 (4.53%)	~~HHG_7~~ SC_1 (girls student survey)
Low level of support from the household (percentage of total sample)[114]	195 (22.06%)*	120 (14.32%)	Household Support Scale (items H2, H3, H4, H5 from girls student survey)
Community support for girls' education			
Gap in attitudes towards girls' and boys' education: - Feel boys should go to primary school but had below average attitudes to girls' education (percentage of woreda officials surveyed)	Woreda officials: 29 (35.8%) *	Woreda officials: 8 (35.0%)	Q1, Q3, and Attitudes towards girls' education Scale (six item-scale from woreda officials survey)
- Feel boys should go to secondary school but had below average attitudes to girls' education (percentage of woreda officials surveyed)	Woreda officials: 32 (38.1%)	Woreda officials: 8 (32.0%)	
School level			
Attendance			
Attends school at least **half the time** (percentage of total sample)[115]	857 (96.95%)	723 (86.28%)	~~PCG_6enr~~ Attendance item (girls survey)
Attends school **less than half time** (percentage of total sample)[116]	27 (3.05%)	115 (13.72%)*	~~PCG_6enr~~ Attendance item from the

[111] High chore burden is understood as girls typically spending half of the day or the whole day doing chores.
[112]This was defined as the percentage of girls responding that their parents do not want them to continue to the next grade.
[113] The relevant percentage of girls in secondary grades will be included in the next evaluation point.
[114] This was determined by the percentage of girls whose score on the household-support scale was one standard deviation or lower than 1.02 on a four-point scale below mean. The household-support scale is based on four items asked to girls. The items included: do your parents or caregivers pay for everything that you need so that you can go to school, do your parents or caregivers encourage you to go to school, do your parents or caregivers decrease your household chores so that you can do your school work, do your parents or caregivers encourage you to marry rather than pursue your education. The scale score for each girl was computed as the mean across these four items; higher scores on the scale correspond with higher levels of household support.
[115] This was determined by the percentage of girls who report attending school at least three days in the past week on the girls student survey.
[116] This was determined by the percentage of girls who report attending school less than three days in the past week on the girls student survey.

	Intervention (baseline)	Comparison (baseline)	Source
			girls student survey
Missed at least one day of school last week (percentage of total transition sample)	G4: 100 (58.51%) G6: 143 (65.35%) G8: 131 (73.15%)	n/a	Transition survey
Does not feel safe at school (percentage of total sample)	54 (6.13%)	36 (4.32%)	CS_W14s
Primary School facilities			
No seats for all students (percentage of total sample)	62 (7.01%)	54 (6.44%)	CS_W5s
Difficult to move around school (percentage of total sample)	43 (4.88%)	30 (3.59%)	CS_W6s
Does not use drinking water facilities (percentage of total sample)	739 (83.60%)*	396 (47.26%)	CS_W7s
Does not use toilet at school (percentage of total sample)	27 (3.05%)	3 (0.36%)	CS_W9s
Does not use areas where children play or socialize (percentage of total sample)	60 (6.79%)	69 (8.23%)	CS_W11s
Primary and Secondary Teachers			
Disagrees teachers make them feel welcome (percentage of total sample)	28 (3.17%) of girls	27 (3.23%) of girls	CS_WA
Agrees teachers treat boys and girls differently in the classroom (percentage of total sample)	665 (75.23%) of girls	661 (79.26%) of girls	CS_1s
Agrees teachers often absent from class (percentage of total sample)	431 (48.81%) of girls	434 (51.91%) of girls	CS_2s
High corporal punishment exercised by the teacher (percentage of total sample)[117]	276 (31.22%) of girls	496 (59.19%)* of girls	Discipline scale includes seven items TQ_6s, TQ7sa, TQ_7sb, TQ_7sc, TQ_7sd, TQ_8s_recode, TQ_9s_recode from the girls student survey

[117] This was determined by the percentage of girls reporting yes" to at least two out of six questions on disciplinary actions by students This scale includes the following six items: Do your teachers discipline or punish students who get things wrong in a lesson, my teachers discipline students with physical punishment, my teachers discipline students with shouting, my teachers discipline students with detention, in the past week did you see a teacher use physical punishment on other students? (includes students who responded "once or twice" or "almost every day"), in the past week did a teacher use physical punishment on you? (includes students who responded "once or twice" or "almost every day"). Of these six items, only one asks about punishment that is not physical or verbally inappropriate (detention). This item is included here because it was part of the scale and does not skew the proportion of girls who reported yes to at least two of the six items because when they responded yes to the detention item, they also responded yes to at least two other punishment items. Therefore, the cutoff of two items does not inadvertently suggest that detention is corporal punishment.

	Intervention (baseline)	Comparison (baseline)	Source
Gaps in teachers self-reported gender perceptions (percentage of total sample)[118]	14 (15.73%) of primary teachers 0% of secondary teachers	13 (14.29%) of primary teachers No data collected from secondary teachers	Gender-perception scale includes seven items, Q9–Q16, from the classroom teachers survey
Gap in attitudes towards girls' and boys' education[119] - Feel boys should go to **primary** school but had below average attitudes to girls' education (percentage of total sample)	20 (23.81%) of primary teachers 0% of secondary teachers	27 (29.67%)	Q1, Q3, and attitudes towards girls' education scale, includes six item-scale from the classroom teachers survey
- Feel boys should go to **secondary** school but had below average attitudes to girls' education (percentage of total sample)	20 (23.82%) of primary teachers 0% of secondary teachers	26 (28.89%)	
Secondary School Guidance Counselors			
Gaps in guidance counselor training in gender sensitivity and mainstreaming[120]	3 (42.86%) of secondary school guidance counselors	N/A, to be collected	GC_Training_1
Gaps in guidance counselors training in socio-emotional learning or para-counseling[121]	5 (71.43%) of secondary school guidance counselors	N/A, to be collected	GC_Training_2
Sample breakdown (boys)			
Home and community level			
Safety			
Does not feel safe traveling to or from school (percentage of the total sample)	3 (1.31%)	7 (3.06%)	CS_W13s
School level			
Attendance			
Does not feel safe at school (percentage of total sample)	1 (0.44%)	3 (1.31%)	CS_W14s
School facilities			
Difficult to move around school (percentage of total sample)	7 (3.06%)	6 (2.62%)	CS_W6s
Does not use toilet at school (percentage of total sample)	1 (0.44%)	1 (0.44%)	CS_W9s

Significance tests were limited to those groups with sufficient n's presenting the characteristic. Significance test results are reported at the 0.05 level.

[118] This was determined by percentage of teachers whose score on the gender-perception scale was one standard deviation below mean (<2.27).
[119] This was determined by the proportion of teachers who support boys going to primary and secondary but have below average attitudes towards girls' education.
[120] Proportion of secondary school guidance counselors who responded "no" to having received training in gender sensitivity and mainstreaming in the past year.
[121] Proportion of secondary school guidance counselors who responded "no" to having received training in socio-emotional learning or para-counseling in the past year.

3.4 Intersection between Key Characteristics and Barriers

The intersection between key girls' characteristics and barriers to education is examined in this section. First, the intersection between key characteristics and barriers that emerged from the qualitative data will be explored, followed by findings from the evaluation surveys.

Examples of barriers to education by characteristic

Table 10 examines dichotomous characteristic and barrier variables as well as the proportion of primary school girls who are categorized within both. In other words, groups of girls who demonstrated both the characteristics and were affected by the barrier are shown in Table 10 as a proportion. Following **Error! Not a valid bookmark self-reference.** is an examination of continuous variables—such as underlying constructs or scales—that cannot be presented as a proportion of students. In this case, it is more appropriate to examine the relationship between variables.

Table 10: Examples of Barriers to Education by Characteristic

Barriers	**Characteristic (percentage of girls in treatment primary schools)[122]**				
	High corporal punishment exercise by the teacher[123]	**Girls' language at home, in school with friends, and with teachers is different from MOI**	**Low levels of household support**	**Low life-skills score**	**Low decision-making score**
Teacher level					
Teacher treats boys and girls differently	35.18%	11.54%	18.21%	12.78%	13.24%
Teacher often absent from class	18.78%	9.05%	14.25%	10.29%	9.62%
School level					
Not using drinking water facilities	26%	12%	18%	13%	15%

The following section examines the intersection of the barriers and girls' characteristics for primary only. Specifically, bivariate correlations between constructs—variables and scales—at the school and grade level were examined.[124] Analyses were not done at the secondary level because data for girls in grade 9 were collected when they were still in primary school (grade 8).

[122] All related figures for this table are drawn from the girls student survey.
[123] This was determined by the percentage of girls reporting "yes" to at least two out of six questions on disciplinary actions by students on the girls student survey.
[124] For example, each school has a grade-level set of scores for all girls student scales, boys student scales and classroom teachers scale in that school or grade, as well as school-level audit scales and woreda-level staff scales
The relationships between barriers to education and girls' characteristics are explored at the aggregate level; data were aggregated to the grade level across girls student surveys and classroom teachers survey to examine relationships between girls and teachers survey responses. Similarly, relationships at the home-community level are explored by examining responses from girls, teachers, and woreda staff at the school and grade level; woreda staff's results are attributed to all grades and schools within their woreda.

Teacher-level

1. **Primary teacher attendance appears to be a challenge, with slightly more than half of the girls (see Table 10) in both the comparison and treatment groups reporting that their teachers are often absent from class.**[125]
    - High rates of teacher absenteeism were associated with lower rates of girls reporting that they will go on to secondary school after grade 8.
    - High rates of teacher absenteeism were associated with higher reports of teachers treating boys and girls differently in class.
    - Schools with high teacher absenteeism, as reported by girls, also had lower scores on school-level scales, including teacher opportunities and motivation, gender policy or mainstreaming, support for gender advisory committees, gender-sensitive curriculum design and implementation, girls support mechanisms, community support, school gender friendliness, and support for girls' clubs.
2. **Primary teachers' reports of frequent disciplinary actions appear to be high overall but lower in treatment schools than in comparison schools.**
    - Twice as many girls in comparison primary schools reported more than two disciplinary actions compared with girls in treatment schools. Despite the lower rates of reported disciplinary actions, one-third of girls in treatment primary schools also reported more than two disciplinary actions.
    - High rates of disciplinary actions were associated with lower rates of girls reporting that they will go to secondary school after grade 8.
    - High rates of teachers' disciplinary actions were associated with higher reports of teachers treating boys and girls differently in class.
    - Teachers' attendance was lowest where disciplinary actions were highest.
    - Schools where teachers exercised the most disciplinary actions were also schools where more students, reported that the MOI was different from the language they spoke at home, at school with friends, and with a teacher.
    - Where teachers' disciplinary actions were highest, gender perceptions among girls were lowest, as were scores on the life-skills and decision-making scales.
    - Where disciplinary actions were highest, teachers also reported poorer levels of gender-sensitive curriculum design and implementation.
3. **Less positive attitudes towards girls' education appear to be a school-level issue than an individual teacher-level issue.**
    - Examining the distribution of primary teachers with below-average scores on the attitudes-towards-girls scale across schools, results show that at least half of the treatment schools had two or more grade-levels where teachers reported below-average attitudes towards girls' education; two treatment schools had teachers with below-average attitudes towards girls' education in all three grade levels. In other words, when less positive attitudes towards girls' education were reported, they were not isolated to specific teachers.
    - While the average score on the attitudes-towards-girls scale was 2.79—a score of 3.00 indicates very positive attitudes towards girl education—the lowest scores for any primary teacher approached 1.00, which indicates negative attitudes towards girls' education.[126] The

[125] This was determined by female students who responded either "agree a lot" or "agree a little" to the statement.
[126] At least two teachers had a score of 1.00 on the scale; 35 teachers had a score of 2.50 or lower on the scale indicating that they agreed, on average, with the six items included in the scale but did not "agree a lot."

influence of teachers who held negative attitudes, and their particular role in the school, should be examined further.

- Almost one-quarter of primary teachers have a gap in their perceptions regarding boys' and girls' education. Specifically, almost one-quarter of teachers reported that they believed boys should go to primary or secondary school but had below-average scores on the attitudes-towards-girls-education scale. In other words, one-quarter of teachers supported boys attending school but did not have a positive attitude towards girls' education.

Home-community level

4. **In all woredas, at least one in three woreda staff members had a score on the attitude-towards-girls-education scale that was lower than 2.68.**

   Specifically, of the 116 woreda officials surveyed, at least one-third had scores on the attitude-towards-girls-education scale that was below the average score across all woredas.[127] This woreda-level gap is in addition to the gap in attitudes seen at the teacher-level.[128]

5. **Gender perceptions across groups are inconsistent.**

   Overall, primary schools that had, on average, positive gender perceptions among girls were schools where boys also held positive gender perceptions.[129] No relationship was observed between girls' gender perception and teacher's gender perceptions. Relationships were examined at both the school and grade level.

6. **When girls do not have high levels of household support, they have lower scores on the student- and school-level constructs of interest.**

   Approximately one in six girls in comparison primary schools and one in four girls in treatment primary schools reported relatively low levels of household support.[130]

   - Schools where girls reported higher levels of household support also had higher scores on the gender policy or mainstreaming, girls' advisory committee, gender-sensitive curriculum design and implementation, girls' support mechanisms, and girls' clubs scales.
   - At the individual level, higher household support was related with intentions to go to secondary school, lower teacher absenteeism, higher teacher focus on equitable classroom practices, higher well-being scores among girls, more positive gender perceptions among girls and boys, and higher life-skills scores among girls.[131]
   - Schools where girls reported higher levels of household support were within woredas where officials reported more positive attitudes towards girls and lower gender perceptions. There was no relationship between girls' reports of household support and teachers' gender perceptions.
   - Examining treatment schools alone, the majority of girls reported that their chores were not decreased by their families in order to allow time for school work.[132] While this indicates a high level of chore burden for girls, it should be interpreted along with their overall household

[127] The attitude-towards-girls-education scale includes six items, some of which are reverse coded. The items include the following: girls should go to primary school, girls should go to secondary school, girls need to go to school every school day of a month, education is more important for boys than for girls (reverse scored), girls who get pregnant while still at school should be allowed back in school, and boys' education should get preference when money is scarce (reverse scored). While the average across all woreda officials was 2.68 on a three-point scale, scores were as low as 1.00, which indicates negative attitudes towards girls' education.
[128] Differences by woreda officials' gender were examined but must be interpreted with caution; only eight treatment woreda officials surveyed were female and only two in comparison the woreda. There were no differences by gender on their attitudes towards girls' education.
[129] However, these schools were in woredas where, on average, woreda officials' gender perceptions were less positive.
[130] Low levels of household support are defined as a mean score on the household-support scale that is one standard deviation below the overall mean.
[131] For example, girls reporting teachers ask questions—including harder questions—to boy and girls equally in the classroom.
[132] This includes girls who reported that their chores were never or sometimes decreased to allow them to do school work.

support.[133] While chore burden affects the majority of girls, it is part of overall household support. Specifically, the data show that girls who had higher levels of household support also reported higher levels of teacher support and had higher scores on the life-skills scale.[134] The relationship of overall household support, therefore, to other desirable outcomes, is as important to underscore as is the high level of chore burden for girls in treatment schools.

- Schools where girls reported having higher household support were also schools where boys had higher scores on the gender-perception, household-support, and life-skills scales.
- Schools where girls reported higher rates of household support were also schools with higher scores on the following school-level scales: gender policy or mainstreaming, curriculum design and implementation, girls support mechanisms, school gender friendliness, and girls' clubs.

3.5 Appropriateness of Project Activities to the Characteristics and Barriers Identified

The following questions, posed in the Baseline report template, are addressed using the analysis of characteristics and barriers presented earlier in this chapter.

1. **Does the prevalence of characteristics or subgroups of characteristics identified by the project as part of their beneficiary mapping correspond with the data collected by the evaluator through the household survey and school survey?**

   The prevalence of characteristics and barriers corresponds with the information on which the project interventions are designed. Beneficiary mapping suggests that girls have high levels of marginalization, particularly if they are pregnant, have disabilities, or are orphans. While data on these specific characteristics are limited, other characteristics suggest that both physically apparent and non-physically apparent challenges are present for the target population. When transition data were collected in phase two, additional triangulation of prevalence of disabilities was conducted.[135] Specifically, the use of the Child Functioning questions was found to be more appropriate but renders the results only applicable to the girls who returned to school in the fall, rather than the complete sample from phase one.

2. **Are there any other characteristics or subgroups revealed through the baseline data collection that may be at risk of educational marginalization that is not mentioned or supported specifically by the project?**

   The characteristics of girls at risk that emerged out of the baseline data that affected large proportions were less physical in nature; most were intangible barriers. While these barriers are mentioned in the theory of change provided by the project, they are worth noting again because they emerged more strongly as barriers than observable characteristics alone.[136, 137]

   - Perceptions and levels of support experienced by girls emerge as a greater source of risk to educational marginalization than any other observable characteristic or resource-based barrier.
   - The aggregate effect of perceptions at the individual-, classroom-, school-, and woreda-level should be specifically identified for additional support.

[133] Household support scale includes parental support to pay for school, encouragement to go to school, reduction in chores, and encouraging girls to marry instead of pursuing their education—provides an important and broader context
[134] Teacher support items include: teacher makes them feel welcome, teacher treats boys and girls equally, teacher asks harder questions to boys and girls equally. Life-skills scale items include 12 items for girls in grade 4 and 18 for girls in grades 6 and 8.
[135] Please see MEL Framework regarding STAGES evaluation approach to household surveys.
[136] For example, addressing low levels of life skills and gaps in perceptions towards girls.
[137] For example, girls experiencing high levels of migration, as well as girls who experience lower levels of support and have lower levels of readiness to learn.

- In addition, qualitative analysis elucidated the need to take some additional characteristics into consideration such as girls who are at-risk, for early marriage, migration, or abduction.

3. **Do the most prevalent barriers identified by the analysis correspond with the projects' theory of change?**

   The most prevalent barriers faced by girls—including gaps in perceptions, corporal punishment, and low levels of well-being and life skills—are within the project's theory of change. The revised Theory of Change addresses many of the gaps identified in baseline phase one; therefore, no additional revisions are suggested to the Theory of Change.

4. **Do the project interventions address the key barriers for the key subgroups?**

   Based on the theory of change, the project's interventions address the key barriers for the subgroups. The analyses presented in this chapter suggest nuances to the barriers; for example, teacher support may need to focus specifically on teacher absenteeism in the classroom as an issue apart from teachers' present on school grounds. Furthermore, all STAGES interventions are being strengthened regarding the inclusion of girls with disabilities and other categories of marginalization.

Box 2: Project's contribution [Link Community Development]

Response to sample characteristics. Overall, we do not feel the characteristics are very different from the original project mapping through the baseline shines more of a light on some.

One difference which slightly surprised us is the finding around the level of corporal punishment as this did not come out strongly in previous analysis and mapping. We will address this across the project, for example within our support to teacher training, as well as with government partners, and community/school structures (PTSA/SIC, Mother and Father Groups and will be specifically addressed in SRGBV awareness raising and training and via the Teachers' Code of Conduct. Safeguarding cuts across all of these trainings and interventions.

The percentage of girls stating some kind of disability was very high at Baseline I compared to other GEC-T projects (22.45 percent). This was not visible in previous project mapping and therefore warranted re-examination. During phase two of the baseline, when transition data was collected December 2018, a more nuanced and detailed set of 'child-functioning' questions were used with children and the percentage of girls stating disability enormously reduced to 6 percent in grade 6 and 4.5 percent in grade 8. Whilst we had begun to make changes across the program based on the initial disability findings, we will not divert from our response as a result of the new findings. We had already observed for example that whilst teacher training materials and delivery were strongly tilted towards gender-responsiveness, they could be strengthened to help meet the needs of girls (and boys) who faced additional or multiple barriers, particularly disability. Providing teachers with the skills to identify children who are struggling to attend, participate and learn, for whatever reason, and to respond to the diverse range of needs of children in the classroom is good practice for any program.

Our approach to gender and inclusion generally is 'twin-track.' It works on broad system and school-wide improvements for all at the same time as providing specific support for learners, for example, learners with a disability or who face other additional barriers, using child-centered pedagogy. It means focusing on the changes that are needed to make education easier for everyone, at the same time as considering the specific changes needed to make sure that girls, orphaned/vulnerable children, and children with disabilities are not excluded.

Theory of Change. Overall, we feel the project's theory of change corresponds with the key barriers. It remains more or less the same, except that it is strengthened across all outputs for inclusion and safeguarding and this includes corporal punishment, evidenced internationally to have a hugely detrimental effect on all aspects of girls' education.

4. Key Outcome Findings

4.1 Learning Outcome

Learning outcome data were gathered using literacy and numeracy assessments for girls in grades 4, 6, and 8 in treatment and comparison schools. The numeracy assessments were conducted in the MOI for each grade. The literacy test was conducted in two languages for girls in grades 4 and 6—English and Wolayttatto—and in one language for girls in grade 8—English.[138]

The literacy assessments included subtasks typically found in the EGRA for primary school students and a newly developed assessment—the SeGRA—for girls in grade 8. Similarly, the numeracy assessment included subtasks typically found in the EGMA for primary school students and a newly developed assessment—the SeGMA—for girls in grade 8.

The total number of subtasks administered and used in aggregate scoring is shown, by grade, in Supplementary Table 8. Total Subtasks The next section discusses the decision to exclude one subtask from both the Wolayttatto-medium EGRA and English-medium EGRA administered in grades 4 and 6, which resulted in a total of eight subtasks included in the required aggregate scoring.

Supplementary Table 8. Total Subtasks Administered to Students by Grade

	Grade 4	Grade 6	Grade 8
Number of literacy subtasks administered	10	10	6
Number of literacy subtasks included in an aggregate score	8	8	6
Number of numeracy subtasks administered	5	6	6
Number of numeracy subtasks included in the aggregate score	5	6	6

Exclusion of the Letter Sound Identification Subtask from Aggregate Scores

The letter sound identification subtasks on the EGRA for grades 4 and 6 were excluded from the aggregate score calculations due to training and data collection issues that were confirmed during data analysis. This affected both English- and Wolayttatto-medium assessments. First, when collecting feedback from education experts and enumerators at the end of each day of training, enumerators noted difficulty while scoring the letter sound subtask. Specifically, they noted that the marking of a correct response to an item often differed by the enumerator and stemmed from disagreements about the actual correct letter sound. Despite an attempt to review letter sounds with the group, limited time did not allow the trainers to arrive at a consensus on the correct response to each item. At the time of data collection, enumerators also noted that there was a further confusion with the subtask instructions among students who are more familiar with providing letter names rather than letter sounds. Because this subtask was not piloted, it was not possible to have identified this issue prior to operational data collection. Instead, evaluators decided that baseline data would be reviewed to identify the quality of the results and examine the relationship between students' performance on this subtask and their overall aggregate score. Ideally, the relationship between the subtask—as with all subtasks—and the aggregate score should be strong and positive.

Baseline data from grade 4 and 6 students show that the relationship between the letter sound subtask score and a student's aggregate score was weak (see Annex 17). Furthermore, exclusion of the subtask yielded a Cronbach's Alpha value that was notably improved—closer to 1.00—for the aggregate score.[139]

[138] Within the Wolaita Zone, the MOI is Wolayttatto in grade 4 and transitions to English in grade 5 onwards.
[139] Cronbach's alpha was computed on the set of subtasks included in the aggregate score; with and without the letter sound subtask. Cronbach's alpha is a measure of reliability and ranges from 0 (poor reliability) to one (good reliability).

This suggests that the subtask did not contribute to the overall score in the same way as all other subtasks. Based on these results, the letter sound identification subtask was removed from consideration in the aggregate score.

Summary of Learning Tests and Scoring

Per GEC-T guidance, aggregate learning scores were computed for each student. First, for each language, subtask scores were calculated as the percentage of items answered correctly; on the oral reading fluency subtask, fluency scores were capped at 100 words per minute (WPM). For example, when scored, results for grade 4 girls on the Wolayttatto-medium EGRA included four subtasks, each with a score range from zero to 100.

Second, an individual's subtask scores for each language were combined to create an aggregate literacy score and a numeracy aggregate score. In each aggregate score, all subtasks were weighted equally.[140] For example, students in grade 4 had three aggregate scores at this point: one for the Wolayttatto-medium EGRA, one for English-medium EGRA, and one for EGMA. Each aggregate score consisted of equally weighted subtasks from that assessment.

Finally, for students in grades 4 and 6, the Wolayttatto- and English-medium EGRAs' aggregate scores were combined into a single aggregate score for literacy. Both language scores were equally weighted. Statistical software—*IBM SPSS Statistics 24*—was used to clean, merge, score, and analyze all data.

Average Baseline Aggregate Scores

The average baseline aggregate scores, by grade and group, are shown in Table 11 and 12. Because the aggregate literacy score for students in grade 4 and 6 include EGRAs in two languages—Wolayttatto and English—these scores should be interpreted as a measure of a student's combined literacy across both languages. Moreover, as English becomes the MOI in grade 5, the relative emphasis of English over Wolayttatto is not accounted for in the equal weighting of languages in the aggregate score. Alternative weighting decisions were considered to emphasize the transition to English. However, the importance of mother language ability and its relationship to second language acquisition cannot be adequately captured in a simple weighting exercise.

Statistically significant differences between groups in each grade were examined. For literacy, there was no statistically significant difference in the mean aggregate scores of the treatment and comparison groups for girls in grades 4 and 6. However, in grade 8, girls in comparison schools outperformed girls in treatment schools. The average aggregate literacy scores for girls in the comparison schools was 50.97 compared to 44.27 for girls in the treatment schools; this represents a small effect size (partial eta squared = 0.024).[141]

For numeracy, there was no statistically significant difference in the mean aggregate scores of the treatment and comparison groups for girls in grades 6 and 8. In grade 4, however, girls in comparison schools outperformed girls in treatment schools. The average aggregate numeracy scores in the comparison schools were 41.76 compared to 38.40 in the treatment schools; this represents a very small effect size (partial eta squared = 0.008).

It is important to highlight the standard deviations in the mean aggregate scores by grade. The large standard deviations convey the widespread of scores among students in the intervention group.

[140] Alternate weighting options were considered, including lower weighting applied to Wolayttatto than English for grades 4 and 6. However, there were no indication from the data or trends that alternate weighting was more appropriate.

[141] Partial eta squared is the estimated effect size obtained from an ANOVA. It is interpreted in the same way an effect size is. Effect sizes convey the magnitude of difference between groups and the meaningfulness of a statistically significant difference between groups.

Table 11. Literacy Aggregate Scores—EGRA and SeGRA, English and Wolayttatto

	Intervention group mean	Comparison group mean[142]	Standard deviation in the intervention group
Grade 4	27.31	29.39	27.12
Grade 6	39.58	37.94	29.77
Grade 8	44.27	50.97*	22.17

Note: Statistically significant differences in mean scores, by group, are indicated with an asterisk (*) next to the mean score of the higher performing group. Differences are reported at the p<0.05 level.

Table 12. Numeracy Aggregate Score—EGMA and SeGMA, English and Wolayttatto

	Intervention group mean	Comparison group mean	Standard deviation in the intervention group
Grade 4	38.40	41.76*	20.46
Grade 6	32.94	35.20	21.28
Grade 8	35.55	38.13	18.37

Note: Statistically significant differences in mean scores, by group, are indicated with an asterisk (*) next to the mean score of the higher performing group. Differences are reported at the p<0.05 level.

Comparability of Treatment and Comparison Woredas

The treatment group includes primary schools from four woredas, and the comparison group includes primary schools from only one woreda. The comparability of these two groups—based on the comparability of the woredas—is examined.

As the data suggest, the baseline literacy and numeracy levels for students in treatment and comparison woredas vary. Female students in the comparison woreda outperformed female students in treatment woredas in grade 4 numeracy and grade 8 literacy. However, there are several differences between the four treatment woredas and the one comparison woreda that are important to note as comparability is central to the analyses:

1. Because of the full saturation of the STAGES project in each of the four treatment woredas, it is impossible to identify comparison schools that were within the same administrative and geographic parameters as treatment schools but did not receive treatment.
2. The comparison woreda—Ofa—had fewer girls affected by barriers than did the treatment woredas:
    a. Ofa has a higher score on the gender-parity index than do the treatment woredas (see Section 2).[143] However, results from the teacher surveys suggest that support for girls' education was comparable in treatment and comparison woredas (see Section 3).
    b. Fewer girls and boys in Ofa reported that their households were unable to meet basic needs than did girls and boys in the treatment woredas—24.89 percent of boys and 23.99 percent of girls in comparison versus 35.37 percent of boys and 39.71 percent of girls in treatment woredas.

[142] As noted in previous sections, the word "control" has been replaced with "comparison" as this more accurately reflects the STAGES sampling approach and quasi-experimental design.
[143] The Federal Democratic Republic of Ethiopia, EMIS, and ICT Directorate and MOE, *Education Statistics Annual Abstract, 2008 E.C. (2015/16)* (Addis Ababa: June 2017).

c. Ofa has fewer girls who report low levels of household support—14.32 percent for comparison versus 22.06 percent for treatment.[144]

d. Girls in Ofa report slightly higher levels of perceived safety traveling to and from school—5.88 percent in Ofa compared to 9.39 percent in treatment woredas.

3. By contrast, fewer girls in the comparison woreda report attending school at least half the time than their peers in the treatment woredas (82.26 percent vs. 96.95 percent, respectively). The barriers girls in treatment and comparison schools faced related to school facilities were comparable except for the use of drinking water facilities. Treatment woredas had twice as many girls who reported not using drinking water facilities than the comparison woreda. The proportion of girls who report high corporal-punishment rates were twice as large in the comparison woreda than in the treatment woredas.

As a result, the comparability of the treatment and comparison woredas is unclear. Characteristics of girls suggest that fewer girls in the treatment woredas face individual-level challenges than girls in the comparison woreda; however, when examining barriers, the opposite seems true. At this point in the evaluation, with no alternative options for a comparison group, a change in the comparison group is not feasible. Instead, it will be important to examine shifts in characteristics and barriers within the two groups at the next evaluation point and determine whether there are indications that comparability is compromised.

Distribution and Floor and Ceiling Effects

Distributions of baseline aggregate scores are presented in Annex 17. Distributions were relatively normal for the aggregate numeracy scores; however, distributions that were not normal were observed for literacy aggregate scores. When examining the distributions for floor or ceiling effects, the aggregate literacy score distribution for grade 4 suggests a floor effect; this effect is less pronounced in grade 6.[145] However, the high proportion of students with zero scores demonstrates that a cluster of students was not able to answer basic foundational reading questions on either the Wolayttatto- or English-medium EGRAs in grade 4; this subgroup of students is notably smaller in grade 6.[146, 147] Therefore, the literacy assessments are deemed appropriate to gauge students' growth from grade 4 to 6 with no concerns of floor effects. By contrast, there was no indication of a ceiling effect at any grade. Additionally, the distribution of scores for all three grades does not suggest floor or ceiling effects to the numeracy scores.[148]

Learning Achievement Bands by Subtask and Grade for Treatment Schools

In addition to aggregate scores, foundational skill score bands for non-learners, emerged learners, established learners, and proficient learners were created. The proportion of female students in intervention schools that fall within each numeracy band is presented in Table 133a, 13b, and 13c.[149] The proportions of female students that fall within each literacy band are presented for grade 4 girls in Wolayttatto and English in 14a and 14b, for grade 6 girls in Wolayttatto and English in

[144] See household-support index results in Section 3.
[145] Floor and ceiling effects are examined to determine the extent to which an assessment captures the full range of respondents' abilities. Floor effects are observed when there are a large proportion of students with low or no score; ceiling effects are observed when there are a large proportion of students with the maximum possible score on the observed variable.
[146] Students with zero scores are students who were unable to answer a single item correctly.
[147] Basic foundational reading subtasks include familiar word and invented word. The letter sound subtask was dropped from the aggregate score so is not included in the floor and ceiling effect analysis.
[148] Both treatment and comparison groups received the highest concentration of scores in the 35-50 percent range with right "tails" comprised of a small number of female students receive scores above 75 percent. Grade 6 appeared to have a similar distribution of scores but with a longer tail to the right. Grade 8 had no zero numeracy scores. Both treatment and comparison groups in grade 8 appeared to have a small number of outliers that received scores notably higher than the mean.
[149] This table has been split into three tables (Table 14a, 14b and 14c) in order to present the relevant results by grade level.

The thresholds follow the proposed bands suggested by GEC guidance using subtask scores in within the following ranges: non-learner (0 percent of items), emergent learner (1 percent–40 percent of items), established learner (41 percent–80 percent of items), and proficient learners (81 percent–100 percent of items). Thresholds for oral reading fluency were as follows: non-reader (0–5 WPM), emergent reader (6–44 WPM), established reader (45–80 WPM), and proficient reader (80 WPM plus).

Table 144c and 14d, and for grade 8 girls in English in

Table 144e.[150]

On the numeracy assessments, more than half of grade 4 female students assessed in intervention schools were classified as established learners or proficient learners on the number identification, quantity discrimination, and addition subtasks. The missing numbers subtask—which assesses a student's ability to detect number patterns—appears to be a difficult subtask for girls in both grades 4 and 6. Specifically, more than half of grade 4 girls and almost half of grade 6 girls were classified as "non-learners" on the subtask while the proportion of non-learners on all other numeracy subtasks was relatively lower. Girls also struggled with the word problems subtask—more than half of grade 6 girls and almost half of grade 8 girls—were classified as non-learners.

On the literacy assessments, more than one-third of grade 4 girls in intervention schools were classified as non-learners on all literacy subtasks except for the letter sound identification subtask; this proportion decreased to approximately one-fourth of grade 6 girls. More than half of grade 8 girls were classified as established or proficient learners on the English familiar word, invented word, and oral reading subtasks. However, slightly less than half of grade 8 girls were classified as non-learners on the reading comprehension subtask. This suggests that although some grade 8 girls are fluent readers, many struggled to understand what they read.

Table 13a: Foundational Numeracy Skills by Learning Achievement Bands—Grade 4 Girls

Grade 4 EGMA—Percentage of Girls' Achievement					
Categories (bands)[151]	Number identification	Quantity discrimination	Missing numbers	Addition	Subtraction
Non-learner	7.96%	15.92%	52.94%	6.57%	13.15%
Emergent learner	22.49%	19.38%	33.56%	35.64%	58.82%
Established learner	59.52%	57.44%	12.8%	54.67%	27.34%
Proficient learner	10.03%	7.27%	0.69%	3.11%	0.69%

[150] The thresholds follow the proposed bands suggested by GEC-T guidance using subtask scores in within the following ranges: non-learner (0 percent of items), emergent learner (1 percent–40 percent of items), established learner (41 percent–80 percent of items), and proficient learners (81 percent–100 percent of items). Oral reading fluency was placed within a separate set of four ranges in accordance with GEC guidance: non-reader (0–5 WPM), emergent reader (6–44 WPM), established reader (45–80 WPM), and proficient reader (80 WPM plus).

[151] The thresholds follow the proposed bands suggested by GEC guidance using subtask scores in within the following ranges: non-learner (0 percent of items), emergent learner (1 percent–40 percent of items), established learner (41 percent–80 percent of items), and proficient learners (81 percent–100 percent of items). Thresholds for oral reading fluency were as follows: non-reader (0–5 WPM), emergent reader (6–44 WPM), established reader (45–80 WPM), and proficient reader (80 WPM plus).

Table 13b: Foundational Numeracy Skills by Learning Achievement Bands—Grade 6 Girls

Grade 6 EGMA—Percentage of Girls' Achievement						
Categories (bands)[152]	Number Identification	Quantity Discrimination	Missing Numbers	Addition	Subtraction	Word Problems
Non-learner	13.72%	24.91%	44.77%	9.39%	9.03%	61.01%
Emergent learner	23.47%	22.74%	37.91%	44.04%	56.68%	21.66%
Established learner	50.9%	42.96%	14.08%	43.32%	33.21%	11.55%
Proficient learner	11.91%	9.39%	3.25%	3.25%	1.08%	5.78%

The thresholds follow the proposed bands suggested by GEC guidance using subtask scores in within the following ranges: non-learner (0 percent of items), emergent learner (1–40 percent of items), established learner (41–80 percent of items), and proficient learners (81–100 percent of items). Thresholds for oral reading fluency were as follows: non-reader (0–5 WPM), emergent reader (6–44 WPM), established reader (45–80 WPM), and proficient reader (80 WPM plus).

Table 13c: Foundational Numeracy Skills by Learning Achievement Bands—Grade 8 Girls

Grade 8 EGMA/SeGMA—Percentage of Girls' Achievement						
					SeGMA	
Categories (bands)[153]	Addition	Subtraction	Word Problems	Geometry and Measurement	Fractions	Multiplication
Non-learner	7.20%	5.68%	47.35%	16.09%	20.31%	19.16%
Emergent learner	38.26%	53.03%	25.38%	33.72%	53.64%	38.7%
Established learner	46.59%	40.91%	15.91%	36.78%	22.22%	28.74%
Proficient learner	7.95%	0.38%	11.36%	13.41%	3.83%	13.41%

The thresholds follow the proposed bands suggested by GEC guidance using subtask scores in within the following ranges: non-learner (0 percent of items), emergent learner (1 percent–40 percent of items), established learner (41 percent–80 percent of items), and proficient learners (81 percent–100 percent of items).

[152] See previous footnote.
[153] See previous footnote.

Table 14a: Foundational Literacy Skills by Learning Achievement Bands—Grade 4 Girls, Wolayttatto

Grade 4 EGRA Wolayttatto—Percentage of Girls' Achievement					
Categories (bands)[154]	Letter Sound Identification	Familiar Word	Invented Word	Oral Reading Fluency	Comprehension
Non-learner	6.57%	35.29%	34.26%	35.29%	46.71%
Emergent learner	19.38%	20.42%	21.8%	49.13%	23.18%
Established learner	32.53%	26.64%	26.3%	14.53%	25.26%
Proficient learner	41.52%	17.65%	17.65%	1.04%	4.84%

The thresholds follow the proposed bands suggested by GEC guidance using subtask scores in within the following ranges: non-learner (0 percent of items), emergent learner (1 percent–40 percent of items), established learner (41 percent–80 percent of items), and proficient learners (81 percent–100 percent of items). Thresholds for oral reading fluency were as follows: non-reader (0–5 WPM), emergent reader (6–44 WPM), established reader (45–80 WPM), and proficient reader (80 WPM plus).

Table 14b: Foundational Literacy Skills by Learning Achievement Bands—Grade 4 Girls, English

Grade 4 EGRA English—Percentage of Girls' Achievement					
Categories (bands)[155]	Letter sound identification	Familiar word	Invented word	Oral reading fluency	Comprehension
Non-learner	14.19%	47.06%	39.10%	46.37%	70.93%
Emergent learner	31.14%	25.61%	9.69%	26.99%	20.76%
Established learner	39.10%	23.18%	29.76%	21.11%	7.96%
Proficient learner	15.57%	4.15%	21.45%	5.54%	0.35%

The thresholds follow the proposed bands suggested by GEC guidance using subtask scores in within the following ranges: non-learner (0 percent of items), emergent learner (1 percent–40 percent of items), established learner (41 percent–80 percent of items), and proficient learners (81 percent–100 percent of items). Thresholds for oral reading fluency were as follows: non-reader (0–5 WPM), emergent reader (6–44 WPM), established reader (45–80 WPM), and proficient reader (80 WPM plus).

[154] See previous footnote.
[155] See previous footnote.

Table 14c: Foundational Literacy Skills by Learning Achievement Bands—Grade 6 Girls, Wolayttatto

Grade 6 EGRA Wolayttatto—Percentage of Girls' Achievement					
Categories (bands)[156]	Letter sound identification	Familiar word	Invented word	Oral reading fluency*	Comprehension
Non-learner	4.33%	22.74%	23.1%	24.19%	29.6%
Emergent learner	11.55%	19.86%	24.55%	46.21%	25.63%
Established learner	43.68%	33.94%	35.02%	22.38%	31.41%
Proficient learner	40.43%	23.47%	17.33%	7.22%	13.36%

The thresholds follow the proposed bands suggested by GEC guidance using subtask scores in within the following ranges: non-learner (0 percent of items), emergent learner (1 percent–40 percent of items), established learner (41 percent–80 percent of items), and proficient learners (81 percent–100 percent of items). Thresholds for oral reading fluency were as follows: non-reader (0–5 WPM), emergent reader (6–44 WPM), established reader (45–80 WPM), and proficient reader (80 WPM plus).

Table 14d: Foundational Literacy Skills by Learning Achievement Bands—Grade 6 Girls, English

Grade 6 EGRA English—Percentage of Girls' Achievement					
Categories (bands)	Letter sound identification	Familiar word	Invented word	Oral reading fluency*	Comprehension
Non-learner	6.86%	27.08%	25.63%	24.19%	54.51%
Emergent learner	19.86%	26.71%	10.11%	29.24%	27.08%
Established learner	50.54%	32.13%	35.02%	23.47%	15.88%
Proficient learner	22.74%	14.08%	29.24%	23.1%	2.53%

The thresholds follow the proposed bands suggested by GEC guidance using subtask scores in within the following ranges: non-learner (0 percent of items), emergent learner (1 percent–40 percent of items), established learner (41 percent–80 percent of items), and proficient learners (81 percent–100 percent of items). Thresholds for oral reading fluency were as follows: non-reader (0–5 WPM), emergent reader (6–44 WPM), established reader (45–80 WPM), and proficient reader (80 WPM plus).

Table 14e: Foundational Literacy Skills by Learning Achievement Bands—Grade 8 Girls, English

Grade 8 EGRA/SeGRA—Percentage of Girls' Achievement							
						SeGRA	
Categories (bands)[157]	Familiar Word Reading	Invented Word Reading	Oral Reading Fluency*	Reading Comprehension	Reading Passage	Fill in the blanks	Revising Sentences
Non-learner	17.05%	18.56%	14.77%	43.18%	9.89%	11.03%	11.79%
Emergent learner	20.83%	11.36%	20.08%	30.68%	34.98%	57.79%	59.32%
Established learner	33.71%	33.33%	32.58%	18.18%	35.74%	30.8%	28.14%
Proficient learner	28.41%	36.74%	32.58%	7.95%	19.39%	0.38%	0.76%

[156] See previous footnote.
[157] See previous footnote.

Alternative Thresholds for Learning Achievement Bands

The thresholds used to establish the achievement bands for each subtask are based on external criteria provided by GEC-T and intended to be applied to datasets across varied contexts. The thresholds currently reported cannot be used to determine where girls may be struggling—as the threshold does not convey where girls **should be** but rather how they are performing against arbitrary cut points. These cut points may not indicate meaningful information regarding performance on subtasks and what that indicates about a student's readiness to being able to read and comprehend. In other words, while the current thresholds are set at a subtask level, they do not necessarily connect with the ultimate goal of literacy: comprehension.

In response to this, the data for girls in the Wolaita Zone were analyzed using a different approach, which yielded a different set of thresholds to characterize girls' performance. The alternative approach, described in this section, has its benefits and drawbacks.

One alternative approach is to set a threshold on each subtask based on those students who correctly answered 80.00 percent of the reading comprehension questions. The primary benefit of this approach is that bands are language and context-specific, and it acknowledges the interconnectedness of the foundational skills to a student's ultimate ability to read and comprehend connected text.

However, if the threshold for proficiency on each subtask was tied to proficiency in reading comprehension, then the threshold carries greater meaning: it suggests the level at which a student should be performing in that subtask if they are to ultimately be able to read and comprehend.

The primary drawback of this approach is that the number of students who achieve 80.00 percent comprehension is typically very low in developing contexts—at times representing less than 5.00 percent of the total sample population. In the baseline data, the proportion of grade 4 female students who were identified as proficient in Wolayttatto, using the alternative approach, was 38 students in grade 4 and 78 students in grade 6; and in English, it was five students in grade 4, 13 students in grade 6, and 34 students in grade 8.

The alternative thresholds for proficient readers in Wolayttatto are presented in Supplementary Table 7.

Supplementary Table 9. Alternative Thresholds to Identify Proficient Female Students—Wolayttatto

Wolayttatto Subtasks	***Grade 4***		***Grade 6***	
	n	Subtask score for girls with 80% reading comprehension correct	n	Subtask score for girls with 80% reading comprehension correct
Letter sounds	38	91.76	78	82.81
Familiar words	38	87.47	78	86.67
Invented words	38	85.47	78	80.82
Oral reading fluency	38	84.99	78	84.72

Note: Suggested thresholds are based on baseline results for all students, including both treatment and comparison schools, to maximize the number of students meeting the threshold.

Supplementary Table 10. Alternative Thresholds to Identify Proficient Female Students—English

English Subtasks	Grade 4		Grade 6		Grade 8	
	n	Subtask score for girls with 80% reading comprehension correct	n	Subtask score for girls with 80% reading comprehension correct	n	Subtask score for girls with 80% reading comprehension correct
Letter sounds	5	85.40	13	85.08	n/a	n/a
Familiar words	5	95.60	13	85.08	34	91.88
Invented words	5	94.00	13	83.69	34	91.88
Oral reading fluency	5	80.48	13	78.66	34	98.75

If alternative thresholds are of interest to track girls' progression in reading across evaluation points, then additional analyses will need to be conducted to identify thresholds for emergent and established readers for each subtask. Furthermore, differences in thresholds across grades and ages would need to be examined further. Additionally, an examination of baseline results against student age and characteristics, as well as revisions at each evaluation point, would be beneficial.

Grade-Level Achieved

This section examines the grade level achieved by girls at baseline. The grade level achieved is established in a conversion grid that maps proficiency levels from the tests to the grade level—end-of-year achievement—that that proficiency level represents according to the national curriculum.

The mapping is based on a review of the national curriculum for girls in grades 4, 6, and 8 at the time the baseline assessments were developed. Additionally, each table shows the proportion of girls meeting the expected level for each grade; these data are based on thresholds for emergent, established, and proficient readers per guidance.

In interpreting the grade-level achieved, it is important to note that at the time the literacy assessments were developed, they were based on a review of the national curriculum as well as textbooks and previous assessments in Ethiopia. However, based on previous assessment results and discussions with relevant zone and woreda stakeholders, the assessments were generally developed at two grade levels below the grade of the students being tested, in terms of the difficulty of passages, in EGRA passages and SeGRA and SeGMA exercises. For example, multiple reading passages of varying levels of difficulty were developed for girls in grade 6 during the pilot. The analyses showed floor effects for passages that were on-grade level, but adequate performance ranges for the passage that was based on grade 4 textbooks and content.

For numeracy, previous assessments indicated that, in previous years, assessments based on content two grade-levels below the grade of enrollment captured the range of girls' performance. As such, numeracy assessments were based on content two grade-levels below the assessed grade but included on-grade level items in the pilot as well. Generally, these on-grade items were dropped after piloting as their item-level performance was poor.

Supplementary Table 11 shows the mapping of proficiency levels to grade levels for Wolayttatto, English, and numeracy accounting for the difficulty level of the assessments.[158] Supplementary Table 12 shows the mapping of proficiency levels to grade levels for English and accounts for the difficulty level of the assessments. Finally, Supplementary Table 13 shows the mapping of proficiency levels to grade levels for numeracy and accounts for the difficulty level of those assessments.

[158] To interpret the table, for example, 40.43 percent of grade 6 girls were achieving proficiency in letter sounds—a criteria for having achieved end-of-grade 1 expectations.

Supplementary Table 11. Conversion of Proficiency Levels to Grades—Literacy (Wolayttatto)

Grade	Wolayttatto Proficiency Levels to Grades			
	Relevant subtasks	Literacy	Proportion of grade 4 girls achieving grade-level proficiency levels	Proportion of grade 6 girls achieving grade-level proficiency levels
Grade 1 achieved	Subtask 1, 2, and 3 (EGRA)	Proficient in letter sound identification, familiar word, invented word	41.52% LS 17.65% FW 17.65% IW	40.43% LS 23.27% FW 17.33% IW
Grade 2 achieved	Subtask 4 (EGRA)	Established in oral reading fluency and comprehension of grade-level text	14.43% ORF 25.26% RC	22.38% ORF 31.41% RC
Grade 3 achieved	Subtask 5 (EGRA)	Proficient in oral reading fluency and comprehension of grade-level text	1.04% ORF 4.84% RC	7.22% ORF 13.36% RC
Grade 4 achieved	Subtask 5 (EGRA)	Established in inferential comprehension of grade-level text	Insufficient data to examine inferential reading comprehension items only	Insufficient data to examine inferential reading comprehension items only
Grade 5 achieved	Subtask 5 (EGRA)	Proficient in inferential comprehension of grade-level text	Insufficient data to examine inferential reading comprehension items only	Insufficient data to examine inferential reading comprehension items only

Note: The term "grade-level text" is used to convey the minimum learning requirements articulated in the national curriculum. Inferential reading comprehension questions were limited (one to at most two questions per passage).

