
GEC-Transition STAGES Endline Report

August 2024

Project Number: 6473

Project Name: Supporting Transition of Adolescent Girls through Enhancing Systems

Implementing Agency: Link Education International

Authors: Parnika Bhatia, Fiona Eichinger, Drew Schmenner

External Evaluator Institution: School-to-School International

Version Number: I

Revised Date:

Acknowledgments

The School-to-School International team would like to thank the Girls' Education Challenge -Transition team for their support and guidance during this endline study of Supporting Transition of Adolescent Girls through Enhancing Systems. We would also like to recognise the strong and indispensable engagement of the Ministry of Education's Wolaita Zone and woreda education officials for serving as true partners and leaders throughout the evaluation. In addition, woreda education supervisors and experts played a crucial role in the data collection process. The research would not have been possible without the participation of all the students, teachers, head teachers and officials who participated in the assessments, surveys and key informant interviews. Finally, this endeavour would not have been possible without the dedication and support of the Link Education teams who lent their continuous, thoughtful support throughout the endline design, implementation and analysis.

The following persons played a key role in the process:

Angela Keenan	Link Education International
Sandra Graham	Link Education International
Samantha Ross	Link Education International
Nicolo Di Marzo	Link Education Ethiopia
Dagne Negash	Link Education Ethiopia
Abeynew Mulat	Link Education Ethiopia
Parnika Bhatia	School-to-School International
Fiona Eichinger	School-to-School International
Drew Schmenner	School-to-School International
Briona Graham-Clayton	School-to-School International
Muhiburrahman Muhib	School-to-School International
Khalilurahman Hanif	School-to-School International
Alemneh Tadele	School-to-School International
Lamirot Abera Ali	School-to-School International
Bekalu Yayeh	School-to-School International
Abiy Aboret	School-to-School International
Hanibal Imbaye	School-to-School International

List of Acronyms

CPD	Continuous Professional Development
GEC	Girls' Education Challenge
GEC-T	Girls' Education Challenge-Transition
GIRP	Gender and Inclusive Responsive Pedagogy
IO	Intermediate Outcome
KII	Key Informant Interview
Link	Link Education International
ML1	Midline 1
ML2	Midline 2
ORF	Oral Reading Fluency
READ M&E	Reading for Ethiopia's Achievement Developed Monitoring and Evaluation
SEL	Social-emotional Learning
SRGBV	School-related Gender-based Violence
STAGES	Supporting Transition of Adolescent Girls Through Enhancing Systems
STS	School-to-School International

Table of Contents

Acknowledgments	2
List of Acronyms.....	3
Executive Summary	7
Background.....	7
Purpose of the Evaluation	7
Research Design	7
Past Reading Outcomes in Ethiopia.....	9
Endline Findings	10
Conclusions	13
Recommendations.....	14
1. Background to Project	16
1.1 Project Theory of Change.....	16
1.2 Key Evaluation Questions and Role of the Endline Evaluation.....	17
1.3 Project Context	17
2. Research Design	18
2.1 Evaluation Approach and Methodology.....	18
2.2 Quantitative Data Collection Tools	19
2.3 Qualitative Data Collection Tools	20
2.4 Sample.....	21
3. Outcomes	25
3.1 Learning Outcomes	25
3.2 Transition Outcomes.....	34
3.3 Sustainability Outcome	38
3.4 Value for Money.....	41
4. Key Intermediate Outcome Findings	43
4.1 Attendance.....	43
4.2 School Governance and Management.....	45
4.3 Quality of Teaching.....	46
4.4 Positive Community Attitudinal Change (IO4)	50
4.5 Gender Perceptions and Life Skills	51
5. Conclusion and Recommendations.....	54
5.1 Conclusions	54
5.2 Recommendations.....	56

List of Figures

Figure 1. Grade 8 Girls Endline Subtask Literacy Outcomes	26
Figure 2. Changes in Subtask Scores from Grade 6 ML2 to 8 Endline (%)	27
Figure 3. Grade 10 Girls Endline Aggregate Literacy Outcomes.....	28
Figure 4. Changes in Subtask Scores from Grade 10 ML2 to Endline (%)	28
Figure 5. Grade 8 Girls Endline Subtask Numeracy Outcomes	30
Figure 6. Changes in Subtask Scores from Grade 10 ML2 to Endline (%)	30
Figure 7. Grade 8 Literacy and Numeracy Outcomes at Endline by Woreda.....	32
Figure 8. Grade 10 Literacy and Numeracy Outcomes at Endline by Woreda.....	32
Figure 10. Proportion of Teachers Receiving Training, Over Time	46