Supplementary Table 12. Conversion of Proficiency Levels to Grades—Literacy (English)

Grade	English Proficiency Levels to Grades				
	Relevant subtasks	Literacy	Proportion of grade 4 girls achieving grade-level proficiency levels	Proportion of grade 6 girls achieving grade-level proficiency levels	Proportion of grade 8 girls achieving grade-level proficiency levels
Grade 1 achieved	Subtask 1, 2, and 3 (EGRA)	Established in letter sound identification, familiar word, invented word	39.10% LS 23.18% FW 29.76% IW	50.54% LS 32.13% FW 35.02% IW	33.71% FW 33.33% IW
Grade 2 achieved	Subtask 4 (EGRA)	Proficient in letter sound identification, familiar word, invented word	15.57% LS 4.15% FW 21.45% IW	22.74% LS 14.08% FW 29.24% IW	28.41% FW 36.74% IW
Grade 3 achieved	Subtask 5 (EGRA)	Established (grade 4), emergent (grade 6 and 8) in oral reading fluency of grade-level text	21.11%	29.24%	
Grade 4 achieved	Subtask 5 (EGRA)	Proficient (grade 4) , established (grades 6 and 8) in oral reading fluency and	5.54% ORF 0.35% RC	23.47%	32.58%

Grade	English Proficiency Levels to Grades				
	Relevant subtasks	Literacy	Proportion of grade 4 girls achieving grade-level proficiency levels	Proportion of grade 6 girls achieving grade-level proficiency levels	Proportion of grade 8 girls achieving grade-level proficiency levels
		comprehension of grade-level text			
Grade 5 achieved	Subtask 5 (EGRA)	Proficient in oral reading fluency and comprehension of grade-level text	-	-	-
Grade 6 achieved	Subtask 5 (EGRA)	Proficient in oral reading fluency and comprehension of grade-level text	-	23.10% ORF 2.53% RC	32.58% ORF 7.95% RC
Grade 7 achieved	Subtask 5 (EGRA)	Proficient in oral reading fluency and comprehension of grade-level text	-	-	-
Grade 8 achieved	Subtask 8 (SeGRA 1)	Established in reading and comprehension of written, grade-level text	Not tested	Not tested	35.74%
Grade 9 achieved	Subtask 8 (SeGRA 1)	Proficient in reading and comprehension of written, grade-level text	Not tested	Not tested	19.39%

Note: The term "grade-level text" is used to convey the minimum learning requirements articulated in the national curriculum. Inferential reading comprehension questions were limited (one to at most two questions per passage).

Supplementary Table 13. Conversion of Proficiency Levels to Grades—Numeracy

Grade	Numeracy Proficiency Levels to Grades				
	Relevant subtasks	Numeracy	Proportion of Grade 4 Girls Achieving grade-Level Proficiency Levels	Proportion of Grade 6 girls Achieving Grade-Level Proficiency Levels	Proportion of Grade 8 girls Achieving Grade-Level Proficiency Levels
Grade 1 achieved	Subtask 1, 2, and 3 EGMA)	Established in number identification and in quantity discrimination	59.52% NI 57.44% QD 12.8% MN	50.90% NI 42.96% QD 14.08% MN	Not tested
Grade 2 achieved	Subtask 1, 2, and 3 EGMA)	Proficient in number identification, quantity discrimination and missing numbers	10.03% NI 7.27% QD 0.69% MN	11.91% NI 9.39% QD 3.25% MN	Not tested
Grade 3 achieved	Subtask 4 and 5 (EGMA)	Established in and additions and subtractions	54.67% ADD 27.34% SUB	43.32% ADD 33.21% SUB	46.59% ADD 40.91% SUB
Grade 4 achieved	Subtask 4 and 5 (EGMA)	Proficient in and additions and subtractions	3.11% ADD 0.69% SUB	3.25% ADD 1.08% SUB	7.95% ADD 0.38% SUB
Grade 5 achieved	Subtask 6 EGMA)	Established in grade-level word problems	Not tested	11.55%	15.91%
Grade 6 achieved	Subtask 6 EGMA)	Proficient in grade-level word problems	Not tested	5.78%	11.36%

Grade	Numeracy Proficiency Levels to Grades				
	Relevant subtasks	Numeracy	Proportion of Grade 4 Girls Achieving grade-Level Proficiency Levels	Proportion of Grade 6 girls Achieving Grade-Level Proficiency Levels	Proportion of Grade 8 girls Achieving Grade-Level Proficiency Levels
Grade 7 achieved	(SeGMA 1, 2, 3)	Emergent in geometry and measurement, fractions and multiplication	Not tested	Not tested	33.72% GEO 53.64% FRAC 38.70% MULT
Grade 8 achieved	(SeGMA 1, 2, 3)	Established in geometry and measurement, fractions and multiplication	Not tested	Not tested	36.72% GEO 22.22% FRAC 28.74% MULT
Grade 9 achieved	(SeGMA 1, 2, 3)	Proficient in geometry and measurement, fractions and multiplication	Not tested	Not tested	13.41% GEO 3.83% FRAC 13.41% MULT

Mapping girls' proficiency levels to grade-level achieved suggest that, on average, girls are performing two grade levels lower than the grade they attend. This aligns with the review of the curriculum and minimum learning competencies for each grade at the time the assessments were developed. For example, in Wolayttatto, one-quarter of grade 4 girls and one-third of grade 6 girls are achieving proficiency in reading comprehension on a grade-level text—the criteria for achieving grade 2 literacy skills. For the purposes of comparison, girls in grade 6 were assessed with the same difficulty passage as girls in grade 4.

In English, very few grade 4 girls achieved grade 4 literacy skills—5.54 percent in oral reading fluency, and 0.35 percent were established in reading comprehension. Approximately one-quarter (23.10 percent) of grade 6 girls were achieving proficiency in oral reading fluency, the criteria for achieving at a grade 6 level. Even fewer girls met the second criteria for grade 6 level achievement: 2.53 percent in reading comprehension.

By contrast, one-third of grade 8 girls met the first criteria for achieving grade 6 level literacy skills—32.58 percent were proficient in oral reading fluency—and one-in-ten met the second criteria for grade 6 level literacy skills—7.95 percent were proficient in reading comprehension. Insufficient data on inferential comprehension questions did not allow further examination of grade 8 girls' abilities to meet grade 7 literacy skills. Approximately one-third of grade 8 girls were achieving grade 8 level literacy skills—35.74 percent were established in reading and comprehension of written, grade-level text; one-in-five grade 8 girls were achieving grade 9 level literacy skills—19.39 percent were proficient in reading and comprehension of written, grade-level text.

For numeracy, the majority of students in both grades 4 and 6 were achieving at one to two grade levels below their enrolled grade. For example, the majority of grade 4 students were established in number identification and quantity discrimination, which meant they were achieving at a grade 1 level for these skills. With addition and subtraction, the majority of girls were achieving at a grade 3 level, with half of grade 4 girls being established in additions and one-quarter in subtractions (54.67 percent and 27.34 percent, respectively). The proportion of grade 6 girls achieving at a grade 3 level in additions and subtractions was higher (43.32 percent and 33.21 percent, respectively).

For students in grade 8, approximately one-third of girls were achieving grade 8 level numeracy skills, and approximately one-in-ten were achieving grade 9 level numeracy skills.

4.2 Subgroup Analysis of Learning Outcome

In this section, trends in learning for key subgroups and trends by barriers are explored. Average literacy and numeracy aggregate scores for girls in each subgroup or affected by each barrier are presented by grade. First, differences by woreda are presented, followed by subgroups and barriers. Note that the subgroup analysis excludes examining results by disability because of the lack of complete data for the full sample of girls in phase one.

A Caution When Interpreting Subgroup Analysis Results

When comparing the learning outcomes of those girls affected by each barrier to the learning outcomes for all girls, the data may suggest that there is only a small difference. Specifically, differences in learning scores for all girls do not vary from the learning scores of some subgroups—including girls with high chore burdens, girls with low levels of household support, girls who report that teachers treat boys and girls differently or are absent from class, or girls who report high levels of corporal punishment. However, it is important to keep in mind the small sample sizes as well as the limitations in the survey data; only one item was asked about chore burden, teachers' treatment of boys and girls, or teacher absenteeism. As such, the relationships between overall learning assessment scores and single items from a survey should be interpreted with caution. Relationships with constructs generated from a set of items—such as household support, life-skills or decision-making—can be interpreted with greater confidence because of higher reliability in the observed data to assess the construct of interest.

For those barriers that have emerged as affecting a large proportion of girls, the next evaluation point should include a larger set of items so that the construct can be more fully explored, and a more reliable measure constructed.

Differences in learning levels across regions

Girls' performance in literacy and numeracy was compared across woredas. Since the woreda variable confounds with the group variable, the differences at the woreda level should be interpreted with caution. The results, by woreda, show that girls in Damot Pulasa and Damot Woide generally tended to have higher performance levels than girls in the other woredas.

Learning scores by subgroups are presented in Tables 15b, 15c, and 15d for girls in grades 4, 6, and 8, respectively. The following subgroups are highlighted because the gap in performance scores for the subgroup compared to all girls were the largest:

- Girls living without both parents had notably lower average literacy aggregate scores; this held true for all three grade levels.
- Girls with poor overall well-being scores had lower average literacy aggregate scores in grades 4 and 8 than girls with high well-being scores in the same grade.
- Girls with poor self-esteem scores had lower average literacy and numeracy aggregate scores in grades 4, 6, and 8 than girls with high self-esteem scores in the same grade.
- The majority of girls in grade 4 and all girls in grades 6 and 8 reported that they speak Wolayttatto at home and with friends; however, the MOI is English beginning in grade 5, which could mean that girls in grades 6 and 8 are faced with a language difficulty barrier due to the transition from Wolayttatto to English.
- In all three grades, girls who were overage for their grade performed lower on both literacy and numeracy than girls who were on-age or underage.
- Girls in grade 6 who had low life-skills scores (controlling for decision-making scores) had lower scores in literacy and numeracy than girls who had high life-skills scores. The interaction between

life-skills and decision-making was statistically significant, suggesting that the effect of life-skills varies by levels of decision-making.[159]

Table 15a: Learning scores of key subgroups—by Woreda

Woreda	Literacy			Numeracy		
	Mean	Standard Deviation	Significant Differences	Mean	Standard Deviation	Significant Differences
Grade 4						
Damot Pulasa	31.66	29.80	-	41.86	21.68	> DS, DW, KK
Damot Sore	19.65	21.07	-	36.19	19.23	< DP
Damot Woide	28.62	27.66	-	37.23	19.59	< DP
Kindo Koisha	28.26	27.74	-	38.55	21.12	< DP
Ofa (comparison)	29.39	27.11	-	41.76	19.72	-
Grade 6						
Damot Pulasa	37.16	28.83	-	29.95	16.91	-
Damot Sore	35.98	27.88	-	31.35	20.44	-
Damot Woide	47.72	30.02	-	40.06	23.60	-
Kindo Koisha	36.89	30.57	-	30.19	21.41	-
Ofa (comparison)	37.94	29.38	-	35.20	21.28	-
Grade 8						
Damot Pulasa	47.10	19.93	-	36.26	16.28	> KK
Damot Sore	46.43	21.30	-	35.85	17.82	-
Damot Woide	47.71	22.34	< Ofa	40.05	22.03	-
Kindo Koisha	38.27	23.10	< Ofa	31.25	15.94	< Ofa
Ofa (comparison)	50.97	20.71		38.13	16.23	-

[159] Life skills will not be measured in subsequent evaluation points.

Table 15b: Learning scores of key subgroups—Grade 4

	Average literacy score (aggregate)		Average numeracy score (aggregate)	
	Score	**n (% of total)**	**Score**	**n (% of total)**
All girls	**27.31**	**264 (100%)**	**38.40**	**264 (100%)**
Living without both parents	17.35	15 (5.47%)	37.92	15 (5.47%)
Household unable to meet basic needs	26.56	120 (45.45%)	38.35	120 (45.45%)
Language difficulties: girl does not speak MOI	29.75	283 (97.25%)	40.71	283 (97.25%)
Poor overall well-being	23.02	82 (31.06%)	35.10	82 (31.06%)
Overage for grade	24.06	62 (22.63%)	37.51	62 (22.63%)

Note: Grade 4 girls were not asked questions regarding disabilities. Tests for significance of differences by subgroups was not conducted due to small n sizes for some subgroups. For "household unable to meet basic needs," an asterisk (*) indicates that the group of girls whose households were unable to meet basic needs had statistically significantly lower scores than the group of girls who reported that their households were able to meet basic needs. Results for girls learning assessment scores for low self-esteem are not included since the data are based on a different sample (phase two) rather than phase one.

Table 15c: Learning scores of key subgroups—Grade 6

	Average literacy score (aggregate)		Average numeracy score (aggregate)	
	Score	**n (% of total)**	**Score**	**n (% of total)**
All girls	**39.58**	**296 (100%)**	**32.94**	**296 (100%)**
Living without both parents	26.86	12 (4.05%)	28.95	12 (4.05%)
Household unable to meet basic needs	34.04	112 (37.84%)	27.86	112 (37.84%)
Vision impairment	not reported[160]			
Hearing impairment				
Mobility impairment				
Cognitive impairment				
Self-care impairment				
Communication impairment	48.99	22 (7.43%)	43.07	22 (7.43%)
Language difficulties: girl does not speak MOI	All girls in grade	296 (100%)	All girls in grade	296 (100%)
Poor overall well-being	39.78	110 (37.16%)	30.87	110 (37.16%)
At least one disability	not reported[161]			
Overage for grade	30.34	49 (16.55%)	27.52	49 (16.55%)

Note: Tests for significance of differences by subgroups were not conducted due to small n sizes for some subgroups. For "household unable to meet basic needs,"' asterisks (*) indicates that the group of girls whose households were unable to meet basic needs had statistically significantly lower scores than the group of girls who reported that their households were able to meet basic needs.

[160] This data should be understood as missing. Results for disability subgroup—barrier 1—were not possible to disaggregate at this time as the Child Functioning set was administered in phase two data collection and has a different denominator than the relevant survey items. This information will be provided starting at midline using the Washington Group Child Functioning questions.
[161] Ibid.

Table 15d: Learning scores of key subgroups—Grade 8

	Average literacy score (aggregate)		Average numeracy score (aggregate)	
	Score	n (% of total)	Score	n (% of total)
All girls	**44.27**	**289 (100%)**	**35.55**	**289 (100%)**
Living without both parents	26.66	15 (5.19%)	27.47	15 (5.19%)
Household unable to meet basic needs	39.73*	75 (25.95%)	31.85*	75 (25.95%)
Vision impairment	not reported[162]			
Hearing impairment				
Mobility impairment				
Cognitive impairment				
Self-care impairment				
Communication impairment				
Language difficulties: girl does not speak MOI	All girls in grade	289 (100%)	All girls in grade	289 (100%)
Poor overall well-being	36.72	45 (15.57%)	31.82	45 (15.57%)
At least one disability	not reported[163]			
Overage for grade	36.97*	37 (12.80%)	31.51*	37 (12.80%)

Tests for significance of differences by subgroups were not conducted due to small n sizes for some subgroups. For "household unable to meet basic needs," an asterisk (*) indicates that the group of girls whose households were unable to meet basic needs had statistically significantly lower scores than the group of girls who reported that their households were able to meet basic needs.

Learning scores aggregated by barrier are presented in Table 16a,16b, and 16c for girls in grades 4, 6, and 8, respectively.[164] The following barriers are highlighted because the gap in performance scores was the largest when comparing girls facing the barrier to those who did not:

- Girls who attended three or fewer days of school last week had critically low literacy and numeracy average aggregate scores
- Girls who have a high chore burden in grade 4 performed lower than all girls in grade 4; this gap became even more pronounced in grades 6 and 8

[162] This data should be understood as missing. Results for disability subgroup—barrier 1—were not possible to disaggregate at this time as the Child Functioning set was administered in phase two data collection and has a different denominator than the relevant survey items. This information will be provided starting at midline using the Washington Group Child Functioning questions.
[163] Ibid.
[164] Barrier analyses are presented in Section 3.

Table 16a: Learning Scores of Key Barriers—Grade 4

	Average literacy score (aggregate)		Average numeracy score (aggregate)	
	Score	n (% of total)	Score	n (% of total)
All girls	**27.31**	**274 (100.00%)**	**38.40**	**274 (100.00%)**
High chore burden	25.65	205 (74.82%)	37.81	205 (74.82%)
Low levels of support from household	25.74	48 (17.52%)	39.96	48 (17.52%)
Woreda officials with a gap in attitudes towards girls' and boys' education	See narrative		See narrative	
Attends school half the time	17.40	34 (12.40%)	31.99	34 (12.40%)
Feels distance to school is not close	24.13	78 (28.46%)	33.79	78 (28.46%)
Agrees teachers treat boys and girls differently in the classroom	26.48	197 (71.90%)	37.89	197 (71.90%)
Agrees teachers often absent from class	27.06	126 (50.73%)	38.63	139 (50.73%)
High corporal punishment exercised by teacher	26.98	118 (43.07%)	36.89	118 (43.07%)
Low life-skills score	23.31	63 (23.36%)	36.45	63 (23.36%)
Low decision-making score	27.91	80 (29.20%)	37.29	80 (29.20%)

Note: Tests of significance were conducted on main effects. Interactions between life skills and decision-making were explored, but no statistically significant interactions were observed. No statistically significant differences were found between the group of girls who face each barrier and the group of girls who did not face the barrier.

Table 16b: Learning Scores of Key Barriers—Grade 6

	Average literacy score (aggregate)		Average numeracy score (aggregate)	
	Score	n (% of total)	Score	n (% of total)
All girls	**39.58**	**296 (100%)**	**32.94**	**296 (100%)**
High chore burden	33.15*	181 (61.15%)	28.83*	181 (61.15%)
Low levels of support from household	30.34*	73 (24.66%)	25.20*	73 (24.66%)
Woreda officials with a gap in attitudes towards girls' and boys' education	See narrative		See narrative	
Attends school half the time	23.63*	24 (8.11%)	22.34*	24 (8.11%)
Feels distance to school is not close	36.92	70 (23.65%)	28.87	70 (23.65%)
Agrees teachers treat boys and girls differently in the classroom	37.57	207 (69.93%)	30.75	207 (69.93%)
Agrees teachers often absent from class	37.29	136 (45.95%)	29.62	136 (45.95%)
High corporal punishment exercised by teacher	40.28	161 (54.39%)	33.34	161 (54.39%)
Low life-skills score	45.12	33 (11.15%)	31.01	33 (11.15%)
Low decision-making score	43.48	39 (13.18%)	34.08	39 (13.18%)

Note: Tests of significance were conducted on main effects. Interactions between life skills and decision-making were explored and reported in the report. An asterisk (*) indicates that the group of girls who face the barrier noted had statistically significantly lower scores than the group of girls who did not face the barrier.

Table 16c: Learning Scores of Key Barriers—Grade 8

	Average literacy score (aggregate)		Average numeracy score (aggregate)	
	Score	n (% of total)	Score	n (% of total)
All girls	**44.27**	**289 (100%)**	**35.55**	**289 (100%)**
High chore burden	41.40*	174 (60.21%)	33.01*	174 (60.21%)
Low levels of support from household	39.01	43 (14.88%)	31.47	43 (14.88%)
Woreda officials with a gap in attitudes towards girls' and boys' education	see narrative		see narrative	
Attends school half the time	38.20	12 (4.15%)	29.48	12 (4.15%)
Feels distance to school is not close	34.92	63 (21.80%)	30.84	63 (21.80%)
Agrees teachers treat boys and girls differently in the classroom	42.36	176 (60.90%)	34.78	176 (60.90%)
Agrees teachers often absent from class	44.84	102 (35.29%)	37.16	102 (35.29%)
High corporal punishment exercised by teacher	41.24	110 (38.06%)	34.18	110 (38.06%)
Low life-skills score	37.96	26 (9.00%)	29.07	26 (9.00%)
Low decision-making score	40.24	23 (7.96%)	32.44	23 (7.96%)

Note: Tests of statistical significance were conducted on main effects. Interactions between barriers were explored to a limited extent, where the relationship between barriers was indicated through high bivariate correlations. An asterisk (*) indicates that the group of girls who face the barrier noted had statistically significantly lower scores than the group of girls who did not face the barrier.

Woreda Officials difference in attitudes towards girls' and boys' education

One barrier of interest is the proportion of woreda officials with a difference in attitudes towards boys' education than towards girls' education. This barrier examined the proportion of woreda officials with below-mean scores on attitudes towards girls' education but high levels of support for boys' education. Scores were aggregated to the woreda level, and the proportion of woreda officials with a gap in perceptions towards boys' and girls' education was layered with girls' learning outcomes to see if girls' performance varied when a greater proportion of woreda officials had gaps in their perceptions. Because several proportions were examined, this measure could not be appropriately presented in the previous tables.

For girls in grades 4 and 6, the relationship between a higher proportion of woreda officials with differences in perceptions of boys' and girls' education and girls' performance in literacy and numeracy was contradictory to what the evaluators expected. In woredas where one-third of officials supported boys going to primary or secondary school but had low perceptions towards girls' education, the average aggregate literacy score for girls in grade 4 was 19.33, and the average numeracy aggregate score was 35.51. When the proportion of woreda officials holding these perceptions increased to almost one-half, girls performed better; the average aggregate literacy score for girls was 31.66, and the average numeracy aggregate score was 41.86. In considering results from grade 6 girls, in woredas where one-third of officials held differing perceptions, the average aggregate literacy score for girls was 35.98, and the average numeracy aggregate score for girls was 31.35. When this proportion increased to almost one-half of officials, the average aggregate score for girls increased to 36.88 for literacy and 29.95 for numeracy.

By contrast, in grade 8, the relationship between the proportion of woreda officials with differences in perceptions of boys' and girls' education and girls' performance in literacy and numeracy was as expected. When the proportion of woreda officials holding these perceptions increased, girls performance dropped. Specifically, where one-third of woreda officials supported boys' education but not girls', the average aggregate literacy score for girls was 46.32, and the average numeracy aggregate score for girls was 35.77. When this proportion increased to almost one-half of officials, girls' performance dropped; the average literacy aggregate score for girls was 38.27, and the average numeracy aggregate score for girls was 31.25.

These data suggest that the influence of woreda officials at lower grades may be smaller than their influence at higher grades. Further exploration of these trends through qualitative study may be warranted to better understand woreda officials' interactions and influence on girls at the primary grade levels.

4.3 Transition Outcome

As noted previously, the external evaluators implemented a two-phased approach to collect baseline data for STAGES. The first phase of the baseline focused on Outcome 1 Learning and Outcome 3 Sustainability and was conducted in April 2018 with cohort girls in grades 4, 6, and 8.[165] The second phase focused on Outcome 2 Transition; data for it was collected in December 2018 when cohort girls were expected to have transitioned into grades 5, 7, and 9.[166] Effectively, this approach to reporting on the transition outcome may differ from other projects because the transition data were collected in the academic cycle following when the baseline was completed, and represents the true transition of girls from the grade they were originally sampled in into the grades they were expected to transition to. This one-year transition was necessary to inform targets and benchmarks, and it serves as context for examining the two-year transition rates reported in subsequent evaluation points.

In addition, the transition pathways defined within the STAGES project exist within the school system only. Therefore, it only follows girls through transitions within the school context and across key primary and lower secondary school transition points. Transition into TVET, employment, or other post-secondary pathways is not part of the STAGES theory of change. Moreover, transition data was collected in interventions schools only; no data was collected in comparison schools. Instead, zone-level EMIS data for the 2017–18 school year is included to provide additional context on the baseline transition rates for the STAGES project.[167]

The external evaluators reviewed the EMIS data to identify relevant trends in the Wolaita Zone related to enrollment, repetition, promotion, drop-out, and survival rates at the primary and secondary level. This information was utilized to inform the setting of transition targets for future evaluation points. While this data provides useful information for ground-truthing baseline trends and informing target setting, it is essential to note some important differences in the collection and reporting of STAGES transition data and EMIS figures. Most pertinently, the EMIS figures provided are zone-level data and inclusive of all 19 woredas within the Wolaita Zone; the STAGES transition data focuses only on the four woredas where the project is implemented. In addition, the process for collecting and reporting the STAGES transition data is based on in-person verification of sampled cohort girls' transition status from one grade to the next; EMIS data focuses more on enrollment rosters and school-based reporting.

[165] This is towards the end of the school academic calendar in the Wolaita Zone context.
[166] This is towards the beginning of the new academic school year in the Wolaita Zone context. While schools may officially start in October, ongoing enrollment during the first month is common. In addition, girls' enrollment at the secondary level for grade 9 is contingent on the dissemination of Grade 8 exam results (i.e. pass rates). Therefore, any delays in this dissemination will also lead to delays in girls' ability to enroll in Grade 9. Therefore, Link and the external evaluator determined that verification of transition and data collection in November/December would provide the most accurate data as it would provide sufficient time for the grade 8 pass rates to be disseminated as well as capture late enrollers.
[167] This data is drawn from a soft-copy of the SNNPR Regional Education Sector Factsheet for the 2017/2018 school year that was made available to the project. This factsheet includes data that will be used to inform the forthcoming Regional EMIS Abstracts for the 2017/2018 school year.

For the STAGES baseline, the transition status of girls was cross-checked against enrollment lists as well as STAGES own monitoring data. Preliminary transition statuses were collected via a transition verification tracker form for cohort girls within the sampled primary schools—those girls moving from grade 4 to grade 5 and from grade 6 to grade 7—the month prior to girls transition survey data collection.[168] However, a similar approach within secondary schools to capture the transition from grade 8 to grade 9 was not feasible.[169] Instead, preliminary verification of the grade 8/9 transition was done in collaboration with primary and secondary school directors as part of the school mobilization workshop held the weekend before data collection. Transition status was then re-checked day of data collection by woreda officials when administering the girls' student survey. For girls who were not present on the day of data collection, their transition status was determined by woreda officials in consultation with the school director and relevant teachers.[170]

During data analysis, several unexpected trends emerged related to the transition rates of grade 8 cohort girls, especially within certain subgroups. Several variables have been identified as potentially providing a more nuanced understanding of the barriers to transitioning into a secondary school. For example, within the Wolaita context, students are required to pass a zone-level grade 8 exit exam in order to enroll in grade 9; in turn, the impact of grade 8 exam pass rates on transition requires further exploration.[171] In addition, distance to school often increases substantially at the secondary level and is likely another important variable that may contribute to girls' ability to transition.[172] Furthermore, an examination of the impact of bursary support as an enabling factor for successful grade 8 to 9 transition for the most marginalized girls was not possible at this time but warrants additional consideration at future evaluation points.

Table 18 defines the transitions for girls in STAGES between the key transition points identified in the theory of change—from grade 4 into grade 5, from grade 6 into grade 7, and from grade 8 into grade 9.

Table 17: Transition pathways

	Baseline point (Baseline Phase 1—spring)	Successful Transition (Baseline Phase 2—fall)	Unsuccessful Transition
Lower primary school	Enrolled in grade 4 in spring	In-school progression to grade 5 in fall	Remains in grade 4 in fall Did not return to grade 4 or 5 in fall in treatment school Transferred to school outside of treatment sample Drops out of school Status of transition unknown
Upper primary—Grade 6	Enrolled in grade 6 in spring	In-school progression to grade 7	Did not return to grade 6 or 7 in fall Remains in grade 6 in fall Transferred to school outside of treatment sample Drops out of school Status of transition unknown

[168] Preliminary transition data was collected in November 2018.
[169] This tracking proved difficult as it required the identification, sharing and cross-comparing of individual student level data across two separate school systems (i.e. the primary and the secondary) and physical locations.
[170] For girls who were no longer enrolled in the school, this should be understood as a proxy measure as the information was provided by the school, not the girls' themselves.
[171] This has the potential to create a substantial barrier to girls' ability to transition into grade 9, as it directly prevents all girls who do not sit for or fail the grade 8 exam from transitioning into secondary school. Unfortunately, it was not possible at the time of data analysis to directly link our individual cohort girl data with their associated grade 8 pass rate. In turn, the evaluators were unable to cross-compare and examine trends related eligibility to transition into secondary school
[172] It was not possible to examine this variable at this time as the current baseline data collection focused mainly on the primary school level, and included items on distance to primary schools only. Future data collection should include an item on the girls' surveys to capture this information.

	Baseline point (Baseline Phase 1—spring)	Successful Transition (Baseline Phase 2—fall)	Unsuccessful Transition
Upper primary—Grade 8	Enrolled in grade 8 in spring	Progression to lower secondary school (grade 9)	Did not return to grade 8 or 9 in fall Remains in grade 8 in fall Transferred to school outside of treatment sample Drops out of school Status of transition unknown
Out of school (age A to B)	n/a	n/a	n/a

These transition pathways reflect a typical education-focused transition pathway for girls at the woreda-level and zone-level. As discussed in the theory of change and in previous sections of this report, transition at the higher grades becomes increasingly difficult for girls as they often face increased household chore burdens, low aspirations, early marriage, and pregnancy. Moreover, transition into the secondary school requires girls to pass a grade 8 leaving exam at the zone-level in order to enroll in grade 9, greatly restricting the number of girls who are even eligible to transition. Furthermore, the number of secondary schools is limited, and distance to the closest secondary school can prove a considerable challenge for girls.

Benchmarking

Benchmarking for transition is based on the transition outcomes for girls in grades 6 and 8 after one year between phase one of baseline and phase two (i.e., transition from grade 6 into grade 7 and from grade 8 into grade 9. Because the benchmark sample also served as the baseline sample, benchmarking transition data for grades 6 and 8 girls is presented in Table 20a.

Table 18: Benchmarking for the Transition Outcome

Benchmark group		
	Benchmark transition pathway	**Transition rates**
See Table 20a		

Transition outcome of cohort girls

Transition outcomes for girls in treatment schools were examined using the barriers described in earlier sections of this report. Table 20b examines the transition rates for girls by each barrier separately, as required in the report template. Table 20a, however, examines the transition rates by a composite variable, or index, of four barriers that were prevalent in the population: (1) household chore burden; (2) parents living with a girl; (3) overall well-being; and (4) academic performance.[173, 174] The index is a mean on the four items, and the mean is then binned into four groups based on the distribution of scores on the composite. The resulting four subgroups on this Student Background Index can be described as follows:

[173] These four barriers were chosen as they were prevalent in the data, and because the external evaluator used the index to pull a purposive sample of parents to survey based on girl's survey responses in the spring. Since the external evaluator did not survey parents in phase one (spring), the girls survey data was used to create profile groups from which girls and their parents/caregivers were sampled (this process is further described in the MEL and in the methodology section in this report).
[174] See the methodology section for further details on the construction of this index and its use in sampling parent/caregivers for surveys.

- Level 1: No household chore burden, both parents living, high well-being, average to high academic performance
- Level 2: Some chore burden, one or both parents living, low/mod well-being, average academic performance
- Level 3: Moderate chore burden and/or living without parents and/or low well-being, low academic performance
- Level 4: High chore burden, no parents, low reported well-being, low academic performance

This Student Background Index is a useful tool to understand the impact of multiple barriers on girls' transition rates.[175] For girls in grades 4 and 6, transition rates seem to be predictable by these four underlying barriers—such that girls who are facing the most barriers (Level 4) have the lowest transition rates, but those in Level 1 have relatively high transition rates. For girls in grade 8, however, the trend is the opposite. One possible interpretation of this is that the barriers included in the index do not explain transition rates for girls in grade 8; another explanation is that girls in Level 4 receive project-provided resources, such as bursaries, that may mitigate the barriers impact.[176]

Table 19a: Transition Pathways for Intervention Group (Girls)

Grade	Transition Rate by Student Background Index							
	Level 1 no household chore burden, both parents living, high well-being, average to high academic performance		Level 2 some chore burden, one or both parents living, low/mod well-being, average academic performance		Level 3 moderate chore burden and/or living without parents and/or low well-being, low academic performance		Level 4 high chore burden, no parents, low reported well-being, low academic performance	
	%	n	%	n	%	n	%	n
Grade 4	87.80%	36	85.00%	85	80.00%	64	73.33%	33
Grade 6	84.62%	33	72.82%	75	67.61%	48	64.10%	25
Grade 8	55.88%	19	57.78%	52	67.27%	37	71.93%	41

Note: Grade 4 n=266 with 41 in Level 1; 100 in Level 2, 80 in level 3 and 45 in level 4. Grade 6 n= 582, with 39 in Level 1, 103 in level 2, 71 in level 3, and 39 in level 4. Grade 8 n=566 with 34 in level 1, 90 in level 2, 55 in level 3 and 57 in level 4.

The transition pathways and rates of transition for girls in the intervention group are shown in Table 20b for grades 4, 6, and 8 girls. Transition data were not collected from girls in the comparison group; instead, transition data at the zone-level from government reported EMIS systems are noted.

The overall transition rate across grades was 66.44 percent. This represents an attrition rate in a six-month period of almost 35 percent, exceeding the assumption made in the sampling approach of a 30 percent attrition rate in a two-year period. This will render the sample size by midline 1 smaller than expected.

In grade 4, the overall transition was 73.40 percent for girls transitioning into grade 5 in the fall. By age, girls who were 1–4 years older than their peers had a higher transition rate than girls who were on age for grade 4 (9–11 years old). Girls in grade 4 who faced the barriers discussed in this report generally had

[175] At future evaluation points the index may also include disability status.
[176] At future evaluation points the impact of bursary support will be examined.

mixed transition rates. Lower than average transition rates were observed among girls who had a high chore burden, were living without parents, had a household unable to meet basic needs, reported experiencing corporal punishment, and had low household support scores. Girls who had a low overall well-being, those who reported their teacher treats girls and boys differently, those who reported their teacher is often absent also had lower than average transition rates.

In grade 6, the overall transition rate was lower than it was for girls in grade 4 at 68.00 percent. Among girls who were on-age for the grade, 79.45 percent transitioned while the transition rate for girls who were 1–2 years older was 62.92 percent; for girls who were more than 2 years older, the rate was 58.33 percent. Girls in grade 6 who faced barriers had mixed transition rates compared to the overall average for the grade. Lower transition rates were observed among girls who were living without both parents, but for all other barriers, transition rates were higher than was average for the grade.

By disability status in grade 6 using the Child-Functioning questions, transition was examined among girls who were re-identified in the fall during phase two. Therefore, the sample size represents the number of girls within phase two respondents. When computing the proportion girls who have difficulty (using the Child-Functioning questions), the sample is one in the same as the phase two transition respondents. As a result, all transition rates for this sample of girls with difficulties are 100.00 percent and cannot be used to make programmatic changes to STAGES interventions.[177]

In grade 8, the overall transition rate—69.05 percent—was similar to that for grade 6. The transition rate for girls who were on-age for the grade (14–15 years old) was 69.50 percent, and the rate was lower for girls who were 1–2 years over-age (53.85 percent). Examining the rate in terms of other barriers, girls who reported that their teacher was often absent from class had lower transition rates; for all other barriers, the transition rates were higher than that of the overall average for the grade. These trends will be further explored to examine whether there is a relationship between girls who received bursaries or other project-provided interventions at the time of the baseline.

When considering girls by disability status in grade 8, all transition rates for this sample are 100.00 percent since Child-Functioning questions were administered in phase two along with the transition data collection.

Table 20b: Transition Pathways for Intervention Group (Girls) by Age

Intervention Group—by Age							
		Transition Pathway					Transition rates
Age	Sample size (#)	In-school progression	Retained in same grade	Moves into secondary school	Drops out of school[178]	Lost from sample (transferred, not registered, status unknown)	Successful transition rate per age (%)
Grade 4 students (enrolled in G4 during baseline phase one, April 2018)							
All G4	297	218	33	n/a	28	18	73.40%
Grade 4 by age							
6–8 yrs	3	1	1	n/a	0	1	33.33%

[177] This will be examined in more detail at the midlines when comparative data is available.
[178] Girl is known to have dropped out from registration information and through verification of Student Transition data on day of data collection.

Intervention Group—by Age							
		Transition Pathway					Transition rates
Age	Sample size (#)	In-school progression	Retained in same grade	Moves into secondary school	Drops out of school[178]	Lost from sample (transferred, not registered, status unknown)	Successful transition rate per age (%)
9–11 yrs	142	102	14	n/a	14	12	71.83%
12–13 yrs	101	79	10	n/a	10	2	78.22%
14–15 yrs	23	19	2	n/a	1	1	82.61%
16–17 yrs	1	1	0	n/a	0	0	100.00%
18–19 yrs	2	0	0	n/a	2	0	0.00%
20+ yrs	2	1	0	n/a	1	0	50.00%
Missing age	23	15	6	n/a	0	2	26.09%
Grade 6 students (enrolled in G6 during baseline phase one, April 2018)							
All G6	300	204	15	n/a	34	47	68.00%
Grade 6 by age							
6–8 yrs	0	0	0	n/a	0	0	--
9–11 yrs	7	5	0	n/a	0	2	71.43%
12–13 yrs	156	124	7	n/a	12	13	79.49%
14–15 yrs	89	56	4	n/a	16	13	62.92%
16–17 yrs	12	7	2	n/a	1	2	58.33%
18–19 yrs	3	2	1	n/a	0	0	66.67%
20+ yrs	4	2	0	n/a	2	0	50.00%
Missing age	29	8	1	n/a	3	17	27.59%
Grade 8 students (enrolled in G8 during baseline phase one, April 2018)							
All G8	294	n/a	13	203	10	68	69.05%
Grade 8 by age							
6–8 yrs	0	n/a	0	0	0	0	--
9–11 yrs	0	n/a	0	0	0	0	--
12–13 yrs	22	n/a	1	18	0	3	81.82%
14–15 yrs	200	n/a	10	139	8	43	69.50%

Intervention Group—by Age							
		Transition Pathway					Transition rates
Age	Sample size (#)	In-school progression	Retained in same grade	Moves into secondary school	Drops out of school[178]	Lost from sample (transferred, not registered, status unknown)	Successful transition rate per age (%)
16–17 yrs	26	n/a	2	14	0	10	53.85%
18–19 yrs	8	n/a	0	7	1	0	87.50%
20+ yrs	5	n/a	0	3	1	1	60.00%
Missing age	33	n/a	0	22	0	11	66.67%
Overall	891	average prevalence of each pathway across all ages (%)					66.44%

Table 21c: Transition Pathways for Intervention group (girls) by Subgroup

Intervention group- by Subgroup							
		Transition pathway					Transition rates
Age	Sample size (#)	In-school progression	Retained in same grade	Moves into secondary school	Drops out of school	Lost from sample (transferred, not registered, status unknown)	Successful transition rate per age (%)
Grade 4 students (enrolled in G4 during baseline phase one, April 2018)							
All G4	297	218	33	n/a	28	18	73.40%
Grade 4/5 by subgroup							
Living without both parents	15	12	0	n/a	3	0	80.00%
High chore burden	205	153	16	n/a	24	12	74.63%
Household unable to meet basic needs	126	96	12	n/a	13	5	76.19%
Poor overall well-being	105	70	16	n/a	9	10	66.67%
Low HH support	49	39	3	n/a	3	4	79.59%
Teacher treats boys and girls differently	204	149	19	n/a	25	11	73.04%
Teacher often absent	142	103	14	n/a	16	9	72.54%

Intervention group- by Subgroup							
		Transition pathway					Transition rates
Age	Sample size (#)	In-school progression	Retained in same grade	Moves into secondary school	Drops out of school	Lost from sample (transferred, not registered, status unknown)	Successful transition rate per age (%)
Student experiences corporal punishment [179]	144	109	15	n/a	13	7	75.69%
Grade 6 students (enrolled in G6 during baseline phase one, April 2018)							
All G6	300	204	15	n/a	34	47	68.00%
Grade 6/7 by subgroup							
Living without both parents	12	6	1	n/a	4	1	50.00%
High chore burden	181	126	11	n/a	25	19	69.61%
Household unable to meet basic needs	112	78	5	n/a	14	15	69.64%
Poor overall well-being	110	77	8	n/a	12	13	70.00%
Low HH support	73	52	7	n/a	7	7	71.23%
Teacher treats boys and girls differently	207	146	11	n/a	27	23	70.53%
Teacher often absent	136	97	7	n/a	18	14	71.32%
Student experiences corporal punishment	161	114	7	n/a	22	18	70.81%
Grade 6/7 by disability (WGQ Child Functioning administered in fall)							

[179] This was determined by the percentage of girls reporting yes" to at least two out of six questions on disciplinary actions by students This scale includes the following six items: Do your teachers discipline or punish students who get things wrong in a lesson, my teachers discipline students with physical punishment, my teachers discipline students with shouting, my teachers discipline students with detention, in the past week did you see a teacher use physical punishment on other students? (includes students who responded "once or twice" or "almost every day"), in the past week did a teacher use physical punishment on you? (includes students who responded "once or twice" or "almost every day"). Of these six items, only one asks about punishment that is not physical or verbally inappropriate (detention). This item is included here because it was part of the scale and does not skew the proportion of girls who reported yes to at least two of the six items because when they responded yes to the detention item, they also responded yes to at least two other punishment items. Therefore, the cutoff of two items does not inadvertently suggest that detention is corporal punishment.

Intervention group- by Subgroup							
		Transition pathway					Transition rates
Age	Sample size (#)	In-school progression	Retained in same grade	Moves into secondary school	Drops out of school	Lost from sample (transferred, not registered, status unknown)	Successful transition rate per age (%)
Difficulty communicating	4	4	0	n/a	0	0	100.00%
Difficulty remembering things or concentrating	6	5	1	n/a	0	0	83.33%
Difficulty hearing	0	0	0	n/a	0	0	--
Difficulty walking	0	0	0	n/a	0	0	--
Difficulty seeing	2	2	0	n/a	0	0	100.00%
Difficulty with self-care	2	2	0	n/a	0	0	100.00%
Student has at least 1 disability	13	12	1	n/a	0	0	91.67%
Grade 8 students (enrolled in G8 during baseline phase one, April 2018)							
All G8	294	n/a	13	203	10	68	69.05%
Grade 8/9 by subgroup							
Living without both parents	17	n/a	1	13	0	3	76.47%
High chore burden	174	n/a	8	120	8	38	68.97%
Household unable to meet basic needs	84	n/a	5	58	4	17	69.05%
Poor overall well-being	83	n/a	3	58	1	21	69.88%
Low HH support	50	n/a	1	37	3	9	74.00%
Teacher treats boys and girls differently	194	n/a	7	138	9	40	71.13%
Teacher often absent	110	n/a	6	69	6	29	62.73%

Intervention group- by Subgroup							
		Transition pathway					Transition rates
Age	Sample size (#)	In-school progression	Retained in same grade	Moves into secondary school	Drops out of school	Lost from sample (transferred, not registered, status unknown)	Successful transition rate per age (%)
Student experiences corporal punishment	150	n/a	2	113	3	32	75.33%
Grade 8/9 by disability (WGQ Child Functioning)							
Difficulty communicating	4	n/a	0	4	0	0	100.00%
Difficulty remembering things or concentrating	3	n/a	0	3	0	0	100.00%
Difficulty hearing	1	n/a	0	1	0	0	100.00%
Difficulty walking	1	n/a	0	1	0	0	100.00%
Difficulty seeing	2	n/a	0	2	0	0	100.00%
Difficulty with self-care	1	n/a	0	1	0	0	100.00%
Student has at least 1 disability	6	n/a	0	6	0	0	100.00%
Overall	891	average prevalence of each pathway (%)					66.44%

Table 22: Transition Pathways for Control group (girls)

Control group (girls)							
		Benchmark transition pathway					Transition rates
Age	Sample size (#)	In-school progression	Moves into secondary school	Enrolled in TVET course	Drops out of school	Lost from Sample	Successful transition rate per age (%)
		Not collected at the time of baseline phase two. See EMIS data reported in Table 22 for additional context.					

4.4 Subgroup analysis of the transition outcome

Since data were collected in two phases in the STAGES baseline, transition rates represent the actual transition of girls from the phase one grade level to the phase two grade level—one grade level progression. Using these data, transition rates were examined for girls in grades 4, 6, and 8 moving into

grades 5, 7, and 9, respectively. As presented in the previous section, transition rates were lowest for girls in higher grades when compared to girls in lower grades. By subgroup, transition rates for the following groups were lower than the overall average for their grade:

- Grade 4 girls who reported low overall well-being.
- Grade 6 girls who reported that they were living without both parents.
- Grade 4 girls and grade 8 girls who reported that their teacher was often absent.

4.5 Cohort tracking and target setting for the transition outcome

Tracking transition at midline will be done in a single-phase approach to examine the two-year transition rates for girls in the baseline grade 4, 6 and 8 cohorts when they reach grades 6, 8, and 10, respectively.

The targets presented in Table 22 represent stretch, but realistic, targets for each subsequent evaluation point based on the actual transition rates observed at the first evaluation point for the key transition. Targets determined by the outcomes spreadsheet are not included as the computation does not apply to the design used in this evaluation.

These alternative targets presented are based on the following assumptions:

1. That the sample will be "re-upped" so that the full target sample at each school will be the starting point;
2. The target for transition will be based on the difference in transition at evaluation point 2 minus evaluation point 1; and
3. The comparison is made for each cohort to its preceding data point, for instances comparing grade 6 transition rate at evaluation point 2 to grade 4 transition rate at evaluation point 1.

Therefore, Table 22 presents targets for the *transition rate* between each neighboring evaluation point.

Table 23: Target setting

	EMIS Wolaita Zone Contextual Data[180]		STAGES Baseline data and evaluation targets			
Cohort	**Repetition Rate for grade at baseline—Girls**	**Completion Rate for grade at baseline—Girls**	**Evaluation Point 1**	**Evaluation point 2**	**Evaluation point 3**	**Evaluation point 4**
Target generated by the outcome spreadsheet	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**	**N/A**
Grade 4 cohort	5.6%[181]	70.4%[182]	73.40% of baseline G4 girls transitioned successfully to G5	49.30% of baseline G4 girls transitioned successfully to G6	59.30% of midline 1 G6 girls transitioned successfully to G8	69.30% of midline 2 G8 girls transitioned successfully to G10
Grade 6 cohort	6.1%[183]	unavailable but estimated at 60.00%[184]	68.00% of baseline G6 girls transitioned	38.50% of baseline G6 girls transitioned	48.50% of midline 1 G8 girls transitioned	n/a

[180] This data is drawn from a soft-copy of the SNNPR Regional Education Sector Factsheet 2010 that was made available to the project. This factsheet includes data that will be used to inform the forthcoming Regional EMIS Abstracts for last school year. Please note that the year 2010 indicates the Ethiopian calendar, it aligns with the 2017/2018 school year according to the Gregorian calendar.
[181] This represents the grade 5 repetition rate for the Wolaita Zone.
[182] This represents the grade 5 completion rate for the Wolaita Zone.
[183] This represents the grade 6 repetition rate for the Wolaita Zone.
[184] Although this data is not available, evaluators estimates this as 60% given the prior and subsequent data points.

			successfully to G7	successfully to G8	successfully to G10	
Grade 8 cohort	15.0%[185]	50.5%[186]	69.05% of baseline G8 girls transitioned successfully to G9	40.60% of baseline G8 girls transitioned successfully to G10	n/a	n/a
Grade 9 to 10 (SeGRA/SeGMA completion)[187]	6.6%[188]	80.0% *(promotion)*[189]	Not collected	To be collected	To be set at midline 1	To be set at midline 1

Note: Relevant zone-level data has been provided to contextualize existing trends at the zone level and the rationale for future evaluation point targets.

4.6 Sustainability Outcome

Activities and outcomes under the STAGES project are geared towards embedding respect and support for girls' education within the target communities and the decentralized education system. Moreover, localized activities like SPAMs and Gender Action Plans will be unique to each school, offering a bottom-up approach to ensure support for girls' education and maximum buy-in from local stakeholders. Considering these points, the indicators for sustainability were selected to cover a wide range of domains—attitudes, support, engagement, and pursuit—and across interventions. As with GEC1, it is expected that different communities will utilize interventions in subtly different ways according to need. To adequately capture this, future iterations of the sustainability outcome will seek to highlight emergent indicators of sustainability. Indicators of sustainability will also be further informed once the definition of sustainability has been vetted by project staff.

Sustainability is assessed at the school, community, and system level according to the sustainability scorecard. Results for treatment schools are reported in Table 21. Since STAGES will work with all communities- and system-level stakeholders in the four target woredas, the community and system outcomes cover all beneficiaries at these levels.

Table 24 summarizes the baseline sustainability score for each level with a brief summary of the baseline status. Qualitative and quantitative data are presented in support of the score at each level.

Table 24: Sustainability Indicators

	Community	School	System
Indicator 1:	Indicator: Proportion of communities who have the capacity and willingness to sustain STAGES activities Baseline status: KIIs did note that STAGES programming was consistent with the larger government approach to girls' education.	Indicator: Percentage of GAP targets/actions have been undertaken Baseline status: no data available as GAP targets/actions were not set as of baseline.	Cost sharing from woredas, communities, and schools in support of STAGES missions (values of in-kind supports borne by MOE, community members, parents, teachers, school directors, and students in the service of primary outcomes) Baseline status: Strong integration of STAGES activities with government systems, as well as alignment with government goals sets the context for high-levels of

[185] This represents the grade 8 repetition rate for the Wolaita Zone.
[186] The represents completion rate for grade 8 for the Wolaita Zone.
[187] Transition from grade 9 to 10 will be captured during learning assessment data collection conducted with grade 10 students at the end of school year.
[188] This represent the grade 9 repetition rate.
[189] The indicates promotion rate at grade 9.

	Community	School	System
			cost-sharing.[190] An estimated cost-saving of $75K USD was achieved through the use of woreda officials as enumerators.[191]
Indicator 2:	Proportion of community stakeholders actively engaged in SPAM, GEAC, GAP activities, and other groups including mothers and fathers' groups and PTSAs Baseline status: All primary schools in the sample have GEACs and active PTSAs, although the proportion of schools in woredas may be fewer.	Number of schools accredited as "girl-friendly." Baseline status: Activities for accreditation of schools as girl-friendly has not begun yet. Gender-friendliness index scores were low while at least one-quarter of teachers reported their school acted in a girl-friendly way.	Proportion of government officials who are actively engaged in the delivery of STAGES activities—such as leading training, coordinating workshops, facilitating activities/events, collecting data Baseline status: At baseline, approximately 150 government officials served as data collectors for all surveys and learning assessments; 16 female teachers served as either facilitators or notetakers for qualitative data collection.
Indicator 3:	Proportion of community stakeholders who report pursuing new initiatives or activities to further support STAGES interventions Baseline status: no respondents were asked directly about this in surveys nor were detailed supporting statements recorded in qualitative data. However, the baseline data collection activity itself represents a new initiative for the government staff, where they engaged with child protection policies (CPP). After data collection, government staff expressed interest in engaging with child protection officers, club coordinators, and teachers in ensuring girls' and boys' safety in schools.	Proportion of schools with regular monitoring of teaching quality Baseline status: woreda officials or school administrators were not asked about classroom monitoring practices at baseline.[192]	n/a
Indicator 4:	n/a	Proportion of school directors, teachers, and PTSAs, GEACs, mothers	n/a

[190] In the latest RAM report submitted by Link, it was noted that "STAGES sustainability strategy that Mother Groups and Gender Action Planning incorporated into National Girl's Education Strategy. It appears (to be fully confirmed) that the MOE are also conducting gender auditing activities as per Link program. These are positive sustainability signs."

[191] It is estimated it would cost a daily rate of $50 USD per enumerator, with two weeks payment for 150 enumerators if external data collectors or enumerators were contracted. Moreover, it's important to recognize that while the project covered transportation costs by providing woreda officials per-diems during data collection, the MOE assumed the regular salary costs for woreda officials during the time of they served as enumerators.

[192] In the latest Review and Adaptation Meeting (RAM) report submitted by Link, Activity 2.5: Teacher coaching/mentoring and monitoring by cluster supervisors (maps to output indicator 2.1 and 2.3), This is ongoing on a monthly basis, and in this quarter Link-Ethiopia have provided the means to conduct the activities (fuel/lubricant for motorbikes) and supported supervisors to monitor key areas of GEC-T implementation (capacity building for school directors and teachers, tutorials for girls, and provision of sanitation rooms/items).

	Community	School	System
		and fathers groups who pursue new initiatives or activities to further support girls' education Baseline status: the presence of PTSAs, GEACs, and other groups were asked about; not about new initiatives.	
Baseline Sustainability Score (0–4)	**1.00** (limited data only available on one of three indicators)	**1.00** (limited data only available on one of four indicators)	**3.00** (data only available on one of two indicators)
Overall Sustainability Score (0–4, an average of the three-level scores)	**Overall score = 1.70** **(based on available data for three out of nine indicators)**		

Systems-level: Cost sharing by community members for STAGES activities were not included in survey instruments and questionnaires. Instead, fundraising activities by the school were inquired about; the majority of respondents (66.70%, n=10) to the school audit survey said "no" and 100.00 percent of comparison school respondents (n=15) said no.

As a precursor to cost-sharing activities, the level of collaboration with the community was asked of woreda officials. One-third said they "disagree a lot" to questions about whether they are encouraged by their superiors to cooperate with communities, schools, parents, and communities. Without a cooperative

> Link's interventions about girls' education are also highly linked to the policies, strategies of the MOE. It is also informed by evidences and best practices, such as the impact of tutorial services, and capacity building supports to the education system, to increase girls' performance.
>
> Zone-level Key Informant

environment, expectations of cost sharing are limited. By contrast, the level of support education officials cited from STAGES was high, including a willingness to participate in all program design activities and lead data collection efforts in the field.

At baseline, the MOE supported data collection through the provision of approximately 150 zonal- and local-level Ministry staff for 2-weeks of training and data collection in schools. The cost of external data collectors is approximated at $50 USD/per person/per day. The total cost shared by the MOE, as a result, is approximately $75,000.

While little was specifically said by respondents in regard to GEACs, GAPs, SIPS, or SPAMS, KIIs did note that STAGES programming was consistent with and expanded on the larger government approach to girls' education.

Figure 6: Zone-level Key Informant Explains Link's Approach

Despite the limited indicator data on the system sustainability, a baseline systems sustainability score of 3 has been determined appropriate by the evaluators for several reasons. First and foremost, the design

and implementation of the STAGES project have sustainability embedded throughout the core of the project with its direct engagement, partnership and capacity building of zone and woreda-level personnel throughout the design, implementation, monitoring, and evaluation of project activities. This is an important distinction for other programmatic approaches that may recruit, hire or train outside personnel to conduct these activities—often creating external or parallel systems or structures that are difficult to sustain beyond the end of the project. In contrast, STAGES supports government personnel to be part of the decision-making, implementation and learning process throughout its activities, which not only supports greater ownership and accountability throughout the process and life of the project, but also situates the knowledge, learnings and best practices within the government personnel and systems—rather than outside of it—so it may be carried out beyond the timeline—and targeted woredas—of the project.

Moreover, STAGES supports the government in activating the systems and structures promoting gender equality and girls' education that already exist. As noted in the STAGES GESI Analysis, there is a strong government policy environment in existence regarding girls' and inclusive education in Ethiopia and the project was designed to support the government in implementing aspects of these policies.[193] Furthermore, at the Zone and woreda-level there are already assigned, existing gender-focused personnel—gender officers—whose Link is working with to build their capacity to be effective in their roles supporting girls' education both now and in the future. With the existence of an enabling policy environment and gender-sensitive personnel structures or frameworks already in place, some of the barriers and obstacles projects often face in supporting girls' education at the systems level have been greatly reduced.

Lastly, while this is a baseline for the STAGES project specifically, it is also important to recognize that it is far from the start or beginning of Link's work and engagement in education systems support in the Wolaita Zone. STAGES is drawing and building on in-depth experience, knowledge and continued relationships that come from working in the zone for the past 10 years, including most recently its heavy engagement and collaborative partnership with zone and woreda level education officials throughout GEC1. This experience, including the relationships and trust developed between Link and government education officials, set both the project and government up for sustainable, system-level success and the scoring reflects this.

While systems-level integration with STAGES activities is strong, the focus on girls with disabilities has yet to permeate the government structure. The government's prioritization of addressing specific barriers to girls' education for the most marginalized was widely mentioned by key informants. Numerous respondents stated that despite the government's specific call to address challenges for the most marginalized girls little had been done. Several key informants did mention that special provisions were made for girls with children, such as allowing them to arrive at school late or leave early. The majority of key informants agreed that more support at all levels was needed to provide for female students with disabilities and that many schools were ill-equipped to provide any additional support to these students. According to one key informant, there had been, "some improvements related to discussing the needs of children with disabilities in the school system, but a lot remains." This sentiment was echoed by several other respondents citing a "critical shortage of teachers and resources for children with disabilities."

Community-level: According to the school audit surveys, 100.00 percent of treatment schools (n=15) and 87.00 percent of comparison schools (n=13) have an active GEAC. This suggests that all treatment schools are well positioned to set GAP targets and incorporate SPAMs into their practices. However, at least four woreda officials noted that, of all schools in their woreda, some did not have an active GEAC. Similarly, at least two woreda officials noted that some schools in their woreda did not have active PTSAs.

[193] Casey McHugh and Ashley Doria, *STAGES GESI Analysis* (Pacifica: School-to-School International, 2017).

KIIs relayed that although community support for girls' education had improved in recent years, more still needed to be done to raise awareness and encourage support of girls' education. Several KIIs stated that community support for girls' education was strong in primary school but tended to decrease starting in early secondary school and all community support "ceased in upper secondary." One key informant stated that in order to address challenges to girls' education, "attempts should be done to raise awareness aimed at bringing behavioral change in the community to help value girls' education." Several respondents hinted at the lack of responsibility families and communities felt towards girls; one respondent citing a common saying in their community, "girls for their husbands, boys for their parents." Additionally, KIIs alluded that community support for girls' education was not only important to shifting social norms but to ensuring long-term sustainability and financial support.

School-level: Scores on the school gender-friendliness index were lower for treatment schools than for comparison schools—1.38 versus 1.63 on a three-point scale. When teachers were asked if their school acts in a girl-friendly way, one-quarter (27.27 percent) of teachers in treatment schools said "always" compared to one-third (34.44 percent) of teachers in comparison schools.

The following subsection and Table 25 were completed by the project.

Table 25: Changes Needed for Sustainability

	Community	School	System
Change: **What change should happen by the end of the implementation period?**	Attitudinal changes and shifts in the status of girls' education and gender and inclusion and safeguarding norms in Wolaita Zone Mobilized gender and inclusion-aware communities demanding high-quality education Positive community attitudinal change for girls' and inclusive education	Improved leadership for inclusive girls' learning at the school level Improved attendance and retention of girls in schools School improvement plans in all schools in the Wolaita Zone that contain girls' education targets Decreased dropout rates and increased transition from grade 4 to 5, grade 8 to 9, and grade 10 to 11 in target communities Increased understanding of SRGBV and improved reporting mechanisms Improved school management and governance in project primary and secondary schools, with increased accountability through community engagement Shorter and safer journeys for girls to less crowded secondary schools due to four new local rural	Improved leadership for girls' learning at woreda, zone, and regional level Government personnel able to implement interventions consistent with GEC-T and GEC1, resulting in continued literacy and numeracy gains Increased capacity of regions, zones, woredas, and schools to monitor and plan for girls' education through a critical mass of motivated leaders GRP becomes part and parcel of pre-service training for all teachers in SNNPR

	Community	School	System
		secondary schools, which will be constructed, staffed, and managed by MOE.	
Activities: **What activities are aimed at this change?**	Awareness raising and campaigns PTSA and SIC and KETB (Kebele Education and Training Board) capacity development SPAMS Mother and father groups Gender clubs Good brother awards	Teacher training and mentoring on gender-responsive pedagogy for primary and secondary teachers Language competency teacher training Learner testing in grades 4, 7, and 9 in all schools Awards for outstanding teachers Teacher coaching, mentoring, and monitoring by cluster supervisors Construction and equipping of four rural secondary schools	Capacity development of partners at woreda level (monitoring and supervision of gender, inclusion-responsive, and safe schools. Cluster- and woreda-level (SPAMS Zonal and Regional Girls' Education Conferences and dialogue Gender mainstreaming training for all SNNPRs zone and woreda heads, woreda gender officers, planning unit performers, and cluster supervisors SRGBV training for schools, zone, and woreda heads and woreda gender officers Roll-out of activities identified by the Region to be particularly valuable (e.g., GRP or English Language Training).
Stakeholders: **Who are the relevant stakeholders?**	Community members, community leaders, PTSA and SIC members, parents, education and training boards at the kebele level Mother and Father Groups Brothers who support their sisters to attend and stay in school	Head teachers, teachers, and students	SNNPRs Regional and Zonal Education Bureaus; Wolaita Zone Education Department; Woreda Education Office staff; and SNNPRs Bureau of Women, Children, and Youth Affairs
Factors: **What factors are hindering or helping achieve changes? (Think of people, systems, social norms, etc.)**	**Helping**: School communities mobilized through GEC1, change already happening **Hindering**: Social norms around gender and additional factors of marginalization, including disability, still require further emphasis	**Helping:** Schools and teachers benefited from GEC1, which achieved positive results for schools and learning outcomes **Hindering:** Improving teaching practice and pedagogy to support girls' education is a process; it also requires changes in	**Helping:** STAGES works directly with the government in designing, implementing, monitoring, and evaluating program strategy and activities. There are very strong relations between Link and the relevant government education authorities,

	Community	School	System
		attitude and social norms, especially around girls with additional factors of marginalization, including disability. Further training emphasis has been implemented to support teachers to identify and meet the needs of all children in the classroom	particularly through the implementation of GEC1. **Hindering**: As with school level, attitudinal change takes time, and continued emphasis will be needed to support the system to respond to the needs of girls, including girls affected by multiple factors of marginalization

The STAGES project aims to transform access to secondary education, establishing locally owned systems to improve the quality of education, and creating sustainable support for girls in the long-term. It aims to bring holistic change about in leadership for girls' learning at all levels, the quality of learning that girls receive in school, access to secondary schools in remote areas, girls' readiness to learn, and community support for girls' education. The changes it aims to bring about are systemic, supporting the system to include all girls and boys, working through the government directly and through community engagement for improved accountability. All project activities are designed to be sustainable in the longer term: capacity development of government offices at all levels; capacity development of teachers and school directors; and formation of formal community-school structures to improve governance and accountability through community engagement. While some elements offer direct support to girls—such as the provision of uniform, sanitary pads, and items such as soap—other may not be provided in the future by government agencies as the learning on the difference that such provision makes to girls' education (attendance, transition, learning), can influence thinking on how such support is provided beyond the program.

The findings against the sustainability measures in the baseline correlate well with the previously discussed factors where measurement has been possible. Specifically, it found that STAGES programming is consistent with the government approach to girls' education; strong integration of STAGES with government systems; alignment with government goals; all primary schools having active PTSAs and established GEACs; and government staff with a growing interest in child protection. Where scores show "low gender-friendliness" and only one-quarter of teachers reporting their school to be acting in a "girl-friendly" manner, this correlates with the point in the table on the need to continue to work around attitudinal change for marginalized girls as inclusion is a process.

5. Key Intermediate Outcome Findings

This section presents the key IOs and baseline results for each outcome. Note that outcomes that focus on changes in stakeholders' perceptions and behaviors (well-being, perceived access, self-esteem) have revised targets that are lower than those presented in the logframe and MEL. These revisions are suggested primarily because of the complexity of relationships between these factors and other barriers, as presented in Section 3. Details on the items used in the reporting of all scores cited in the results can be found in Annex 18.

5.1 Attendance

Intermediate outcome and indicator selection and measurement

As highlighted in the STAGES' theory of change, attendance is a key focus—and a required IO—of the project's activities. Poor attendance is one of the challenges girls encounter in being "ready to learn." Moreover, attendance is a prerequisite for improved learning, transition, and sustainability of the interventions. Therefore, the project aims to reduce this challenge by supporting greater quality and engagement for girls when they do come to school, as well as addressing barriers girls face that currently prevent them from attending school.

Under the IO on attendance, the project aims to reduce the following barriers to education: poor access to secondary schools in extreme and remote areas; limited opportunities and access to academic; and limited social-emotional, basic needs, life skills, and hygiene inputs in school.[194]

IO Indicator 1.a: Percentage improvement in attendance rates

The percentage of improvement in attendance rates was selected as an indicator for this IO as it allows tracking of actual attendance of all girls in treatment schools over time. In the MEL framework, attendance tracking intended to use school registers at all evaluation points, on a quarterly basis, and through spot checks.[195] However, due to constraints in enumerators' time in sampled schools during the baseline administration and uncertainty in the reliability of attendance monitoring, attendance using school registers was not captured for all girls at baseline.[196] Therefore, the data that are used to report on attendance at baseline include the following measures:

1. The proportion of sampled girls who were not present on the day of surveys; and
2. Self-reported attendance of sampled girls for the number of days they were present at school in the past week (phase one) or absent in the past week (phase two).

Several steps will be implemented to track attendance quarterly and during spot checks for future M&E points. The external evaluators will work with Link to determine the reliability of school registers as well as the supports required to improve the quality of attendance data available in project schools. During quarterly monitoring visits, attendance data for all girls in a sample of schools will be collected by Link. Monitoring instrument may include

- **Girls' self-reported attendance**. Girls from the cohort sample responding to an identical baseline survey question on "how many days were you absent, or did you miss school in the past school week?"
- **Review of attendance data from the school register**.[197] Discrepancies will be evaluated against other sources with the teacher and school director. Monitoring instruments will record the

[194] Refer to the STAGES' theory of change for specific activities.
[195] The attendance indicator does not include reporting on attendance for the comparison school cohorts.
[196] School visits were also only feasible for the 15 treatment schools included in the baseline evaluation.
[197] The list of female students sampled at baseline will be utilized as the basis for the girls expected to be present in school.

discrepancies identified and use this information to improve the quality of attendance data maintained by the school.

- **Teachers' assessment of girls' level of attendance**. Class enrollment lists may be used to obtain a proxy measure of girls' attendance over the past five days.

As a result of these steps, the source of attendance data tracking—and the specific indicators—may need to be adjusted in the project's logframe. For example, sources may need to be changed from school registers to self-reported surveys. Since comparisons of girls' attendance rates will be made at each evaluation point and compared to the prior evaluation points as well as the monitoring data, consistency in sources and reconciling discrepancies between sources will be an important consideration in any revisions.

IO Indicator 1.b: Percentage of students with improved perceptions of access

Perception of access will be captured through the girls student survey at all evaluation points. At baseline, girls were asked four questions to assess their perception of access during phase one:

- How long does it usually take you to get to school?
- Do you feel safe traveling to and from school?
- Is it reasonably easy for you to get to school and back?
- Do you feel that the distance to your school is very close, somewhat close, somewhat far, or very far?

Because these questions were administered in the baseline instruments to sampled primary school girls, the results can be generalized to the population of treatment girls participating in the STAGES project at the primary school level.[198]

Findings

The baseline levels for each of the attendance indicators are reported in the following sections. The reported data constitute the data available from 15 treatment schools and the girls who were in attendance at the time of baseline. In subsequent reports, the evaluators will attempt to identify appropriate sources and data to report on the indicators as stated in the logframe.

Indicator 1.a1: The proportion of sampled girls who were absent on the day of surveys

The sample of primary school girls was randomly drawn from the enrollment lists for the targeted grade levels in the sampled treatment schools.[199] As such, the random sample represents the population of girls from which the schools and students were drawn; the proportion of replacement girls required on the day of assessment can be used to examine attendance.[200]

At the time of the phase one surveys, 8.80 percent of girls in treatment primary schools (n=78) were not present and had to be replaced in the survey data collection.[201] By grade, replacements in treatment schools were highest in grades 4 and 6 and lowest in grade 8.

[198] Results at future evaluation points will be disaggregated, at a minimum, by primary and secondary school level.
[199] The enrollment list was also verified by school directors and woreda quality assurance officers prior to the beginning of data collection for evaluation surveys.
[200] Less than two weeks had passed between the time of student list verification and administration of the survey. These data do not account for or differentiate girls who may have dropped out in that period from girls who were absent.
[201] In comparison primary schools, 14.10 percent of randomly selected female students were not present on the day of evaluation survey data collection.

Supplementary Table 14. Percentage of Sampled Girls Absent on the Day of Learning Assessments IO 1.a.1—Baseline and Target Figures

Cohort	Baseline (2018)		Midline 1 (2020)	Midline 2 (2022)	Endline (2024)
Treatment	***Actual***	***n***	***Targets***		
Grade 4 cohort	10.37%	31	7.00%	5.00%	4.00%
Grade 6 cohort	10.14%	30	7.00%	5.00%	-
Grade 8 cohort	5.88%	17	4.00%	-	-
Average (G4, G6, G8)	**8.80%**	**78**	-	-	-
Grade 10[202]	TBD	TBD	TBD	TBD	TBD

Note: Treatment n=884. Grade 4 treatment n=299; grade 6 treatment n=296; grade 8 treatment n=289.

Supplementary Table 15. Percentage of Sampled Girls Absent on the Day of Assessment at Baseline by Subgroup IO 1.a.1

Grade Level	Grade 4 cohort		Grade 6 cohort		Grade 8 cohort	
Subgroup	***%***	***N***	***%***	***N***	***%***	***N***
Girls with at least one disability	N/A	N/A	not reported[203]			
Girls who are overage for grade	11.29%	7	14.29%	7	0.00%	0
Girls in households unable to meet basic needs	3.17%	4	7.14%	8	2.38%	2
Girls without high levels of self-reported well-being	3.66%	3	6.17%	5	6.00%	3
Girls with low levels of support from household	4.08%	2	4.11%	3	2.00%	1
Girls who report teachers treat boys and girls different in the classroom	7.35%	15	8.70%	18	5.15%	10
Girls who report teachers are often absent from class	8.45%	12	9.56%	13	5.45%	6
Girls who report high corporal punishment exercised by the teacher	8.26%	10	8.33%	11	5.13%	6

Note: Treatment n=884. Grade 4 treatment n=299; grade 6 treatment n=296; grade 8 treatment n=289.

Indicator IO 1.a2: Self-reported attendance of sampled girls—number of days they were in attendance in the past week

Of all the girls in treatment schools who were surveyed during phase one of the baseline—including replacements—approximately two-thirds reported attending all five days of school in the past week. The targets for subsequent evaluation points are the same as those presented in the logframe, and no adjustments are suggested. A similar question was included in the girls' transition survey during phase two of the baseline to examine the attendance rate of girls who successfully transitioned to the next grade; however, this question asked how many days girls had been absent in the past school week as opposed to asking about how many days attended. While the proportion of grade 8 girls with high levels of attendance was similar across phase one and two, there was considerable variation for the grade 4/5

[202] Grade 10 attendance data will be collected as part of monitoring. The baseline measure for grade 10 will be based on those data and targets set accordingly.

[203] This data should be understood as missing. Results for disability subgroup—barrier 1—were not possible to disaggregate at this time as the Child Functioning set was administered in phase two data collection and has a different denominator than the relevant survey items. This information will be provided starting at midline using the Washington Group Child Functioning questions.

and 6/7 cohorts. Moreover, while the questions both focus on the same concept, it remains unclear to what extent responses were impacted by the structure of the question, the time of year data was collected, or the grade level. Given these limitations, the findings are included for additional context at this time but may be examined in more detail at future evaluation points.

Supplementary Table 16. Self-reported Attendance of Sampled Girls who Attended All Five Days in the Past Week IO 1.a.2—Baseline and Target Figures

Grade Level	Baseline Phase 1 (Apr 2018)		Baseline Phase 2[204] (Dec 2018)		Midline 1 (2020)	Midline 2 (2022)	Endline (2024)
Treatment	***Actual***	***N***	***Actual***	***N***	***Targets (+% over phase one Actual)***		
Grade 4 cohort	67.22%	201	*59.51%*	*147*	+1%	+3%	+5%
Grade 6 cohort	75.68%	224	*66.36%*	*144*	+1%	+3%	-
Grade 8 cohort	75.09%	217	*74.16%*	*132*	+1%	-	-
Average (G4, G6, G8)	**72.62%**	**642**	*65.88%*	*423*	+1%	+3%	+5%
Context: Grade 9	74.16%[205]	132	*n/a*	*n/a*	n/a	n/a	n/a

Note: Baseline phase 1 n=884. Grade 4 treatment n=299; grade 6 treatment n=296; grade 8 treatment n=289. Baseline phase 2 n=642. Grade 4/5 treatment n=247; grade 6/7 treatment n=217; grade 8/9 treatment n=178.

The proportions of girls who attended all five days across the three grades were lowest among girls who had low levels of support from their household and girls who reported high levels of corporal punishment from teachers.

Supplementary Table 17. Self-reported Attendance of Sampled Treatment Girls Who Attended All Five Days in the Past Week at Baseline by Subgroup IO 1.a.2

Grade Level	Grade 4		Grade 6		Grade 8	
Subgroup by Barriers[206]	***%***	***N***	***%***	***N***	***%***	***N***
Girls with at least one disability	N/A	N/A	not reported[207]			
Girls who are overage for grade	72.58%	45	71.43%	35	74.36%	29
Girls in households unable to meet basic needs	60.32%	76	75.89%	85	67.86%	57
Girls without high levels of self-reported well-being	50.00%	41	50.91%	56	39.76%	33
Girls with low levels of support from household	42.86%	21	64.38%	47	64.00%	32
Girls who report teachers treat boys and girls different in the classroom	65.20%	133	78.74%	163	76.29%	148
Girls who report teachers are often absent from class	61.97%	88	70.59%	96	75.45%	83
Girls who report high corporal punishment exercised by the teacher	63.64%	77	60.87%	98	58.67%	88

Note: For this item, n sizes are as follows: grade 4 treatment n= 274; grade 6 treatment n=271; grade 8 treatment n=261; for all other subgroup analyses: grade 4 treatment n=267; grade 6 treatment n=296; grade 8 treatment n=289.

[204] Findings from phase two of the baseline are provided as additional context.
[205] Results from Grade 8/9 cohort from phase two of the baseline is serving as a proxy for grade 10 data until it grade 10 becomes available at midline.
[206] This data should be understood as missing. Results for disability subgroup—barrier 1—were not possible to disaggregate at this time as the Child Functioning set was administered in phase two data collection and has a different denominator than the relevant survey items. This information will be provided starting at midline using the Washington Group Child Functioning questions.
[207] Ibid.

IO Indicator 1.b: Percentage of students with improved perceptions of access

In order to determine girls' perceptions of access, a perception-of-access index was developed and computed. Details on the girls' student survey items that were included, as well as coding and scoring of the index, are presented in Annex 18.[208]

For the purposes of baseline report, the proportion of girls reporting "excellent" perceptions of access are shown in the subsequent tables. For subsequent evaluation points, the proportion with improved perceptions of access will be reported. In other words, at subsequent evaluation points, the Supplementary Table 16 will be updated to include students who have moved from "very poor" to "poor," from "poor" to "adequate," from "adequate" to "good," and from "good" to "excellent."

At baseline, more than half of all primary girls in treatment schools surveyed reported having excellent perception of access as demonstrated by responding yes to all four items included in this index. In comparison primary schools, 52.27 percent of girls had an excellent perception of access, compared to 59.84 percent in treatment schools. Perceptions of access were more positive among grade 8 girls in treatment schools, followed by grade 6 and grade 4 girls. The proportions of girls with very poor, poor adequate, and good perceptions of access in treatment schools were 2.15 percent, 8.60 percent, 11.43 percent, and 17.99 percent, respectively.

The targets noted are revised from those in the logframe and will be based on the proportion of students with improved perceptions and not just excellent perceptions of access.[209]

Supplementary Table 18. "Excellent" Perception of Access IO 1.b—Baseline and Target Figures

Cohort	Baseline (2018)		Midline 1 (2020)	Midline 2 (2022)	Endline (2024)
Treatment	***Actual***	***n***	***Targets***		
Grade 4 cohort	55.18 of girls with excellent perception of access	165	10.00% with improved perception; 60.00% of girls with excellent perception of access	20.00% with improved perception; 65.00% of girls with excellent perception of access	30.00% with improved perception; 75.00% of girls with excellent perception of access
Grade 6 cohort	61.15% of girls with excellent perception of access	181	10.00% with improved perception; 65.00% of girls with excellent perception of access	20.00% with improved perception; 70.00% of girls with excellent perception of access	-
Grade 8 cohort	63.32% of girls with excellent perception of access	183	10.00% with improved perception; 65.00% of girls with excellent perception of access	-	-
Average (G4, G6, G8)	**59.83%**	**529**			

Note: Total n =1722. Comparison n=838; treatment n=884. Results reported for treatment schools only. Grade 4 treatment n=299; grade 6 treatment n=296; grade 8 treatment n=289. Baseline data only shows the proportion of girls with excellent perception. Subsequent time points will show the proportion of girls that move from very poor to poor, from poor to adequate, from adequate to good, and from good to excellent.

[208] Original items included the following four questions: How long does it usually take you to get to school; do you feel safe traveling to and from school; is it reasonably easy for you to get to school and back; is it reasonably easy for you to get to school and back; do you feel that the distance to your school is very close, somewhat close, somewhat far, very far?
[209] Targets noted in the logframe were 20 percent, 40 percent, and 60 percent students with improved perceptions.

When considering the data by subgroups, the lowest proportions of girls who had an excellent perception of access were those who reported low levels of well-being. In subsequent time points, results will be reported against the proportion of girls with improved perceptions along with proportion with an excellent perception of access.

Supplementary Table 19. "Excellent" Perception of Access at Baseline by Subgroup IO 1.b

Grade Level	Grade 4		Grade 6		Grade 8	
Subgroup	***%***	***N***	***%***	***N***	***%***	***N***
Girls with at least one disability	N/A	N/A	not reported[210]			
Girls who are overage for grade	45.16%	28	55.10%	27	61.54%	24
Girls in households unable to meet basic needs	42.06%	53	46.43%	52	51.19%	43
Girls without high levels of self-reported well-being	36.59%	30	43.21%	35	34.00%	17
Girls with low levels of support from household	42.86%	21	45.21%	33	44.00%	22
Girls who report teachers treat boys and girls different in the classroom	54.41%	111	61.35%	127	63.40%	123
Girls who report teachers are often absent from class	61.97%	88	63.24%	86	69.09%	76
Girls who report high corporal punishment exercised by the teacher	43.80%	53	62.12%	82	59.83%	70

Note: For this item, n sizes are as follows, grade 4 treatment n= 274; grade 6 treatment n=271; grade 8 treatment n=261; for all other subgroup analyses: grade 4 treatment n=299; grade 6 treatment n=296; grade 8 treatment n=289.

In grades 4 and 6, girls who reported attending school all five days and those with higher perceptions of access also had higher learning scores on select literacy and numeracy subtasks—correlation coefficients between 0.12 and 0.34. In grade 8, higher perceptions of access were related to higher reports of attendance—correlation coefficients between 0.14 and 0.29.

Baseline data suggest that attendance is an important IO for learning and for transition because the primary vehicle of the STAGES interventions is in-school supports. Reducing barriers girls face in coming to school is a purposeful area for STAGES and has been identified in the theory of change as a critical issue requiring resolution so that girls can fully take advantage of resources being provided to them in school. The levels of attendance reported at baseline are critically low for specific subgroups of girls. This information can help identify areas of targeted intervention and support. As such, achieving this IO, particularly in the higher grades, will be important to learning and transition outcomes.

Interpretation and Reflections

Attendance data will be collected through each evaluation point as well as quarterly monitoring visits on a sample of girls. As such, the logframe has been updated to reflect a sampled approach in both the monitoring and evaluation point data collections for approval from the fund manager.

The barriers that seem to present the greatest challenge to improving girls' attendance include support in the home, well-being, and corporal punishment. This, however, affects subgroups of girls differently. The overall level of attendance is between two-thirds to three-quarters in grades 4, 6, and 8. The evidence suggests that attendance is lowest in the lower grades; however, it should be noted that among girls in

[210] This data should be understood as missing. Results for disability subgroup—barrier 1—were not possible to disaggregate at this time as the Child Functioning set was administered in phase two data collection and has a different denominator than the relevant survey items. This information will be provided starting at midline using the Washington Group Child Functioning questions.

grades 6 and 8, those individuals facing the greatest challenges may have already dropped out of school. With the project's focus on retaining more girls in higher grades, this attendance rate may drop as girls with greater challenges enroll and are retained in higher grades. As such, it will be important to interpret subsequent data points in that context.[211]

Furthermore, when comparing attendance between baseline phase one and phase two, considerable variations emerge. For example, while the proportion of girls with a high-level of attendance remains similar for the grade 8 cohort, a considerable decrease was observed for the grade 4 and grade 6 cohorts.[212] However, it is not possible at this time to disentangle these trends since the data represent different times of the academic year and grade level.

Similar to the quantitative analysis, qualitative findings identified girls' home responsibilities as a major barrier to their ability to attend school consistently. Respondents across the focus groups and KIIs described an uneven division of household labor that results in girls completing the majority of household chores and participating in income-generating activities. Teachers and parents mentioned that it was common for girls to come to school late due to their morning chores, and as a result, girls often came to school tired. Teachers also reported an increase in girls' absences on market days; many girls stated they regularly participate in market days instead of attending school. Respondents alluded that girls who were married, orphaned, had children, or came from very poor households would likely face even greater challenges in attending school due to increased home responsibilities.

Menstruation was also mentioned by respondents as a hindrance to girls' ability to attend school consistently. Female upper primary students reported varying methods for addressing their monthly menstruation. Several female students stated they were able to effectively plan for their menstruation, so they did not need to miss school, but others stated they would consistently miss school for the duration of their menstruation. Parents also identified menstruation as a challenge but stated it had become less of a hindrance since the provision of sanitary napkins by Link. The day-to-day challenges that girls expressed in attending school—such as missing school due to household responsibilities, being unable to be fully attentive in the school setting due to exhaustion from chores, absenteeism due to income-generating activities, and missing school due to menstruation—should all be understood as continued barriers that require targeted interventions to address. Furthermore, the fact that these challenges would likely be exacerbated for girls who are married, orphaned, mothers, or from poor households should also be taken into consideration.

5.2 School Management and Governance

Intermediate outcome and indicator selection and measurement

This IO was chosen because the project's interventions assume that improved school management and governance is one of the prerequisites for better learning, transition, and sustainability outcomes for girls in the Wolaita Zone. This IO will be measured and reported for all woredas participating in the STAGES project. Capacity to support the STAGES interventions and outcomes will be examined at each of the four evaluation points scheduled during the life of the project. Woreda staff surveys and monitoring data will be used to generate descriptions of woredas' with high and low levels of capacity and support.

Through this IO, STAGES aims to reduce three barriers to girls' education: lack of community awareness and support, gaps in perceptions of girls' education, and gaps in support structures at the woreda-, community-, and school-level for girls to persist and learn.

Indicators for the IO are

- IO Indicator 2.a: Percentage of GAP targets or actions undertaken

[211] Attendance and retention rates should both be tracked, as feasible.
[212] High attendance in this context relates to either attending school 5 days in a row or having zero absences 5 days in a row.

- IO Indicator 2.b: Level of incorporation of SPAM into school or community practices as determined by case studies or meeting minutes

Data will be captured through school audits and woreda-official surveys at each evaluation point. In addition, case studies, meeting minutes, and KIIs will capture levels of incorporation and enthusiasm for SPAMs at the community and school level at each evaluation point.

As of baseline, questions regarding GAP targets and SPAM were not included for woreda staff surveys; these interventions had not started and, therefore, would not be appropriate for officials to reflect upon. Instead, school audit surveys included questions to school directors regarding the GEAC, which is the body that completes GAP actions and holds SPAMs.

Findings

Baseline data show that 100.00 percent of treatment schools (n=15) and 87.00 percent of comparison schools (n=13) have active GEACs. This suggests that all treatment schools are well positioned to set GAP targets and incorporate SPAMs into their practices.

Indicator 2a. Percentage of GAP targets or actions have been undertaken

The baseline level for this indicator is assumed to be 0.00 percent because GAP targets have not been set as part of the STAGES project. Furthermore, all treatment schools report having active GEACs, which enables them to engage in setting GAP targets. At this time, no revisions to the targets identified in the logframe are suggested.

Indicator 2b. Level of incorporation of SPAM into practices of school or community, as determined by case studies or meeting minutes

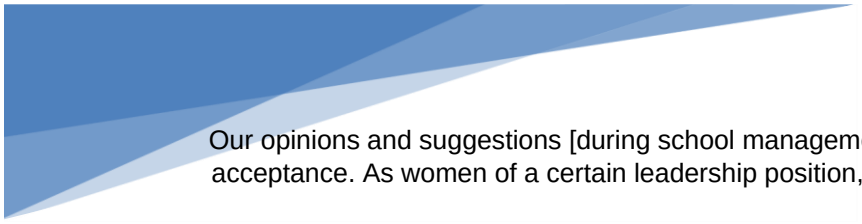
The baseline level for this indicator is assumed to be 0.00 percent since SPAMs have not yet been initiated as part of the STAGES project. This indicator will also be tracked using primarily qualitative data sources. No direct data is available at this time to examine a correlation between this IO and learning or transition outcomes.

Interpretation and reflections

Because all treatment schools report having active GEACs, but no GAP targets have been set, it was not possible to collect this data at baseline. The presence of GEACs in all schools is a promising start for monitoring GAP targets. As a next step, the STAGES project will engage with GEACs to set targets that can be monitored at the next evaluation point. Explicit questions around GEACs, GAPs, SIPs, and SPAMS will be added to subsequent qualitative research to ensure that relevant activities are thoroughly investigated. The relationship between the completion of these targets and student learning and transition outcomes will be explored at the midline.

While little was specifically said by qualitative participants about GEACs, GAPs, SIPs, or SPAMS, some important contextual background can be gleaned regarding school management. Namely, while the majority of female teachers in the focus groups reported being involved in school management and leadership activities, some questioned the extent to which their participation mattered. Several female teachers voiced concerns related to underlying gender norms and power dynamics in school management activities.

Figure 7: Quotes from Female Teachers



Our opinions and suggestions [during school management activities] usually fail to win acceptance. As women of a certain leadership position, no one is willing to listen to us and encourage us.
Respondent A

As a representative of female teachers, nobody pays attention to my presence in a meeting. A woman who excels men does not get recognition.
Respondent B

Some females in leadership positions are put under duress amidst people who are even not yet fully convinced in the equality of men and women. All priority is given to men.
Respondent C

Their comments highlight the need to address gender norms and power dynamics beyond the student population and throughout the whole school.

5.3 Quality of teaching

Intermediate outcome and indicator selection and measurement

This IO was chosen because the STAGES project theorizes that if the quality of teaching is improved, then girls will be more likely to attend school consistently, improve their performance, and increase their transition rates into secondary school.

The project aims to reduce three barriers through this IO: lack of teachers adequately prepared with subject-matter knowledge, inappropriate instructional methodologies, and inadequate classroom management techniques. Although not required, focus on these areas at the primary level were demonstrably important to improving student learning outcomes and will be carried over to the secondary level.

The indicators that were chosen for the IO include

- IO Indicator 3.a: Percentage of teachers with improved subject knowledge disaggregated by gender. *Secondary required, primary optional.*
- IO Indicator 3.b: Percentage of teachers with improved methodology disaggregated by gender. *Secondary required, primary optional.*
- IO Indicator 3.c: Percentage of teachers with improved classroom management disaggregated by gender. *Secondary required, primary optional.*

These indicators were selected because of the link observed during GEC1 at the primary level between teachers' support in these areas and girls' learning outcomes. Interventions under STAGES are aimed at improving mathematics and literacy teachers' content knowledge, classroom methodology, and classroom management skills.

In baseline phase one, data were only collected from primary teachers; in baseline phase two, data were collected from secondary teachers. Data reported below include teachers in grades 4, 6, and 8 in 15 treatment schools and 15 comparison schools. Please note teacher surveys were completed in both primary and secondary treatment schools, whereas classroom observations were conducted in primary

schools only. A total of 89 teachers from treatment schools and 91 teachers from comparison schools were surveyed at the primary level. At the secondary level, 16 teachers from treatment schools and no teachers from comparison schools were surveyed. Supplementary Table 20. Teachers Surveyed and Their Level and Subject Taughtshows the teacher sample by level, gender, and subject matter taught.

Supplementary Table 20. Teachers Surveyed and Their Level and Subject Taught

Subject	Treatment		Comparison	
	n	%	n	%
Primary, Male				
English	22	36.67%	22	37.29%
Math	11	18.33%	11	18.64%
Science	27	45.00%	26	44.07%
Primary, Female				
English	15	51.72%	18	56.25%
Math	4	13.79%	2	6.25%
Science	10	34.48%	12	37.50%
Secondary, Male				
English	7	53.85%	0	0.00%
Math	6	46.15%	0	0.00%
Science	0	0.00%	0	0.00%
Secondary, Female				
English	1	33.33%	0	0.00%
Math	2	66.67%	0	0.00%
Science	0	0.00%	0	0.00%

Findings

Indicator 3a: Percentage of teachers with improved subject knowledge disaggregated by gender—secondary required, primary optional

Primary or secondary teachers' improved subject-matter knowledge was not examined at baseline. Instead, teachers' subject-matter knowledge will be examined as part of the teacher training using pre- and post-tests administered to participating teachers. Due to the sensitivity of teacher examinations, assessing teachers' subject-matter knowledge will be difficult during evaluation points. For the baseline report, self-reported data from primary and secondary teachers on attendance at training within the past year is used as a proxy for the indicator. The data show that about half of primary school teachers—with comparable proportions for both male and female teachers—reported attending a training in their subject matter. At the secondary level, the data show that a lower proportion—one-third of secondary teachers—reported having attended training in their subject matter within the past year.

IO Indicator 3b: Percentage of teachers with improved methodology disaggregated by gender—secondary required, primary optional

For teachers at the primary and secondary school levels, the IO used self-reported attendance at a training in gender-responsive pedagogy in the past year. On average, two-thirds of primary teachers—60.00 percent of male teachers and 68.97 percent of female teachers—reported attended a gender-responsive pedagogy training. At the secondary level, only one male teacher—7.69 percent of male teachers—and no female teachers reported attending a gender-responsive pedagogy training in the last year.

IO Indicator 3c: Percentage of teachers with improved classroom management disaggregated by gender—secondary required, primary optional

Teachers self-report of their participation in classroom management trainings are reported for primary and secondary. On average, half of all primary school teachers surveyed reported having attended classroom management trainings, and one-in-three secondary school teachers reported the same.

Additionally, classroom observation data from primary level teachers on 10 items, which comprise the classroom-management index, was used for this IO. Higher scores on this index indicate higher levels of classroom management; the scale ranged from 0 to 3. On average, male and female teachers in primary schools had classroom management scores that were slightly higher than the midpoint of the scale. Secondary school teachers were not observed during phase two of the baseline. Moving forward, primary and secondary school teachers will be observed as part of ongoing monitoring by Link. Reporting at subsequent evaluation points will include those monitoring observation data and teachers' self-reported data on attending trainings.

Supplementary Table 21 includes targets and baseline results for the three indicators related to the quality of teaching for secondary school teachers; targets for the baseline primary teachers are optional and therefore not included.

Supplementary Table 21. Quality in Teaching IO 3.a, IO 3.b and IO 3.c—Baseline and Target Figures

Indicator	Level	Baseline (2018)		Midline 1 (2020)	Midline 2 (2022)	Endline (2024)
Treatment		***Actual***	***n***	***Targets***		
IO 3.a. Percentage of teachers with improved subject matter knowledge	Secondary	T: 37.50% M: 38.46% F: 33.33%	T: 6 M: 5 F: 1	40.00% teachers report receiving training; 50.00% secondary teachers with improved pre- to post-training test scores in literacy and numeracy [213]	60.00% teachers report receiving training; 55.00% secondary teachers with improved pre- to post-training test scores in literacy and numeracy	75.00% teachers report receiving training; 60.00% secondary teachers with improved pre- to post-training test scores in literacy and numeracy
	Primary	T: 51.69% M: 51.67% F: 51.72%	T: 46 M: 31 F: 15	n/a	n/a	n/a
IO 3.b. Percentage of teachers with improved methodology	Secondary	T: 6.25% M: 7.69% F: 0.00%	T:1 M: 1 F: 0	40.00% teachers report receiving training on gender-responsive pedagogy; 30.00% secondary teachers with average scores above the midpoint on a scale for gender-responsive pedagogy portion of monitoring classroom observation	60.00% teachers report receiving training on gender-responsive pedagogy; 40.00% secondary teachers with average scores above the midpoint on a scale for gender-responsive pedagogy portion	75.00% teachers report receiving training on gender-responsive pedagogy; 50.00% secondary teachers with average scores above the midpoint on a scale for gender-responsive pedagogy

[213] This was increased from the original logframe target as the baseline figure exceeds original target of 35%.

Indicator	Level	Baseline (2018)		Midline 1 (2020)	Midline 2 (2022)	Endline (2024)
Treatment		***Actual***	***n***	***Targets***		
					of monitoring classroom observation	portion of monitoring classroom observation
	Primary	T: 62.92% M: 60.00% F: 68.97%	T: 56 M: 36 F: 20	n/a	n/a	n/a
IO 3.c. Percentage of teachers with improved classroom management	Secondary	T: 12.50% M: 15.38% F: 0.00%	T: 2 M: 2 F: 0	40.00% received training in classroom management; 30.00% secondary teachers with average scores above mid-point on a scale for classroom management portion of classroom observation	60.00% received training in classroom management; 40.00% secondary teachers with average scores above mid-point on a scale for classroom management portion of classroom observation	75.00% received training in classroom management; 50.00% secondary teachers with average scores above mid-point on a scale for classroom management portion of classroom observation
	Primary-training received	T: 46.07% M: 43.33% F: 51.72%	T:41 M: 26 F: 15	n/a	n/a	n/a
	Primary-classroom observation	T: 1.74 M: 1.75 F: 1.72	T: 64 M: 49 F: 16	n/a	n/a	n/a

Note: Total n = 15 treatment primary schools and 8 treatment secondary schools; at baseline, a total of 90 primary teachers in treatment schools were surveyed, one each in grades 4, 6, and 8 in each school.[214] A total of 16 secondary school teachers in treatment schools.

At the primary level, there was not a statistically significant relationship between teachers' classroom-management scores and girls' literacy and numeracy performance. However, several inverse relationships were observed. First, in primary schools where the proportion of female teachers was low, the classroom-management score was high. Additionally, in primary schools where the average years of teaching was high or teachers' average age was high, the mean classroom-management score was low. Correlation coefficients ranged from -0.28 and -0.13 for these correlations indicating weak relationships. Finally, when primary school teachers had higher classroom-management scores, they also had more positive gender-perception scores—correlation coefficient of 0.11.

For girls who were in grade 8 during phase one, relationships with secondary school teachers in phase two were examined. There was a high correlation between a positive attitude towards girls' education and positive gender perceptions as well as teachers' perceptions of the support from schools and communities. However, there was no statistically significant relationship between girls' literacy or numeracy scores and teachers' attitudes towards girls' education, gender perceptions, or school and community support indices.

[214] For IO 3.c, a total of 65 treatment classroom observations were completed at baseline across grades 4, 6, and 8 and across math, science, English, and other subjects. No classroom observations were completed in secondary schools during the baseline.

Taken together, these data suggest that although the relationship between classroom management and girls' academic performance was not observed to be statistically significant at baseline in either primary or secondary schools, teachers' classroom management in primary matters. Higher levels of classroom management scores were associated with teachers also having more positive gender perceptions, higher self-perceptions of English competence, and were in schools where boys also had higher gender-perception scores.

Interpretation and reflections

The role of teachers in the classroom can matter. It is related to gender perceptions among primary school boys, teachers' motivation, and school and community support for the classroom. However, what occurs in the classroom can only affect those girls who actually arrive at school and are ready to learn. Therefore, the project's dual focuses on supporting girls to ensure girls enroll and stay in school is as important as focusing on improving the quality of teaching girls within the school environment.

The qualitative analysis provided additional insights for understanding teaching quality in treatment schools. The majority of female primary school teachers in the FGDs described feeling well prepared to teach the content in their classroom. Similarly, the majority of upper primary school students stated that their teachers provided adequate educational support and were able to answer questions in class. Moreover, numerous students reported that teachers were willing to provide learning supplies, extra tutorials, or make-up lessons. However, several female teachers also reported feeling ill-equipped to teach the assigned subjects as they had originally trained for other content areas; other teachers cited numerous environmental challenges, such as a lack of reference materials and restrooms. Furthermore, classroom challenges appeared to be exacerbated by large class numbers, shortages of books and teaching materials, and inadequate school facilities. Lastly, interviews with zone- and woreda-level officials, as well as school management, revealed that a lack of quality education was seen as a major challenge to girls' education.

Given these findings, quality of teaching appears to include areas to strengthen—teachers' commitment and dedication—as well as areas requiring more targeted capacity building and support—mastery of subject-matter knowledge and access to reference materials.

5.4 Community-based attitudes and behavior change

Intermediate outcome and indicator selection and measurement

This IO was chosen due to the STAGES project's focus is on improving support for girls' education in the community, household, and school environments. Specifically, the project aims to reduce barriers to girls' education, including lack of community awareness to problems facing girls, lack of community support, and gaps in perceptions between the importance of girls' and boys' education among all stakeholders. Because the focus is at the secondary level, the indicators of this IO examining attitudinal and behavior changes at the secondary level. The following indicators were chosen for the IO:

- IO Indicator 4.a: Teachers reporting positive changes in gender perceptions and gender-sensitive teaching. *Secondary required, primary optional.*
- IO Indicator 4.b: Teachers and school directors reporting positive attitudinal change towards girls' education and learning. *Secondary required, primary optional.*
- IO Indicator 4.c: Percentage of parents—disaggregated by gender—reporting greater support for secondary education, especially for girls.
- IO Indicator 4.d: Percentage of boys—disaggregated by level—reporting a positive perception of the value of girls' education.

These indicators allow the evaluators to track changes in attitudes and behaviors among key stakeholders, including teachers, school directors, parents, and boys. At the time of the phase one of the

baseline, data was only collected and reported on from primary schools. As such, IO 4.a and IO 4.d were addressed at the primary level only during the original submission of the baseline report. Upon completion of data collection and analysis during phase two, this section has been revised to include additional information from secondary schools for IO 4.a as well as parents for IO 4.c. Furthermore, while IO 4.b will serve primarily as a qualitative indicator moving forward, some additional quantitative data from primary and secondary school teachers, as well as guidance counselors, have been incorporated for additional background and context.

This IO also connects to IO 3—school management and governance—because the purpose of the SPAMs is to bring together all stakeholders in the community around the needs of girls to enhance community engagement and promote change in community attitudes. As such, progress on IO 3 indicators will also be tied to progress on IO 4 indicators.

Findings

IO Indicator 4.a: Teachers reporting positive changes in gender perceptions and gender-sensitive teaching—secondary required, primary optional

Results for IO 4.a are reported at baseline as the proportion of respondents with higher gender perceptions. Generally, two-thirds of primary school teachers in treatment schools had gender-perception scores above average for their group, compared to half of secondary school teachers.[215] Subsequent time points will use these same absolute scores to examine the proportion of teachers with improved gender perceptions.

Targets are not suggested at the primary level because the primary level is an optional reporting requirement for IO 4.a. Instead, only targets for secondary level teachers are provided. The proposed targets are in the form of percentages rather than raw n's, as was noted in the log frame.

Supplementary Table 22. Teachers Reporting Positive Changes in Gender Perceptions and Gender-sensitive Teaching IO 4.a—Baseline and Target Figures

Level	Baseline (2018)		Midline 1 (2020)	Midline 2 (2022)	Endline (2024)
Treatment	***Actual***	***n***	***Targets***		
Primary (grades 4, 6, 8)[216]	67.42% of teachers with gender perception scale score above 2.67 on a 3-point scale	60	-	-	-
Secondary (grade 9/10)[217]	50.00% of teachers with gender perception scale score above 2.67 on a 3-point scale	8	60.00% of teachers with gender perception scale score above 2.67 on a 3-point scale	70.00% of teachers with gender perception scale score above 2.67 on a 3-point scale	80.00% of teachers with gender perception scale score above 2.67 on a 3-point scale

Note: Total n=458 teachers from treatment schools and n=16 teachers for treatment secondary schools.

[215] Results are reported by grade, and for those teachers who taught more than one grade of interest—grade 4, 6, or 8—they are reported separately under multi-grade.
[216] Proportion of teachers whose gender-attitude toward girls' education index score is above the mean (2.79).
[217] Original targets were as follows for midline 1, midline 2 and endline: 65, 195, 375. We assume these were set based on the number of teachers planned for training under STAGES.

Supplementary Table 23. Teachers—Disaggregated by Gender—Reporting Positive Changes in Gender Perceptions and Gender-sensitive Teaching at Baseline by Subgroup IO 4.a

School Level	Primary Schools								Secondary School	
Grade Level[218]	Grade 4		Grade 6		Grade 8		Multi-grade		Grade 9/10	
Subgroup	***%***	***N***	***%***	***N***	***%***	***N***	***%***	***N***	***%***	***N***
Male teachers	78.95%	15	40.00%	2	77.78%	14	55.56%	10	46.15%	6
Female teachers	57.14%	4	66.67%	4	85.71%	6	55.56%	5	66.67%	2
All teachers, by grade	73.08%	19	54.55%	6	80.00%	20	55.56%	15	50.00%	8

Note: Grade 4 only n=26; grade 6 only n=11; grade 8 only n=25; multi-grade n=27; grade 9/10 n=16. The average score on the gender-perception scale for teachers was 2.67 on a three-point scale. Subsequent time points will use these same absolute scores to examine the proportion of teachers with improved gender perceptions.

Among grade 8 girls, there was no relationship between teachers' gender-perception scores and literacy and numeracy outcomes for girls. Although gender perceptions among teachers were not correlated with girls' literacy and numeracy performance, primary school teachers' gender perceptions were associated with girls' attendance. Specifically, when more primary school teachers had lower gender perceptions in the school, the average number of days primary school girls attended school was lower (correlation coefficient = -0.13).

IO Indicator 4.b: Teachers and school directors' reporting positive attitudinal change towards girls' education and learning—secondary required, primary optional

While IO 4.b is noted in the log frame as a qualitative indicator, baseline data were collected using teacher surveys since extensive qualitative data could not be collected from teachers. For future evaluation points, qualitative data will be collected. The survey results suggest that attitudes among teachers—as captured by the index—are already high and the index may not capture nuanced changes in teachers' attitudes toward girls' education.[219] Attitude questions were not asked of primary or secondary school directors at baseline; instead, surveys were administered with guidance counselors at the secondary level. Guidance counselors were deemed an important stakeholder in establishing positive gender perceptions in the school as they will also be receiving considerable engagement and support from STAGES over the course of the project.

The findings indicate similar results for primary and secondary school teachers, with 71.91 percent of primary school teachers in treatment schools reporting gender-attitudes scores above average for their group, as compared to 75.00 percent of secondary school teachers.[220] In addition, 71.43 percent of secondary school guidance counselors reported above-average gender attitude scores. Please note no targets are established at this time as the indicator will be understood as qualitative and examined through case studies with teachers and school directors at the secondary level moving forward.

[218] Teachers were grouped by the grade they reported they taught; for primary school teachers who reported teaching more than one grade, they were classified as "multi-grade."
[219] Proportion of teachers whose attitude index score is above the mean (2.79).
[220] Results are reported by grade, and for those teachers who taught more than one grade of interest—grade 4, 6, or 8—they are reported separately under multi-grade.

Supplementary Table 24. Teachers and Guidance Counselors with Positive Attitudes towards Girls' Education IO 4.b—Baseline Figures

Level	Baseline (2018)		Midline 1 (2020)	Midline 2 (2022)	Endline (2024)
Treatment	***Actual***	***n***	***Targets***		
Primary teachers (grades 4, 6, 8)[221]	71.91%	64	qualitative		
Secondary teachers (grade 9/10)	75.00%	12	qualitative		
Guidance Counselor (secondary)	71.43%	5	qualitative		

Note: Total n=458 teachers from treatment primary schools; n=16 teachers for treatment secondary schools; n=7 guidance counselors from treatment secondary schools. The average score on the teachers' attitudes scale was 2.79.

Supplementary Table 25. Percentage of Teachers and Guidance Counselors at Baseline—Disaggregated by Gender—Reporting Positive Attitudes Towards Girls' Education IO 4.b

Level	Primary teachers *(G4, 6, 8)*		Secondary teachers *(G9/10)*		Guidance counselors *(secondary)*	
Treatment	**%**	**n**	**%**	**n**	**%**	**n**
Male	71.67%	43	69.23%	9	50.00%	1
Female	72.41%	21	100.00%	3	40.00%	2
Total	71.91%	64	75.00%	12	42.86%	3

Note: Total n=458 teachers from treatment primary schools and n=16 teachers for treatment secondary schools and n=7 guidance counselors.

IO Indicator 4.c Percentage of parents—disaggregated by gender—reporting greater support for secondary education, especially for girls

Results for IO 4.c are reported at baseline as the percentage of parents reporting a high level of support for secondary education, especially for girls. This included a targeted subsample of parents and caregivers of grade 4 and 6 girls identified within treatment schools for the baseline. The data was collected via a parent/caregiver survey during phase two of the baseline in December 2018.

In order to understand parents' level of support for girls' secondary education, a 10-point scale was utilized that examined a number of items related to the following: quality of schooling and teaching the girl receives, difficulties in costs/affordability for the girl to go to school, commitment to investing in girl's education even when funds were limited, safety for girls to travel to and from school, need for support for girls' secondary education in their community, and the level of schooling parent/caregiver's want their girls to receive. The average score on the support scale for parents was 7.83, with similar means for grade 4/5 parents and grade 6/7 parents—7.62 and 7.83, respectively. While there was no difference between grade 4/5 parents by respondent sex, for grade 6/7 parents, female respondents had a lower score, 7.62, on the scale than did the male respondent, 7.64. The high mean scores were mainly driven by parents' desire for girls to pursue higher levels of education. More specifically, 94.34 percent of grade 4/5 parents and 93.67 percent of grade 6/7 parents reported they want their girl to go to college or university. In

[221] Proportion of teachers whose gender-perception index score is above the mean (2.67).

addition, parents cited costs as the most common reasons why girls may not attend secondary schools in their community.

Nearly six in ten primary school parents within the targeted subgroup sample reported support scores above average for their group.[222] Targets from the original logframe are provided, and no revisions are suggested at this time. Subsequent time points will use these same absolute scores to examine the proportion of parents with greater support for secondary education, especially for girls.

Supplementary Table 26. Percentage of Parents/Caregivers Reporting High Levels of Support for Secondary Education, Especially for Girls IO 4.c—Baseline and Target Figures

Level	Baseline (2018)		Midline 1 (2020)	Midline 2 (2022)	Endline (2024)
Parents	***Actual***	***n***	***Targets***		
Grade 4/5 parents	56.41%	44	-	-	-
Grade 6/7 parents	60.38%	64	-	-	-
All parents	58.70% of grade 4 and grade 6 parents with secondary support scale scores above 7.83 on a 10-pt scale	108	63.00% of parents with secondary support scale scores above 7.83 on a 10-pt scale	70.00% of parents with secondary support scale scores above 7.83 on a 10-pt scale	75.00% of parents with secondary support scale scores above 7.83 on a 10-pt scale

Note: total n=184 parent/caregivers from treatment primary schools.106 with daughters in cohort Grade 4/5 (m=25, f=81) and 78 with daughters in cohort grade 6/7 (m=24, f=54). The average score on the support scale for parents was 7.83 on a ten-point scale. Subsequent time points will use these same absolute scores to examine the proportion of parents with greater support for secondary education, especially girls.

Supplementary Table 27. Percentage of Parents—Disaggregated by Gender—Reporting Greater Support for Secondary Education, Especially for Girls IO 4.c

Level	Grade 4/5 parents		Grade 6/7 parents		Total	
Treatment	**%**	**n**	**%**	**n**	**%**	**n**
Male	66.67%	16	60.00%	15	63.27%	31
Female	51.85%	28	60.49%	49	57.04%	77
Total	56.41%	44	56.41%	44	58.70%	108

Note: total n=184 parent/caregivers from treatment primary schools.106 with daughters in grade 4/5 (m=25, f=81) and 78 with daughters in grade 6/7 (m=24, f=54).

IO Indicator 4.d: Percentage of Boys—Disaggregated by Level—Reporting a Positive Perception of the Value of Girls' Education

Results for IO 4.d are reported at baseline as the proportion of respondents with higher gender perceptions. Almost two-thirds of boys in primary treatment schools had gender-perception scores above average for their group.[223] For boys, the average was 2.51. Subsequent time points will use these same absolute scores to examine the proportion of boys with improved gender perceptions. Although baseline data were not collected for boys at the secondary level, targets are revised for all boys based on the

[222] Proportion of parents whose support for secondary school index score is above the mean (7.83). Note that results are reported by daughters' grade-level.
[223] Results are reported by grade, and for those teachers who taught more than one grade of interest—grade 4, 6, or 8—they are reported separately under multi-grade.

baseline data for boys in primary school. Baseline figures for grade 10 boys will be established at midline 1; separate targets for secondary boys may also be established then.

Supplementary Table 28. Percentage of Boys' Reporting Positive Perception of the Value of Girls' Education IO 4.d—Baseline and Target Figures

Level	Baseline (2018)		Midline 1 (2020)	Midline 2 (2022)	Endline (2024)
Treatment	***Actual***	***n***	***Targets***		
Primary (grades 4, 6, 8)[224, 225]	65.50%	150	-	-	-
Secondary (grade 10)	n/a	n/a	-	-	-
Total, boys	65.50% of boys with gender perception scale scores above 2.51 on a 3-point scale	150	70.00% of boys with gender perception scale scores above 2.51 on a 3-point scale	75.00% of boys with gender perception scale scores above 2.51 on a 3-point scale	80.00% of boys with gender perception scale scores above 2.51 on a 3-point scale

Note: Total n=229 boys in treatment primary schools. For boys, the average score on the gender-perception scale was 2.51 on a three-point scale. Subsequent time points will use these same absolute scores to examine the proportion of boys with improved gender perceptions.

Supplementary Table 29. Percentage of Boys' Reporting Positive Perception of the Value of Girls' Education at Baseline by Subgroup IO 4.d

School Level	Primary						Secondary	
Grade Level	Grade 4		Grade 6		Grade 8		Grade 10	
Subgroup	***%***	***N***	***%***	***N***	***%***	***N***	***%***	***N***
Boys who are overage for grade	61.11%	11	81.82%	9	80.00%	12	TBD	TBD
Boys who live without both parents	58.97%	46	72.97%	54	64.00%	48	TBD	TBD
Boys with high chore burden	59.21%	45	74.65%	53	64.38%	47	TBD	TBD
All boys, by grade	58.97%	46	73.68%	56	64.00%	48	TBD	TBD

Note: Total n=229 boys in treatment primary schools. Grade 4 treatment n=78; grade 6 treatment n=76; grade 8 treatment n=75. For boys, the average score on the gender-perception scale was 2.51 on a three-point scale. Subsequent time points will use these same absolute scores to examine the proportion of boys with improved gender perceptions.

Conversely, when the average gender-perception score was higher among boys in primary schools and grades, so too were the average literacy and numeracy scores among girls. Gender perceptions among woreda officials were also examined. As with boys' gender-perception scores, when woreda officials had higher gender-perception scores, the average literacy and numeracy scores among primary school girls in that woreda were also higher; correlation coefficients were between 0.16 and 0.22.

[224] Proportion of primary school boys whose gender-perception index score is above the mean (2.67).
[225] Given the cohort approach, no data will be collected or reported at the primary school level at endline.

Interpretation and reflections

The IO indicators adequately capture a range of gender perceptions among key stakeholders in the community. Initial trends between girls' performance and gender perceptions suggest that the opinions of stakeholders in the community can influence girls' performance in the classroom.

The baseline data for this IO underscore the importance of gender perceptions among teachers, parents, community members, and boys on the attendance and performance of girls. Positive gender perceptions are associated with higher performance in class. Data on gender perceptions in subsequent time points will have to be interpreted within the context of changes in girls' attendance rates. Improvements in attendance and changes in teachers' gender perceptions may not occur at the same pace over subsequent evaluation points. More specifically, in upper primary and secondary grades, as the proportion of girls attending school increases—per IO 1—then the relationship between teachers' gender perceptions and girls' performance may remain weak as the baseline data indicate that where primary school teacher's gender perceptions were already low, girls' attendance was low.

The data also indicates that primary school teachers in treatment schools have more positive gender-perceptions and self-reported gender-sensitive teaching practices. Given STAGES considerable engagement and support within these primary schools over the course of GEC1—including gender-sensitive pedagogy training—this finding is unsurprising. It also indicates there is considerable room for growth and improvement in this area for teachers at the secondary level. Interestingly, when examining positive attitudinal change towards girls' education and learning, the findings between primary and secondary school teachers were more similar. In addition, when analyzing and comparing the teacher data, it is important to take into consideration the substantial variation in sample size across the two school levels, namely the inclusion of 458 primary teachers as compared to only 16 secondary school teachers.

Furthermore, the data from parents indicate that overall, primary school parents have educational aspirations for their primary school girls that go beyond the scope of the project—an overwhelming majority of parents cited their desire for their girls to reach college or university levels. However, as illustrated in this baseline report, despite girls' and their parents' high aspirations, there are numerous barriers—such as costs—that prevent girls from transitioning to secondary level education, let alone college or university.

Qualitative data also provides important context and depth to the quantitative data around other community members' attitudes towards girls' education. Some key informants reported that although community support for girls' education had improved in recent years, more needed to be done to raise awareness and encourage support of girls' education. Several respondents stated that community support for girls' education was strong at the primary-school level but tended to decrease starting in early secondary school and cease at the upper-secondary level. One key informant stated that in order to address challenges to girls' education, "attempts should be done to raise awareness aimed at bringing behavioral change in the community to help value girls' education." Several respondents hinted at the lack of responsibility families and communities felt towards girls' education; one respondent even quoted "girls for their husbands, boys for their parents" as a common saying in their community. Additionally, respondents noted that community support for girls' education was not only important to shifting social norms but also to ensuring long-term sustainability and financial support.

5.5 Girls' Self-esteem

Intermediate outcome and indicator selection and measurement

This IO was chosen because the project's interventions assume that supporting the self-esteem of marginalized girls is one of the prerequisites for better learning, transition, and sustainability outcomes. Girls' sense of well-being was also selected as an indicator because well-being leads to success in both school and work. While a proxy measure was utilized in phase one for overall girls' life-skills, the decision

was made to examine girls' self-esteem in phase two. Going forward, self-esteem will be measured as the key intermediate outcome. As a result, the following indicators are reported for the IO:

- IO Indicator 5.a: Percentage of girls reporting improved well-being.
- IO Indicator 5.b: Percentage of girls reporting improved self-esteem.[226]

These indicators were chosen because they provide an important measure of girls' "readiness to learn." Girls' self-reported feelings of well-being were asked on the girls student survey during phase one of baseline data collection and are reported here. Once the data were collected, survey items were examined to identify underlying well-being construct. During phase two of the baseline, girls' self-reported feelings on self-esteem[227] were asked as part of the girls transition survey.[228] Once the data were collected, survey items were examined to identify underlying self-esteem construct.[229]

Findings

Baseline levels for IO 5.a are reported as the proportion of girls who reported high levels of well-being. For subsequent evaluation points, the proportion of girls with improved well-being will be reported. Specifically, the proportion of girls who go from answering "yes" on zero questions to one question, from one question to two questions, and from two questions to three questions. Baseline levels for IO 5.b are reported as the proportion of girls who reported high levels of self-esteem. Specifically, the proportion of girls with an average self-esteem score that is two or higher on a four-point scale.[230] At baseline, on average, almost three-quarters of girls reported a high sense of well-being while slightly less than half of the girls surveyed reported a high sense of self-esteem. The targets presented are for the proportions of girls with an improved sense of well-being and an improved sense of self-esteem. The targets proposed for well-being and self-esteem are lower than those in the logframe.[231]

[226] The original indicator read 'due to SEL' and has been removed in this revision. Self-esteem will be driven by teachers teaching in a more gender-responsive, child-friendly manner for example, or through parent and community valuing of girls being in school and learning, or through girls not having the practical challenges and stigma associated with menstruation. In other words, it will not only be due to SEL.

[227] The original baseline report submission reported on IO Indicator 5.b using a girls' life-skills construct rather than a self-esteem construct. However, it was later determined that self-esteem was the most appropriate construct for reporting on IO Indicator 5.b for the STAGES project, and therefore an additional measure of self-esteem was required, and an expanded set of self-esteem items were included and examined as part of the phase two baseline data collection. These self-esteem items and associated construct will be utilized for comparison at midline 1. Analysis and findings from the originally reported life-skills construct have been removed from the body of the report but can be found in Annex 21 as an additional reference. The rationale for this decision includes the following: in phase one data collection, items for self-esteem were asked as part of a set of questions on life-skills. When self-esteem survey items were examined to identify underlying construct—specifically, whether the items expected to measure self-esteem suggested a single factor and whether items measuring life-skills suggested a different factor– the results indicated that life-skills encompassed self-esteem items. In other words, a single factor emerged out of the data analysis instead of two distinct factors. As such, the results in the original baseline submission were reported for life-skills—the more encompassing factor—instead of self-esteem and a recommendation was made to update the indicator to life-skills or to add additional measure to self-esteem. As detailed previously, the final decision was to return to self-esteem by adding an additional measure.

[228] Self-efficacy items were also included in the girls' transition survey.

[229] The self-efficacy construct was also examined but it was determined less relevant, and therefore will not be included as part of IO5.b indicator reporting. Details on the associated self-efficacy construct and findings can be found in Annex 21.

[230] The self-esteem construct represents the mean score taken across 10 items, with higher scores indicating higher self-esteem. Five negatively worded items were reverse coded (Self-esteem 2, 5, 6, 8, 9).

[231] Targets noted in the logframe (approved prior to baseline) for both well-being and self-esteem were 20 percent, 40 percent and 60 percent of students with improved self-esteem (or well-being) at midline 1, midline 2 and endline, respectively.

IO Indicator 5a: Percentage of girls reporting improved well-being

Supplementary Table 30. Percentage of Girls Reporting High Sense of Well-being IO 5.a—Baseline and Target Figures[232]

Cohort	Baseline (2018)		Midline 1 (2020)	Midline 2 (2022)	Endline (2024)
Treatment	***Actual***	***n***	***Targets***		
Grade 4	71.24% of girls report a high sense of well-being	213	10.00% who report improved well-being; 75.00% of girls report a high sense of well-being	20.00% who report improved well-being; 80.00% of girls report a high sense of well-being	30.00% who report improved well-being; 85.00% of girls report a high sense of well-being
Grade 6	71.28% of girls report a high sense of well-being	211	10.00% who report improved well-being; 75.00% of girls report a high sense of well-being	30.00% who report improved well-being; 85.00% of girls report a high sense of well-being	-
Grade 8	79.93% of girls report a high sense of well-being	231	10.00% who report improved well-being; 85.00% of girls report a high sense of well-being	-	-
Average (G4, G6, G8)	74.10%	655	**-**	**-**	**-**

Note: Treatment n=884 girls. Grade 4 treatment n=299; grade 6 treatment n=296; grade 8 treatment n=289. Baseline levels are reported as the proportion of girls who reported high levels of well-being. For subsequent evaluation points, the proportion of girls with improved well-being will be reported. Specifically, the proportion of girls who go from answering "yes" on zero questions to one question, from one question to two questions, and from two questions to three questions.

Supplementary Table 31. Percentage of Sampled Girls Reporting High Sense of Well-being at Baseline by Subgroup IO 5.a

Grade Level	Grade 4		Grade 6		Grade 8	
Subgroup	***%***	***N***	***%***	***N***	***%***	***N***
Girls with at least one disability	N/A	N/A	not reported[233]			
Girls who are overage for grade	17.19%	11	15.00%	6	17.24	5
Girls in households unable to meet basic needs	61.11%	77	60.71%	68	70.24%	59
Girls who do not attend five days of school per week	54.44%	49	43.75%	42	46.81%	44
Girls with low levels of support from household	48.98%	24	57.53%	42	56.00%	28

[232] High sense of well-being means the girl responded "yes" to all three questions on well-being.
[233] This data should be understood as missing. Results for disability subgroup—barrier 1—were not possible to disaggregate at this time as the Child Functioning set was administered in phase two data collection and has a different denominator than the relevant survey items. This information will be provided starting at midline using the Washington Group Child Functioning questions.

Grade Level	Grade 4		Grade 6		Grade 8	
Subgroup	***%***	***N***	***%***	***N***	***%***	***N***
Girls who report teachers treat boys and girls different in a classroom	73.53%	150	71.50%	148	77.84%	151
Girls who report teachers are often absent from class	66.20%	94	69.12%	94	75.45%	83
Girls who report high corporal punishment exercised by the teacher	69.42%	84	56.52%	91	60.67%	91

Note: Grade 4 treatment n=299; grade 6 treatment n=296; grade 8 treatment n=289. Baseline levels are reported as the proportion of girls who reported high levels of well-being. For subsequent evaluation points, the proportion of girls with improved well-being will be reported. Specifically, the proportion of girls who go from answering yes on zero questions to one question, from one question to two questions, and from two questions to three questions.

IO Indicator 5b: Percentage of girls reporting improved Self-esteem

Supplementary Table 32. Percentage of Girls Reporting High Sense of Self-esteem IO 5.b—Baseline and Target Figures [234]

Cohort	Baseline (2018)		Midline 1 (2020)	Midline 2 (2022)	Endline (2024)
Treatment	***Actual***	***n***	***Targets***		
Grade 4 cohort	47.47% of girls report high levels of self-esteem	117	10.00% who report improved self-esteem; 50.00% of girls report high levels of self-esteem	20.00% who report improved self-esteem; 55.00% of girls report high levels of self-esteem	30.00% who report improved self-esteem; 60.00% of girls report high levels of self-esteem
Grade 6 cohort	42.40% of girls report high levels of self-esteem	92	10.00% who report improved self-esteem; 50.00% of girls report high levels of self-esteem	30.00% who report improved self-esteem; 60.00% of girls report high levels of self-esteem	-
Grade 8 cohort	51.69 of girls report high levels of self-esteem	92	10.00% who report improved self-esteem; 50.00% of girls report high levels of self-esteem	-	-
Average (G4/5, G6/7, G8/9)	**46.88% of** girls report high levels of self-esteem	**301**	**-**	**-**	**-**

[234] High sense of self-esteem means the girls' average self-esteem score is two or higher on a four-point scale. The construct includes 10 items.

Cohort	Baseline (2018)		Midline 1 (2020)	Midline 2 (2022)	Endline (2024)
Treatment	***Actual***	***n***	***Targets***		

Note: Treatment n=642 girls. Grade 4/5 treatment n=247; grade 6/7 treatment n=217; grade 8/9 treatment n=178. Baseline levels are reported as the proportion of girls who reported high levels of self-esteem. For subsequent evaluation points, the proportion of girls with improved self-esteem will be reported. Specifically, the proportion of girls with an average self-esteem score that is two or higher on a four-point scale.

Supplementary Table 33. Percentage of Sampled Girls Reporting High Sense of Self-esteem by Subgroup—Baseline Scores IO 5.b

Grade Level[235]	Grade 4/5		Grade 6/7		Grade 8/9	
Subgroup	***%***	***N***	***%***	***N***	***%***	***N***
Girls with at least one disability	N/A	N/A	not reported[236]			
Girls who are overage for grade	42.31%	22	38.71%	12	47.83%	11
Girls in households unable to meet basic needs	39.62%	42	37.18%	29	41.07%	23
Girls who do not report high levels of well-being[237]	44.71%	38	42.31%	33	38.46%	20
Girls who do not attend five days of school per week	50.66%	77	42.57	63	54.03	67
Girls with low levels of support from household	51.22%	21	35.29%	18	42.86%	12
Girls who report teachers treat boys and girls different in the classroom	47.59%	79	43.06%	62	47.41%	55
Girls who report teachers are often absent from class	45.22%	52	40.86%	38	40.32%	25
Girls who report high corporal punishment exercised by the teacher	47.54%	58	42.20%	46	55.32%	52

Note: Treatment n=642 girls. Grade 4/5 treatment n=247; grade 6/7 treatment n=217; grade 8/9 treatment n=178. Baseline levels are reported as the proportion of girls who reported high levels of self-esteem. For subsequent evaluation points, the proportion of girls with improved self-esteem will be reported. Specifically, the proportion of girls with an average self-esteem score that is two or higher on a four-point scale.

Additionally, higher well-being scores among girls were associated with higher scores for girls on three other scales of interest: self-esteem, decision making, and gender perceptions. This held true for all schools—treatment and comparison.[238] The relationship between these three scales with literacy and numeracy scores were weak but statistically significant in all grades—correlation coefficients between 0.10 and 0.17. Additionally, schools where girls reported higher levels of well-being were also schools where boys reported higher levels of well-being. Finally, when girls reported higher levels of well-being, schools also had higher scores on three school- or community-focused scales of interest: curriculum design and implementation, girls' support mechanisms, and community support.[239]

[235] Grade levels are grouped as the data was collected from cohort of girls, which include both girls who successfully transition to the next grade since phase one data collection as well as repeaters.
[236] This data should be understood as missing. Results for disability subgroup—barrier 1—were not possible to disaggregate at this time as the Child Functioning set was administered in phase two data collection and has a different denominator than the relevant survey items. This information will be provided starting at midline using the Washington Group Child Functioning questions.
[237] Girl did not report 'yes' to all three well-being items asked.
[238] Self-esteem scale was only asked of treatment girls in phase two: 10 items were used to compute the self-esteem scale mean score. Decision-making scale: in grade 4, six items were used and grades 6 and 8, seven items. Gender-perceptions scale: In grades 4, 6, and 8, seven items were used. See Annex 15 for individual items results and scale score means.
[239] Gender-sensitive curriculum design and implementation scale includes nine items; girls support mechanisms includes five items; and community-support scale includes eight items. See Annex 15 for individual items results and scale score means.

Interpretation and reflections

Indicators 5.a and 5.b were measured and reported at baseline. These IO indicators is to measure whether girls are ready to learn when they come to school, and the correlations to student learning outcomes suggest a weak but present relationship.

The evidence suggests that there is a need to focus not only on the supports provided to girls in school but also on their well-being and self-esteem. These indicators demonstrate that students with high levels of well-being and self-esteem are more likely to succeed academically than their peers who have lower levels.

Project Checks on Intermediate Outcomes

6. Conclusion and Recommendations

6.1 Conclusions

The baseline evaluation of the STAGES project was conducted by STS, an external evaluator, and the results of that evaluation are presented in the body of the report and summarized here.

Scope of the Baseline Report

The baseline data were collected in two phases. During phase one, surveys, KIIs, and FGDs were conducted with primary school girls, boys, teachers, parents, zone- and woreda-levels officials. Approximately one month later, learning assessments were conducted with girls in grades 4, 6, and 8. The original baseline report - drawing on phase one data - was submitted during the summer of 2018.

As noted in the MEL framework, phase one of the baseline evaluation of the STAGES project was not designed to capture household-level data, data on the transition outcome, or data on girls in secondary schools. Instead, the following phase two baseline activities were completed in the fall of 2018, after the baseline report was updated and resubmitted in May 2019:

1. Benchmarking learning outcomes for girls in grade 10.
2. Transition tracking for sampled girls in grades 4, 6, and 8.
3. Attendance monitoring checks for all girls in treatment schools in grades 4, 6, and 8.
4. Household surveys with a purposive subsample of girls selected from the baseline learning sample, including data from their head of households, caregivers, and siblings; the sample of girls will be selected from within each profile group.[240]

Therefore, this revised report includes the following additional results from phase two:

1. Methodology for data collection in secondary schools (Section 2).
2. Characteristics of the sample for girls in grade 10, heads of household and caregivers, or siblings (Section 3).
3. Transition outcomes, subgroup analysis of transition outcome, or target setting for transition outcome (Section 4).
4. IO 1.a and IO 1.b for secondary girls, IO 2.a and IO 2.b; IO 3.a–c and IO 4.a–d for secondary teachers, parents, and boys; and IO 5.a–b for secondary girls (Section 5).[241]

Profile of Project's Beneficiaries and Relationship with Outcomes

The gender-equity measures for the four target woredas—Damot Pulasa, Damot Sore, Damot Woide, and Kindo Koisha—demonstrate the need for interventions such as those provided under the STAGES project.[242] The consistently low gender-equity measures—including the gender-parity index, the proportion of female students, and the proportion of female teachers—broadly convey the barriers faced by women and girls in these woredas. The manifestation of these barriers in communities and classrooms served by STAGES is partly captured in the survey results presented in Section 3. Supplementary Table

[240] Profile groups were be generated using baseline survey data and specifically, items regarding girls' intentions to transition, their demographics and barriers.

[241] IO1: Percentage improvement in attendance rates, percentage of students with improved perceptions of access; IO2: percentage of teachers with improved subject knowledge, percentage of teachers with improved methodology, percentage of teachers with improved classroom management. IO3: percentage gap targets/actions taken and level of incorporation of SPAM into practices of school/community. IO4: teachers reporting positive changes in gender perceptions and gender-sensitive teaching, percentage boys reporting positive perception of value of girls' education; IO5: girls with improved levels of wellbeing and life skills.

[242] The Federal Democratic Republic of Ethiopia, EMIS, and ICT Directorate and MOE, *Education Statistics Annual Abstract, 2008 E.C. (2015/16)* (Addis Ababa: 2017).

27 highlights some profiles of subgroups of girls in grades 4, 6, and 8 as well as their relationships with learning and transition outcomes.

Supplementary Table 34. Highlighted Results for Barriers Faced by Subgroups of Girls and Relationship with Learning Outcomes

Subgroup description	*Relationship of barrier with learning outcomes*
Girls who are overage for their grade are performing lower than other girls. One-in-five girls are overage for their grade. The proportion of overage girls was slightly higher in grade 4 than in grades 6 and 8.	Girls who were overage for their grade performed lower on literacy and numeracy aggregate scores at baseline than girls who were on age or underage.
About 5–6 percent of girls surveyed from the grades 6 and 8 cohorts reported having at least one disability.[243] For the population of girls served by STAGES, this means that approximately 294 of grade 6 girls and 184 of grade 8 girls in intervention schools have disabilities.[244]	Since child functioning questions were asked in phase two, the sample of girls who responded does not constitute the complete baseline sample. Qualitative data from phase one suggest that the lack of resources, accommodation, or specialized training to support girls with disabilities was seen as a barrier.
The number of girls who are orphans is estimated in the baseline data based on school-level surveys. These data show that five primary schools (of 15) reported having 10 or fewer girls who are orphans and the remaining 10 schools reporting having between 11–20 girls who are orphans. Also, as a proxy, girls were asked who lives in their household; 50 girls (5.66%) and two boys (0.87%) reported they lived without both parents. The variability in these data shows that the estimate of orphans in treatment woredas is unclear.	Relationships with learning outcomes were not explored due to the qualitative nature of the data and the limited sample size. Qualitative data suggest that girls who are orphans are likely to face greater challenges to attend school and receive interventions or support. Additionally, respondents noted that **girls who are orphans** were less likely to attend secondary school and received limited, targeted interventions or support.
The proportion of girls faced with high chore burdens was high. Having high chore burden and being from poor households was prominently noted as a barrier to girls' education in surveys, interviews, and focus groups.	**Grade 6 and 8 girls with a high chore burden performed lower in literacy and numeracy than those with a low chore burden**. The relationship between chore burden and a girl's ability to successfully engage in schoolwork at home was explored under GEC1, where the majority of girls in the treatment woredas worked an average of five hours per day on household chores.[245] **Chore burden has a compounding effect on girls' performance when they face other barriers.** Girls in

[244] Computed based on total numbers of girls enrolled in each grade in the previous school year (these numbers are shown in the sampling frame). A total of 6,670 girls in grade 4, 4,915 girls in grade 6 and 4,104 girls in grade 8 were reported in the sampling frame. The following proportions are estimated based on the baseline data: 5.99 percent of grade 6 girls and 4.49 percent of grade 8 girls.

[245] At the conclusion of the GEC1 project, the evaluation reported that, over the course of GEC1 parents support of girls' education increased, noting "reduced chores, [parents] provided school materials including lamp oil to aid evening study". In qualitative findings, there was evidence that high chore burden is increasingly recognized by families and accepted as an issue facing girls' education. While there was evidence in GEC1 that some accommodations have been made, the evaluation noted that work remains to be done. Qualitative data in the current baseline suggested that reducing chore burden for girls does not remove the need to complete the chores; if the girls don't do it or have less time to do it—who will? Mothers may be more resistant to reducing chores or overburdening the girls—as the assumption is that chores would be picked up by mothers (as opposed to say girls sharing the chores more equally with their brothers or fathers).

Subgroup description	*Relationship of barrier with learning outcomes*
	grades 6 and 8 who faced multiple barriers—such as high chore burden, living without both parents, and being overage for their grade—performed statistically significantly lower than girls who did not face these barriers. **Girls with low levels of household support, which includes chore burden, performed lower in literacy and numeracy than girls with higher levels of household support.** The relationship between household support and other outcomes of interest underscore the importance of support at home as well as in school. Girls who had lower levels of household support also had lower levels of school support, reported higher absenteeism among teachers and were in schools with lower average scores on the gender-perception scale among girls and boys. When girls reported they had missed at least one day of school in the last week, they were asked the reason for the absence; the most frequently cited reason was household chores.
There is a disparity between the language girls use with friends, family, and teachers and the MOI. In all grades, girls primarily spoke Wolayttatto—their mother language—with friends, family members, and teachers. However, the MOI is English beginning in grade 5.	Girls report using Wolayttatto at home in all grades. As a result, all girls face the same language gap. Evaluators are unable to determine how this transition impacts girls' literacy and numeracy outcomes.
One-quarter of girls did not report high levels of well-being, and less than half of girls report high levels of self-esteem; both of these were associated with other outcomes of interest. Specifically, high well-being was associated with having higher decision-making, and gender-perception scores, on average.	Girls who reported low levels of well-being and low levels of self-esteem had lower literacy and numeracy scores than girls who reported high levels of well-being and higher levels of self-esteem.
Physical safety for girls is a challenge. Abduction and SRGBV were noted in interviews and focus group discussions, although the extent to which these affect girls needs to be further explored. Corporal punishment by teachers was reported by almost one-third of girls in treatment schools.	Individuals may be hesitant to discuss or disclose information regarding SRGBV due to stigma or concern of retaliation, or punishment. In addition, some forms of SRGBV, such as corporal punishment or harassment may, be considered commonplace and internalized as acceptable.[246] Moreover, analysis from the evaluation surveys indicates that one form of SRGBV—corporal punishment—persists within the school environments. Relationships of other forms of physical punishment with learning outcomes were not explored due to the qualitative nature of the data and the limited sample size.

[246] Jenny Parkes et al., *Addressing School-Related Gender-Based Violence in Côte d'Ivoire, Togo, Zambia and Ethiopia: A Cross-country Report* (New York: UCL Institute of Education, August 2017), http://discovery.ucl.ac.uk/10027909/7/Parkes_Addressing%20SRGBV%20Cross%20Country%20report_2017.pdf.

Baseline Literacy and Numeracy Levels of Beneficiary Girls

At the baseline, girls in the comparison woreda outperformed girls in the treatment woreda on assessments of grade 4 numeracy and grade 8 literacy. However, there was not a statistically significant difference in performance on other assessments, including grade 4 literacy, grade 6 literacy or numeracy, or grade 8 numeracy.

The proportion of girls in treatment schools classified as proficient are shown in Figure 87 and 8. Red cells indicate lower relative proportions of girls meeting proficiency in that subtask; green indicates higher proportions of students meeting proficiency.[247] In evaluations of Wolayttatto literacy, grades 4 and 6 girls in treatment schools struggled most with fluency and reading comprehension and, to a lesser extent, with familiar words and invented words. At least half mastered letter sound identification. In evaluations of English literacy, grade 4 and 6 girls in treatment schools struggled with all foundational skills. The lowest proportion of girls reached proficiency in reading comprehension and in familiar words in grade 4; in grade 6, the lowest proportion of girls reached proficiency in reading comprehension. In grade 8, the lowest proportion of girls reached proficiency in written SeGRA subtasks and the EGRA reading comprehension subtask. Although higher, only one-third of grade 8 girls reached proficiency in invented words and oral reading fluency.

Figure 8: Girls' Literacy Proficiency by Language and Grade

	Literacy subtasks							
							SeGRA	
	Letter sound identification	Familiar word	Invented word	Oral reading fluency	Reading comprehension	Reading passage	Fill in the blank	Revising sentences
Grade 4—Wolayttatto	41.52%	17.65%	17.65%	1.04%	4.84%			
Grade 6—Wolayttatto	40.43%	23.47%	17.33%	7.22%	13.36%			
Grade 4—English	15.57%	4.15%	21.45%	5.54%	0.35%			
Grade 6—English	22.74%	14.08%	29.24%	23.10%	2.53%			
Grade 8—English		28.41%	36.74%	32.58%	7.95%	19.39%	0.38%	0.76%

In numeracy, girls in grades 4 and 6 at treatment schools struggled with all foundational skills; only one-in-ten girls reached proficiency in number identification. In grade 8, one-in-ten girls reached proficiency in word problems and slightly more reached proficiency in geometry. Surprisingly, grade 8 had the fewest girls to reach proficiency in subtraction of all three grade levels tested.

Figure 9: Girls' Numeracy Proficiency by Grade

	Numeracy subtasks								
							SeGMA		
	Number Identification	Quantity Discrimination	Missing Number	Addition	Subtraction	Word Problems	Geometry	Fractions	Multiplication
Grade 4—Numeracy	10.03%	7.27%	0.69%	3.11%	0.69%				

[247] No absolute criteria are imposed on the color scheme selected. Instead, lower relative proportions are indicated in red while higher relative proportions are indicated in green. These color choices are intended to indicate relative proportions, not identify expected or absolute proportions.

Grade 6—Numeracy	11.91%	9.39%	3.25%	3.25%	1.08%	5.78%			
Grade 8—Numeracy				7.95%	0.38%	11.36%	13.41%	3.83%	13.41%

Comparability of treatment and comparison girls

There are several differences between the four treatment woredas and the one comparison woreda that are important to highlight, as the issue of comparability is central to the analyses presented:

1. Full saturation of the STAGES project to all schools in each of the four treatment woredas meant that no schools within the same administrative and geographic parameters as treatment schools could be selected as comparison schools.
2. The comparison woreda—Ofa—averaged fewer girls affected by barriers examined in Section 3 than did the treatment woredas.[248]

Baseline Transition Levels of Beneficiary Girls

In phase one, the majority of girls in all grades expressed intentions to go on to the next grade level. However, actual transition rates obtained during phase two show a high rate of attrition—one that exceeds the assumptions made in the sampling strategy. Successful transition was defined as the proportion of girls who move into the next grade at the beginning of the following academic cycle. One-quarter of grade 4 girls did not successfully transition into grade 5, one-third of grade 6 girls did not successfully transition into grade 7, and slightly more than one-third of grade 8 girls did not successfully transition into grade 9.

Figure 10: Transition Rates at Key Transition Grades by Cohort

Key Transition	Total Sample	Overall successful transition rate
Grade 4 into Grade 5	297	73.40%
Grade 6 into Grade 7	300	68.00%
Grade 8 into Grade 9	294	69.05%
Overall	891	66.44%

Baseline Sustainability Scores

The project's baseline sustainability scores at the community level and at the school level were 1.00; these scores reflect the limited nature of data available on the indicators included in the MEL framework. The baseline sustainability score at the system level is 3.00, reflecting the strong integration of the project in government systems as a platform for long-term engagement and adoption of STAGES activities. Based on the data available, the linkages with the government system and Stages' activities are most likely to support further sustainability of activities. The factors most likely to hinder the sustainability of the project's activities, however, are the barriers identified in Section 3.

The removal of some of these barriers—such as chore burdens, gender perceptions, classroom management, and household support—requires a shift in perceptions and behaviors among a critical

[248] For example, Ofa has a higher score on the gender-parity index than do the treatment woredas. However, results from the teacher surveys suggest that support for girls' education was comparable in treatment and comparison woredas. The girls and boys in Ofa reported facing lower rates of poverty than girls and boys in the treatment woredas—24.89 percent of boys and 23.99 percent of girls in comparisons versus 35.37 percent of boys and 39.71 percent of girls in treatment woredas. Ofa has fewer students who report low levels of household support—14.32 percent for comparison versus 22.06 percent for treatment. Girls in Ofa report slightly higher levels of perceived safety traveling to and from school—5.88 percent compared to 9.39 percent in treatment woredas.

mass of stakeholders. Such changes in human behavior take time and occur incrementally. The stated IO measures, therefore, may capture changes in these perceptions and behaviors, but the longer-term impact of these IOs on girls' performance may not be observed in the timeframe of data collection. Therefore, conclusions about sustainability should be made carefully, considering that changes in perceptions and behaviors from the school- to system-level are central to the project's activities.

Baseline Levels of Intermediate Outcome Indicators

The following table summarizes each IO and the data reported against it at baseline, as well as the quantitative and qualitative findings.

Supplementary Table 35. Intermediate Outcome Baseline Findings for Primary Girls in Treatment Schools

Indicator	*Baseline data reported for primary girls*	*Quantitative findings for primary girls*	*Qualitative findings for primary girls*
IO Indicator 1.a Percentage improvement in attendance rates	Proportion of girls not present on the day of surveys Self-reported attendance of sampled girls in the past week	8.80% of girls in treatment primary schools were not present and were replaced on the day of the survey Of all the girls in treatment primary schools who were surveyed—including replacements—approximately two-thirds reported attending all five days of school in the past week.	The barriers highlighted in the quantitative data include support in the home, well-being, and corporal punishment. This, however, affects subgroups of girls differently. Respondents across the focus groups and KIIs described an uneven division of household labor that results in girls completing the majority of household chores and participating in income-generating activities. Menstruation was also mentioned by respondents as a hindrance to girls' ability to attend school consistently. Moreover, qualitative findings indicated the barriers were especially challenging for girls within the poorest households, as well as exacerbated at the secondary school level as girls become older, face increased household chore burdens, expectations to providing income for their households (including through migration), as well as greater pressure to get married and begin child-bearing.
IO Indicator 1.b Percentage of students with improved perceptions of access	Perception-of-access scale (four items)	Overall, more than half of all primary girls in treatment schools surveyed reported having "excellent" perception of access as demonstrated by responding yes to all four items included in this index.	
IO Indicator 2.a Percentage of GAP targets or actions undertaken	As of baseline, questions regarding GAP targets and SPAM were not included for woreda staff surveys; these interventions had not started and, therefore, would not	Baseline data show that 100.00 percent of treatment primary schools have active GEACs. This suggests that all treatment primary schools are well positioned to set GAP targets and	Little was specifically discussed by qualitative participants about GEACs, GAPs, SIPs, or SPAMS; some important contextual background can be gleaned regarding school management. Namely, while the majority of female teachers in the focus groups reported being involved in school management and leadership activities, some questioned the extent to which their participation mattered. Several female teachers voiced concerns

Indicator	*Baseline data reported for primary girls*	*Quantitative findings for primary girls*	*Qualitative findings for primary girls*
	be appropriate for officials to reflect upon. Instead, school audit surveys included questions to school directors about the presence and role of the GEAC; this committee completes GAP actions and holds SPAMs.	incorporate SPAMs into their practices.	related to underlying gender norms and power dynamics in school management activities.
IO 2.b Level of incorporation of school performance appraisal meeting (SPAM) into school or community practices determined by case studies or meeting minutes		The baseline level for this indicator is assumed to be 0.00 percent because GAP targets have not been set as part of the STAGES project.	n/a
IO Indicator 3.a Percentage of teachers with improved subject knowledge disaggregated by gender—secondary required, primary optional	Teachers' improved subject-matter knowledge was not examined at baseline; instead, teachers' subject-matter knowledge will be examined as part of the teacher training. Self-reported data from primary school teachers on attendance at training is used as a proxy for the optional indicator.	The data show that about half of primary school teachers and one-third of secondary school teachers—with comparable proportions for both male and female teachers at the primary level but not at the secondary level—reported attending a training in their subject matter.	n/a
IO Indicator 3.b Percentage of teachers with improved methodology disaggregated by gender—secondary required, primary optional	As with the previous IO, improved methodologies among teachers at the secondary level cannot be reported at baseline. For teachers at the primary school level, the IO data reported at baseline used self-reported attendance at a training in gender-responsive pedagogy.	On average, two-thirds of primary teachers—60.00 percent of male teachers and 68.97 percent of female teachers—reported attending the training. At the secondary level, less than one in ten teachers reported attending gender-responsive pedagogy training—which included only one of the female teachers surveyed.	n/a
IO Indicator 3.c Percentage of teachers with	Self-reported data from primary level	At the primary level, average classroom-management index	

Indicator	*Baseline data reported for primary girls*	*Quantitative findings for primary girls*	*Qualitative findings for primary girls*
improved classroom management disaggregated by gender—secondary required, primary optional	teachers on 10 items, which comprise the classroom-management index, was used for this IO. Higher scores on this index indicate higher levels of classroom management.	score was slightly lower than the mid-point (2.00, on a four-point scale) for both male and female primary teachers. The proportion of teachers reporting that they received training in classroom management was one in ten teachers at the secondary level and half of all teachers at the primary level.	
IO Indicator 4.a Teachers reporting positive changes in gender perceptions and gender-sensitive teaching—secondary required, primary optional	Proportion of teachers holding high gender perceptions on the gender-perception scale.	Almost two-thirds of primary teachers and one-half of secondary teachers had high gender-perceptions. Among grade 8 girls, there was no relationship between teachers' gender-perception scores and literacy and numeracy outcomes for girls. Although gender perceptions among teachers were not correlated with girls' literacy and numeracy performance, teachers' gender perceptions were associated with girls' attendance. Specifically, when more teachers had lower gender perceptions in the school, the average number of days girls attended school was lower	Some key informants reported that although community support for girls' education had improved in recent years, more needed to be done to raise awareness and encourage support of girls' education. Several respondents stated that community support for girls' education was strong at the primary-school level but tended to decrease starting in early secondary school and cease at the upper-secondary level.
IO Indicator 4.b Teachers and school directors' reporting positive attitudinal change towards girls' education and learning—secondary	Proportion of primary and secondary teachers and secondary guidance counselors with positive	Almost three-quarters of teachers and counselors surveyed reported high attitudes towards girls' education.	n/a at baseline; however, future evaluation points will report using qualitative data on this indicator.

Indicator	*Baseline data reported for primary girls*	*Quantitative findings for primary girls*	*Qualitative findings for primary girls*
required, primary optional	attitudes towards girls education		
IO Indicator 4.c Percentage of parents—disaggregated by gender—reporting greater support for secondary education, especially for girls	Proportion of parents reporting high levels of support for secondary education (data reported on the subsample of parents of girls in grades 4 and 6 at baseline)	Two-thirds of a subsample of parents of girls in grades 4 and 6 at baseline had scores above 7.83 on a 10-pt scale (mean).	n/a
IO Indicator 4.d Percentage of boys—disaggregated by level—reporting positive perception of the value of girls' education	Proportion of teachers holding high gender perceptions on the gender-perception scale.	Almost two-thirds of boys in treatment primary schools had high gender perceptions. When the average gender-perception score was higher among boys in a school and grade, so too were the average literacy and numeracy scores among girls.	
IO Indicator 5.a Percentage of girls reporting improved well-being	Girls' scores on the well-being index (3 items).	Almost three-quarters of girls report a high sense of well-being.	
IO Indicator 5.b Percentage of girls reporting improved self-esteem	Proportion of girls with high levels of self-esteem as measured by a self-esteem scale.	Almost one-half of girls reported high levels of self-esteem.	

Project's Approach to Addressing Gender Inequalities

The STAGES project fulfills the requirements of "gender sensitive" on the GEC GESI Continuum as it meets the GEC GESI minimum standards and includes both GESI accommodating and transformative practices and activities. The STAGES project takes a holistic approach to gender equality in girls' education—addressing both the immediate and practical needs that girls face in their education and addressing girls' more long-term strategic needs—with the aim of transforming girls' status in the home, school, and community. While the main beneficiaries are primary and secondary school girls, STAGES employs a multi-pronged approach to promoting gender equality by engaging and supporting a wide range of stakeholders and beneficiaries in project activities and capacity-building.[249] Furthermore, the project's theory of change highlights STAGES' commitment to promoting gender equality in a number of ways, such as the inclusion of positive community attitude change as an IO, gender-aware communities

[249] This includes female and male students, mothers and fathers, female and male teachers, school directors, woreda officials—including gender officers—as well as male and female community members.

demanding high quality education, and improved leadership for girls' learning at school, woreda, zone, and regional levels as outputs.

The project promotes gender equality through its interventions in numerous ways. For example, gender inequality in educational access and performance has been connected to the low status of girls' education, the burden of domestic chores, and early marriage in the Wolaita Zone.[250] In order to address these challenges, the project will work with parents and community members through mothers' groups and fathers' groups, SPAMS, school improvement committees, and PTSAs to reflect upon, challenge, and subvert these negative practices. Their aim is to transform the roles of girls in their community—especially in regards to education. In addition, gender inequality in the classroom has been noted as an issue within the school setting. In order to transform this inequality in the long term for all children, STAGES is providing intensive training and mentoring for all teachers on gender-sensitive pedagogy. For additional details on the project's approach to gender equality and gender transformative activities, please reference the STAGES GESI report submitted to GEC-T in July 2017.[251]

Moreover, the inclusion of boys—as well as other male and female family and community members—in evaluation surveys and focus groups will be critical throughout the evaluation points to help gauge how successfully gender transformation is being integrated, accepted, and embedded within the broader school, home, and communities as well as ensuring any potential backlash against the girl-centered project design and implementation can be monitored, flagged, and addressed as needed. In addition, while the project has taken good initiative in already completing a risk assessment of girls' traveling to and from secondary schools or staying in lodging outside their familial home to attend secondary school, this remains a commonly known risk for SRGBV, especially for the most marginalized girls.[252] Therefore continued close monitoring of the risks and employment of mitigation strategies is recommended.

6.2 Recommendations

This section provides recommendations and reflections on revisions to instrumentation and indicators, beneficiary numbers, and project implementation and activities.

Recommendations for Revisions to Instruments

1. Include questions in the next evaluation point on girls' orphan status, marital status, and pregnancy status in lieu of collecting this information from parents of all surveyed girls.
2. Identify a means to surveying parents of girls who have dropped out as well as the girls themselves.

Recommendations for design and implementation

While the external evaluator has not found evidence that the overall design of the STAGES interventions requires major changes to improve relevance, there are some areas for consideration that may benefit from additional emphasis to achieve the learning, transition, or sustainability outcomes for girls in the four target woredas:

- ***Focus on subject-matter teachers' English language competency in grades 5 and above—both as a subject and as an MOI***

  Teachers and students can equally benefit from further support in English in grade 5 and onwards. Activities focused on teachers' English language competency should also engage with teachers' methodology of teaching English to students and building upon existing knowledge in

[250] Casey McHugh and Ashley Doria, *STAGES GESI Analysis* (Pacifica: School-to-School International, 2017).
[251] Ibid.
[252] Jenny Parkes et al., *Addressing School-Related Gender-Based Violence in Côte d'Ivoire, Togo, Zambia and Ethiopia: A Cross-country Report* (New York: UCL Institute of Education, August 2017), http://discovery.ucl.ac.uk/10027909/7/Parkes_Addressing%20SRGBV%20Cross%20Country%20report_2017.pdf, 12.

Wolayttatto. Additional analyses conducted by the external evaluator reveal a relationship between girls' dropout and academic performance in literacy and numeracy in the grade prior to dropout. The role of teachers' level of confidence and use of English may moderate this finding.

- ***Addressing the complex but critical role of IOs and their relationship with literacy and numeracy performance***

  This complex relationship came out throughout the analysis of IOs and will be an important issue to address both in the measurement of indicators as well as interventions. Specific focus on subgroups where the barriers are most present—and most likely to inhibit girls' literacy and numeracy improvements—should be a focus first for project interventions. In the interim between the first draft of the baseline report and this current version, alignment between outputs, intermediate outcomes, and primary outcomes is clearer from both a programmatic perspective as well as an evaluation perspective. For example, the focus of the project on literacy and numeracy, instead of lesser levels of support to more subject-matters, creates a tighter linkage between the outputs that are monitored, the intermediate outcomes of interest and the primary outcome of interest.

- ***Addressing early marriage***

  Qualitative data showed that as girls age and move through upper-primary and into secondary school, the pressure to marry and begin childbearing increases—and may be prioritized over a girl's education. Given the widespread practice of early marriages, along with its negative impact on girls' education, the evaluators recommend best practices and lessons learned within the Ethiopian context be examined and explicitly integrated into STAGES interventions.[253] Moreover, it is essential that these activities are not only targeted at the girls, but increase sensitization for mothers, fathers, boys, and other school and community members as well. In regards to addressing early marriage, a policy brief produced by Young Lives highlighted how "winning hearts and minds by involving girls, their parents, boyfriends and prospective husbands, community and religious leaders, as well as schools and youth and women's groups is likely to be more effective than strict legal enforcement and punishment for offenders."[254] Additional best practices in regards to addressing early marriage in Ethiopia can be drawn from research and reports from UNICEF, Population Council, and the International Center for Research on Women.[255, 256, 257]

- ***Addressing migration***

  The qualitative findings indicated that migration could serve as a major barrier for girls continuing in their education, especially in transitioning to secondary education. Therefore, in order to strengthen transition outcomes, girls' clubs, work with PTSAs, SICs, mothers' and father's groups, and life-skills and community campaign/role modeling activities should address risks and opportunities around domestic and international migration—including trafficking. Both the Girl Effect and Population Council have recently come out with studies on migration and adolescent

[253] Depending on the context, Link may also consider examining abduction in parallel or in conjunction with early marriage as well.
[254] Young Lives, "Child Marriage and Female Circumcision (FGM/C): Evidence from Ethiopia," *Young Lives Policy Brief 21* (December 2014), https://www.younglives.org.uk/sites/www.younglives.org.uk/files/YL-PolicyBrief-21_Child%20Marriage%20and%20FGM%20in%20Ethiopia.pdf.
[255] Nicola Jones et al., *Surprising trends in child marriage in Ethiopia* (s.l.: Overseas Development Institute (ODI) and UNICEF, March 2016), https://www.unicef.org/ethiopia/Briefing.pdf.
[256] Eunice M. Karei and Annabel S. Erulkar, *Building Programs to Address Child Marriage the Berhane Hewan Experience in Ethiopia* (Addis Ababa: Population Council, 2010), http://www.popcouncil.org/uploads/pdfs/2010PGY_BerhaneHewanReport.pdf.
[257] Anju Malhotra et al., *Solutions to End Child Marriage What the Evidence Shows* (s.l.: International Center for Research on Women (ICRW), 2011), https://www.icrw.org/wp-content/uploads/2016/10/Solutions-to-End-Child-Marriage.pdf.

girls that may be of relevance for the project.[258, 259] For example, the Girl Effect study on *Time to Look at Girls: Adolescent Girls Migration in Ethiopia*, recommends tackling the drivers for girls' migration—especially job opportunities—promoting safer migration of girls and raising awareness about gender, sexuality, and sexual violence.[260]

- ***Addressing SRGBV***

  Approaches to understanding and combatting SRGBV are important in supporting a positive school environment where girls can thrive in their studies and education. The evaluators recommend Link continue to implement strategies to prevent, address, and understand SRGBV—including corporal punishment, bullying, sexual harassment, child abuse, and intimate partner violence.[261] In addition, a number of key lessons learned can be drawn from a recent UNICEF cross-country report on SRGBV. For example, evidence suggests it is the children who are most marginalized, such as children with disabilities or from the poorest households, who are often the most vulnerable to SRGBV.[262] Another example relates to some key lessons learned on supporting the prevention of SRGBV—such as corporal punishment, something noted as prevalent in the baseline study. A study in Ethiopia highlighted that

  > while teachers may be aware of and accept the principle of the ban [on corporal punishment], lack of knowledge or support in implementing alternative discipline approaches, particularly in large classes, hindered teachers' capabilities to implement the ban… the challenge for all countries is how to create the condition for child-friendly, nonviolent pedagogies to be used in schools with alternative forms of punishment and classroom management.[263]

  Given this, it will be crucial that Link incorporate nonviolent pedagogies into their teacher training and equip teachers with the skills and knowledge of alternative discipline methods and classroom management skills in order to prevent SRGBV in the school setting. In addition, the report highlights how in order to respond to SRGBV incidents, "support for schools is needed to strengthen these systems with clear guidance for all numbers of school communities on responsibilities and actions to take following SRGBV," something that the Link design already incorporates and should be continued.[264]

  Furthermore, the report highlights the Ethiopian MOE's education management information system has introduced a Violence Reporting Template "to be completed each term in each school, detailing number of cases of different forms of violence, perpetrators, and outcome of cases." This may be an areas Link to potentially lend their support in terms of implementation to

[258] Dr. Marina de Regt, *Time to Look at Girls: Adolescent Girls' Migration in Ethiopia* (s.l.: Swiss Network for International Studies (SNIS) and Girl Effect Ethiopia, May 2016), https://research.vu.nl/ws/portalfiles/portal/16412107.
[259] Annabel Erulkar, PhD, Girmay Medhin, PhD and Lemi Negeri, *The Journey of Out-of-School Girls in Ethiopia: Examining Migration, Livelihoods, and HIV* (Addis Ababa: Population Council, August 2017).
[260] Dr. Marina de Regt, *Time to Look at Girls: Adolescent Girls' Migration in Ethiopia* (s.l.: Swiss Network for International Studies (SNIS) and Girl Effect Ethiopia, May 2016), https://research.vu.nl/ws/portalfiles/portal/16412107.
[261] For example, the United States Agency for International Development (USAID) funded Literacy Achievement and Retention Activity project has developed the "Journeys" approach which promotes guided reflections and dialogue to create positive violence-free school environments. RTI International's blog notes Journeys as an "innovative approach to eliminating violence in schools, which was informed by research on school climate and social and emotional learning; known mediators of SRGBV, such as cultural norms related to gender and power relations; adult experiential learning models; Ugandan educationist views and expertise; SRGBV prevention and intervention case studies; and the Massachusetts Institute of Technology's Presencing Institute U-Model(link is external)." https://shared.rti.org/content/journeys-through-uganda-usaiduganda-literacy-achievement-and-retention-activity%E2%80%99s-lara.
[262] Jenny Parkes et al., *Addressing School-Related Gender-Based Violence in Côte d'Ivoire, Togo, Zambia and Ethiopia: A Cross-country Report* (New York: UCL Institute of Education, August 2017), http://discovery.ucl.ac.uk/10027909/7/Parkes_Addressing%20SRGBV%20Cross%20Country%20report_2017.pdf.
[263] Jenny Parkes et al., *Addressing School-Related Gender-Based Violence in Côte d'Ivoire, Togo, Zambia and Ethiopia: A Cross-country Report* (New York: UCL Institute of Education, August 2017), http://discovery.ucl.ac.uk/10027909/7/Parkes_Addressing%20SRGBV%20Cross%20Country%20report_2017.pdf, 13.
[264] Ibid., 5.

MOE at the local level, as well as providing supplemental prevention monitoring support in their target woredas "such as whether schools have in place Comprehensive Sexuality Education lessons, SRGBV Code of conduct, girls/gender clubs, teachers trained in gender response pedagogy and separate functioning toilets."[265] For additional recommendations, the evaluators recommend Link refer to the full report for information regarding how to support SRGBV work linked to laws and policy, work in schools, work in communities, and use of data and evidence.[266]

The evaluators recommend Link take a similar approach to align their approach to MOE's policy for students with special needs and inclusivity as it did to aligning activities to gender policies. For example, policies for review and consideration may include the MOE's Special Needs/Inclusive Education strategy,[267] and implementation guidance,[268] the Master Plan for Special Needs Education/Inclusion 2016–25,[269] and the Guideline for Curriculum Differentiation and Individual Educational Program.[270]

Reflection on Beneficiary Numbers Calculations

At this time, the evaluator does not have any specific comment or recommendations on the way direct and indirect beneficiary numbers have been calculated, as based on the available evidence, the proposed beneficiary numbers look reliable. However, it has been determined the proxy measures used to estimate the prevalence of certain subgroups within the sample and the broader population of STAGES schools–namely, girls who were orphans or mothers—were not sufficient. Therefore, more direct measures for calculating the prevalence of beneficiary numbers for these targeted subgroups will be collected and monitored at future evaluation points.

Scalability and Sustainability

At this time, the baseline findings on sustainability are limited due to insufficient data. However, as this is only the baseline stage, the evaluators find that the STAGES project is on track toward achieving sustainability outcomes over the next seven years. Multiple promising practices have emerged that the evaluators recommend Link continue to pursue and build upon—including STAGES' strong alignment with overall MOE policy as well as Wolaita Zone education priorities. In addition, Link's should continue to build on their work from GEC1 to work effectively within the instituted school and community structures to improve norms, attitudes, and practices related to girls' education and transition in the Wolaita Zone.

Recent communication with Link indicates that lessons learned from the previous GEC1 project are also being recognized as best practices and scaled up by the MOE. These include the STAGES intervention support on mothers' groups being incorporated into the MOE's Revised National Ethiopian Girls' Education Strategy and demonstrate the need to continue documenting and sharing project practices to enable scaling successful activities where appropriate.

---

[265] Ibid., 32.
[266] Jenny Parkes et al., *Addressing School-Related Gender-Based Violence in Côte d'Ivoire, Togo, Zambia and Ethiopia: A Cross-country Report* (New York: UCL Institute of Education, August 2017), http://discovery.ucl.ac.uk/10027909/7/Parkes_Addressing%20SRGBV%20Cross%20Country%20report_2017.pdf.
[267] The Federal Democratic Republic of Ethiopia, MOE, *Special Needs/Inclusive Education Strategy* (s.l.: July 2012), http://www.moe.gov.et/documents/20182/42694/Special+Needs-+Inclusive+Education+Strategy_English.pdf/083ac0ab-af9a-4f2a-ba00-d99c4260b727?version=1.0.
[268] The Federal Democratic Republic of Ethiopia, MOE, *Special Needs/Inclusive Education Strategy Implementation Guideline* (s.l.: July 2012), http://www.moe.gov.et/documents/20182/42694/Special+Needs+-+Inclusive+Education+Strategy+Implementation+Guideline.pdf/6ce58030-a391-448a-8f90-a91f37e21092?version=1.0.
[269] The Federal Democratic Republic of Ethiopia, MOE, *A Master Plan for Special Needs Education/Inclusive Education in Ethiopia 2016-2025* (s.l.: October 2016), http://www.moe.gov.et/documents/20182/42694/A+Master+Plan+for+Special+Needs+Education+-+Inclusive+Education+in+Ethiopia+2016-2025.pdf/d97b378a-6c68-4d4f-ac20-b0f648696939?version=1.0.
[270] The Federal Democratic Republic of Ethiopia, MOE, *Guideline for Curriculum Differentiation and Individual Education Programme* (s.l.: 2012), http://www.moe.gov.et/documents/20182/42694/Guideline+for+Curriculum+Differentiation+and+Individual+Educational+Programme.pdf/74ccf20d-a211-4b96-a75e-810fe3e0e5c5?version=1.0.

Importantly, the project's deep engagement with local education officials in the baseline data collection should be continued.[271] It not only demonstrates the local education office's commitment to cost sharing—as evidenced in their cost sharing of the Zonal Girls Education Conference along with data collection—but also lends towards a participatory evaluation approach. Such an approach enables sustainability as local education officials gain first-hand exposure to the realities impacting girls' learning and transition outcomes, as well as individual, school, and community attitudes towards girls' education. Finally, scalability and sustainability are critically contingent on shifts in gender norms, attitudes, and perception shifts among stakeholders at the school to system level throughout the woredas. As an early indicator of broader scalability, there is interest among stakeholders to scale the STAGES interventions regionally and nationally. For example, the Regional Girls Education Conference held in 2018 created a platform for sharing with stakeholders in other regions of Ethiopia. The presentation by the project staff and external evaluator to the MOE and DFID generated interest to learn more about the SPAM and School Performance Review processes currently being used under STAGES. The magnitude of the change required to ultimately see changes in girls' learning outcomes and leverage the focus and intensity of STAGES activities to affect these complex issues should not be underestimated—and the value of early interest among regional and national stakeholders should be leveraged.

Project contribution: Response to conclusions and recommendations

The recommendations above have been provided by the external evaluator. The project response to the external evaluator's recommendations– in light of the conclusions of the Baseline Evaluation Report—can be found in Annex 13 along with the project response to the evaluator's comments on Link's gender approach used and how well gender is integrated through the project.

[271] This includes collaborating with zone-level official on adapting the learning assessment, working with woreda-level experts and supervisors as data collectors for the learning assessment and evaluation surveys, as well as female teachers and woreda gender officers support on the qualitative data collection.

Girls' Education Challenge

UKaid from the British people

GEC-T STAGES Baseline Report Annexes

July 2019

Project Number: 6473

Project Name: Supporting Transition of Adolescent Girls through Enhancing Systems (STAGES)

Implementing Agency: Link Community Development (Link)

Author: School-to-School International (under consultancy agreement with Link) and Link

Version Number: V3.

Date: July 3, 2019

Contents

Annexes

Annex 1: Logframe

Please click the icon below for the latest version of the project logframe in Excel along with targets, achieved outputs and outcomes, updated in May 2019.

Logframe_STAGES_L
CD_Ethiopia_052319.

Annex 2: Outcomes Spreadsheet

Please click the icon below for the latest version of the Outcomes Spreadsheet as an Excel document, updated in May 2019.

GEC-T Outcomes Spreadsheet_LCDSTA

Annex 3: Key Findings on Output Indicators[1]

This annex should be completed by the project.

Table 1: Output indicators

Logframe Output Indicator	Means of verification or sources	Collection frequency
Number and Indicator wording	*List all sources used.* *NOTE: Blue font indicates new means of verification as of May 2019*	*E.g., monthly, quarterly, annually. NB: For indicators without data collection to date, please indicate when data collection will take place.*
Output 1: Improved leadership for girls' learning at school, woreda, zone, and regional level		
Output 1.1: # of school & community members, with increased capacity in SRGBV	Pre and post training evaluation tool Monitoring tool criteria to indicate SRGBV Capacity Action plans to prevent or address SRGBV with community contribution (GAPs)	Annual
Output 1.2: # of Cluster Supervisors and Woreda Experts with increased capacity to monitor teaching quality and school improvement	Monitoring tool to monitor increased capacity	Quarterly
Output 1.3: # of HTs & Deputy Head Teachers with increased capacity in school instructional leadership	Pre & post training evaluation tool Monitoring tool to monitor improved school instructional leadership.	Quarterly
Output 1.4 # of active primary school Girls' Education Advisory Committees (GEACs)	GEAC meeting minutes, attendance, reviewed annual plans Monitoring tool with criteria for 'active' Training report	Quarterly
Output 1.5 # of Primary and Secondary Schools with established mechanisms to report and respond to abuse cases.	Monitoring tool with criteria for 'established' Training attendance records Numbers of CP Committees established Number of reporting boxes installed in schools Records of cases identified and reported	Annual
Output 2: Improved Quality of Learning		
Output 2.1: % of teachers practicing Gender and Inclusion Responsive Pedagogy (GRP)	Lesson observation tool	Quarterly
Output 2.2: % increase in average scores of girls (and boys) in literacy and numeracy. Percentage increase in average scores of girls (and boys) in English language and numeracy	Learner core subject test results in grades 4, 7, and 9	Annual

[1] This annex was revised in May 2019.

Logframe Output Indicator	Means of verification or sources	Collection frequency
Output 2.3: % of teachers demonstrating language competency and language teaching competency (English and Wolaytatto)	Pre and post training teacher tests for competency, Lesson observation tool for language teaching competence (gender/inclusion/safeguarding responsive methodology and classroom management)	Quarterly
Output 2.4: % of teachers demonstrating competency in teaching numeracy	Pre and post training teacher tests Numeracy lesson observation tool	Quarterly
Output 3: Better access to secondary schools in extreme and remote areas		
Output 3.1: # of new inclusive secondary schools in project target areas	Constructed schools; Photographs	Per term (until construction compete)
Output 3.2: # of inclusive separate girls' toilets blocks in secondary schools in project area (12 current and four new schools)	Upgraded separate girls' toilets photographs	Per Term (until construction complete)
Output 4: Girls 'ready to learn'		
Output 4.1: # of girls have access to sanitary and hygiene service (sanitary pad, towels and soap).	Signed receipt or distribution records of items	Annual for sanitary pads Soap provided quarterly
Output 4.2: # of extremely vulnerable girls receiving a bursary enrolling & staying in school (primary and secondary)	Disbursement of bursaries Attendance of recipients in school through finger print readers and school records	Per Term
Output 4.3: # of girls regularly attending literacy and numeracy tutorials	Literacy and numeracy tutorial Attendance Sheets Records of tutorial hours	Per Term
Output 5. Mobilized, gender-aware communities demanding high-quality education		
Output 5.1: # of community members, disaggregated by gender, participating in School Performance Appraisal Meetings	SPAM attendance, disaggregated	Annual
Output 5.2: # of Gender and Safeguarding Action Plans produced and fully operationalized	Monitoring tool with criteria for 'operationalised' Gender and Safeguarding Action Plan Monitoring of implementation reports	Annual
Output 5.3: # of role model and community awareness-raising campaigns	Attendance of stakeholders at campaign activities Campaign material developed Video clips or photographic evidence	Annual
Output 5.4: # of community school structures (KETB/PTSAs and SIC) actively supporting girls (attendance, learning and transition).	Monitoring tool to capture roles and actions of KETB/PTSA and SIC in actively supporting girls	Annual

Table 2: Baseline status of output indicators

Logframe Output Indicator	Baseline status or Baseline values Relevance of the indicator for the project ToC	Baseline status or Baseline values
Number and Indicator wording	*What is the contribution of this indicator for the project ToC, IOs, and Outcomes? What does the baseline value or status mean for your activities? Is the indicator measuring the right things? Should a revision be considered? Provide short narrative.*	*What is the baseline value or status of this indicator? Provide short narrative.*
Output 1: Improved leadership for girls' learning at school, woreda, zone, and regional level		
Output 1.1: # of school & woreda staff (including community members) with increased capacity in SRGBV	The relevance of the indicator for the project ToC—it is hoped that sensitized school community concerning SRGBV would create a conducive and safe learning environment for girls' education, in which violence is not acceptable, including corporal punishment.	The training for school staff, in which woreda staff were involved, was conducted in March 2019 but the monitoring tool for measuring increased capacity had not yet been finalized. Values will be established in year 3. Total = 426 of 810 trained in year 2 on SRGBV - 352 school staff & Community members - 74 woreda staff
Output 1.2 # of Cluster Supervisors and Woreda Experts with increased capacity to monitor teaching quality and school improvement	Teacher training alone will not provide robust enough support to teachers to bring about real and lasting change in teaching practice and content. Ongoing and good quality monitoring, mentoring and coaching support related to training received will ensure that new areas of practice are fully embedded.	This is a new activity and Cluster Supervisors and Woreda Experts have not yet received specific training on their roles in monitoring, mentoring and coaching teachers, or in monitoring school improvement. However, Cluster Supervisors and Woreda Experts are involved in all STAGES interventions as LCDE works in partnership with government and aligns interventions with government policy and aims. Values will be established in year 3
Output 1.3: # of HTs & DHTs with increased capacity in school instructional leadership	This indicator is new. Improved leadership at school level is critical to improvements in quality of teaching and learning, as well as to all other outcome and core indicators. The leaders of the school have much influence on whether the school as a whole is gender, inclusion and safeguarding responsive.	The first school leadership training took place in quarter 6 of year 2. Both School Directors and their Deputies participated. However, the monitoring tool for measuring increased capacity had not yet been finalized. Values will be established in year 3 Total = 295 HTs and DHTs were trained in year 2 on instructional leadership - 140 head teachers/school directors - 155 deputy head teachers
Output 1.4 # of active Primary school Girls' Education Advisory Committees (GEACs)	Active GEACs will support improvement of girls' enrolment, attendance, transition and attainment in their education as well as impact on creating a conducive and safe learning environment for girls' education. They are an existing structure at school level in Ethiopia and therefore the activity adds to the sustainability of this support.	This activity started in quarter 7 of STAGES (Oct-Dec 2018) and continued in quarter 8 (Jan-Mar 2019). All GEACs are activated with plans to support girls in place. However, the monitoring tool for measuring 'active' had not yet been finalized. Values will be established in year 3. Total= 122 of 136 GEACs trained in Yr 2 - 127 primary schools - 9 secondary schools
Output 1.5 # of Primary and Secondary Schools with established	This indicator is critical to the strength of STAGES support to safeguarding and child protection. With reporting mechanisms in place and in use, and an orientation on these	136 schools trained on mechanisms to report abuse and letter link boxes installed in year 2. This is all primary and secondary schools minus the 4 under construction.

Logframe Output Indicator	Baseline status or Baseline values Relevance of the indicator for the project ToC	Baseline status or Baseline values
mechanisms to report and respond to abuse cases.	for children, school staff and local government, girls and boys will have a channel through which to report incidents of bullying, abuse, and any form of harassment. Through strengthening case management, girls who are affected by such incidents will be supported to recover.	Year 2 Value = 136 schools with reporting mechanisms in Yr 2 (2018/2019) - 127 primary schools - 9 secondary schools Letter-link boxes for reporting abuse at school level were installed in all schools in year 2 of STAGES, and key stakeholders including children oriented on how to use them. The success of this mechanisms will be judged through internal monitoring.
Output 2: Improved Quality of Learning		
Output 2.1: % of teachers practicing Gender and Inclusion Responsive Pedagogy (GRP)	The wording of this indicator has changed to add 'inclusion'. STAGES aims to reach the most marginalized girls, some of whom face additional or multiple barriers to attending, participating and learning in school. The wording is strengthened to ensure that children for example who are overage for their grade, are young mothers, have a disability or other factors, are not excluded in training approaches, content and methodology, or in project monitoring.	This training was provided for more than 2,000 teachers in year 2 of STAGES. Content added included the use of differentiated teaching methods to help teachers understand and identify children who may be struggling in the classroom, and devise ways to respond and meet the needs of all children in the class. However, the monitoring tools for measuring teacher practices had not yet been finalized. Values will be established in year 3 Total= 2,811 teachers trained on GRP in Yr 2 - 2,329 primary teachers - 482 secondary teachers
Output 2.2: % increase in average scores of girls (and boys) in literacy and numeracy.	This indicator has been reworded to capture an adaptation made to the programme. Initially STAGES was supporting the separate performance monitoring testing of children in grade 4, 7 and 9 in all core subjects. However, STAGES will be judged on improvements only in literacy and numeracy. The activity has been revised to access the tests that government are already implementing instead of testing separately, and in literacy and numeracy only. Investing resources into analysis and presentation (in SPAMs) of literacy and numeracy performance only will help children to improve in the subjects on which they will be judged. Improved literacy and numeracy will still help children to access other core subjects.	**Year 1 (Baseline) Values** • Average English (Literacy) Score (2018) ○ Grade 4: 35.75=f; 38.98=m ○ Grade 7: 33.08=f; 35.26=m ○ Grade 9: 35.46=f; 39.33=m • Average Maths (Numeracy) Score (2018) ○ Grade 4: 30.87=f; 35.57=m ○ Grade 7: 34.88=f; 36.48=m ○ Grade 9: 38.95=f; 41.89=m In January 2019 STAGES implemented performance monitoring testing for all G4, 7 and 9 children in all core subjects. The data is still being analysed and increase in average scores of girls will be added to Year 2 in the logframe when finalized by comparing 2018 and 2019 data. In future, the results of the tests that government implement already as per their mandate will be the results used for analysis and presentation at SPAMs.
Output 2.3: % of teachers demonstrating language competence (English and Wolaytatto)	This indicator is key to whether children understand their lessons or not, and therefore to our learning outcome and transition outcomes. It also contributes to intermediate outcomes. We have added an activity to also strengthen teachers of Wolaytatto (Grades 1-4) as their teaching competence is also key to	English language competency teacher training was conducted in November 2018 for more than 600 teachers. Classroom observations in a small sample of schools suggests that teachers are making efforts to implement what they learned in training. Wolaytatto teacher training is planned for year 3. Please note the monitoring tools for measuring teacher demonstrating language competence have

Logframe Output Indicator	Baseline status or Baseline values Relevance of the indicator for the project ToC	Baseline status or Baseline values
	children's overall competency in literacy. It is reflected in the new wording for this indicator	not yet been fully implemented. Values will be established in year 3. Total= 938 teachers trained English language competency in Yr 2
Output 2.4: % of teachers demonstrating competency in teaching numeracy	New indicator. Baseline response. This indicator focuses on change in teacher's own numeracy skills linked to the curriculum and their ability to teach it. Numeracy results at baseline 1 were very low in all grades. Numeracy is a key core outcome of STAGES	Some numeracy training has been provided in year 2 to Tutors as well as other core subjects, but not all teachers become Tutors. This training will reach all teachers of numeracy in supported schools. Please note the monitoring via pre and post training teacher tests and numeracy lesson observation tool has not yet been initiated. Values will be established in year 3 No teachers trained specifically in numeracy in year 2.
Output 3: Better access to secondary schools in extreme and remote areas		
Output 3.1: # of new inclusive secondary schools in project target areas	This indicator is key to outcomes related to attendance, learning and transition. Girls will have access to secondary schools closer to their homes in 4 'black hole' areas where girls are dropping out after primary due to difficult secondary access.	Year 1 (Baseline) value=0 Year 2 value = 0 School 1 will be finished in year 3 July Schools 2-4 will be finished in year 4 The construction of the first school in Kindo Koisha Woreda is well underway with the foundations now laid for all blocks. Plans are underway for the 3 remaining schools to be built. A tender will be advertised in the coming months for contractors for the remaining schools. The design of the schools considers children who may have difficulty with access to the school building, to classrooms and to toilets.
Output 3.2: # of inclusive separate girls' toilets blocks in secondary schools in project area (13 current and four new schools)	This indicator is key to improvement in girls attendance and transition as well as learning. Without access to separate toilets for boys and girls, girls may move out of school, or put themselves at risk by using bushes in scrub or wasteland.	Year 1 (Baseline) value=0 Year 2 value = 0 The toilets for construction/upgrading have all been assessed, with recommendations for improvement documented.
Output 4: Girls 'ready to learn'		
Output 4.1: # of girls have access to sanitary and hygiene service (sanitary pad, towels and soap).	Linked to outcome indicators of attendance and retention, and self-esteem. Without access to these items, girls will not attend school regularly and may drop out altogether	Year 1 (Baseline) value = 18,487 number of girls received sanitary and hygiene service in Yr 1 (2017/2018) - 16,616 number of primary girls - 1,871 number of secondary girls Year 2 value = 20,124 number of girls received sanitary and hygiene service in Yr 2 (2018/2019) - 16,390 number of primary girls - 3,734 number of secondary girls

Logframe Output Indicator	Baseline status or Baseline values Relevance of the indicator for the project ToC	Baseline status or Baseline values
		Sanitary items and soap have been disbursed on a regular basis over the life of the project so far.
Output 4.2: # of extremely vulnerable primary and secondary girls receiving a bursary enrolling & staying in school	Bursary provision will add to outcome indicators of attendance and retention and to self-esteem and well-being of girls. The bursaries are focused on particularly marginalized girls including girls with disabilities, orphaned girls or young mothers.	Year 1 (Baseline) Value = 0 girls received bursaries and remained in school in Yr 1 (2017/2018) - 0 number of primary girls - 0 number of secondary girls Year 2 Value = 195 girls received bursaries and remained in school in Yr 2 (2018/2019) - 0 primary girls - 195 secondary girls 2 rounds of bursaries have been provided in year 2
Output 4.3: # of girls regularly attending literacy and numeracy tutorials	The tutorials link to intermediate outcomes 1, 2 and 5, and to improved literacy and numeracy overall. They focus on girls who are struggling academically. They previously focused on all core subjects but from year 3 on will intensify support only on literacy and numeracy towards project outcomes and to address baseline numeracy findings.	Year 1 (Baseline) Value = 9378 attended tutorials in Yr 1 (2017/2018) - 8719 number of primary girls - 659 number of secondary girls Year 2 Value = 9852 girls attended tutorials in Yr 2 (2017/2018) - 9172 number of primary girls - 680 number of secondary girls Tutorials have been ongoing throughout years 1 and 2 focused on all core subjects. They will continue, focused only on literacy and numeracy.
Output 5. Mobilized, gender-aware communities demanding high-quality education		
Output 5.1 # of community members, disaggregated by gender, participating in School Improvement processes, specifically SPAMs	School performance appraisal meetings (SPAMs) are effectively a school performance review based on the collection of statutory data on 4 statutory domains of school improvement in Ethiopia (teaching and learning, school environment, community participation, and school leadership), as well as the performance monitoring data analysed for each school and presented to school communities. This activity is linked to almost all outcome and core indicators as it helps to highlight gaps in schools in relation to academic performance of students, and the 4 domains mentioned above. The community are involved in action planning based on gaps identified, strengthening voice and accountability at school community level/	Year 1 (Baseline) Value = 0 community members participating in SPAMS (2017/2018) Year 2 Value =-14,571 participated in SPAM in Year 2 (2018/2019) School SPAM was held in year 2 and the results from each school summarized for cluster and woreda level SPAMs. The Zonal Conference held in March 2019 provided a forum for presentation and discussion of Woreda level results.
Output 5.2: # of Gender and Safeguarding Action	Action plans are developed annually in all schools related to school improvement plans and school grants and based on the 4 domains described under 5.2. Link has supported the development of gender and	Year 1 (Baseline) Value = 0 Action plans produced (2017/2018)

Logframe Output Indicator	Baseline status or Baseline values Relevance of the indicator for the project ToC	Baseline status or Baseline values
Plans produced and fully operationalized	safeguarding indicators which it is hoped will be incorporated under the 4 domains as relevant. This activity is critical to all outcomes as more gender responsive, safe schools will attract girls to enroll, attend, remain, and learn.	Year 2 Value = 127 actions plans produced in Year 2 (2018/2019) These action plans were developed following the school SPAM in year 2
Output 5.3: # of role modelling and community awareness-raising campaigns	Community awareness campaigns are also linked across outcome areas. More gender aware communities demanding quality education for both boys and girls will support attendance, retention, teaching and learning, and transition. They will also add to sustainability beyond the project. Role model activities are added into this activity.	Year 1 (Baseline) Value = 0 campaigns in Yr 1 (2017/2018) Year 2 Value = 0 campaigns in Yr 2 (2018/2019) This activity is yet to be implemented.
Output 5.4: # of community school structures (KETB/PTSAs and SIC) supporting girls learning and transition.	This activity has been reinstated after being removed at the time of proposal. LCD felt that it is critical to work with these structures as they are the existing statutory structures which exist in Ethiopia to link the community with the school. Whilst the wider community may participate, these structures are involved in decision-making about how school and community resources are used through school improvement planning. They therefore have decision making power about how resources might be used to support girls, and particularly marginalized girls (and boys). It is key across outcomes of attendance, quality of teaching, improved school management and governance and positive community attitudinal change.	Year 1 (Baseline) Value = 0 community school structures supported in Yr 1 (2017/2018) Year 2 Value = 0 community school structures supported in Yr 2 (2018/2019) This training is yet to be implemented

Table 3: Output indicator issues

Logframe Output Indicator	Issues with the means of verification or sources and the collection frequency, or the indicator in general?	Changes or additions
Number and Indicator wording	Inappropriate wording, irrelevant sources, or wrong assumptions. Was data collection too frequent or too far between? Or no issues?	Change wording, add or remove sources, increase or decrease the frequency of data collection; or leave as is.
Output 1: Improved leadership for girls' learning at school, woreda, zone, and regional level		
Output 1.1: # of School staff, communities and relevant woreda officials of 4 LCDE supported woredas with increased capacity in SRGBV	There were 3 indicators for this activity previously focused on too many structures/bodies. One for school level plus community, and two for higher level capacity change (Woreda/Zone/Region). LCDE have no control over the extent to which and the quality with which this training is rolled out in other zones and woredas of the region, so we are focusing internal monitoring on school and community level and the 4 woredas that we work with. Means of verification will be an internal monitoring tool developed to indicate criteria for increased capacity based on capacity gaps identified and training content delivered. Training pre and post assessments will also be used to assess knowledge change.	Removed indicators for zonal and regional rollout
Output 1.2: # of Cluster Supervisors and Woreda Experts with increased capacity to monitor teaching quality and school improvement	No issue. New indicator for which a tool will be developed to monitor change in capacity	New indicator
Output 1.3: # of HTs & DHTs with increased capacity in school instructional leadership	New indicator as this training is critical to Output 1 – improved leadership for girls' education MoV is a monitoring tool developed with criteria which reflect increased capacity expected as a result of this training	This is an additional indicator
Output 1.4 # of active Primary school Girls' Education Advisory Committees (GEACs)	MoV: Monitoring tool developed to indicate 'active'.	
Output 1.5 # of Primary and Secondary Schools with established mechanisms to report and respond to abuse cases.	No issues. Monitoring tool developed to measure	
Output 2: Improved Quality of Learning		
Output 2.1:	Teacher monitoring/mentoring reports;	Changed to quarterly

Logframe Output Indicator	Issues with the means of verification or sources and the collection frequency, or the indicator in general?	Changes or additions
% of teachers practicing Gender and inclusion-Responsive Pedagogy (GRP)	Lesson observation tool (lesson plan, participation, seating arrangement and inclusion) ~~Gender Action Plans~~	
Output 2.2: % increase in average scores of girls (and boys) in literacy and numeracy. Percentage increase in average scores of girls (and boys) in English language and numeracy	No issue	
Output 2.3: % of teachers demonstrating language and teaching competence in English and Wolaytatto	Monitoring tool developed to indicate expected change in English and Wolaytatto language and teaching competence. Classroom observation to use the tool	Changed from annual to quarterly
Output 2.4: % of teachers demonstrating competency in teaching numeracy	New indicator, added Pre and post training teacher tests and Numeracy lesson observation tool	New indicator
Output 3: Better access to secondary schools in extreme and remote areas		
Output 3.1: # of new inclusive secondary schools in project target areas	No issues Added inclusive	None
Output 3.2: # of separate inclusive girls' toilets blocks in secondary schools in the project area (12 current and four new schools)	No issues Added inclusive	None
Output 3.3: **3.3** % of girls reporting improved quality of secondary school infrastructure	Indicator not appropriate or robust in measuring 'better access' Requires setting up interviews with girls and qualitative data collection	Removed
Output 4: Girls "ready to learn"		
Output 4.1: # of girls that have access to sanitary and hygiene service (sanitary pad, towels and soap).	There is no total of girls who will receive from which to derive a percentage.	Changed from percentage to number of girls that have access Reusable sanitary pads distribution to be monitored annually Soap distribution to be monitored quarterly
Output 4.2: # of extremely vulnerable primary and secondary	Added use of school attendance records in addition to finger print readers as a back-up Disbursement of bursaries	To be monitored Per term

Logframe Output Indicator	Issues with the means of verification or sources and the collection frequency, or the indicator in general?	Changes or additions
girls receiving a bursary enrolling & staying in school		
Output 4.3: # of girls regularly attending literacy and numeracy tutorials	Added actual tutorial hours as accurate records of these more easily available	
Output 4.4 Number of girls regularly participating in life skills, financial literacy and career advice?	Indicator dropped. Had to prioritise related to budget, staff capacity to collect data, and importance of collecting the data related to intermediate outcomes	Indicator Dropped
Output 5. Mobilized, gender-aware communities demanding high-quality education		
Output 5.1: # of community members, disaggregated by gender, participating in School Improvement processes, specifically SPAMs	No issue	none
Output 5.2: # of Gender and Safeguarding Action Plans produced and fully operationalized	No issue	Added annually to monitoring frequency
Output 5.3: # of role modelling and community awareness-raising campaigns	Indicator changed according to change of activity.	Role modelling activity (previously output 4) combined with community awareness raising campaigns, which is where role models identified will speak and interact with wider community. Frequency now annual
Output 5.4: # of community school structures (KETB/PTSAs and SIC) supporting girls learning and transition.	New indicator added as explained in table 2	Key community school structures which have the potential to contribute enormously to core outcomes of STAGES. Re-instated as removed at proposal stage to reduce budget.

Annex 4: Beneficiary Tables[2]

This annex should be completed by the project.

Table 4: Direct beneficiaries

Beneficiary type	Total project number	Total number of girls targeted for learning outcomes that the project has reached by endline	Comments
Direct learning beneficiaries (girls)— *girls in the intervention group who are specifically expected to achieve learning outcomes in line with targets. If relevant, please disaggregate girls with disabilities in this overall number.*	*[This should align with the total beneficiary numbers reported in the outcomes spreadsheet]*	*[This may equal the total project number in the outcomes spreadsheet and in the column to the left, or may be less if you have a staggered approach]*	*[Projects should provide additional information on who they are and the methodology used. If the numbers have changed since baseline, an explanation should be provided]*
Learning beneficiaries (girls)	Total: **61,345** Girls in Grades 1-9 present at baseline during 2017/2018 academic year G1: 10,761 G2: 8,758 G3: 7,605 G4: 6,670 G5: 5,468 G6: 4,915 G7: 4,397 G8: 4,104 G9: 2,061 [this cohort of G10 girls will be counted as 'Broader student beneficiaries' See Table 2 below] The source of this data is Education Management Information System from the	**61,345** (this includes 6,606 girls who are estimated to join from outside project woredas over the life of the project).	LCDE's operational definition for the direct project beneficiaries is girls **from Grade 1 to Grade 10.** Girls from Grades 1 to 10 in STAGES schools over the life of the project stand to benefit from better quality, safe and more gender and inclusion responsive pedagogy, improved literacy and numeracy teaching, a more supportive system at local government, zonal and regional levels, more supportive communities and community school structures, higher self-esteem and belief that they can achieve, better chances of further study or employment as they leave school, and direct support where particularly vulnerable. From year 2 onwards Grade 10 girls (the same girls who would have been in Grade 9 during the 2017/2018 academic year) will be project beneficiaries. The new learning beneficiary number of **61,345** includes Grade 1 students at the time of baseline as they stand to benefit from STAGES activities over the life of the project. An additional 6,606 girls are estimated to join

[2] This was revised in May 2019.

Beneficiary type	Total project number	Total number of girls targeted for learning outcomes that the project has reached by endline	Comments
	region for 2017/2018. Girls expected to join STAGES supported schools over life of project: 6,606 (see explanation in comments)		STAGES supported schools from neighbouring kebeles, woredas and zones over the life of the project. This number is based on the calculation of (6 students * 8 Grade levels * 127 primary schools) + (60 students * 4 lagre secondary schools) + 30 students * 9 small secondary schools). The previous beneficiary count did not include these groups. It does not include 2,227 Grade 10 girls enrolled during the 2017/2018 academic year as they were about to finish lower secondary school and would have experienced minimal benefits from the project activities following the baseliine.

Table 5: Other beneficiaries

Beneficiary type	Number	Comments
Learning beneficiaries (boys)—*as above, but specifically counting boys who will get the same exposure and therefore be expected to also achieve learning gains, if applicable.*	68,784 G1: 12,057 G2: 9,896 G3: 8,805 G4: 7,499 G5: 6,354 G6: 5,435 G7: 4,828 G8: 4,651 G9: 2,653 Boys expected to join STAGES supported schools over life of project: 6,606	Apart from interventions designed to address the identified barriers affecting female students' education, all other STAGES project interventions would provide similar exposure to all boys. For example, while boys benefit from general school improvement and learning interventions such as teacher training and school management and governance activities, boys will not receive tutorial support or the targeted direct support such as soap, sanitary pads, socio-emotional learning, awards, bursary, uniform. An additional 6,606 boys are estimated to join STAGES supported schools from neighbouring kebeles, woredas and zones over the life of the project. This number is based on the calculation of (6 students * 8 Grade levels * 127 primary schools) + (60 students * 4 lagre secondary schools) + 30 students * 9 small secondary schools).
Broader student beneficiaries (boys)—*boys who will benefit from the interventions in a less direct way, and therefore may benefit from aspects such as attitudinal change but not necessarily achieve improvements in learning outcomes.*	68,784	STAGES project interventions would provide similar exposure to all boys.
Broader student beneficiaries (girls)—*girls who will benefit from the interventions in a less direct way, and therefore may benefit from aspects such as attitudinal change but not necessarily achieve improvements in learning outcomes.*	2,227	The girls who will benefit from the interventions in a less direct way are those who are not direct beneficiaries of STAGES project, namely, the 2,227 grade 10 girls enrolled during year 1 of the project that would have stood to benefit from activities which took place in the 2017–18 academic year. Grade 10 students (those in Grade 9 in 2017/2018) become direct beneficiaries in year 2 of the intervention during the 2018 – 2019 academic year.
Teacher beneficiaries—*number of teachers who benefit from training or related interventions. If possible or applicable, please disaggregate by gender and type of training, with the comments box used to describe the type of training provided.*	2,487 Male = 1,829 Female = 658	Gender responsive training, Literacy and numeracy training, Active learning, or others Note this includes all primary and secondary teachers[3] only—not school directors or counselors. Also, grade 9 and 10 teachers are the same teachers.
Broader community beneficiaries (adults)—*adults who benefit from broader interventions, such as community messaging or dialogues,*	16,680 Est. male = 10,008 (60.00%)	School community who would take part in SPAM and all other awareness raising activities related to GEAC campaigns, school construction, or others

[3] Please note this is lower secondary only; namely, grades 9 and 10.

Beneficiary type	Number	Comments
community advocacy, economic empowerment interventions	Est. Female = 6,672 (40.00%)	This calculation depends on the school and how many community members attend or participate in school SPAMs. This number was calculated as 120 adults per school, with 60 percent men and 40 percent women. Note that 120 is the average number determined from past SPAM attendance lists. This was then multiplied by 139 total schools—primary and secondary—with 127 primary schools officially, but new schools opening often within the four woredas—and 13 secondary schools.

Table 6: Target groups—by school

School Age	Project definition of target group (Tick where appropriate)	Number targeted through project interventions	Sample size of target group at baseline
Lower primary (grade 1–4)	✓	33,794 (includes grade 1)	G4: 299 (33.82%)
Upper primary (grade 5–8)	✓	18,884	585 G6: 296 (33.48%) G8: 289 (32.69%)
Lower secondary (grade 9–10)	✓	4,288 (minus 2,227 G10 students not direct beneficiaries of the project)	N/A[4]
Upper secondary	X	x	N/A
Total:		**54,739** (does not include estimate of those joining schools)	**884**

[4] This will be included in the sample starting in midline 1.

Table 7: Target groups—by age

Age Groups	Project definition of target group	Number targeted through project interventions	Sample size of target group at baseline
Aged 6–8 (percentage aged 6–8)[5]	✓	Est. 19,519 (35.66%)	3 (0.34%)
Aged 9–11 (percentage aged 9–11)[6]	✓	Est. 14,275 (26.08%)	157 (17.76%)
Aged 12–13 (percentage aged 12–13)[7]	✓	Est. 10,383 (18.97%)	309 (34.95%)
Aged 14–15 (percentage aged 14–15)[8]	✓	Est. 8,501 (25.53%)	344 (38.91%)
Aged 16–17 (percentage aged 16–17)[9]	✓	Est. 2,061 (3.77%)	43 (4.86%)
Aged 18–19 (percentage aged 18–19)	X	x	14 (1.58%)
Aged 20+ (percentage aged 20 and over)	X	x	14 (1.58%)
Total:		**54,739 (does not include estimate of those joining schools)**	**884**

Table 8: Target groups—by subgroup

Social Groups	Project definition of target group	Number targeted through project interventions	Sample size of target group at baseline
Girls with disabilities	✓	Est. 2,912[10]	Est. 47 (5.32%)[11]
Orphaned girls	✓	TBD[12]	To be collected at future evaluation points[13]

[5] Number targeted through project intervention assumes, on average, students in grades 1 and 2 fall within the ages 6-8, for example. It draws on gross enrolment figures for the corresponding grade levels in the 2017/2018 academic year.
[6] Number targeted relates to grades 3 and 4.
[7] Number targeted relates to grades 5 and 6.
[8] Number targeted relates to grades 7 and 8.
[9] Number targeted relates to grades 9 and 10.
[10] At this time Link is only supporting girls with disability directly in grade 9, so only know numbers of girls with disabilities directly receiving support (bursaries). However, if extrapolating the proportion of girls with disabilities found in baseline sample, the project estimates it will reach approximately 2,912 girls with disabilities through its general interventions and activities.
[11] The disability prevalence figure is based on the Washington Group – Child Functioning set that were administered during phase 2 of the baseline as part of the girls transition survey in December 2018 with the cohort sample of grade 6/7 and 8/9 female students only. The sample size is estimated by extrapolating the prevalence rate to the total sample size – including grade 4 girls.
[12] At this time, only supporting orphaned girls directly in grade 9.
[13] Due to the sensitive nature of this question, especially with younger girls, orphanhood status was not directly asked of girls during the baseline. Instead, girls and other stakeholders were asked about the prevalence rates within their schools and communities. In addition, for a subsample of girls who were determined to be at high risk of drop-out, a subsample of parent/caregivers were asked

Social Groups	Project definition of target group	Number targeted through project interventions	Sample size of target group at baseline
Girls who are pregnant or have children	✓	TBD	To be collected at future evaluation points[14]
Pastoralist girls	N/A	N/A	N/A
Child laborers	N/A	N/A	N/A
Poor girls[15]	✓	**54,739**	**884[16]**
Total:		**54,739 (does not include estimate of those joining schools)**	**884**

Table 9: Target groups—by school status

Educational subgroups	Project definition of target group	Number targeted through project interventions	Sample size of target group at baseline
Out-of-school girls: have never attended school	N/A	N/A	N/A
Out-of-school girls: have attended school, but dropped out	N/A	N/A	N/A
Girls in-school	✓	**54,739**	**Total: 884** G4: 299 (33.82%) G6: 296 (33.48%) G8: 289 (32.69%)

questions about the girls' orphanhood status. However, during baseline analysis it was determined proxy measures were insufficient to inform estimates, and therefore, at future evaluation points, girls will be asked directly about their orphanhood status.

[14] Due to the sensitive nature of this question, motherhood and pregnancy status were not directly asked of girls during the baseline. Instead, girls and other stakeholders were asked about the prevalence rates within their schools and communities. However, during baseline analysis it was determined proxy measures were insufficient to inform estimates, and therefore, at future evaluation points, girls will be asked directly about their motherhood status, but not directly about pregnancy status.

[15] Please note all girls within the targeted Woreda schools are considered poor.

[16] Based on Link's definition, all girls included in the sample should be considered poor.

Annex 5: MEL Framework

Please click the icon below to view a slightly revised version of the MEL Framework updated in May 2019.[17] This MEL Framework requires review and approval from the Evaluation Manager.

Annex 5_MEL Framework_revised_M

[17] The previous MEL Framework document was approved by the Evaluation Manager as of October 17, 2017 and can be provided upon request.

Annex 6: External Evaluator's Inception Report

Please click the icon below to view the External Evaluator's Inception Report. This report was submitted to the implementing agency, Link Community Development, on January 10th, 2018.

Annex 6_Inception Report.docx

Annex 7: Data Collection Tools Used for Baseline[18]

The external evaluator, STS, utilized several different assessments, evaluation surveys, and qualitative instrument guides as part of the baseline data collection. Additional details on the tools utilized during the study and included in the annex can be found below:

1. Learning Assessments (Phase 1 – Spring 2018)

Grade 4 EGRA/EGMA

LA_Grade 4 EGRA EGMA.docx

For the purpose of this annex, STS has provided the paper version of the Grade 4 EGRA/EGMA for easy reference. Please note that the assessment was administered to grade 4 female students electronically via tablets using the Tangerine platform for this study. The paper version served only as a back-up in case technology failed but was not used in any field administrations. The Grade 4 EGRA subtasks were conducted separately in Wolayttatto and English. The Grade 4 EGMA subtasks were conducted in Wolayttatto. No Washington Group Questions were administered to grade 4 students.

Grade 6 EGRA/EGMA

LA_Grade 6 EGRA EGMA.docx

For the purpose of this annex, STS has provided the paper version of the Grade 6 EGRA/EGMA for easy reference. Please note that the assessment was administered to grade 6 female students electronically via tablets using the Tangerine platform for this study. The paper version served only as a back-up in case technology failed but was not used in any field administrations. The Grade 6 EGRA subtasks were conducted separately in Wolayttatto and English. The Grade 6 EGMA subtasks were conducted in English only. Washington Group Questions – Short Set were provided in English and Amharic.[19]

Grade 8 EGRA/EGMA

LA_Grade 8 EGRA EGMA.docx

For the purpose of this annex, STS has provided the paper version of the Grade 8 EGRA/EGMA for easy reference. Please note that the assessment was administered to grade 6 female students electronically via tablets using the Tangerine platform for this study. The paper version served only as a back-up in case technology failed but was not used in any field administrations. The Grade 8 EGRA and EGMA subtasks were conducted in English only. Washington Group Questions – Short Set were provided in English and Amharic.[20]

[18] This annex was updated in May 2019.
[19] Enumerators were allowed to translate questions into Wolayttatto as needed.
[20] Enumerators were allowed to translate questions into Wolayttatto as needed.

Grade 4 EGRA/EGMA Stimuli

LA_Grade 4 EGRA EGMA Stimuli.docx

STS has provided the stimuli used by grade 4 female students during the administration of the Grade 4 EGRA/EGMA.

Grade 6 EGRA/EGMA Stimuli

LA_Grade 6 EGRA EGMA Stimuli.docx

STS has provided the stimuli used by grade 6 female students during the administration of the Grade 6 EGRA/EGMA.

Grade 8 EGRA/EGMA Stimuli

LA_Grade 8 EGRA EGMA Stimuli.docx

STS has provided the stimuli used by grade 8 female students during the administration of the Grade 8 EGRA/EGMA.

Grade 8 SeGRA

LA_Grade 8 SeGRA.docx

STS has provided the written assessment form completed by grade 8 female students during the administration of the Grade 8 SeGRA.

Grade 8 SeGMA

LA_Grade 8 SeGMA.docx

STS has provided the written assessment form completed by grade 8 female students during the administration of the Grade 8 SeGMA.

2a. Evaluation Surveys (Phase 1 – Spring 2018)

Girls Student Survey

EL_Girls Student Survey.docx

For the purpose of this annex, STS has provided the paper version of the Girls Student Survey in English and Amharic for easy reference.[21] Please note that the Amharic version of the survey was administered to grade 4, 6, and 8 female students electronically via tablets using the SurveyCTO platform for this study.[22] The paper version served only as a back-up in case technology failed but was not used in any field administrations. The English version is provided for the readers' benefit.

Boys Student Survey

EL_Boys Student Survey.docx

For the purpose of this annex, STS has provided the paper version of the Boys Student Survey in English and Amharic for easy reference. Please note that the Amharic version of the survey was administered to grade 4, 6, and 8 male students electronically via tablets using the SurveyCTO platform for this study.[23] The paper version served only as a back-up in case technology failed but was not used in any field administrations. The English version is provided for the readers' benefit.

Classroom Teachers Survey (Primary-level)

EL_Classroom Teachers Survey.docx

For the purpose of this annex, STS has provided the paper version of the Classroom Teachers Survey in English and Amharic for easy reference. Please note that the Amharic version of the survey was administered to grade 4, 6, and 8 classroom teachers electronically via tablets using the SurveyCTO platform for this study. The paper version served only as a back-up in case technology failed but was not used in any field administrations. The English version is provided for the readers' benefit.

School Audit (Primary-level)

EL_School Audit.docx

For the purpose of this annex, STS has provided the paper version of the School Audit Survey in English and Amharic for easy reference. Please note that the Amharic version of the survey was administered to school directors[24] electronically via tablets using the SurveyCTO platform for this study. The paper version served only as a back-up in case technology failed but was not used in any field administrations. The English version is provided for the readers' benefit.

[21] Please note the Washington Group Questions – Short Set are not included in the Girls Student Survey. Instead, these were administered as part of the Grade 6 EGRA/EGMA and Grade 8 EGRA/EGMA and can be found in the associated tools.
[22] Enumerators were allowed to translate the Amharic questions and response options into Wolayttatto as needed.
[23] Enumerators were allowed to translate the Amharic questions and response options into Wolayttatto as needed.
[24] Within the Wolaita Zone in Ethiopia, the terms school director, principal, and head teacher are used interchangeably.

Classroom Observation (Primary-level)

EL_Classroom Observation.docx

For the purpose of this annex, STS has provided the paper version of the Classroom Observation in English and Amharic for easy reference. Please note that the Amharic version of the Classroom Observation was conducted in grade 4, 6, and 8 classrooms. The English version is provided for the readers' benefit.

Woreda Staff Survey

EL_Woreda Staff Survey.docx

For the purpose of this annex, STS has provided the paper version of the Woreda Staff Survey in English and Amharic for easy reference. Please note that the Amharic version of the survey was self-administered by woreda staff via tablets using the SurveyCTO platform for this study. The paper version served only as a back-up in case technology failed but was not used in any field administrations. The English version is provided for the readers' benefit.

2b. Evaluation Surveys (Phase 2 – Fall 2019)

Girls Transition Survey

Girls Transition Survey ENGLISH.pdf

Girls Transition Survey AMHARIC.pdf

For the purpose of this annex, STS has provided the SurveyCTO printable version of the Girls Transition Survey in English and Amharic for easy reference.[25] Please note that the Amharic version[26] of the survey was administered electronically via tablets using the SurveyCTO platform with cohort girls expected to have transitioned into grades 5, 7, and 9. Due to the successful electronic administration of surveys during phase one of the baseline, no paper versions of the tool were developed or utilized during data collection. The English version is provided for the readers' benefit.

Parent/Caregiver Survey

Parent Caregiver Survey ENGLISH.pdf

Parent Caregiver Survey AMHARIC.pdf

For the purpose of this annex, STS has provided SurveyCTO printable version of the Parent/Caregiver Survey in English and Amharic for easy reference. Please note that the Amharic version[27] of the survey was administered to a sub-sample of grade 4 and 6 cohort girls' parents electronically via tablets using the SurveyCTO platform for this study. Due to the successful electronic administration of surveys during phase one of the baseline, no paper versions of the tool were developed or utilized during data collection. The English version is provided for the readers' benefit.

[25] Please note the Washington Group Questions – Child Functioning Set were included in the Girls Transition Survey for girls in grades 6 and above.
[26] Enumerators were allowed to translate the Amharic questions and response options into Wolayttatto as needed.
[27] Ibid.

Guidance Counsellor Survey (Secondary-level)

Guidance Counselor Survey ENGLISH.pdf

Guidance Counselor Survey AMHARIC.pdf

For the purpose of this annex, STS has provided SurveyCTO printable version of the Guidance Counselor Survey in English and Amharic for easy reference. Please note that the Amharic version[28] of the survey was administered to secondary school guidance counselors in treatment schools electronically via tablets using the SurveyCTO platform for this study. Due to the successful electronic administration of surveys during phase one of the baseline, no paper versions of the tool were developed or utilized during data collection. The English version is provided for the readers' benefit.

Classroom Teachers Survey (Secondary-level)

Secondary School Teacher Survey ENGL

Secondary School Teacher Survey AMH/

For the purpose of this annex, STS has provided the SurveyCTO printable version of the Classroom Teachers Survey in English and Amharic for easy reference. This is the same tool administered to teachers in primary schools. Please note that the Amharic version[29] of the survey was administered to lower secondary school teachers (grade 9 and 10) electronically via tablets using the SurveyCTO platform for this study. Due to the successful electronic administration of surveys during phase one of the baseline, no paper versions of the tool were developed or utilized during data collection. The English version is provided for the readers' benefit.

School Audit (Secondary-level)

Secondary School Audit ENGLISH.pdf

Secondary School Audit AMHARIC.pdf

For the purpose of this annex, STS has provided the SurveyCTO printable version of the Secondary School Audit Survey in English and Amharic for easy reference. This is the same tool administered to school directors in primary schools. Please note that the Amharic version[30] of the survey was administered to secondary school directors[31] electronically via tablets using the SurveyCTO platform for this study. Due to the successful electronic administration of surveys during phase one of the baseline, no paper versions of the tool were developed or utilized during data collection. The English version is provided for the readers' benefit.

[28] Ibid.
[29] Ibid.
[30] Ibid.
[31] Within the Wolaita Zone in Ethiopia, the terms school director, principal, and head teacher are used interchangeably.

2. Qualitative Instruments (Phase 1 – Spring 2018)

Upper-primary Girls Focus Group Discussion Guide

QU_Upper-primary Girls Focus Group Dis

For the purpose of this annex, STS has provided the guiding questions for the Upper-primary Girls Focus Group Discussion Guide in English and Amharic. Please note that the focus group was facilitated by trained, female teachers in a mix of Amharic and Wolayttatto with female students from grades 6 and 8. The English version is provided for the readers' benefit.

Upper-primary Boys Focus Group Discussion Guide

QU_Upper-primary Boys Focus Group Dis

For the purpose of this annex, STS has provided the guiding questions for the Upper-primary Boys Focus Group Discussion Guide in English and Amharic. Please note that the focus group was facilitated by trained, female teachers in a mix of Amharic and Wolayttatto with male students from grades 6 and 8. The English version is provided for the readers' benefit.

Female Teachers Focus Group Discussion Guide

QU_Female Teachers Focus Group Discussi

For the purpose of this annex, STS has provided the guiding questions for the Female Teachers Focus Group Discussion Guide in English and Amharic. Please note that the focus group was facilitated by trained, female teachers in a mix of Amharic and Wolayttatto with female primary school teachers. The English version is provided for the readers' benefit.

Parents Focus Group Discussion Guide

QU_Parents Focus Group Discussion Gui

For the purpose of this annex, STS has provided the guiding questions for the Parents Focus Group Discussion Guide in English and Amharic. Please note that the focus group was facilitated by trained, female teachers in a mix of Amharic and Wolayttatto with mothers and fathers of female primary school students. The English version is provided for the readers' benefit.

School Management Key Informant Interview Guide

QU_School Management Key Infc

For the purpose of this annex, STS has provided the guiding questions for the School Management Key Informant Interview Guide in English only. Please note that interview was conducted by STS's qualitative consultant in Amharic. The English version is provided for the readers' benefit.

Zone and Woreda Education Officials Key Informant Interview Guide

QU_Zone and Woreda Education Of

For the purpose of this annex, STS has provided the guiding questions for the Zone and Woreda Education Officials Key Informant Interview Guide in English and Amharic. Please note that interview was conducted by STS's qualitative consultant in Amharic. The English version is provided for the readers' benefit.

Annex 8: Datasets, Codebooks, and Programs

Quantitative datasets and codebooks. All cleaned and labeled quantitative datasets and associated codebooks—including evaluation surveys and learning test data—have been submitted separately. The datasets were fully anonymized before submission. All datasets were cleaned and clearly labeled so individuals, and school or communities can be matched across datasets. Quantitative data sets were provided in SPSS formats and associated codebooks provided in Excel. The files submitted, and associated format type and content are detailed in the following table.

File submitted	Format	Description
Datasets		
Grade 4 dataset	SPSS	Included Grade 4 EGRA/EGMA data and the associated girl student survey data from Phase 1 baseline, student transition data from Phase 2 baseline, as well as aggregate data from boy student surveys, classroom teacher surveys, school audit survey, woreda official survey and parent/caregiver survey
Grade 6 dataset	SPSS	Included Grade 6 EGRA/EGMA data and the associated girl student survey data from Phase 1 baseline, student transition data from Phase 2 baseline, as well as aggregate data from boy student surveys, classroom teacher surveys, school audit survey, woreda official survey and parent/caregiver survey
Grade 8 dataset	SPSS	Included Grade 6 EGRA/EGMA ddata and the associated girl student survey data from Phase 1 baseline, student transition data from Phase 2 baseline, as well as aggregate data from boy student surveys, classroom teacher surveys, school audit survey, woreda official survey and parent/caregiver survey
Datasets		
Grade 4 codebook	Excel	Included codebook for Grade 4 EGRA/EGMA data and the associated girl student survey data from Phase 1 baseline, student transition data from Phase 2 baseline, as well as aggregate data from boy student surveys, classroom teacher surveys, school audit survey, woreda official survey and parent/caregiver survey
Grade 6 codebook	Excel	Included for Grade 6 EGRA/EGMA data and the associated girl student survey data from Phase 1 baseline, student transition data from Phase 2 baseline, as well as aggregate data from boy student surveys, classroom teacher surveys, school audit survey, woreda official survey and parent/caregiver survey
Grade 8 codebook	Excel	Included codebook for Grade 6 EGRA/EGMA data and the associated girl student survey data from Phase 1 baseline, student transition data from Phase 2 baseline, as well as aggregate data from boy student surveys, classroom teacher surveys, school audit survey, woreda official survey and parent/caregiver survey

Qualitative codebook. The qualitative codebook was developed utilizing NVivo and exported into a word document format. Please click the icon below to view the qualitative codebook.

QU_Qualitative Codebook.docx

Annex 9: Learning Test Pilot and Calibration

This annex includes a summary of pre-baseline activities for preparing the learning assessments. The assessment items retained at the end of the pilot were those used at baseline.

Design of Learning Tests

The learning tests were designed based on content from the following:

1. Previous assessments conducted under GEC-1
2. Assessments for girls in secondary school in Ethiopia shared by Young Lives
3. STS' experience in developing assessments for students in over 30 countries

The following table shows each subtask as it was administered at the pilot and at baseline. The table shows several difficulty levels were administered for each reading passage; for all other subtasks, the number of items piloted included items of varying ranges of difficulty. At baseline, the final tools only included those items that showed good item statistics (discussed in the next table).

The table shows the design of each subtask in the draft tools, followed by the tools that were piloted and finally, the tools that were used at baseline. Autostop indicates the item at which a student, if they did not answer any questions correctly, would be automatically stopped in the subtask and moved on to the next subtask (this is a function of Tangerine and is programmed to reduce test frustration for students unable to answer the first few items of a subtask). The nudge rule is the rule used in training enumerators to ensure they are moving students along in each subtask and not allowing students to "hang" at any single item for a long period of time. Again, this rule is enforced to reduce test-taking frustration.

Subtask Details—Pilot and Baseline

	Draft Tool		Tools that were Piloted			Tools that were used at baseline			
	Time Assumed	No. Items Developed	Time	Autostop	Nudge Rule	No. Items	Time	Autostop	Nudge Rule
EGRA									
Letter Sound	120 seconds	None	n/a	n/a	n/a	100 letters	120 seconds	10 letters	3 seconds
Familiar Word—English and Wolayttatto	120 seconds	75 items each language	180 seconds (75 items)	10 words	3 seconds	50 items	120 seconds	10 words	3 seconds
Nonword—English and Wolayttatto	120 seconds	75 items each language	180 seconds (75 items)	10 words	3 seconds	50 items	120 seconds	10 words	3 seconds
Reading A English		None	180 seconds	1 word before end 1st Q	3 seconds	Using	60 seconds	Follow EGRA guidance	3 seconds
Reading B English		None	180 seconds	1 word before end 1st Q	3 seconds	Not using			
Reading C English		None	180 seconds	1 word before end 1st Q	3 seconds	Not using			
Reading A Wolayttatto		None	180 seconds	1 word before end 1st Q	3 seconds	Not using			
Reading B Wolayttatto		None	180 seconds	1 word before end 1st Q	3 seconds	Using	60 seconds	Follow EGRA guidance	3 seconds
Total time estimated						English: eight minutes maximum (allow 10 for estimate) Wolayttatto: eight minutes maximum (allow 10 for estimate)			

	Draft Tool		Tools that were Piloted			Tools that were used at baseline			
	Time Assumed	No. Items Developed	Time	Autostop	Nudge Rule	No. Items	Time	Autostop	Nudge Rule
EGMA									
Number Identification	120 seconds	4 numbers	Not piloted	Not piloted	Not piloted	20 items	120 seconds	5 items	5 seconds
Quantity Discrimination	Not timed	25 problems	Not timed	5 items	5 seconds	10 items	Not timed	5 items	5 seconds
Missing Number	Not timed	25 problems	Not timed	5 items	5 seconds	10 items	Not timed	5 items	5 seconds
Addition	180 seconds	11 problems (level A), 14 problems (level B)	180 seconds	11 items (level A)	5 seconds	25 items (20 items level A, 5 items level B)	120 seconds	Level A	5 seconds
Subtraction	180 seconds	14 problems (level A), 10 problems (level B)	180 seconds	14 items (level A)	5 seconds	25 items (20 items level A, 5 items level B)	120 seconds	Level A	5 seconds
Word Problems—G6 (level A) and G8 (level B)		10 problems (level A), 10 problems (level B)	Not timed	10 items (level A)	15 seconds	6 items	Not timed	3 items	15 seconds
Total time estimated						Math: 10 minutes maximum using the nudge rule consistently			

Calibration

The pilot instruments were not piloted in full form; this would have presented an individual student with a high number of assessment subtasks that would take extremely long to finish and certainly end in test fatigue. Instead, since items in two languages needed to be piloted, the subtasks were designed to include additional items (rather than multiple subtasks) and were then selected on an item-level. This process of selection is described below.

Rules for selecting items from Pilot Instruments

- Exclude all items that
    - Have a point-biserial correlation with the total score (task level) less than 0.25
    - Have a p-value less than 0.08 and greater than 0.92
- Retain items will be classified into the following groups

	Low difficulty (p value > 0.75 and less than 0.92)	Medium difficulty (p-values between 0.25 and 0.75)	High difficulty (p-values lower than 0.25 and higher than 0.08)
Low discrimination (correlation of less than 0.45)	Type A	Type B	Type C
High discrimination (correlation of more than 0.45)	Type D	Type E	Type F

The items with highest relative quality are those of Type E. To ensure comparability and assuming there are items in all the categories, the retained items will be shared across forms as follows:

Step	Baseline Form	Midline 1 Form	Midline 2 Form	Endline Form
Step 1	¼ Type E items	¼ Type E	¼ Type E	¼ Type E
Step 2	2 Type F and 2 Type D	2 Type F and 2 Type D	2 Type F and 2 Type D	2 Type F and 2 Type D
Step 3	Complete the form using Type B items first, and then balancing Type A and C			
Step 4		Add all items Type E from baseline		
Step 5		Fill in with the best items left (from either baseline form or pilot data)		
Step 6			Add as many Type E items as possible from the midline 1 form	
Step 7			Fill in with the best items left (from either baseline form or pilot data) or make minor modifications to other baseline items	

Step	Baseline Form	Midline 1 Form	Midline 2 Form	Endline Form
Step 8				Add as many Type E items as possible from the midline 2 form
Step 9				Fill in with the best items left (from either baseline form or pilot data) or make minor modifications to other midline 1 items

The resulting selection of items is shown in the tables at the end of this annex, by grade, language, and test.

Implications from the pilot and final test

- Uncertain about the level of difficulty of grade 8 subtasks for grade 10. There were no ceiling effects suggesting the instrument will be too easy for grade 10 girls; further development of SeGRA or SeGMA subtasks may be required before piloting for grade 10.
- Review item-level results from baseline instruments and determine if any items in subsequent test points need to be revised as a result.
- Consider piloting the midline 1 instrument when piloting grade 10 instrument to address any equating required.

Grade 8 SeGRA Pilot Tables

G8_Pilot_Tables_SEGRA.xlsx

Grade 8 SeGMA Pilot Tables

G8_Pilot_Tables_SEGMA.xlsx

Grade 8 EGRA Pilot Tables

G8_Pilot_Tables_EGRA.xlsx

Grade 8 EGMA Pilot Tables

G8_Pilot_Tables_EGMA.xlsx

Grade 6 EGRA Pilot Tables

G6_Pilot_Tables_EGRA.xlsx

Grade 6 EGMA Pilot Tables

G6_Pilot_Tables_EGMA.xlsx

Annex 10: Sampling Framework

Please click the icon below for the updated and final Sampling Framework in Excel.

Sampling Framework.xlsx

Annex 11: Control Group Approach Validation

The content reported in this annex is similar to that include in the body of the report (Section 4) since the concerns regarding comparability of treatment and comparison schools are of concern to the External Evaluator.

Because the schools selected for comparison come from a geographically distinct area—all comparison schools come from a separate woreda rather than from within the same woredas as the treatment schools—the group is referred to as the **comparison** group and not the **control** group throughout this report.

The treatment group includes primary schools from four woredas, and the comparison group includes primary schools from only one woreda. The comparability of these two groups—based on the comparability of the woredas—is examined below.

As the data suggest, the baseline literacy and numeracy levels for students in treatment and comparison woredas vary. Female students in the comparison woreda outperformed female students in treatment woredas in grade 4 numeracy and grade 8 literacy. However, there are several differences between the four treatment woredas and the one comparison woreda that are important to note as comparability is central to the analyses:

1. Full saturation of the STAGES project in each of the four treatment woredas rendered it impossible to identify comparison schools that were within the same administrative and geographic parameters as treatment schools, but minus treatment.
2. The comparison woreda—Ofa—had fewer girls affected by barriers than did the treatment woredas:
    a. Ofa has higher scores on the gender-parity index than do the treatment woredas (see Section 2).[32] However, results from the teacher surveys suggest that support for girls' education was comparable in treatment and comparison woredas (see Section 3).
    b. Fewer girls and boys in Ofa reported their households were unable to meet basic needs than girls and boys in the treatment woredas—24.89 percent of boys and 23.99 percent of girls in comparison versus 35.37 percent of boys and 39.71 percent of girls in treatment woredas.
    c. Ofa has fewer girls who report low levels of household support—14.32 percent for comparison versus 22.06 percent for treatment.[33]
    d. Girls in Ofa report slightly higher levels of perceived safety traveling to and from school (5.88 percent in Ofa compared with 9.39 percent in treatment woredas).
    e. Twice as many girls in comparison schools reported more than two disciplinary actions compared with girls in treatment schools. Despite the lower rates of reported disciplinary actions, one-third of girls in treatment schools also reported more than two disciplinary actions.

[32] Wolayta Education Sector—Education Management Information System. 2017. Education Statistics Annual Abstract 2008
[33] See household support index results in Section 3.

3. By contrast, fewer girls in the comparison woreda report attending school at least half the time than their peers in the treatment woredas (82.26 percent vs. 96.95 percent, respectively). The barriers girls in treatment and comparison schools faced related to school facilities were comparable except for the use of drinking water facilities. Treatment woredas had twice as many girls who reported not using drinking water facilities than the comparison woreda. The proportion of girls who report high corporal-punishment rates were twice as large in the comparison woreda than in the treatment woredas.

As a result, the comparability of the treatment and comparison woredas is unclear. Characteristics of girls suggest that fewer girls in the treatment woredas face individual-level challenges than girls in the comparison woreda. When examining barriers, the opposite seems true. At this point in the evaluation, with no alternative options for a comparison group, a change in the comparison group is not feasible. Instead, it will be important to examine shifts in characteristics and barriers within the two groups at the next evaluation point and determine whether there are indications that comparability is compromised.

Annex 12: External Evaluator Declaration

Name of Project:	STAGES—Supporting Transition of Adolescent Girls through Enhancing Systems
Name of External Evaluator:	School-to-School International
Contact Information for External Evaluator:	Casey McHugh (cmchugh@sts-international.org) Hetal Thukral (hthukral@sts-international.org)

Names of all members of the evaluation team:	• Hetal Thukral
	• Casey McHugh
	• Randy Tarnowski
	• Ami Kanani
	• Allison Born
	• Ashley Doria
	• Melyssa Sibal
	• Drew Schmenner
	• Kayla Nachtshiem
	• Consultants
	  ○ Alemneh Tadele
	  ○ Zewdu Gebrekidan
	  ○ Bekalu Yayeh
	  ○ Erkyhun Desta
	  ○ Girma Demissie
	  ○ Yehualashet Desalegn

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

Hetal Thukral	School-to-School International	May 30, 2019
(Name)	(Company)	(Date)

Annex 13: Project Management Response[34]

This annex should be completed by the project.

The project finds the recommendations of both stages of the baseline – phase 1 focused on learning and sustainability outcomes and phase 2 focused on transition outcomes - useful and relevant in relation to the baseline findings.

The EGRA/EGMA/SEGRA/SEGMA test results provide a strong rationale for investing project resources in literacy and numeracy only and reducing or removing investment in other subjects. With resources spread too thinly across additional subject areas, there is a risk of reducing impact on core outcome-3, improved literacy and numeracy skills. The test results also provide a rationale for adding specific training for teachers of mathematics (content, methodology and classroom management) in the same way that STAGES provides training in English Language Competency teaching training, and for adding in training for grade 1-4 Wolaytatto teachers too as a step towards improved literacy.

The test results present many challenges for STAGES and government partners to address in pursuing programme goals. One of these is the clear drop evidenced in children's learning outcomes following the switch to English as the medium of instruction at grade 5. This is a complex and sensitive issue and STAGES will balance its support, working within existing education system structures and policy whilst also sharing learning from the project to influence thinking on how children's, particularly girls learning can be improved in the longer term.

The more nuanced analysis of additional areas of marginalization that girls face—for examples orphaned girls, girls affected by disability and young mothers—is helpful. It provides the project with the rationale to largely strengthen existing activities to ensure that these girls are included—present in school, participating in classroom activities and achieving. The project will continue to strengthen its inclusive education 'twin-track' approach, working to strengthen the education system to include marginalized girls whilst further targeting girls identified to experience additional factors of marginalization. Reaching girls and boys affected by disability, who are orphans, or who are pregnant or young mothers can be achieved through strengthening all program activities at all levels - system, school, community/family and child. Transition data collected in December 2018 from the actual transition of girls provides strong rationale for strengthening girls' retention and return to school in activities such as Mother and Father Groups, community campaigning and role modelling, and PTSA/SIC/KETB activities. Community structures and structures which link community and school can take action and have an impact on girls who drop out of school early. The project will also consider the provision of some basic need items to particularly vulnerable girls at primary level. Improving the quality and relevance of education received by girls in school (teaching, learning, leadership, environment, life-skills & financial literacy, social and emotional learning), will also attract girls to remain in school, and attract their parents to keep them there.

Link welcomes the suggestion of considering and learning from international and national best practice around addressing early marriage, migration and further addressing school-related gender-based violence. This is one of Link's strengths, channeling information on best practice into country programmes as well as learning and sharing from the contexts in which we work. It is particularly useful that the baseline has found corporal punishment to be an issue which had not formerly been emphasized

[34] This annex was updated in May 2019.

or identified to this extent. The links and references suggested by the evaluator will be useful in addition to those already used by the program. In light of baseline findings, more recent global events around safeguarding, and in response to recent DFID/Fund Manager requirements, Link has updated and very much strengthened safeguarding across the organization and within programme interventions.

The recommendation on working effectively with instituted community-school structures is in line with Link reflections, and PTSA/SIC activities removed from the budget earlier have been reinstated during the recent budget re-profiling exercise. Link intends to strengthen its Community Engagement Strategy overall, to bring all components under 'improved community support' together in a more strategic, sustainable and integrated manner, linked to other outputs. Community engagement is key when considering the particular barriers which are faced by girls with disability, girls who are orphans, who may be at risk of abduction, migration or early marriage, or who may be young unmarried mothers. It is through working with communities, instituted community-school structures and community leaders (traditional or religious) that great gains can be made in addressing the social, attitudinal and cultural barriers that these girls face. Continuing to strengthen the SPAMs and ensuring a truly diverse and representative group of participants at school level SPAMs will also encourage more diverse and inclusive planning and action for excluded or particularly marginalized girls.

Link does not underestimate the magnitude of change required to see changes in girls' learning outcomes through STAGES, and the complexity involved. Continuing deep engagement with local, zonal and regional education officials will be absolutely central in all activities and as is recommended in further evaluations of STAGES. The baseline findings around the difference in woreda officials' attitudes towards boys' and girls' education also highlight that, despite these strong relations, there is still much work to be done. Link will provide further capacity development for key partners, particularly at Woreda level, on gender, inclusion and safeguarding-responsive monitoring and supervision of schools. Regional and Zonal level government will be key partners in delivering this.

The baseline findings around teacher attendance are also interesting and significant, and this is an area that will need to be planned for as STAGES moves forward. There was interesting work done around community voice and accountability in Nigeria which can be drawn upon, where School-Based Management Committee monitoring of school improvement plans and everyday management brought positive change around teacher attendance and punctuality. STAGES will consider the role that PTSAs, School Improvement Committees and Kebele Education Training Boards might play in supporting teacher attendance through interventions designed to develop capacity on gender and inclusive education, and safeguarding. The findings on female teachers are also particularly interesting, and the project will have to find ways to address limiting gender norms at all levels and associated power dynamics.

Project's response to evaluators' comments on project's gender approach and integration.

As in the previous section, the project is happy with the comments on the project's gender approach and integration and will consider these comments in further strengthening the integration of gender throughout the project. Link has also done much thinking around inclusion and how this can be better reflected throughout the program to ensure that particularly marginalized girls are not left out. Inclusion will be embedded, as will gender and safeguarding across all programme activities. This has required some research into the policy framework supporting, for example, girls with disability, and how Link can work around this to support the government to deliver on its commitments. Already teacher training on Gender Responsive Pedagogy (GRP) has been adapted to ensure that teachers are better able to meet the

commonly diverse needs of students in a classroom - girls who face additional and/or multiple barriers to participation and learning, such as girls (and boys) with disability or specific needs in the classroom. An inclusive education approach reflected through all program activities would not as is often feared necessarily mean much additional cost, or the mass provision of specific aids and equipment.

It will be good to have a more nuanced discussion on how key barriers and characteristics interplay to impact on girls' attendance and learning, and transition when we can measure it.

Project proposed changes to the logframe.

Based on all of the above, the theory of change has been updated slightly, mainly through strengthening it across all levels for inclusion. Internal monitoring output indicators and targets for which the project is responsible have been reviewed, revised and updated as per Annex 3, and beneficiary numbers revised as per annex 4. Transition targets have been entered and all other targets at intermediate outcome level reviewed. The project is strengthening the internal monitoring of attendance for intermediate outcome 1, to support this at evaluation points. All targets will be subject to review at midline.

Supplementary Annex 14: Number of Records from Stage 1 (Evaluation Surveys) to Complete Records (Stage 5)

		Stage 1: Sample from Evaluation Survey Results	Stage 2: Learning Assessments Collected (based on completed School Visit Forms)				Stage 3: Electronic Data Capture Merge		Stage 4: SeGRA/SeGMA Merge		Stage 5: Evaluation Survey Merge		Complete Cases	
		Count	Students from evaluation survey	Replacements	Unable to be replaced	Total	Dropped/ Missing	Viable	Missing SeGRA and/or SeGMA	Contain all learning assessments	Unable to Merge*	Successfully Matched and Merged	n	% (based on original sample #s)
Damot Pulasa	Grade 4	60	60	0	0	60	2	58			0	58	58	96.67%
	Grade 6	60	57	3	0	60	2	58			4	50	54	90.00%
	Grade 8	60	59	1	0	60	4	56	0	56	4	48	52	86.67%
Damot Sore	Grade 4	60	57	3	0	60	2	58			10	38	48	80.00%
	Grade 6	60	58	2	0	60	3	57			5	47	52	86.67%
	Grade 8	60	57	3	0	60	6	54	1	53	9	36	45	75.00%
Damot Woide	Grade 4	80	80	0	2	78	3	75			3	69	72	90.00%
	Grade 6	80	79	1	0	80	8	72			5	62	67	83.75%
	Grade 8	80	76	4	0	80	11	69	0	69	6	57	63	78.75%
Kindo Koisha	Grade 4	100	95	5	0	100	2	98			11	76	87	87.00%
	Grade 6	100	93	7	0	100	10	90			10	70	80	80.00%
	Grade 8	96	92	4	2	96	11	85	2	83	8	69	77	80.21%
Ofa	Grade 4	295	278	14	0	292	18	274			53	168	221	74.92%
	Grade 6	294	285	9	9	294	40	254			34	186	220	74.83%
	Grade 8	273	261	13	1	273	25	248	2	246	27	194	221	80.95%
Grand Total		**1758**	**1687**	**69**	**14**	**1753**	**147**	**1606**			**189**	**1228**	**1417**	**80.60%**

Supplementary Annex 15: Girls' Characteristics Scale and Barriers to Education Scales[35]

Supplementary Table 1. Girls' Characteristics Scale

Scale	Items	Range	Reverse Coded[36]
Gender-perceptions scale	7 items	0 (disagree a lot) to 3 (agree a lot)	4 items were reverse coded
Well-being scale	3 items	0 (responded 'no' to all three items) to 3 (responded 'yes' to all three items)	n/a
Discipline scale	7 items	0 (responded 'no' to all seven items) to 7 (responded 'yes' to all seven items)	n/a
Household-support scale	4 items	0 (never) to 2 (always)	n/a
Life-skills scale	Grade 4: 12 items Grade 6 and 8: 18 items	0 (strongly disagree) to 4 (strongly agree)	n/a
Decision-making scale	Grade 4: 6 items Grade 6 and 8: 7 items	0 (my family makes decisions for me) to 3 (I decide)	n/a
Self-esteem scale	Grade 4, 6 and 8: 10 items	0 (not true) to 3 (completely true)	5 items were reverse coded

Supplementary Table 2. Barriers to Education Scale

Scale	Sources	Items	Range	Reverse Coded
Gender-perceptions scale	Classroom teachers survey, woreda staff survey, boys student survey,	7 items	0 (disagree a lot) to 3 (agree a lot)	4 items were reverse coded

[35] This annex was updated in May 2019.

[36] Note that higher scores on each scale are the more desirable scores—for example, a higher gender perception score corresponds to a more positive gender perception. To facilitate this, some items are reverse coded, such as disagreeing with the statement corresponded with a more positive gender perception.

Scale	Sources	Items	Range	Reverse Coded
	parent/caregiver survey			
Attitudes towards girls' education scale	Classroom teachers survey, woreda staff survey, parent/caregiver survey	6 items	0 (disagree a lot) to 3 (agree a lot)	2 items were reverse coded
Classroom-management scale	Classroom observation	10 items	0 (never) to 3 (always)	n/a
Community-gender perceptions scale	Woreda staff survey	16 items	0 (disagree a lot) to 3 (agree a lot)	n/a
Support for girls' education scale	Woreda staff survey	8 items	0 (never) to 3 (always)	n/a
Support for secondary education scale	Parent/caregiver survey	7 items	Varies	Scale computed by adding six items but subtracting one (difficulty in affording girls education)
In-class practice scale	Classroom teachers survey	3 items	0 (disagree a lot) to 3 (agree a lot)	n/a
School and community empowerment scale	Classroom teachers survey	3 items	0 (never) to 3 (always)	n/a
English language competence and confidence scale	Classroom teachers survey	4 items	3 items - 0 (disagree a lot) to 3 (agree a lot); 1 item—0 (none of the class) to 3 (all of the class)	n/a
Support for gender policy and mainstreaming scale	School Audit	10 items	0 (no), 1 (partially), 2 (yes)	n/a

Scale	Sources	Items	Range	Reverse Coded
Support for Girls' Education Advisory Committee (GEAC) scale	School Audit	10 items	0 (no), 1 (partially), 2 (yes)	n/a
Gender-sensitive curriculum design and implementation scale	School Audit	11 items	0 (no), 1 (partially), 2 (yes)	n/a
Girls' support mechanisms scale	School Audit	5 items	0 (no), 1 (partially), 2 (yes)	n/a
Community support scale	School Audit	8 items	0 (no), 1 (partially), 2 (yes)	n/a
School gender friendliness scale	School Audit	11 items	0 (no), 1 (partially), 2 (yes)	n/a
Support for Girls' Clubs scale	School Audit	5 items	0 (no), 1 (partially), 2 (yes)	n/a

All Evaluation Survey Results

Survey Results_ALL.xlsx

This excel file includes all survey demographics, survey scales, and item results for the STAGES baseline study from all evaluation surveys. This includes the girls student survey, boys student survey, classroom teacher survey, classroom observation, school audit, and woreda official survey.

Supplementary Annex 16: Expanded Qualitative Findings on Girls' Barriers to Education

This annex provides more detailed, in-depth findings and reflections from the qualitative analysis of the STAGES baseline study's key informant interviews and focus group discussions. These findings are a compilation of the qualitative data. More specifically, it highlights the key barriers to girls' education in the Wolaita that emerged during the study. The majority of the barriers discussed within the qualitative data were situated within the context of girls' families and communities.

While the degree to which high burden of household chores impedes a girls' ability to attend school and thrive in her studies may vary, it was clear across the focus group discussions and key informant interviews that this continues to be a major challenge for primary school girls in the Wolaita zone that often increases with a girls' age. Within the qualitative data, gendered division of labor within the family or household was named as one of the main barriers to girls' education across respondent types. Both female teachers and parents mentioned that girls were likely to have an obligation to complete household chores in addition to attending school and studying. Parents across focus groups noted parental attitudes and behaviors within the Wolaita zone that supported an unequal distribution of household chores between female and male children, with the bulk of responsibilities falling to girls. Parents in several focus groups also mentioned that they believed their girls were sometimes too tired after completing their morning duties of cooking, fetching water, and other to attend school regularly and that household chores interrupted their daughters' studies. Moreover, parents noted a connection between their daughters' low academic performance and inability to study at home due to household chores. The unequal distribution of household responsibilities was echoed in male and female student focus groups with several male students citing their sisters were responsible for a higher chore burden. Tardiness was also associated with girls' household duties—with several key informants stated that it was common for girls to be late to school or unable to study due to household chores. These findings align with the key learnings from the project activities and evaluation of Link's previous GEC1 in the Wolaita zone. This also fits within the findings of a recent country-wide evidence review on adolescent girls, where Stavropoulou and Gupta-Archer (2017) found that in Ethiopia, "parental pressures to contribute to household chores or earn an income affect girls' school attendance and study time, with nearly half of young adolescent girls having to spend a minimum of 28 hours weekly on housework compared with 35 percent of boys. These pressures are acuter in low-income households and households with younger siblings and less educated parents."[37]

It's also important to note, given the commonly accepted gendered division of labor and responsibilities within the household setting, within the discussions mothers were often discussed as playing a critical role in preventing girls from attending school to assist with household chores. Both parents and female students described instances of mothers keeping daughters at home to complete chores or mothers getting upset when daughters would study after school—instead of doing chores. One parent described how female children are unable to study because they expected to do even more than their mothers in the household. In addition to daily chores, several female students reported having to miss school to look after a sick family member or younger siblings. Although most respondents cited household chores as a

[37] Maria Stavropoulou and Nandini Gupta-Archer, *Adolescent Girls' Capacities in Ethiopia: The State of the Evidence (Gender and Adolescence: Global Evidence, December 2017)* https://www.gage.odi.org/sites/default/files/2018-02/Ethiopia%20Capabilities%20Report.pdf, V.

barrier to girls' education, several girls mentioned that their families would not keep them from attending school to complete chores and that education was valued in their family.

Key informant interview responses highlighted girls' participation in income-generating activity having a negative impact on girls' access to education and attendance, especially for girls as they entered secondary school as well as girls in living in a high poverty context. Key informants also noted girls' participation in income generating activities tended to increase as girls entered secondary school. More specifically, respondents across FGDs linked income-generating activity with girls' absenteeism, such as. girls miss school to participate in income generating activities for their families—with girls' participation in market days cited as the most common income generating activity across respondents. Several teachers stated that girls engaged in income-generating activities were likely to miss school on market days.

Early marriage—and its association with high levels of school drop-out—was another major barrier to girls' education noted by across qualitative respondents, with numerous key informants suggesting that the practice was a widespread phenomenon throughout the Wolaita zone. Despite the legal minimum age of marriage being 18 within Ethiopia, participants in the upper primary female students focus group stated that most girls in their communities are married between the ages of about 15 and 17 years old. Again, these findings are unsurprising when taking into consideration the broader literature and evidence of the prevalence of early marriage in the Ethiopian context as noted by Stavropoulou and Gupta-Archer:

> *"The latest Ethiopia Demographic and Health Survey (EDHS) data shows that 40 percent of women aged 20 to 24 married before the age of 18 and 6 percent of girls aged 15 to 19 married before age 15. On the other hand, men on average marry seven years later than women. Poverty, lack of education and discriminatory social norms about girls' virginity are the key drivers of the practice."*[38]

Early marriage was widely discussed by parents in focus groups as a serious barrier to girls' education given that community social norms often prioritized marriage over a girls' schooling—with girls from families with limited financial resources being the most affected. For example, several parents stated that parents in dire financial situations often subjected their daughters to early marriage to receive a dowry. Parents also alluded that girls themselves also viewed marriage as a means to improve their status, stating that it was common for girls to have perceptions that marriage would make their lives easier as opposed to education.

The most direct barrier to education related to early marriage was girls' drop-out.[39] Several female students and parents mentioned that if a girl was married, it was unlikely her husband would allow her to attend school because of her responsibilities to her family unless he was educated; for example, one parent stated, "no husband sends his wife to school." However, a handful of upper primary female students who participated in the focus groups noted that they were married, were continuing in their schooling, and stated their husbands were supportive of their education.

[38] Maria Stavropoulou and Nandini Gupta-Archer, *Adolescent Girls' Capacities in Ethiopia: The State of the Evidence (Gender and Adolescence: Global Evidence, December 2017)* https://www.gage.odi.org/sites/default/files/2018-02/Ethiopia%20Capabilities%20Report.pdf, IV.

[39] Upper primary male students commented that it was likely for girls to drop out of school due to early marriage.

It's also interesting to note that within this context**, childbearing** was for the most part discussed within the context of marriage. Stavropoulou and Gupta-Archer also highlight how high rates of early marriage and norms are "favoring early childbearing, 13 percent of 17-year-old girls and 28 percent of 19-year-olds have already begun childbearing in 2016. Overall, rural residence, limited education, and poverty increase the likelihood of early motherhood. Early pregnancy and childbirth is a leading cause of mortality amount Ethiopian adolescent girls."[40]

Moreover, the perceptions of it and how likely girls and young women who are pregnant or have children varied across respondent types. Most participants in the upper primary female focus group noted that many girls do not continue to go to school after they have children because of their domestic responsibilities of child upbringing and home care; however, some did note that if she is very committed and marries an educated man, he will be more willing to let her continue to learn. Moreover, several key informant interview respondents at the Zone level noted that some young mothers continue in the regular school system—but the degree to which specific schools or communities provided a supportive and flexible approach to young mothers continued education appears to vary considerably, with a range of more lenient to punitive approach to mothers' tardiness or absenteeism.

Girls' struggles to consistently engage in their education due to menstruation was discussed in varying degrees by respondents. Key informants, teachers, and parents all pointed to menstruation in their discussions of barriers to girls' education. Although some reported a decrease in menstruation related challenges since the provision of sanitary pads provided by Link. One male key informant mentioned that although there had been much progress around gender issues, "…much is still left with regards to gender separated toilets and provision of sanitary towels without the support of Link." Reports of absences due to menstruation were mixed among female students. Some female students reported regularly missing school while others stated they planned accordingly for their monthly menstruation. One upper primary female student described how she and her classmates continued attending class during their menstruation."

The majority of respondent types cited rural to urban as well as international migration—especially at the end of primary school and the beginning of secondary- as a barrier to girls' education as well as a mechanism to improve financial status. Migration of girls from rural to urban areas as well as abroad Arab countries for marriage or work were frequently discussed by respondents. According to one male key informant, "The rate of migration is ever increasing. Since they [girls] see others in their villages coming with money, new clothes, and others, they are distracted and obviously become interested in more to other areas outside of the Wolaita zone." Several key informants stated that migration was most common among girls at the end of primary school or towards the beginning of secondary school. Subsequent evaluation points will include further investigation into the role that migration plays in girls' access to education.[41]

Key informants more readily identified by name social and cultural norms rooted in gender inequality as a major barrier to girls' education than the rest of respondents. Responses included mention of prioritization of boys' education, lack of utility given to educating girls, and the devaluing of women or girls in society. While teachers and parents did not identify gender inequality by name, several

[40] Ibid.
[41] Within migration, an additional area of consideration includes the role and impact of child trafficking. For a summary of evidence on labor and sex trafficking of adolescent girls in Ethiopia, please see the *Adolescent Girls' Capacities in Ethiopia: The State of the Evidence* report.

respondents described difficulties girls face in navigating traditional cultural norms and accessing education. Parents mentioned how preference of schooling was given to male children because girls have the option to better themselves through marriage. Male students also described how girls were more likely to prefer marriage than education and all of its "worries." Respondents in one parent FGD described how education was more important for males because without opportunities created by education their sons would face the dangers of migration or become a nuisance to their community.[42] Key- informants also mentioned that instances of gender inequality were exacerbated in times of financial hardship with parents choosing to use their limited resources to educate sons over daughters. Several key informants also stated that communities were generally supportive of primary education but the utility of secondary education was often questioned. One male respondent from a school management key informant interview explained further saying,

> *The community's perception on the value of secondary education is even low. They think that completion of primary education is adequate and tend to encourage their engagement in employment or other livelihood activities [rather] than secondary education. Especially since there are a number of youth graduated from universities, colleges and TVETs, and those who finished high school, who have no job. It is serving as a discouraging factor.*

Despite parents' overwhelming admission of support for girls' education, even less utility was given to girls attending secondary school by the larger community. Girls were also most likely to report receiving support from male family members than females. Statements from female teachers also highlighted the role of fathers as gatekeepers with multiple suggestions that fathers' support created the greatest likelihood for girls to access education. Additionally, both female teachers and students referenced the importance of having an educated male family member with several mentions of educated fathers, brothers, and husbands. Although girls reported receiving varying degrees of support from their husbands, brothers, and fathers, numerous female students stated they had received little to no support from their community regarding their education. This aligns with other qualitative findings in Ethiopia, including a study in Amhara which highlighted the "significant role of supportive fathers and brothers as well as husbands in enabling unmarried and married girls to continue their education."[43]

Although parents and teachers stated that there had been progress in cultural norms around girls' education, more work needs to be done to change community attitudes.

While limited data and mention of SRGBV directly were noted in the qualitative data, this should not be interpreted as a lack of SRGBV issues within the context, and additional follow-up is still required. For example, the MOE's (2014) National Gender Strategy for Education and Training Sector notes key challenges related to SRGBV as including the "prevalence of corporal punishment, psychological and sexual violence and abuse from peers, teachers, and others are common setbacks for girls' well-being, school attendance, and educational attainment."[44] Few respondents mentioned safety as concern regarding girls' access to education. A handful of respondents stated that safety used to be more

[42] However, the majority of male students did not believe they were given any special preference when it came to accessing education. In fact, several male students stated that female students received extra support for school and that support should be expanded to include boys.

[43] Maria Stavropoulou and Nandini Gupta-Archer, *Adolescent Girls' Capacities in Ethiopia: The State of the Evidence (Gender and Adolescence: Global Evidence, December 2017)* https://www.gage.odi.org/sites/default/files/2018-02/Ethiopia%20Capabilities%20Report.pdf, 36.

[44] The Federal Democratic Republic of Ethiopia, Ministry of Education, *Gender Strategy for the Education and Training Sector*, (s.l.: October 2014), http://www.moe.gov.et/documents/20182/36315/GENDER+STRATEGY.pdf/b9e68a15-bc9e-4930-a5d2-1c1981ca264c, 32.

of an issue in the past but concerns about safety had dwindled in recent years. Despite little mention of safety concerns by FGD participants, several key informants stated their offices still received reports of safety-related issues. Given the sensitive nature of this topic, the external evaluators recommend a more purposeful, sensitive study could be using tools such as the Education in Crisis and Conflict Network's ECCN Safer-Learning Environment Toolkit.[45]

Abduction was mentioned by several teachers as an issue when discussing barriers to girls' education. Participants described several instances of female students being harassed by their male counterparts citing occasions when the harassment resulted in the male student abducting the female student. One teacher recalled a case of a 7th-grade female student who was the top performer in her class being abducted by a male classmate, which ended her studies. Parents and students did not discuss abduction widely although one parent did mention that abduction was no longer a threat to girls. Similarly, a key informant stated that compared with the past reports of girls being targeted were almost nonexistent. Questions explicitly around abduction will be included in subsequent evaluation points to investigate this issue further.[46]

While the distance to secondary schools was noted as a challenge by key informants, this not discussed as a major concern at the primary level within FGDs. Little was mentioned in terms of distance to school as a barrier to girls' education. The majority of FGD respondents including female students reported feeling safe on their way to school. Parents stated that girls used to have issues with safety on their walks to school, but these had decreased in recent years. It should be noted, however, that the FGDs were situated more within the primary school context. KIIs, on the other hand, were more likely to note the limited number of secondary schools and the associated geographic distance as challenges for girls' education as girls move into grade 9.

Lack of resources, accommodation or specialized training to support girls with disabilities was also noted. Teachers mentioned only being able to provide limited accommodations for students with disabilities some of which included teaching aids, additional time, and advice. One female teacher described an instance with a former student:

> *There was one deaf female student in my class last year. She somehow completed a year under the all-encompassing program for there was no trained personnel in special needs education and joined Shanto elementary school where there are teachers trained in special needs education.*

Qualitative respondents mentioned girls with several categories of disabilities within their schools and communities. The most common disabilities mentioned were students who had varying degrees of physical disabilities, deaf or hard of hearing and blind or low vision.

While the MOE recognizes the "multiple disparities against girls with special educational needs,"[47] there continues to be a gap in key evidence and research on the topic in Ethiopia.[48] Key informants also described the lack of resources available for students with disabilities, "…schools in rural areas have no means to

[45] https://eccnetwork.net/resources/sle-qualitative-toolkit/
[46] Additionally, the strategy notes abduction, along with early marriage and female genital mutilation—as harmful traditional practices impeding girls' education.
[47] The Federal Democratic Republic of Ethiopia, Ministry of Education, *Gender Strategy for the Education and Training Sector*, (s.l.: October 2014), http://www.moe.gov.et/documents/20182/36315/GENDER+STRATEGY.pdf/b9e68a15-bc9e-4930-a5d2-1c1981ca264c 16.
[48] Gender & Adolescence Global Evidence. (2017) Adolescent girls' capabilities in Ethiopia: The state of the evidence Maria Stavropoulou and Nandini Gupta-Archer December 2017. P.VI https://www.gage.odi.org/sites/default/files/2018-02/Ethiopia%20Capabilities%20Report.pdf

teach the ones with critical disability issues. Some children with minor to moderate level of disability, are attending school, without small, often insignificant supports from us." The majority of key informants agreed that additional resources needed to be dedicated to serving children with disabilities. One zone-level education official mentioned that although "there are some improvements related to discussing the needs of children with disabilities in the school system, a lot remains.

Girls who are orphans were noted as less likely to attend secondary school and to receive limited, targeted interventions or support. Respondents acknowledged the existence of girls who were orphaned and attended school but did not describe their situations in detail. One female teacher mentioned that orphaned students existed in their classes, but not much was done to support them outside of Link efforts, "There are parent-less children who had been supported by Link. Nothing much has been done to identify parent-less children this year, and there could be more if investigated." Several key informants stated that girls who did not have parents were less likely to attend secondary school.

Lack of resources and poor school infrastructure were also commonly cited by parents and teachers as barriers to girls' education. The majority of teachers stated that school infrastructure and resources were lacking. Teachers cited inadequate facilities, lack of learning or teaching materials, and inconsistent funding as major challenges to providing quality education. Several teachers also mentioned that they often arrived at school exhausted because they had to travel great distances due to lack of accommodations. Numerous teachers also cited the absence of restrooms as an inadequacy many teachers and students faced. Insufficient school facilities and lack of learning materials were again cited by female and male upper primary students. Some female students described not attending school due to shortages of exercise books and pens. Key informants from the local governments also cited lack of "educational supplies" as a key barrier to girls' consistent engagement in education.

Supplementary Annex 17: Additional Key Outcome Tables and Figures

Supplementary Table 3. Lists of Literacy and Numeracy Subtasks by Language and Grade Assessed

Type of Subtask	EGRA Subtasks	Language of Subtask	Grade 4	Grade 6	Grade 8
Literacy Subtasks (Wolayttatto)	Letter Sound Identification	Wolayttatto	√	√	
Literacy Subtasks (Wolayttatto)	Familiar Word Reading	Wolayttatto	√	√	
Literacy Subtasks (Wolayttatto)	Invented Word Reading	Wolayttatto	√	√	
Literacy Subtasks (Wolayttatto)	Reading Passage	Wolayttatto	√	√	
Literacy Subtasks (Wolayttatto)	Reading Comprehension	Wolayttatto	√	√	
Literacy Subtasks (English)	Letter Sound Identification	English	√	√	
Literacy Subtasks (English)	Familiar Word Reading	English	√	√	√
Literacy Subtasks (English)	Invented Word Reading	English	√	√	√
Literacy Subtasks (English)	Reading Passage	English	√	√	√
Literacy Subtasks (English)	Reading Comprehension	English	√	√	√
Literacy Subtasks (English)	SeGRA: Reading Passage	English			√
Literacy Subtasks (English)	SeGRA: Fill in the blanks	English			√
Literacy Subtasks (English)	SeGRA: Revising Sentences	English			√
Math Subtasks	Number Identification	n/a	√	√	
Math Subtasks	Quantity discrimination	n/a	√	√	
Math Subtasks	Missing Numbers	n/a	√	√	

Type of Subtask	EGRA Subtasks	Language of Subtask	Grade 4	Grade 6	Grade 8
Math Subtasks	Addition	n/a	√	√	√
Math Subtasks	Subtraction	n/a	√	√	√
Math Subtasks	Word Problems	n/a		√	√
Math Subtasks	SeGMA: Geometry and Measurement	n/a			√
Math Subtasks	SeGMA: Fractions	n/a			√
Math Subtasks	SeGMA: Multiplication	n/a			√

Supplementary Table 4. Item Total Correlations and Cronbach's Alpha for Literacy Subtasks, by Language and Grade

Literacy Subtasks	Grade 4		Grade 6	
	Corrected Item-Total Correlation	*Cronbach's Alpha if Item Deleted*	*Corrected Item-Total Correlation*	*Cronbach's Alpha if Item Deleted*
English EGRA				
Letter Sounds - English	.662	.954	.670	.950
Familiar Words - English	.880	.945	.911	.940
Nonwords - English	.840	.947	.812	.945
Reading Comprehension-English	.661	.954	.687	.950
Oral Reading Fluency - English	.861	.946	.890	.941
Wolayttatto EGRA				
Letter Sounds - Wolayttatto	.674	.953	.517	.955
Familiar Words - Wolayttatto	.902	.944	.888	.941
Nonwords - Wolayttatto	.894	.944	.858	.942
Reading Comprehension - Wolayttatto	.837	.947	.881	.956
Oral Reading Fluency - Wolayttatto	.829	.949	.863	.958

Literacy Aggregate Scores by Group (Histograms)

Figure 1. Literacy Aggregate scores by group (histograms), Grade 4 treatment

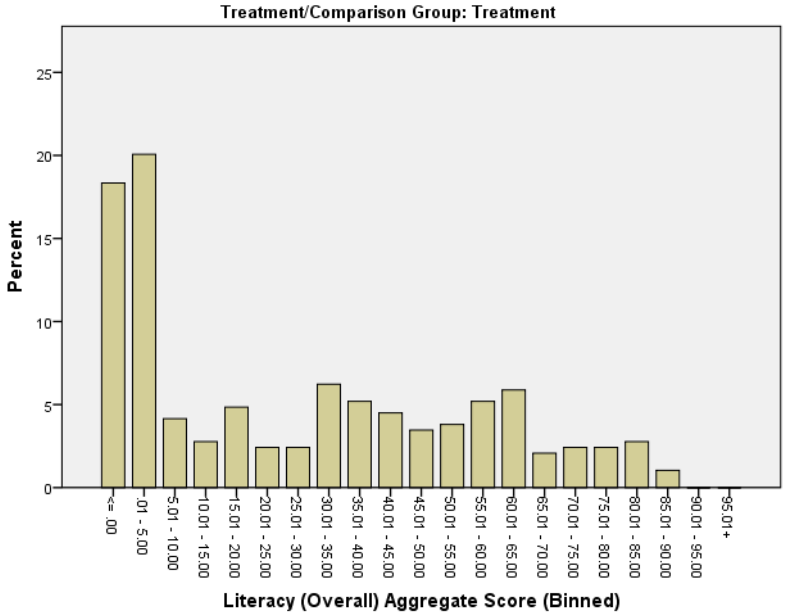


Figure 2. Literacy Aggregate scores by group (histograms), Grade 4 comparison

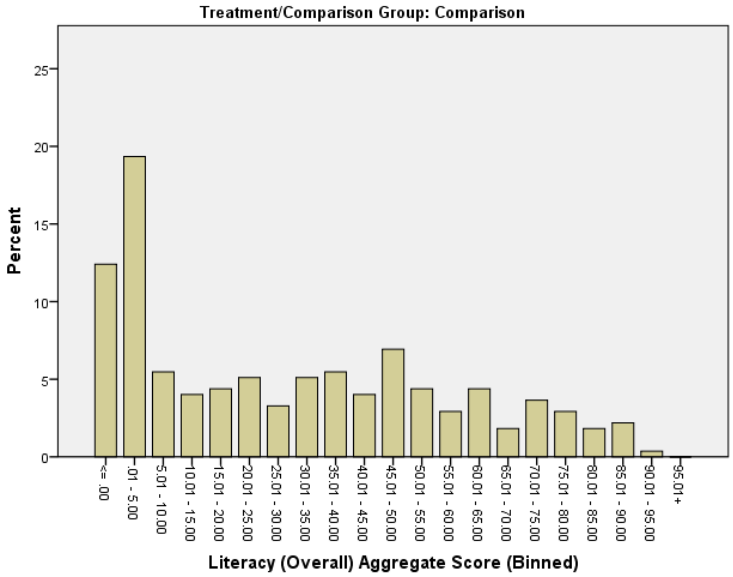


Figure 3. Literacy Aggregate scores by group (histograms), Grade 6 treatment

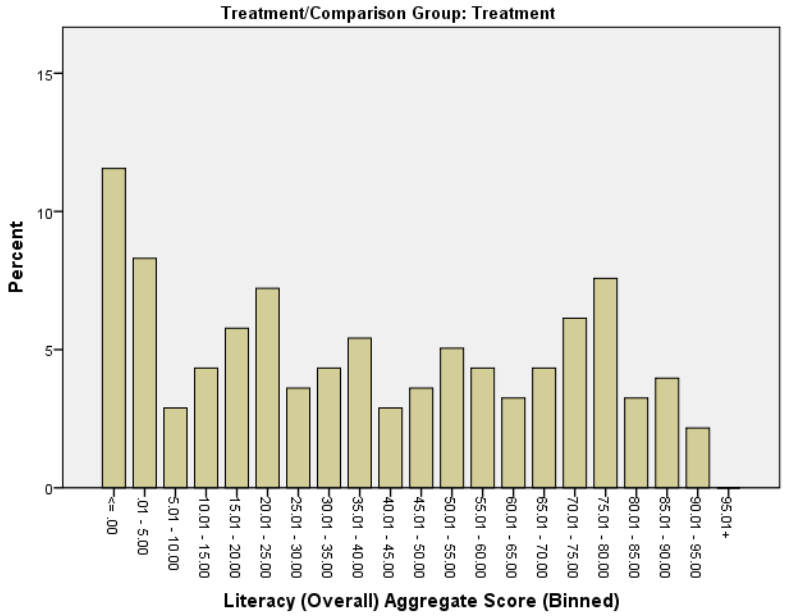


Figure 4. Literacy Aggregate scores by group (histograms), Grade 6 comparison

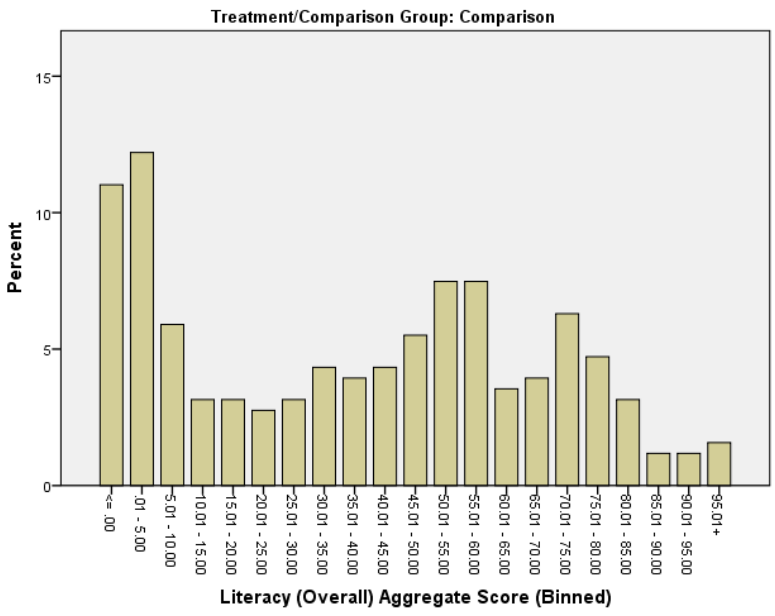


Figure 5. Literacy Aggregate scores by group (histograms), Grade 8 treatment

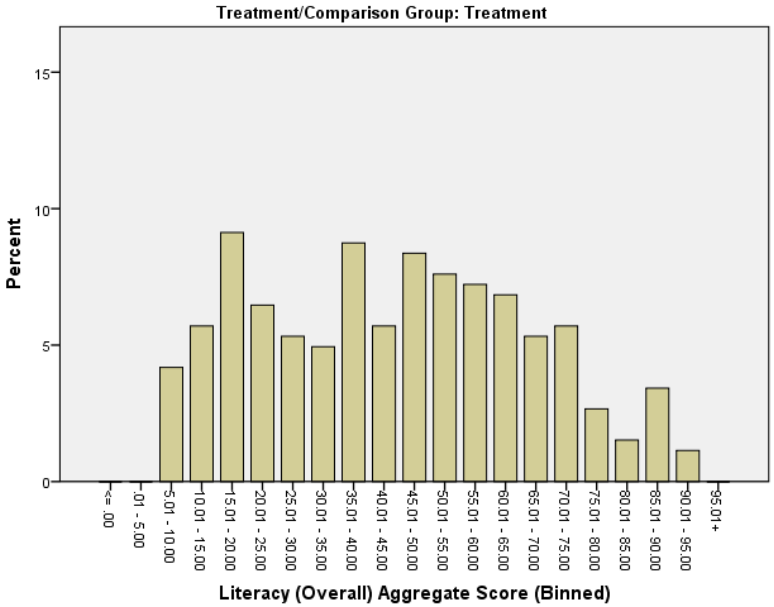
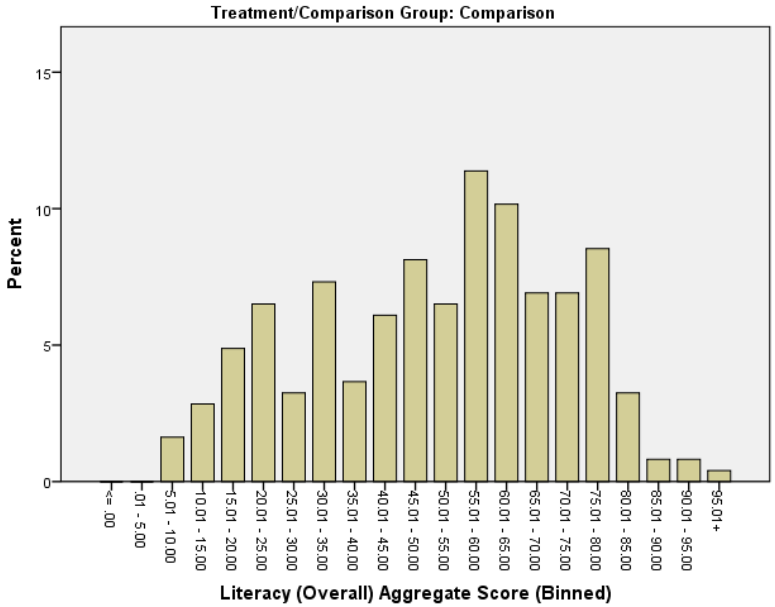


Figure 6. Literacy Aggregate scores by group (histograms), Grade 8 comparison



Numeracy Aggregate Scores by Group (Histograms)

Figure 7. Numeracy Aggregate scores by group (histograms), Grade 4 treatment

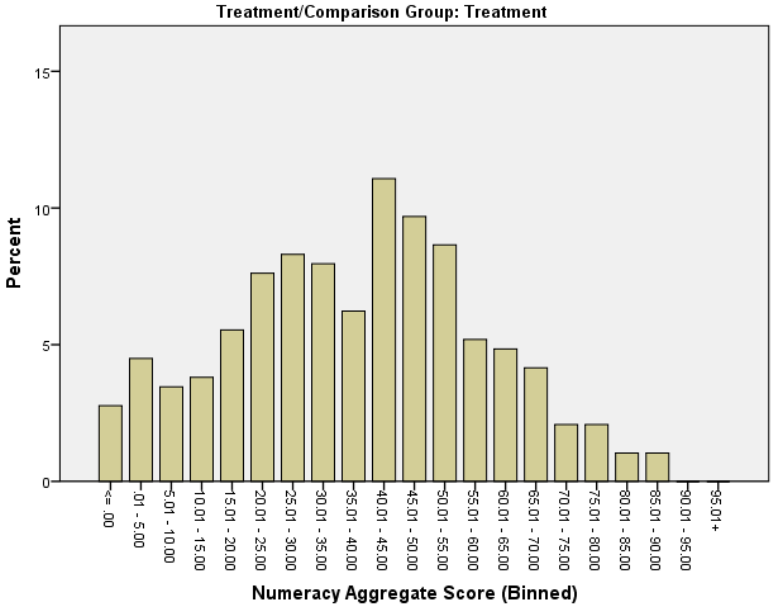


Figure 8. Numeracy Aggregate scores by group (histograms), Grade 4 comparison

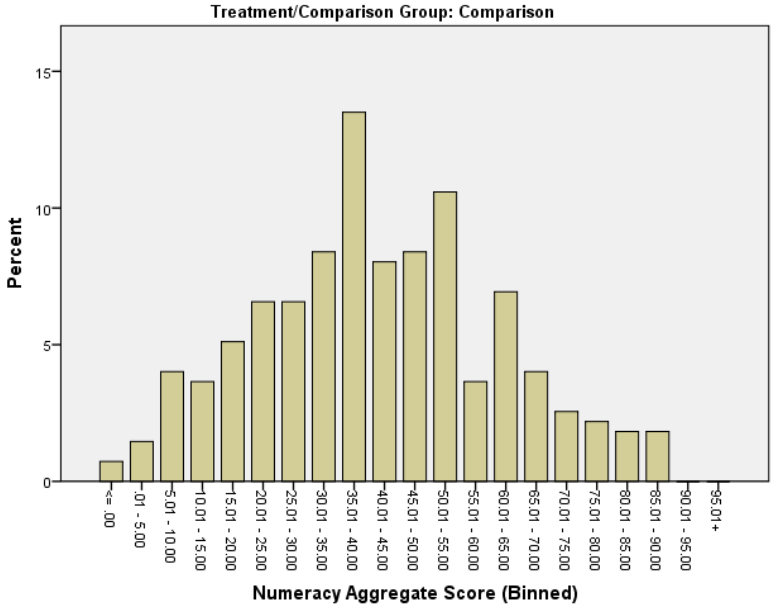


Figure 9. Numeracy Aggregate scores by group (histograms), Grade 6 treatment

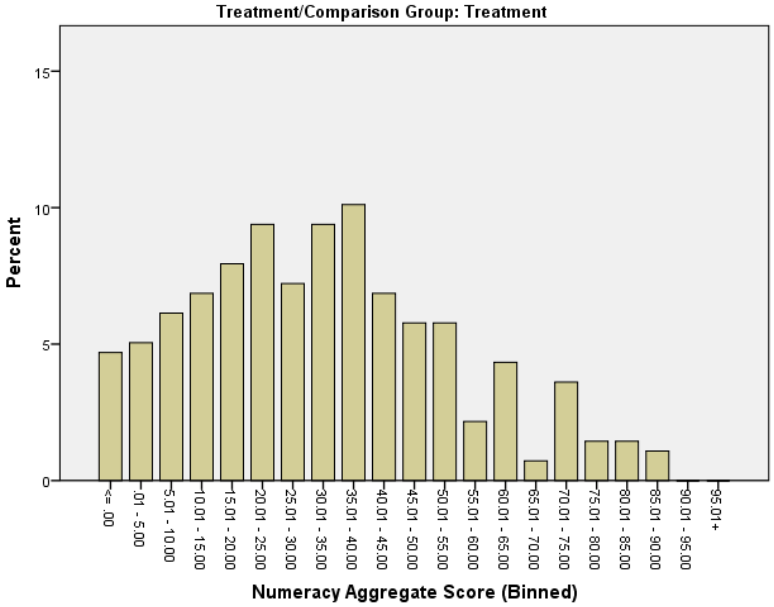


Figure 10. Numeracy Aggregate scores by group (histograms), Grade 6 comparison

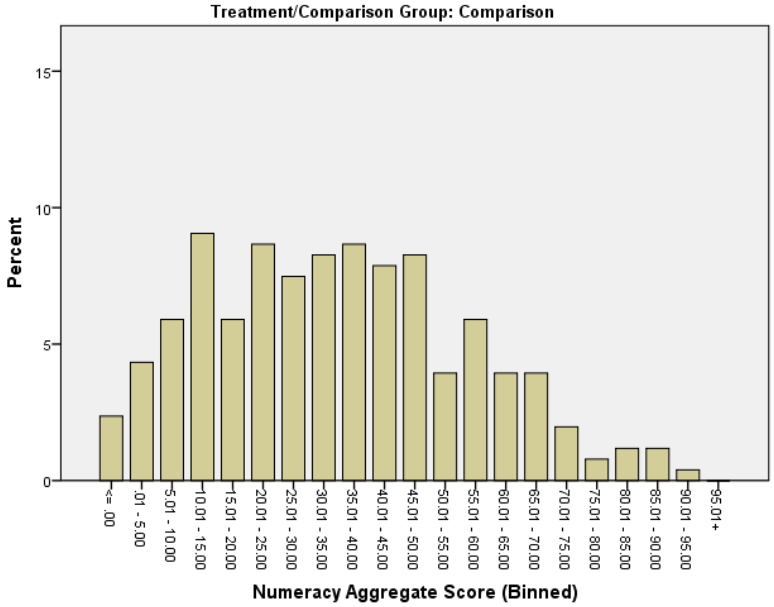


Figure 11. Numeracy Aggregate scores by group (histograms), Grade 8 treatment

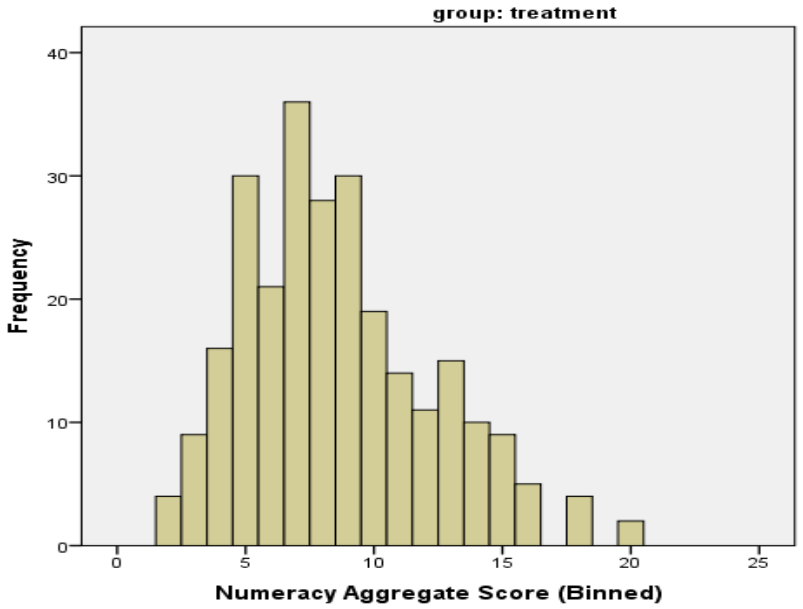
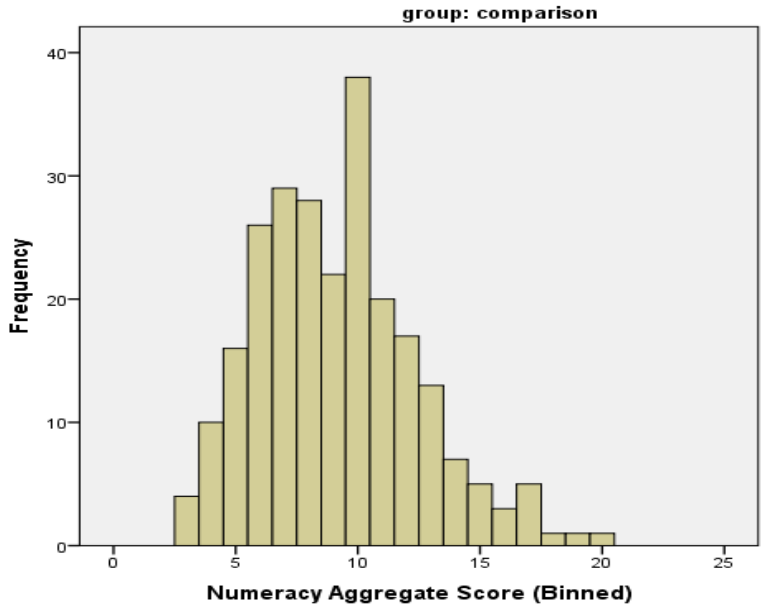


Figure 12. Numeracy Aggregate scores by group (histograms), Grade 8 comparison



Supplementary Annex 18. Intermediate Outcome Analysis Index and Source Tables[49]

Supplementary Table 1. Perceptions of access index

Original Item	Coding used in Index	Scoring for Perception of Access Index
Time to get to school (How long does it usually take you to get to school)?	1=30 minutes or less 0=more than 30 minutes	Count '1' value
Do you feel safe traveling to and from school?	1=yes 0=no	Count '1' value
Is it reasonably easy for you to get to school and back?	1=yes 0=no	Count '1' value
Do you feel that the distance to your school is: Very close; Somewhat close; Somewhat far; Very far	1=very close or somewhat close 0=very far or somewhat far	Count '1' value
Perception of Access Index		4=excellent perception of access 3=good perception of access 2=adequate perception of access 1=poor perception of access 0=very poor perception of access

Supplementary Table 2. Source Data for IO 3: Quality of Teaching

School Level	Index	Survey Item	Source	Scoring
IO Indicator 3a: Percentage of teachers with improved subject matter knowledge				
Secondary (required)		In the past year, have you received training in your subject (math, science, language)?	Classroom teacher survey	Yes=1 No=0

[49] This annex was revised in May 2019.

School Level	Index	Survey Item	Source	Scoring
Primary (optional)	n/a	In the past year, have you received training in your subject (math, science, language)?	Classroom teacher survey	Yes=1 No=0
IO Indicator 3b: Percentage of teachers with improved methodology				
Secondary (required)		In the past year, have you received training in gender-responsive pedagogy?	Classroom teacher survey	Yes=1 No=0
Primary (optional)	n/a	In the past year, have you received training in gender-responsive pedagogy?	Classroom teacher survey	Yes=1 No=0
IO Indicator 3c: Percentage of teachers with improved classroom management				
Secondary (required)		In the past year, have you received training in classroom management?	Classroom teacher survey	Yes=1 No=0
	Classroom Management Index	Teacher uses appropriate teaching materials or aides	Classroom Observation	3=always 2=frequently 1=sometimes 0=never include in index mean
		All students know what to do in a given task	Classroom Observation	3=always 2=frequently 1=sometimes 0=never include in index mean
		Teacher's attention is on students learning	Classroom Observation	3=always 2=frequently 1=sometimes 0=never include in index mean
		All students are on-task	Classroom Observation	3=always 2=frequently 1=sometimes 0=never include in index mean

School Level	Index	Survey Item	Source	Scoring
		Students raise their hands to talk	Classroom Observation	3=always 2=frequently 1=sometimes 0=never include in index mean
		Students remain in their seats unless permitted to move	Classroom Observation	3=always 2=frequently 1=sometimes 0=never include in index mean
		Students are quiet while the teacher or someone else is talking	Classroom Observation	3=always 2=frequently 1=sometimes 0=never include in index mean
		Girls are disciplined physically in class.	Classroom Observation	3=always 2=frequently 1=sometimes 0=never include in index mean
		Boys are disciplined physically in class.	Classroom Observation	3=always 2=frequently 1=sometimes 0=never include in index mean
		Girls and boys have EQUAL access to desks, learning materials, and other materials, including the same amount of sharing of books and desks	Classroom Observation	3=always 2=frequently 1=sometimes 0=never include in index mean
	Classroom Management Index			Average score across 10 observation items; Index retains the same scale as individual items but represents the average 3=always

School Level	Index	Survey Item	Source	Scoring
				2=frequently 1=sometimes 0=never

Supplementary Table 3. Source Data for IO 4: Positive Community Attitudinal Change

School Level	Index	Survey Item	Source	Scoring
IO Indicator 4a: Teachers reporting positive changes in gender perceptions and gender-sensitive teaching				
Secondary (required), Primary (optional)	Gender Perception Index	Education is more important for boys than for girls (REVERSED)	Classroom teacher survey	0=agree a lot, 1=agree, 2=disagree, 3=disagree a lot include in index mean
		A woman's role is to do household jobs and raise children (REVERSED)	Classroom teacher survey	0=agree a lot, 1=agree, 2=disagree, 3=disagree a lot include in index mean
		Men should share household duties	Classroom teacher survey	0=disagree a lot, 1=disagree, 2=agree, 3=agree a lot Include in index mean
		Boys are more naturally skilled than girls at reading and writing (REVERSED)	Classroom teacher survey	0=agree a lot, 1=agree, 2=disagree, 3=disagree a lot
		Boys are more naturally skilled than girls at mathematics (REVERSED)	Classroom teacher survey	0=agree a lot, 1=agree, 2=disagree, 3=disagree a lot
	Gender Perception Index			Average across five items; resulting score binned by mean, one standard deviation above and one standard deviation below. Count of cases above the mean (2.67 on a three-point scale) 3=strong and positive gender perception 2=positive gender perception

School Level	Index	Survey Item	Source	Scoring
				1=negative gender perception 0=strong and negative gender perception
IO Indicator 4b: Teachers and school directors' reporting positive attitudinal change towards girls' education and learning (case studies approach)				
Secondary (required)	n/a	n/a at baseline		
Primary (optional)	n/a	n/a at baseline		
IO Indicator 4c: Percentage of parents, disaggregated by gender, reporting greater support for secondary education, especially for girls				
Primary (optional)- parents of G4 and G6 girls	Support for Secondary School Index	Quality of teaching that girl receives is: very good	Parent/Caregiver survey	1=selected
		In last 12 months, quality of teaching that girl receives has changed: improved	Parent/Caregiver survey	1=selected
		Strongly agree that even when funds are limited, it is worth investing in girls' education	Parent/Caregiver survey	1=selected
		Very safe for girls to travel to schools in this area	Parent/Caregiver survey	1=selected
		Level of schooling you would like girl to receive	Parent/Caregiver survey	0=none, 1=primary, 2=lower secondary, 3=upper secondary, 4=college/univ
		Girls in my community do not need hep to get to secondary school	Parent/Caregiver survey	1=selected

School Level	Index	Survey Item	Source	Scoring
		It is not difficult for girls to go to school	Parent/Caregiver survey	1=selected
IO Indicator 4d: Percentage of boys' reporting positive perception of the value of girls' education				
Primary (required)				
Primary (optional)	Gender Perception Index	It is important for girls to go to school	Boys student survey	0=disagree a lot, 1=disagree, 2=agree, 3=agree a lot Include in index mean
		Education is more important for boys than for girls (REVERSED)	Boys student survey	0=agree a lot, 1=agree, 2=disagree, 3=disagree a lot include in index mean
		Girls learn the same at school as boys	Boys student survey	0=disagree a lot, 1=disagree, 2=agree, 3=agree a lot Include in index mean
		A woman's role is to do household jobs and raise children (REVERSED)	Boys student survey	0=agree a lot, 1=agree, 2=disagree, 3=disagree a lot include in index mean
		Men should share household duties	Boys student survey	0=disagree a lot, 1=disagree, 2=agree, 3=agree a lot Include in index mean
		Boys are more naturally skilled than girls at reading and writing (REVERSED)	Boys student survey	0=agree a lot, 1=agree, 2=disagree, 3=disagree a lot
		Boys are more naturally skilled than girls at mathematics (REVERSED)	Boys student survey	0=agree a lot, 1=agree, 2=disagree, 3=disagree a lot
Secondary (required)		n/a at baseline		
	Gender Perception Index			Average across seven items; resulting score binned by mean, one SD above and on SD below.

School Level	Index	Survey Item	Source	Scoring
				Count of cases above the mean (2.67 on a three-point scale) 3=strong and positive gender perception 2=positive gender perception 1=negative gender perception 0=strong and negative gender perception

Supplementary Table 4. Source Data for IO 5: Improved well-being and life-skills

School Level	Index	Survey Item	Source	Scoring
IO Indicator 5a: Percentage of girls reporting improved well-being				
Girls (required)		Did you smile or laugh the last time you were at school?	Girls student survey	1=yes 0=no
		Did you learn or do something interesting the last time you were at school?	Girls student survey	1=yes 0=no
		Did you have enough energy to get things done the last time you were at school?	Girls student survey	1=yes 0=no
	Well-being Index			Count of cases responding 'yes' to all three well-being items
Boys (optional)	*Well-being Index*		Boys student survey	Same as girls
IO Indicator 5b: Percentage of girls reporting improved self-esteem due to Social and Emotional Learning (SEL)				
	Self-esteem Index Grades 4, 6, 8	Generally, I am satisfied with myself.	Girls Transition Survey	0=not true, 1=slightly true, 2=mostly true, 3=completely true
		I feel that I have a lot of good qualities.	Girls Transition Survey	0=not true, 1=slightly true, 2=mostly true, 3=completely true

School Level	Index	Survey Item	Source	Scoring
		I can do things as well as most other girls my age.	Girls Transition Survey	0=not true, 1=slightly true, 2=mostly true, 3=completely true
		I feel that I am just as important as anybody else.	Girls Transition Survey	0=not true, 1=slightly true, 2=mostly true, 3=completely true
		I feel positively about myself.	Girls Transition Survey	0=not true, 1=slightly true, 2=mostly true, 3=completely true
		reversed: At times, I think I am no good at all	Girls Transition Survey	0=not true, 1=slightly true, 2=mostly true, 3=completely true
		reversed: I feel I do not have much to be proud of	Girls Transition Survey	0=not true, 1=slightly true, 2=mostly true, 3=completely true
		reversed: I certainly feel useless at times	Girls Transition Survey	0=not true, 1=slightly true, 2=mostly true, 3=completely true
		reversed: I wish I could have more respect for myself.	Girls Transition Survey	0=not true, 1=slightly true, 2=mostly true, 3=completely true
		reversed: I am afraid that I will fail	Girls Transition Survey	0=not true, 1=slightly true, 2=mostly true, 3=completely true
Girls (required)	*Life-skills Index grades 6, 8 (18 items)*	I am able to do things as well as my friends	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I want to do well in school	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I get nervous when I have to read in front of others REVERSED	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean

School Level	Index	Survey Item	Source	Scoring
		I get nervous when I have to do math in front of others REVERSED	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I feel confident answering questions in class	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I can stay focused on a goal despite things getting in the way	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I would like to continue studying or attending school after this year	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I can put a plan in place and stick with it	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I recognize when choices I make today about my studies can affect my life in the future	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I can describe my thoughts to others when I speak	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		If someone does not understand me, I try to find a different way of saying what is on my mind	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean

School Level	Index	Survey Item	Source	Scoring
		When others talk, I pay attention to their body language, gestures, and facial expressions	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I can work well in a group with other people	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		When I have the opportunity, I can organize my peers or friends to do an activity	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I often feel lonely at school REVERSED	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I ask the teacher if I don't understand something	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		When I succeed at school, it is because I worked hard	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		If I do well in a test, it is because I am lucky REVERSED	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
	Life-skills Index—grade 6, 8			Mean of all items 4=strongly agree on life-skills items; 3=agree on life-skills items, 2=neither, 1=disagree on life-skills

School Level	Index	Survey Item	Source	Scoring
				items, 0=strongly disagree on life-skills items
	Life-skills Index grade 4 (12 items)	I am able to do things as well as my friends	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I want to do well in school	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I get nervous when I have to read in front of others REVERSED	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I get nervous when I have to do math in front of others REVERSED	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I feel confident answering questions in class	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I would like to continue studying or attending school after this year	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I can describe my thoughts to others when I speak	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I can work well in a group with other people	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree

School Level	Index	Survey Item	Source	Scoring
				0=strongly disagree include in index mean
		When I have the opportunity, I can organize my peers or friends to do an activity	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I ask the teacher if I don't understand something	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		When I succeed at school, it is because I worked hard	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		If I do well in a test, it is because I am lucky REVERSED	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
	Life-skills Index—grade 4			Mean of all items 4=strongly agree on life-skills items; 3=agree on life-skills items, 2=neither, 1=disagree on life-skills items, 0=strongly disagree on life-skills items
Boys (optional)	Life-skills Index—grade 6, 8			Same as girls

Supplementary Table 5. Source Data for IO 5: Improved well-being and life-skills

School Level	Index	Survey Item	Source	Scoring
IO Indicator 5a: Percentage of girls reporting improved well-being				

School Level	Index	Survey Item	Source	Scoring
Girls (required)		Did you smile or laugh the last time you were at school?	Girls student survey	1=yes 0=no
		Did you learn or do something interesting the last time you were at school?	Girls student survey	1=yes 0=no
		Did you have enough energy to get things done the last time you were at school?	Girls student survey	1=yes 0=no
	Well-being Index			Count of cases responding 'yes' to all three well-being items
Boys (optional)	*Well-being Index*		Boys student survey	Same as girls
IO Indicator 5b: Percentage of girls reporting improved self-esteem due to Social and Emotional Learning (SEL)				
Girls (required)	*Life-skills Index grades 6, 8 (18 items)*	I am able to do things as well as my friends	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I want to do well in school	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I get nervous when I have to read in front of others REVERSED	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I get nervous when I have to do math in front of others REVERSED	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I feel confident answering questions in class	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree

School Level	Index	Survey Item	Source	Scoring
				0=strongly disagree include in index mean
		I can stay focused on a goal despite things getting in the way	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I would like to continue studying or attending school after this year	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I can put a plan in place and stick with it	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I recognize when choices I make today about my studies can affect my life in the future	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I can describe my thoughts to others when I speak	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		If someone does not understand me, I try to find a different way of saying what is on my mind	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		When others talk, I pay attention to their body language, gestures, and facial expressions	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I can work well in a group with other people	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree

School Level	Index	Survey Item	Source	Scoring
				0=strongly disagree include in index mean
		When I have the opportunity, I can organize my peers or friends to do an activity	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I often feel lonely at school REVERSED	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I ask the teacher if I don't understand something	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		When I succeed at school, it is because I worked hard	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		If I do well in a test, it is because I am lucky REVERSED	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
	Life-skills Index—grade 6, 8			Mean of all items 4=strongly agree on life-skills items; 3=agree on life-skills items, 2=neither, 1=disagree on life-skills items, 0=strongly disagree on life-skills items
	Life-skills Index grade 4 (12 items)	I am able to do things as well as my friends	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean

School Level	Index	Survey Item	Source	Scoring
		I want to do well in school	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I get nervous when I have to read in front of others REVERSED	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I get nervous when I have to do math in front of others REVERSED	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I feel confident answering questions in class	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I would like to continue studying or attending school after this year	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I can describe my thoughts to others when I speak	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		I can work well in a group with other people	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		When I have the opportunity, I can organize my peers or friends to do an activity	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean

School Level	Index	Survey Item	Source	Scoring
		I ask the teacher if I don't understand something	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		When I succeed at school, it is because I worked hard	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
		If I do well in a test, it is because I am lucky REVERSED	Girls student survey	4=strongly agree, 3=agree 2=neither, 1=disagree 0=strongly disagree include in index mean
	Life-skills Index—grade 4			Mean of all items 4=strongly agree on life-skills items; 3=agree on life-skills items, 2=neither, 1=disagree on life-skills items, 0=strongly disagree on life-skills items
Boys (optional)	Life-skills Index—grade 6, 8			Same as girls

Supplementary Annex 19. Washington Group Analysis

Purpose. The original baseline report included disaggregations on disability prevalence based on the Washington Group Short-Set questions. However, concerns over the accuracy of the data were raised and therefore, a different Washington Group Question – the Child Functioning Set – were administered during phase two of the baseline and determined to be a more appropriate measure of disability prevalence for the revised baseline report. This annex documents the evaluators' approach to administering and reporting on the Washington Group questions conducted under the advisement and consultation of GEC-T disability experts, as well as provides a comparative analysis of the short set and child functioning set data, and approaches to resolving potential coding issues.

Summary of STS approach to addressing issues and concerns on the disability data during STAGES baseline. Following the *GEC-T Household survey template* and GEC-T guidance, the external evaluator collected data using the *Washington Group - Short Set* (six questions) with girls in Grade 6 and above during phase one of the STAGES baseline (spring 2018) in order to enable disaggregation by disability as well as understand disability prevalence within the STAGES beneficiary population. Upon completing the analysis and reporting according to GEC-T guidance, a 22.45 percent prevalence rate of disability among surveyed girls (grade 6 and above) emerged and was included in the original STAGES baseline report.

As the prevalence seemed high, concerns emerged related to its accuracy; so the evaluators connected with the Evaluation Manager and GEC-T's disability experts to discuss the high prevalence rate, potential reasons for this figure, its implications and options to check the rates moving forward. Potential contributing factors discussed included issues related to the translation of the WG questions from English to Amharic (formal - completed by LCD translator), Amharic to Wolayttatto (informal - done in real-time by enumerators) as well as issues related to enumerator training and/or administration of the questions.

In consultation with GEC-T disability expert, the evaluators determined it was best to administer the Washington Group - Child Functioning (24 questions) to the girls in Grade 6 and above as part of the phase two baseline data collection via the *Girls Transition survey* collected in December 2018, as a way to cross-reference the prevalence rate. Furthermore, additional cross-checks on quality of translations from English to Amharic were conducted on WG questions (and the WG guidance on translations re-emphasized). Additional emphasis on WG questions and sensitivity in administration included in the Enumerator Training was also included, and the Washington Group - Child Function questions were also administered to a smaller subgroup of parents as part of the P*arent/Caregiver survey* to serve as additional potential data point for cross-comparison.

Findings from the comparative analysis (Spring Short Set versus Fall Child Functioning). Once the second set of child functioning questions were administered and data analysed, the evaluators compared the findings of the two data sets using two research questions. The findings are detailed below.

Research Question 1: How did disability indicators differ overall during Phase 1-SS and Phase 2-CF? (within a domain and overall)

- **FINDING:** Overall decrease in % of girls with disabilities
    - G6 = 23.0% Short Set → 6.0% Child Functioning

- G8 = 23.0% Short Set → 4.5% Child Functioning

Table 10. RQ1: How did disability indicators differ overall between *Phase 1-WG Short Set* and *Phase 2-WG Child Functioning*? (within a domain and overall) [50]

RQ1: % of **Grade 6/7** children with disability identifier by domain							
	Phase 1 (Short Set)			Phase 2 (Child Functioning)			Overall
Category	**Variable**	**n**	**% flagged**	**Variable**	**n**	**% flagged**	**% difference**
Communication	D_communication	19	8.8%	D_comm_ss_P2	4	1.8%	7.0%
Cognition	D_cognition	14	6.5%	D_cognition_ss_P2	6	2.8%	3.7%
Hearing	D_hear	10	4.6%	D_hear_ss_P2	0	0.0%	4.6%
Mobility	D_walk	12	5.5%	D_walk_ss_P2	0	0.0%	5.5%
Seeing	D_sight	11	5.1%	D_sight_ss_P2	2	0.9%	4.2%
Self-care	D_selfcare	9	4.1%	D_selfcare_ss_P2	2	0.9%	3.2%
Overall (per individual, not cumulative)	D_numDisabilities	50	23.0%	D_numDisabilities_P2	13	6.0%	17.0%

RQ1: % of **Grade 8/9** children with disability identifier by domain							
	Phase 1 (Short Set)			Phase 2 (Child Functioning)			Overall
Category	**Variable**	**n**	**% flagged**	**Variable**	**n**	**% flagged**	**% difference**
Communication	D_communicate_g8	7	3.9%	D_comm_ss_P2	4	2.2%	1.7%
Cognition	D_cognitive_g8	17	9.6%	D_cognition_ss_P2	3	1.7%	7.9%
Hearing	D_hear_g8	10	5.6%	D_hear_ss_P2	1	0.6%	5.0%
Mobility	D_walk_g8	11	6.2%	D_walk_ss_P2	1	0.6%	5.6%
Seeing	D_sight_g8	12	6.7%	D_sight_ss_P2	2	1.1%	5.6%
Self-care	D_selfcare_g8	11	6.2%	D_selfcare_ss_P2	1	0.6%	5.6%
Overall (per individual, not cumulative)	D_numDisabilities_g8	41	23.0%	D_numDisabilities_P2	8	4.5%	18.5%

[50] The Short Set were administered with Grade 6 and 8 sample girls in April 2018. The Child Functioning set were administered to the same girls in December 2019 who either repeated (Grade 6 and 8) or transitioned into Grade 7 or 9. The data being compared only includes sampled girls in the four treatment woredas who completed both the short set and child functioning set of questions.

Research Question 2: How did responses differ between Phase 1/Short Set and Phase 2/Child Functioning within each case?

Table 2 - RQ2: % Agreement within each case by domain

	Grade 6/7				Grade 8/9			
Category	**n with different disability indicator (Phase 1/SS to phase 2/CF) (G6)**	**% with different answer (G6)**	**n moved from non-functioning (Phase 1/SS) to functioning (Phase 2/CF)**	**n moved from functioning to non-functioning**	**n with different disability indicator (phase 1/SS to phase 2/CF) (G8)**	**% with different answer (G8)**	**n moved from non-functioning (Phase 1/SS) to functioning (Phase 2/CF)**	**n moved from functioning to non-functioning**
Communication	23	10.6%	19	4	11	6.2%	7	4
Cognition	18	8.3%	13	5	18	10.1%	16	2
Hearing	10	4.6%	10	0	11	6.2%	10	1
Mobility	12	5.5%	12	0	12	6.7%	11	1
Seeing	13	6.0%	11	2	14	7.9%	12	2
Selfcare	11	5.1%	9	2	12	6.7%	11	1
Overall *(per individual, not cumulative)*	*59*	*27%*	*48*	*11*	*41*	*23.0%*	*37*	*4*

Approach to baseline reporting revision and reconciliation of findings from the analysis (Spring Short Set versus Fall Child Functioning). In consultation with the GEC-T Disability experts, it was determined the Child Functioning findings will be used for STAGES baseline reporting and disaggregation moving forward. In order to fill in the gap of missing comparison data on child functioning, STS will extrapolate the disability prevalence rates as a proxy measure for baseline reporting and administer the child functioning set with the comparison group at midline.

Coding issues identified during comparative analysis and approaches to resolving. Two coding issues were identified when completing the comparative analysis of the short set and child functioning Washington group questions data. The logic of the external evaluator approach to addressing these issues was shared and confirmed with GEC disability expert, in March 2019, and documented here for reference.

- **Issue 1: Adjusting Coding for Skip Logic Variables.** The evaluator's coding differed from the recommended coding provided by GEC (Table CF.2 below). This applies to seeing, hearing, and walking.
    - For Seeing, CF1 (0="no", 1="yes"). This is a skip logic question: If students wear glasses (CF1="yes"), they are sent to CF2 ("WHEN WEARING GLASSES OR CONTACT LENSES, DO YOU HAVE DIFFICULTY SEEING?"), while if they are not wearing glasses (CF1=0/no), they are sent to CF3 ("DO YOU HAVE DIFFICULTY SEEING?")
    - The recommended syntax in Plan 2 was based on different coding – the evaluator's adjusted it accordingly:
        - ~~If CF1=1 AND (CF2=3 OR CF2=4)~~ IF CF1=1 AND (CF2=3 OR CF2=4)
        - ~~If CF1=2 AND (CF3=3 OR CF3=4)~~ IF CF1=0 AND (CF3=3 OR CF3=4)
    - RESPONSE: As this diverged from the prescribed syntax, the evaluators shared their approach with the GEC disability expert for their review and she confirmed and approved the logic

- **Issue 2: Aligning Cognition with Phase 1/Short Set.** Cognition is the only domain that is not 1:1 from phase 1/Short Set. Phase 2/Child Functioning approach splits cognitive into: remembering/concentrating/learning as compared to remembering and concentrating in Phase 1/SS. To construct this, the evaluator combined all three into one.
    - Phase 1: Cognition = Do you have difficulty remembering things or concentrating?
    - Phase 2: Cognition
        - 1. COMPARED WITH CHILDREN YOUR AGE, DO YOU HAVE DIFFICULTY LEARNING THINGS?
        - 2. COMPARED WITH CHILDREN YOUR AGE, DO YOU HAVE DIFFICULTY REMEMBERING THINGS?
        - 3. DO YOU HAVE DIFFICULTY CONCENTRATING ON AN ACTIVITY THAT YOU ENJOY DOING?
    - RESPONSE: the evaluators shared with the GEC disability expert for their review and she confirmed this was an acceptable approach.

Table CF.2: Child functioning for children aged 5 to 17. Responses to questions CF1-CF24 are used to determine whether children have functional difficulty in the reported domains. For indicators generated from the questionnaire for children aged 5-17, the denominator should be confined to all children aged 5-17 including those with missing data. Levels of functional difficulty are tabulated for each functional domain. Refer to Plan 2 below:

Plan 2: Tabulations for prevalence of functional difficulty among children aged 5-17

Functional domains	Functional difficulty if the following is true:
Seeing	If CF1=1 AND (CF2=3 OR CF2=4) **OR** If CF1=2 AND (CF3=3 OR CF3=4)
Hearing	If CF4=1 AND (CF5=3 OR CF5=4) **OR** If CF4=2 AND (CF6=3 OR CF6=4)
Walking	If CF7=1 AND (CF8=3 OR CF8=4) OR (CF9=3 OR CF9=4) **OR** If CF7=2 AND (CF12=3 OR CF12=4) OR (CF13=3 OR CF13=4)
Self-care	CF14=3 OR CF14=4
Communication (being understood inside or outside the household)	CF15=3 OR CF15=4 **OR** CF16=3 OR CF16=4
Learning	CF17=3 OR CF17=4
Remembering	CF18=3 OR CF18=4
Concentrating	CF19=3 OR CF19=4
Accepting Change	CF20=3 OR CF20=4

Controlling Behaviour	CF21=3 OR CF21=4
Making Friends	CF22=3 OR CF22=4
Anxiety	CF23=1
Depression	CF24=1

The percentage of children aged 5-17 years with functional difficulty in at least one domain are those children for whom **at least one domain** is coded 3 or 4 [1 for Anxiety or Depression] (true) as tabulated according to **Plan 2** above.

Supplementary Annex 20. Grade 10 Benchmarking

Objective. A separate benchmark sample of girls in Grade 10 were assessed to establish expected literacy and numeracy learning levels when cohorts of girls move into Grade 10. For example, the Grade 10 benchmark data will allow for comparison at midline for the cohort of baseline Grade 8 girls who will be in Grade 10 at midline.

Grade 10 benchmarking sample for learning outcomes. Benchmarking for Grade 10 was completed in the fall 2018 with a sample of 162 girls in four secondary schools across two woredas.[51] Since girls in current project (and sample) primary schools will matriculate into one of 16 total secondary schools across the four woredas, four schools were selected to be representative of the population of secondary schools. Both town[52] and rural based secondary schools were included in the sample.

Table 11. Sample Size by woreda and school name

Woreda	School Name	# Grade 10 Girls to be assessed
Damot Pulasa	Shanto Secondary School	40
Damot Pulasa	Suke (Galcha) Secondary School	40
Damot Woide	Bedessa Secondary School	40
Damot Woide	Sake Secondary School	42[53]
Total		**162**

Grade 10 benchmarking data collection. The Grade 10 Learning Assessment Benchmarking Exercise utilized the paper-based SeGRA/SeGMA Learning Assessment that was previously administered to Grade 8 female students during the STAGES baseline data collection in April 2018. In order to collect the necessary data, the benchmarking exercise was conducted by two data collection teams over the course of a single day in October 2018. Each team visited two secondary schools and included a member from the STS team[54], at least one Woreda official from the associated woreda, and at least one LCD-E team members. In order to ensure proper administration and oversight during the exercise, STS provided team members with a short overview of purpose and content of the instrument, as well as administration process and protocols in the morning prior to data collection.[55]

Each secondary school visit took approximately one hour.[56] The assessment took less than an hour to administered (approximately 30-40 minutes to administer – with 15 minutes for the SeGRA, 15 minutes SeGMA, and a 5 minutes break in between). After the assessments are completed, STS entered the data

[51] The secondary schools were in Damot Pulasa and Damot Woide.
[52] No larger urban areas are present in the STAGES target woredas, therefore, "town" is a more appropriate term to indicate more centrally, densely-populated areas.
[53] While the original target sample was 40 per school, two additional girls participated in this school.
[54] This included the STS Technical Manager and Project Director.
[55] Note the Woreda and LCD staff members were already deeply familiar with the administration of the SeGRA/SeGMA tools from their participation in the spring 2018 baseline enumerator training and data collection. Therefore, a short refresher was deemed most appropriate to prepare them to serve as invigilators.
[56] This included arrival and introductions at the school, arranging for an appropriate space for conducting the exercise, identifying Grade 10 girls for participation, providing instructions (in Wolayttatto and Amharic), as well as distributing and collecting completed assessments.

into an excel template and conducted a 10 percent double data entry check. Data cleaning and analysis was conducted in SPSS.

Grade 10 benchmarking findings. The following tables provide the literacy and numeracy scores for the Grade 10 benchmarking sample and the proportion girls who are in each proficiency category. The range of scores among Grade 10 girls in the benchmark sample do not suggest that there are floor or ceiling effects, and therefore, the use of the same difficulty tasks as those given to Grade 8 girls are appropriate.

The literacy score for Grade 10 girls is based on three subtasks as part of SeGRA and the numeracy score is based on three subtasks as part of the SeGMA.

Table 12. Grade 10 Benchmarking Scores

	Total Score on 3 SeGRA subtasks	Literacy Aggregate Score *(average of 3 subtasks)*	Total Score on 3 SeGMA subtasks	Math Aggregate Score *(average of 3 SeGMA subtasks)*
Number of Cases	162	162	162	162
Mean	134.49	44.83	120.16	40.05
Median	126.67	42.22	116.67	38.89
Std. Deviation	51.83	17.28	54.84	18.28
Range	251.11	83.70	283.33	94.44
Minimum	38.89	12.96	16.67	5.56
Maximum	290.00	96.67	300.00	100.00

Note: Each subtask, as per guidance, was on a 100-point scale. Across 3 subtasks, the maximum total score was 300 points for Literacy and 300 points for Numeracy.

Table 3 shows the proportion of girls in each proficiency category by subtask. The highest proportion of girls were proficient in the SeGRA reading passage; however, this still only represented under one-third of Grade 10 girls. Additionally, the subtask with the highest proportion of non-learners was fill-in-the-blanks in SeGRA and Geometry in SeGMA. Still, there was good variability in scores, with girls in the proficient category in both subtasks (albeit low proportions in Fill-in-the-blank).

Table 13: Literacy and Numeracy Skills by Learning Achievement Bands—Grade 10 Girls, English

Grade 10 EGRA/SeGRA—Percentage of Girls' Achievement						
Categories	SeGRA			SeGMA		
	Reading Passage	Fill in the blanks	Revising Sentences	Geometry	Fractions	Multiplication
Non-learner	1.85%	1.23%	3.70%	3.70%	16.05%	10.49%
Emergent learner	30.86%	59.26%	58.64%	40.74%	54.32%	38.27%
Established learner	38.27%	37.65%	29.63%	43.83%	25.31%	38.89%
Proficient learner	29.01%	1.85%	8.02%	11.73%	4.32%	12.35%

Based on these analyses, the same assessment used in Grade 8 for Grade 10 is deemed appropriate. Additionally, these data were entered into the outcomes spreadsheet as benchmarks for girls moving into Grade 10.

Supplementary Annex 21. Additional Key Intermediate Outcome Analysis – Girls' Life-skills

Girls' life-skills was reported on for Intermediate Outcome 5.5 Self-esteem in the original baseline report; however, the life-skills measure was replaced in the revised baseline report when additional data from phase two became available that more directly examined and linked to a self-esteem construct. The original findings for the life-skills measure are included here in the annex as reference.

IO5. Girls' Life-skills—Originally Reported in Place of Self-esteem

Intermediate outcome and indicator selection and measurement

This IO was chosen because the project's interventions assume that support for the well-being and self-esteem of marginalized girls is one of the prerequisites for better learning, transition, and sustainability outcomes. Although the identified IO in the log frame was on girls' self-esteem, the data collected at baseline originally reported a scale for life-skills may have been more appropriate to report.

Along with life-skills, girls' sense of well-being was also selected as an indicator, because well-being leads to success in both school and work and that measuring students' perceptions of their life-skills can predict their success in the future.[57] The following indicators were chosen for the IO:

- IO Indicator 5.a: Percentage of girls reporting improved well-being
- IO Indicator 5.b: Percentage of girls reporting improved self-esteem

These indicators were chosen because they provide an important measure of girls "readiness to learn." During phase one of data collection, girls' self-reported feelings of well-being and self-esteem were asked on the girls student survey; items for self-esteem were asked as part of a set of questions on life-skills. Once the data were collected, survey items were examined to identify underlying constructs—specifically, whether the items expected to measure self-esteem suggest a single factor and whether items measuring life-skills suggested a different factor.

The phase one baseline results suggested that life-skills encompassed self-esteem items; in other words, a single factor emerged out of the data analysis instead of two distinct factors. As such, the results were originally reported for life-skills—the more encompassing factor—instead of self-esteem for IO Indicator 5.b. However, it was later determined a separate measure of self-esteem was required by the FM. Therefore, additional items were added to the girls' transition survey in phase two of the baseline and underlying self-esteem construct examined. Within the revised baseline report, the self-esteem measure and findings have now replaced life skills for reporting on IO Indicator 5.b., but the original findings for the life-skills measure are included here in the annex as reference.

Findings

Percentage of girls reporting high levels of Life skills

[57] Gallup, *Principal Reflections on Student Engagement: Using the Gallup Student Poll*, (January 2014), http://www.gallup.com/file/services/176756/Principal%20Reflections%20on%20Student%20Engagement%20--%20Using%20the%20Gallup%20Student%20Poll.pdf.

Supplementary Table 5. Percentage of Girls Reporting High Levels of Life skills—Baseline Figures for IO 5.b[58]

Grade Level	Baseline (proportion reporting "high" life skills (2018)	
Treatment	***Actual***	***n***
Grade 4	30.10%	90
Grade 6	30.07%	89
Grade 8	35.64%	103
Average (G4, G6, G8)	31.90%	282

Note: Treatment n=884 girls. Grade 4 treatment n=299; grade 6 treatment n=296; grade 8 treatment n=289. The proportion of girls with improved life skills in subsequent evaluation points will be reported as the proportion of girls whose responses, on average, go from strongly disagree to disagree, from disagree to agree, and from agree to strongly agree. Furthermore, note that although IO 5.b includes an index for life skills in place of an index for self-esteem.

Supplementary Table 6. Percentage of Girls Reporting High Levels of Life Skills by Subgroup—Baseline Scores IO 5.b[59]

Grade Level	Grade 4		Grade 6		Grade 8	
Subgroup	***%***	***N***	***%***	***N***	***%***	***N***
Girls with at least one disability[60]	N/A	N/A	Not reported			
Girls who are overage for grade	20.73%	17	16.05%	13	24.00%	12
Girls in households unable to meet basic needs	76.19%	96	77.68%	87	85.71%	72
Girls who do not attend five days of school per week	75.56%	68	88.54%	85	87.23%	82
Girls with low levels of support from household	61.22%	30	75.34%	55	72.00%	36
Girls who report teachers treat boys and girls different in classroom	84.31%	172	87.44%	181	90.72%	176
Girls who report teachers are often absent from class	80.99%	115	84.56%	115	87.27%	96
Girls who report high corporal punishment exercised by teacher	79.34%	96	83.85%	135	92.00%	138

Note: Grade 4 treatment n=299; grade 6 treatment n=296; grade 8 treatment n=289. The proportion of girls with improved life skills in subsequent evaluation points will be reported as the proportion of girls whose responses, on average, go from strongly disagree to disagree, from disagree to agree, and from agree to strongly agree. Furthermore, note that although IO 5.b includes an index for life skills in place of an index for self-esteem.

Among girls in grade 8, those who had higher scores on the life-skills index also had higher scores on their overall literacy and numeracy scores. The relationship was statistically significant but weak for life-

[58] High level of life-skills means the student had an average response of 3.74 or higher on the items included in the scale; to achieve a 3.74 or higher on the scale, the student must have responded "strongly agree" to some items and "agree" to most.
[59] High level of life-skills means the student had an average response of 3.74 or higher on the items included in the scale; to achieve a 3.74 or higher on the scale, the student must have responded "strongly agree" to some items and "agree" to most.
[60] This data should be understood as missing. Results for disability subgroup—barrier 1—were not possible to disaggregate at this time as the Child Functioning set was administered in phase two data collection and has a different denominator than the relevant survey items. This information will be provided starting at midline using the Washington Group Child Functioning questions.

skills and well-being scores with the reading fluency and comprehension subtasks as well as with word problems and SeGMA subtasks—correlation coefficients between 0.10 and 0.14.

Interpretation and reflections

Life-skills was measured and reported in the original baseline report. The importance of this IO indicators was to measure whether girls are ready to learn when they come to school, and the correlations to student learning outcomes suggest a weak but present relationship.

The evidence suggests that there is a need to focus not only on the supports provided to girls in school but also on their life skills. These indicators demonstrate, that girls who come to school with high levels of life skills are more likely to succeed in school than their peers who have low levels. However, life-skills was dropped in the revised baseline report in favor of a more focused measurement of self-esteem which was more clearly aligned with the original indicator and logframe.

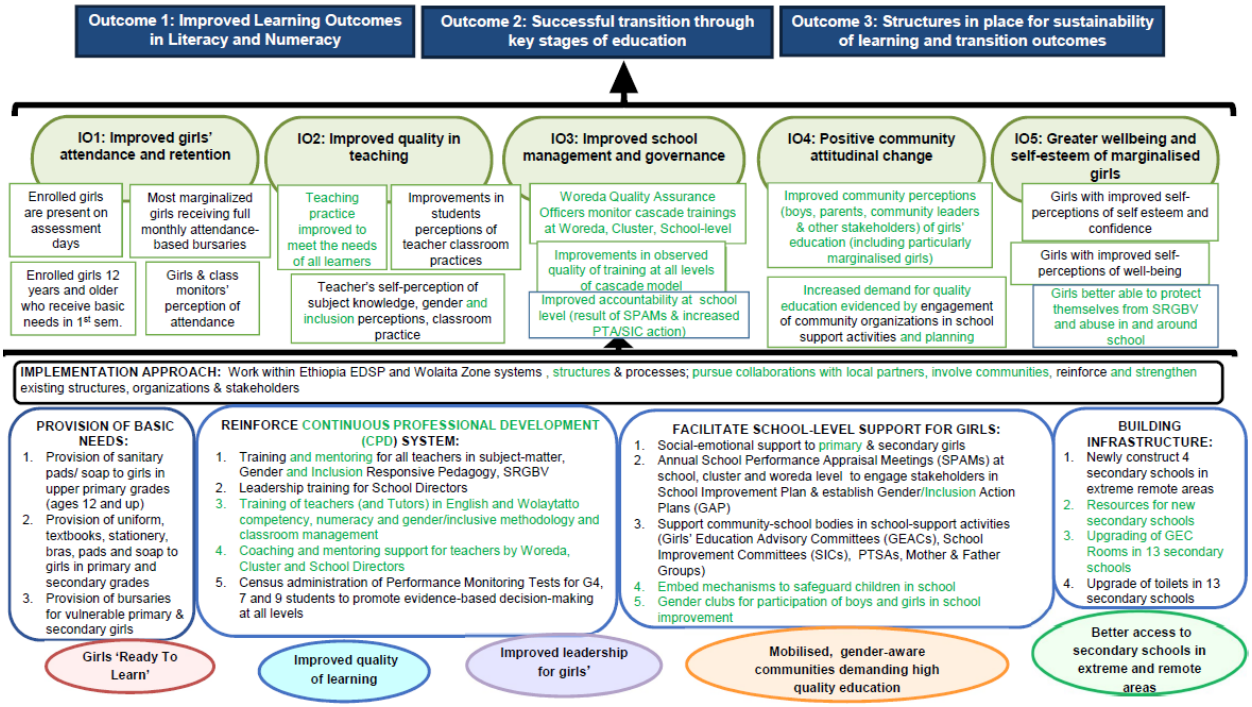
Supplementary Annex 22. Revised STAGES Theory of Change

Revised STAGES Theory of Change

Revised_TheoryofChangeSTAGES_LCD_Ethi

Based on the baseline findings and the most up to date, streamlined approach to STAGES' activities, Link has made minor revisions to the Theory of Change. The PDF version of the revised Theory of Change is provided by clicking the icon; an image is also provided here for easy reference. Green font indicates new content not previously included in the Theory of change.

Figure 13. STAGES Revised Theory of Change, May 2019



Supplementary Annex 23. Link Ethiopia: Fund Manager Feedback and Responses on STAGES Baseline Report Part I

LINK Ethiopia

Fund Manager feedback on baseline report

Date: 09/07/2018

STS & LCD Response Date: 20/07/2018

Executive Summary

Top four areas the project needs to address

PROJECT RESPONSE: See additions and clarifications sent by LCDI below.

1. Project response for girls with disabilities given the large proportion identified in the sample.
2. Project response on how to address/reduce corporal punishment rates in schools and additional support for teacher to teach English in the classroom.
3. Project response for prevalent barriers identified such as girls overage for their grade.
4. Further details on how the project will regularly collect data against the IO's given the large gap between baseline and midline 1.

Top three areas the external evaluator needs to address

1. Clearer steer on where characteristics and barriers data can be aggregated up to the sample/beneficiary population.
STS response: Only data on orphans cannot be aggregated up to the sample/beneficiary level (this was not collected at the individual level).

2. Details on how the research approach/tools were or will be adapted for girls with disabilities.
STS response: edits have been made in the report to address this.

3. Revise scoring for the sustainability indicators based on FM feedback.
STS response: edits have been made in the report to address this.

1. Essential to address for sign-off

The following section outlines essential areas that must be addressed for the Fund Manager to be able to sign-off your baseline report. The following areas must be addressed and your baseline report updated accordingly. Please submit a revised report, to the Fund Manager, within 1 week of receipt of this feedback which responds to the issues in section 1.

PROJECT
LCDI responses to each of the items below is in Annex 19.

Characteristics and barriers:

Disability: The disability numbers from the sample are very high (22.5%) The project's response for adaptations for a group this large requires more work. The FM recognises more time is required to consider adaptations and so will follow up with the project team on producing a separate paper to address this issue.

Corporal punishment: Table 9 shows that 31% of girls are facing some sort of corporal punishment by teachers in the intervention school. Please respond to how this barrier will be addressed or is currently being addressed if that's the case.

Teacher English-language competency: The report mentions that teachers have poor English language competency. Please respond to how this barrier will be addressed or is currently being addressed if that's the case.

Learning:

Less than 1% of girls can subtract but 13% can multiply and 4% can do fractions (G8). What's the reason for this? Is there an issue with the way subtraction is being taught in schools? Is it a testing issue?

Intermediate Outcomes:

What other data is being collected by the project for IO2 and IO3 on a quarterly basis by the project? It seems a bit long to assess this at only evaluation points?

Could project staff/woreda officials collect the type of data the EE is supposed to collect for the evaluation more frequently given the larger amount of time between baseline and midline?

EXTERNAL EVALUATOR

EE to provide some example write ups of the interviews in place of qualitative transcripts.

- ***STS Response:*** *Excerpts of qualitative field note write-ups - one focus group discussion and one key informant interview – have been provided. Please note these are excerpted sections for illustrative purposes only, and do not include all questions and responses recorded in the full field note write-ups. All demographic information and a limited number of question responses have been removed to ensure confidentiality. These excerpts are not for distribution or to be shared outside PwC's Evaluation Advisor and Qualitative Expert.*

Disability: There is no information on how the methodology for collecting data has been tailored to reach girls with disabilities. Was this done? If not – will tools be altered for midline given almost a quarter of girls are girls with disability?

- ***STS Response:*** *At the time of the baseline design there was limited available information regarding the size, scope and characteristics of the population of girls with disabilities within the STAGES project. Given this, a more in-depth methodology for collecting data on girls with disabilities was put on hold until after the baseline and Washington Group data collected and analysed. Several types of accommodations, however, were incorporated into the baseline study learning assessment design and administration in recognition of some of the challenges students with disability may encounter. For example, given that some students may have low or limited vision, all learning assessment student stimuli for EGRA and EGMA subtasks were developed using large print (at least font 16 or above) and spacing between assessment items. In addition, while it is common practice for timed EGRA and EGMA subtasks to be limited to 60 seconds each, for the baseline study the timed EGRA subtasks for letter sound identification, nonword and familiar word, as well as the EGMA subtasks for number identification, addition and subtraction, were extended to 120 seconds to enable extra time for students to complete the subtask (please note, however, the reading passages were limited to 60 seconds per GEC-T guidance). Furthermore, supervisors and enumerators were advised to conduct assessments with students in environments with limited noise and distractions as much as feasible.*

  Looking forward, STS will continue to utilize the previously mentioned accommodations as well as explore other low-cost strategies to support disability-sensitive assessment and data collection practise throughout the remaining evaluation points as well as look to draw on STAGES' learning about the context and prevalence of girls with disabilities at the community, school and project level. It is important to recognize though, that while the findings from the baseline data Washington Group questions potentially provide important insights into perhaps a less 'visible' population of girls with disability in the Wolaita zone primary schools, this data is not meant to serve as a diagnostic tool and the findings must be understood with this in mind. Additional project-level research and follow-up will be essential to better contextual these findings.

  STS is particularly interested to hear more from other GEC-T project implementers and evaluators' experience with their baseline evaluations and learn about low-cost best practices and lessons learned on disability-sensitive evaluation practices that could be incorporated into our midline design and data collection approach. In addition, midline qualitative research will include a more targeted and purposeful sampling of girls with disabilities and their parents for key informant interviews and focus groups discussions to better understand the challenges, opportunities and unique circumstances these girls face in their learning and transition within the Wolaita zone. At this time, however, STS would also like to note it does not anticipate it will be feasible to create or track more individually-based or administered assessment modifications at future evaluation points given the large sample size. Moreover, we do not anticipate being able develop more in-depth or whole scale modifications to learning assessments– such as adapting a braille version of the learning assessment - given evaluation cost and time constraints.

  Relevant information above has been incorporated into section ***2.3 Baseline Evaluation Methodology,*** *under subsection Incorporation of gender and social inclusion minimum standards.*

2.1 Evaluation methodology

**Some of the information has been added*

Characteristics and barriers:

'In the school audit survey, respondents were asked if there were girls who are orphaned in their school. All respondents—one school director per school—said "yes;" respondents were asked to give an approximate number of orphaned girls in the school.' These data are aggregated to provide an estimate of the total number of orphaned girls in the sampled schools. The figure of 429 girls (48.5%) is unclear. Is this an estimate of the number in the sample or total in the schools? It would be useful if the report could make it clear which data in the tables can be used to aggregate up to the sample/beneficiary population and which data cannot.

- ***STS Response:*** *These data can only be used as an estimate as the response depended on the school audit survey respondent's knowledge of all girls in the school. As noted above, all data except that on orphans, can be aggregated to the sample/beneficiary population.*

Table 8 and 9: Make it clear where the difference between the treatment and control group is statistically significant by adding a * or some other way of identifying this.

- ***STS Response:*** *Significance tests were limited to those characteristics and barriers with sufficient n's to compare. An asterisk has been added where tests were appropriate to run and were significant.*

Page 33/34 – The report mentions that none of the girls surveyed in grade 6 and 8 report using the MOI – English- at school with their teachers However table 8 shows that 100% of grade 6 and 8 speak English MOI with teachers, at home and with friends? Is this a mistake in the table?

- ***STS Response:*** *This was an error. It should read "100% of grade 6 and 8 speak Wolayttatto with teachers, at home and with friends".*

Does the EE plan to collect qualitative data with GWD for midline 1 given the large proportion identified in the sample? Did the EE collect qualitative data on this group at baseline? If so can more details be included within the main report in the characteristics and barriers section?

- ***STS Response:*** *While girls with disabilities were noted as one of the targeted profile groups that should be included in the recruitment of female student focus group participants for the baseline study, this demographic information was not highlighted or captured in the focus group respondent types data and girls with disabilities did not appear to be included. This oversight is one of the evaluators' lessons learned from the baseline study and clearer questions to include this type of data capture will be included in future qualitative evaluation tool. Midline qualitative research will also include a more targeted and purposeful sampling of girls with disabilities and their parents for key informant interviews and focus groups discussions to better understand the challenges, opportunities and unique circumstances these girls face in their learning and transition within the Wolaita zone at both the primary and secondary school level.*

Sustainability:

System level sustainability – a score of 3 seems quite high for baseline based on support for evaluation data collection only. If data isn't available for indicators because the activity hasn't started yet, score this

as a 0 rather than excluding from the overall sustainability score. This brings some of the scores down for baseline.

- ***STS Response:*** *Additional rationale for supporting a score of 3 for systems level sustainability has been added to the report and below.*

Despite the limited indicator data on the system sustainability, a baseline systems sustainability score of 3 has been determined appropriate by the evaluators for several reasons. First and foremost, the design and implementation of the STAGES project has sustainability embedded throughout the core of the project with its direct engagement, partnership and capacity building of zone and woreda-level personnel throughout the design, implementation, monitoring and evaluation of project activities. This is an important distinction for other programmatic approaches that may recruit, hire or train outside personnel to conduct these activities - often creating external or parallel systems or structures that are difficult to sustain beyond the end of the project. In contrast, STAGES supports government personnel to be part of the decision-making, implementation and learning process throughout its activities, which not only supports greater ownership and accountability throughout the process and life of the project, but also situates the knowledge, learnings and best practices within the government personnel and systems – rather than outside of it - so it may be carried out beyond the timeline – and targeted woredas – of the project.

Moreover, STAGES supports the government in activating existing systems and structures that promote gender equality and girls' education. As noted in the STAGES GESI Analysis,[61] there is already a strong government policy environment present for girls' education in Ethiopia and the project was designed to support the government in implementing aspects of that policy. Furthermore, at the Zone and woreda-level there are already assigned, existing gender-focused personnel - gender officers – who Link works with directly to build their capacity. With the existence of an enabling policy environment and gender-sensitive personnel structures or frameworks already in place, some of the barriers projects often face in supporting girls' education at the systems level have been greatly reduced.

Lastly, while this is a baseline for the STAGES project specifically, it is also important to recognize that it is far from the start or beginning of Link's work and engagement in education systems support in the Wolaita zone. STAGES is drawing and building on in-depth experience, knowledge and continued relationships that come from working in the zone for the past 10 years, including most recently its heavy engagement and collaborative partnership with zone and woreda level education officials throughout GEC1. This experience, including the relationships and trust developed between Link and government education officials - set both the project and government up for sustainable, system-level success and the scoring reflects this.

The EE feels confident that a rating of 3 in systems level sustainability reflects the integration, by design, of the STAGES project in the government structure. The EE will look for evidence of sustainability at the community and school level in subsequent evaluation points, as well as evidence of further established systems-level integration.

School level indicators 3 and 4 – why was information not collected against these indicators at baseline? These are areas where baseline data should have been available. Will the EE collect data at the time at which transition data is collected or does the project have a way of reporting against these in the interim?

[61] Casey McHugh and Ashley Doria, *STAGES GESI Analysis* (Pacifica: School-to-School International, 2017).

- ***STS Response:*** *In the Woreda officials interview protocol, we included one question in which we intended to probe further if any evidence towards indicators 3 or 4 were collected. That question was:* Has the MOE been working LINK to establish plans for sustainability?
  Given that we didn't get any reasonable responses on this item and probing was not done, there was limited data on which we could generate a response. The EE will explore ways to include this data collection in the fall with transition data; although it does represent an additional data collection with respondent groups we had not previously expected to interview in the fall (community members). We will also explore ways to report on this indicator through project monitoring, as interaction with community groups is frequent as part of the STAGES model.

Intermediate outcomes:

Influence of Woreda officials' perceptions on primary school girls learning in the opposite direction to expected – was any qualitative data collected at baseline to help explain these findings or will this become an area for midline 1?

- ***STS Response:*** *We do not have any qualitative data on this finding. However, we are also confident that the influence of Woreda officials' perceptions on schools is indirect at best; the level of direct engagement and oversight, or classroom-level involvement, is low. That said, we also had a limited sample of Woreda officials so this may also be a limitation of the correlation. We will explore this further in midline by including a qualitative component as well as more Woreda officials per Woreda.*

Teaching: When the EE goes in to collect secondary school benchmark data in September and transition data, could baseline observations/surveys be done for secondary school teachers? Otherwise a baseline at midline 1 in 2020 is quite far off.

- ***STS Response:*** *We are happy to explore this option. However, note that this presents additional data collection and analysis which was not part of STS' budget or plan. Depending on timing and availability of enumerators/government officials, we will explore this option with Link when planning the fall data collection.*

Life skills: EE recommends the indicator should be revised to measure improvement in life-skills, or, if a separate measure of self-esteem is required, additional items should be added to the instruments. - Happy for the logframe indicator to be changed to life skills.

- ***STS Response:*** *Noted. We will include a separate self-esteem scale if there is continued focus from Link on self-esteem separately from life skills.*

The targets proposed for well-being and for life skills are lower than those in the logframe – the FM is happy with the rationale and for these to be changed in the logframe.

- ***STS Response:*** *Noted*

Can the EE (or project) explain why primary school reporting is optional when the youngest beneficiaries are G4 and have 4 years of primary left to complete. We don't remember agreeing this but also aware there has been a lot of back and forth since log frames were agreed so may have simply missed this.

- ***STS Response:*** *The projects' focus on secondary level was the primary driver of the indicators selected. At the time the target for secondary was noted (and primary was optional) we had not*

yet discussed the removal of grade 10 from baseline. As such, it does leave a gap of baseline data. We have reported the primary as that is all that is available at the time of baseline and will continue to do so until primary-level data are collected.

P.79 – It says that 100% of treatment schools have GEACs HOWEVER on P.70 – It says that 4 woreda officials noted that some schools did not have active GEAC or active PTA. Assume this is because of the distinction between a GEAC being set up and being active. IO Indicator 2.a is percentage of GAP targets or actions undertaken. How will data be triangulated to ensure the correct information on level of activity is being collected?

- ***STS Response:*** *The school audit survey asked about the presence of GEACs as did the Woreda official survey. We relied primarily on the school audit survey because it was reported on a school-by-school basis by respondents at each school. By contrast, the Woreda officials were responding generally across all schools in their Woreda. As such, we surmise that there are likely schools in the woredas without active GEACs (or GEACs at all), but within the sample of treatment schools, school-based survey respondents noted that all schools have GEACs. During monitoring, these data will be collected from all treatment schools and will be used to triangulate these findings.*

Conclusion:

The conclusion is really well summarised. Please also include some details to indicate where the project falls on the GESI spectrum (and why).

- ***STS Response:*** *It has been noted in the conclusion that the project fulfils the requirements of "gender sensitive" on the GEC GESI Continuum as it meets the GEC GESI Minimum Standards and includes both GESI accommodating and transformative practices and activities.*

Data (EE)

Exclusion of the Letter Sound Identification Subtask from Aggregate Scores – FM is in agreement with this decision.

- ***STS Response:*** *Noted.*

Weighting between the English and Wolayttatto subtasks to be confirmed by FM at a later point – this requires more thought to get the balance right for reporting purposes.

- ***STS Response:*** *Noted.*

Replication of standardised scores per subtask (out of 100): The FM has been unable to replicate the data for standardising the subtask score out of 100. Please provide further information or syntax to support these calculations, as some of the standardised scores seem to be larger than 100.

- ***STS Response:*** *Syntax has been shared under separate cover.*

Future data collection: Given the baseline data collected was delayed as a result of issues with obtaining the tablet, does the EE plan to collect midline 1 data exactly 2 years after midline or can midline 1 data be collected earlier, e.g. 3-4 months earlier in line with original dates? What are the implications of this?

- ***STS Response:*** *We are expecting to collect the midline 1 data around the same time (early part of the year) – preferably in February/March. This timing works well for the school year, but more importantly, does not overlap with core learning assessments conducted by the project or beginning-of-year monitoring data collection. This will also keep the midline data collection timeframe close to that of the baseline, to ensure comparability.*

2. Recommended to address

The following section outlines areas that are recommended to be addressed to improve the quality of the report.

Executive summary: This is a very good summary. Would have liked to see a statement on whether the TOC still holds along with key recommendations.

- ***STS Response:*** *The ToC still holds, although specification of activities focused on girls with disabilities, may be warranted. STS will work with Link to review project activities focused on girls with disabilities and update the ToC accordingly. STS anticipates the timeframe for this review to be during the fall 2018.*

Background: Well written and concise. The EE has done a great job in integrating Link's M&E findings to further highlight the challenges in the area of implementation. To strengthen the section, we suggest including figures for girl's completion rates and learning outcomes to emphasize the contextual challenges.

- ***STS Response:*** *Gender disaggregated primary completion rates for grade 5 and 8 in the Wolaita Zone have been added to the report. A more thorough, in-depth discussion would be required if government level Wolaita zone learning outcome data – such as Grade 8 or 10 exam pass rates – were to be accurately and appropriately presented and contextualized. In order to maintain the flow and concise language of this section, this information has not been added at this time.*

Supplementary Annex 24. Link Ethiopia: Fund Manager Feedback and Responses on STAGES Baseline Report Part II

LINK Ethiopia

Fund Manager feedback on baseline report part II

Date FM Feedback Shared: 12/06/2019

Project and EE Response Date: 3/07/2019

3. Essential to address for sign-off

The following section outlines essential areas that must be addressed for the Fund Manager to be able to sign-off your baseline report. The following areas must be addressed and your baseline report updated accordingly. Please submit a revised report, to the Fund Manager, within 1 week of receipt of this feedback which responds to the issues in section 1.

PROJECT

Beneficiary numbers:

The direct beneficiary number has changed from the previous baseline when it was 41,917 across these grades. Last report also said data was not from EMIS but enrollment data therefore should be more accurate. How have the figures changed so much given this? An estimated 6,606 are estimated to join from other woredas/zones but will this not be balanced by some girls in the project areas moving out? In addition to this, are new grade 1 girls (after the previous grade 1 last year moved into grade 2 in September 2018) included in the figure – which has resulted in an increase in the number? What happens when a new G1 starts next year etc. Will the beneficiary numbers continue to grow?

LCD Response:

a) **The direct beneficiary number has changed from the previous baseline when it was 41,917 across these grades. Last report also said data was not from EMIS, but enrollment data therefore should be more accurate. How have the figures changed so much given this?**
These figures did not include children in grade 1 or those estimated to join the schools during the life of the project. The original beneficiary number was 43,978 and included girls enrolled in Grades 2 through 9. The number of 41,917 was provided to show the number girls enrolled in primary school only (Grades 2 through 8). We have thoroughly reviewed how our beneficiary number is defined and therefore who is included. We have now established that the source of the data is EMIS data, but figures used in the proposal were based on October 2016 EMIS returns, the baseline and all subsequent revisions have been based on EMIS data from the start of the

2017/2018 academic year (exact month not known). While there may still be inaccuracies with this data, it was determined to be the best available source at the time.

b) **An estimated 6,606 are estimated to join from other woredas/zones but will this not be balanced by some girls in the project areas moving out?** Children who leave the project areas will still have benefited from the project's systemic and school-based interventions and so are still considered as beneficiaries.

c) **In addition to this, are new grade 1 girls (after the previous grade 1 last year moved into grade 2 in September 2018) included in the figure – which has resulted in an increase in the number?** No, it only includes those that were in Grade 1 in 2017/2018 academic year – see answer below for an explanation of what happens beyond September 2018.

d) **What happens when a new G1 starts next year etc. Will the beneficiary numbers continue to grow?** No, they will not continue to increase. Our operational definition, however includes the girls passing through Grades 1 – 10 in each year for the duration of the project. Within this operational definition, the new Grade 1 girls who enter primary school in each year also benefit from project interventions. However, the evaluation design is based on the requirement of a tracked cohort methodology and that we would not have the additional resources needed to assess specific outcomes for these future beneficiaries. By the nature of the project's revised design and the evaluation design, the approach we have decided on for calculating our beneficiary number is effectively provides a sample of the wider group of girls who will benefit over by the life of this systemic and long-term project.

EXTERNAL EVALUATOR

P31: 'To ensure that tracking girls into secondary will be feasible, the evaluator has excluded schools from the sampling frame where girls matriculate into a secondary school outside of the treatment woreda.' So, in a sense – do we expect these girls to have one less barrier to transition? i.e. there is a secondary school within their woreda. However, assume the same was applied to schools in the comparison group even though they aren't tracked for transition? It may be worth expanding on any implication of this approach to transition trends and representativeness of findings based on this.

- This is purely a logistic decision affecting one primary school in the treatment group. All primary treatment schools within the project woredas have a secondary school to which students can matriculate within their woreda (as does the comparison primary schools). For only one of the primary schools, however, the CLOSEST secondary school is located in a separate woreda outside the project's target geographic area. As a result, students from this primary school typically attend that secondary school and not the one within the woreda (and within the projects target area).

P89: Transition trend for girls in G8. Is there not a variable in the dataset for which girls received a bursary to explore this at this phase? Also – are G8 girls the only grade to receive a bursary given the critical transition point or do other grades receive this too?

- The dataset of girls who have received bursaries was not available to the external evaluators during the baseline data analysis, so it was not possible to examine the impact on transition in the revised baseline report. However, given the potential impact of bursaries on transition for the most marginalised girls, the EE will work with Link to ensure that the relevant dataset is made available at future analysis points. Link is currently in the process so strengthening their monitoring systems, which should support a more seamless sharing of the bursary recipient dataset by student ID number at future evaluation points.
- As additional contextual note on bursaries within the project, according to Link's monitoring data, no bursaries were provided in year 1 of the project (2017/2018), and 195 secondary

school girls received bursaries and remained in school in Year 2 of the project (2018/2019) during which two rounds of bursaries were provided. These criteria for bursaries was extremely vulnerable secondary school girls. Going forward the criteria will be girls who are extremely vulnerable and otherwise likely not to attend (girls with disabilities, girls who are orphaned, girls who are young mothers) at both primary and secondary levels. Bursaries are expected to provide items such as uniform, school registration fees, textbooks, and in some cases, house rent where children have to travel long distances to school and cannot return daily to their homes.

P89: Reduction in sample size based on attrition. Replacement feasible for learning sample so this is fine. For transition, given the definition of success for the project is in school progression only – is there a way to broaden out the transition exercise beyond the sample of learning girls when looking at the register or would this be unfeasible?

- The proportion of girls from the baseline sample that transition successfully in school at midline (i.e., two grade levels above where they were at baseline) will be reported at midline. In addition, using Link's data on girls who receive bursaries or sanitary pads may serve as another potential data point to track girls who successfully transition each year because: (1) girls receive sanitary pads in grades 6 and higher at the beginning of the school year (1 month into the academic year, in approx. October) and (2) the criteria for receipt is that the girl is enrolled in school and present to receive the pads. However, the ability to match these data year over year is uncertain – i.e., can girls' data from year 2 of the project be matched to the same girls' data in year 3 and onward? The EE will explore this possibility with Link at midline.

P89: 'By age, girls who were 1–4 years older than their peers had a higher transition rate than girls who were on age for grade 4 (9–11 years old).' This trend then reverses <u>for G6 and G8 i.e. overage = lower transition rate</u>. Any evidence to suggest this may explain the learning trend – girls' overage for their grade have lower learning scores? i.e. schools are trying to push older girls through the system faster at younger grades even if they don't have the right learning levels? Or they start with some basic levels that are higher than early grade girls but then struggle to increase these levels in line with peers? Any project response suggestions to address/explore this trend further?

- The EE cannot determine whether the dropout rate drives learning outcomes or the other way around; the latter is explored in Table 20a (with the Student Background Index). When taking into account four common barriers – one of which is academic performance – we find that the transition rate is lower for girls with more barriers in grades 4 and 6; but in grade 8, girls who face these barriers have likely already dropped out. In terms of the policy – and to explore what effect transition policies have on learning outcomes - the EE can include questions at midline to explore the role that learning outcomes play in promotion decisions in primary grades compared to secondary.

P96: Table 22. Benchmark for G4 into G5 completion at Evaluation point 1 is completion rate for G5. Won't this be lower than for example enrolment into G5 vs then those girls that make it to completion? Is there any data on G4 completion rate and G5 enrolment?

Phase 2 of the baseline collected transition data for girls at the beginning of G5, part of which was verifying their enrolment (i.e., a girl who was not present on the day of the data collection but was enrolled was marked as having successfully transitioned into G5). The completion rate for girls at baseline shown in Table 22 is based on population information reported in EMIS and is used as a reference point. The

actual targets for each 2-year transition period are based on the baseline transition rates after 1 year, using the completion rate as contextual information