List of Tables

Table 1. Cohorts Sampled by Evaluation Timepoint.....	8
Table 2. Change in Literacy Mean Scores from Grade 6 ML2 to Grade 8 Endline	10
Table 3. Change in Numeracy Mean Scores from Grade 6 ML2 to Grade 8 Endline	11
Table 4. Grade 10 Aggregate Literacy Outcomes at ML2 and Endline (%)	11
Table 5. Grade 10 ML2 and Endline Numeracy Outcomes (%)	11
Table 6. Effects of Pandemic on Learning	18
Table 7. Learning Assessment Tools.....	19
Table 8. Quantitative Data Collection Tools.....	20
Table 9. Qualitative Tools, Sample and Corresponding Outcomes and Intermediate Outcomes	20
Table 10. Quantitative Tools, Sample and Corresponding Outcomes and Intermediate Outcomes.....	21
Table 11. Quantitative Sample of Girls	22
Table 12. Quantitative Sample of Boys.....	22
Table 13. Quantitative Sample by Age of Girls	22
Table 14. Quantitative Sample by Age of Boys	23
Table 15. Endline Sample by Disability.....	23
Table 16. Girls' and Boys' Characteristics	24
Table 17. Literacy from EGRA/SeGRA Scores.....	25
Table 18. Change in Literacy Mean Scores from Grade 6 ML2 to Grade 8 Endline	26
Table 19. Grade 10 Aggregate Literacy Outcomes at Endline (%).....	27
Table 20. Grade 10 Aggregate Literacy Outcomes at Endline (%).....	28
Table 21. Numeracy based on EGMA/SeGMA Scores	29
Table 22. Numeracy Scores from ML2 to Endline	29
Table 23. Grade 6 ML2 Subtask Numeracy Outcomes (%)	29
Table 26. Literacy and Numeracy Outcomes at Endline by Woreda, Grades 8 and 10.....	31
Table 27. Learning Scores of Key Subgroups by Grade	33
Table 28. Learning Scores, Given of Key Barriers, by Grade	34
Table 29. Girls Survey, Percentage of Grade 10 Girls Who Want to Continue to Grade 11.....	35
Table 30. Factors Preventing Girls from Transitioning to Secondary School, Grade 10 Girls	35
Table 31. Grade 10 Girls Socio-emotional Composites.....	36
Table 32. Factors that Have Helped Grade 10 Girls Stay in School at ML2 and Endline	37
Table 33. Reported Reasons for Grade 10 Girls Considering Dropping Out at Endline	37
Table 34. Grade 10 Girls' Average Number of Absences at ML2 and Endline	43
Table 35. Grade 10 Girls' Perceptions of School Access and Safety.....	44

Table 36. Gender Perception Composite Scores Among Teachers.....	48
Table 37. Classroom Observation of Teacher Behaviour at ML2 and Endline.....	48
Table 38. Teacher Gender Perception Composite at ML2 and Endline	50
Table 39. Boys' Gender Perception Composite Score at Endline	50
Table 40. Learning Outcomes: Life Skills Composite and Gender Perception for Grade 10 girls.....	52
Table 41. Gender Perception Composite Grade 10 by Subgroup	52
Table 42. Life Skills Composite Grade 10 by Subgroup	53

Executive Summary

Background

The Supporting Transition of Adolescent Girls through Enhancing Systems (STAGES) project was part of the United Kingdom's Foreign, Commonwealth and Development Office's Girls' Education Challenge-Transition (GEC-T). It was implemented by Link Education International (Link) in Ethiopia's Wolaita Zone of the South Ethiopia Regional State where the primary local language is Wolaytatto. STAGES targeted 144 primary and secondary schools in 4 of the 22 zone's woredas—Damot Pulasa, Damot Sore, Damot Woide and Kindo Koisha—from 2017 to 2024. The project's theory of change (ToC) contended that improvements in attendance, quality in teaching and school leadership, management and governance, coupled with embedded positive community support for girls' education and support for the well-being and self-esteem of marginalised girls, will lead to better learning, transition and sustainability outcomes for these students. Over the life of the project, STAGES aimed to improve the literacy and numeracy skills of 63,644 girls.

Purpose of the Evaluation

The overall purpose of the series of STAGES evaluations is to assess the effectiveness, impact and value for money (VfM) of the STAGES project. The purpose of the endline evaluation is to collect empirical, mixed-methods information that allows STAGES to i) evaluate progress or changes since midline 2 (ML2) in 2022 via the project's evaluation questions, stated below and ii) to assess the efficiency, effectiveness, impact, sustainability and VfM of the project. The evaluation design for the STAGES project is intended to facilitate the measurement of primary and intermediate outcomes (IOs) to answer the following primary research questions:

- Evaluation Question 1: Was the STAGES project successfully designed and implemented? Was the STAGES project good value for money?
- Evaluation Question 2: What has been the change in terms of learning outcomes among girls enrolled in schools served by the STAGES project?
- Evaluation Question 3: What has been the change in transition of marginalised girls enrolled in schools served by the STAGES project?
- Evaluation Question 4: What parts of the intervention work to facilitate the transition of marginalised girls through education stages and increase their learning?

The endline evaluation will build on data collected in the prior evaluation time points to address these 4 evaluation questions, including conducting learning assessments to measure both literacy and numeracy competencies among girls in sampled schools.

Additionally, the analysis will thoroughly answer the overall logframe of indicators for the project. Qualitative data will inform a light-touch approach to VfM by highlighting areas perceived to be the most cost-effective. Qualitative data will also be used to understand the drivers and mechanisms behind the relationships established with the quantitative data.

Research Design

The evaluation used mixed methods, repeated measures, and a cross-sectional and non-experimental design. This design is most appropriate because the STAGES interventions reached all schools and girls within the selected grades in target woredas.

Due to the COVID-19 pandemic and a lack of access to comparison schools, the evaluation design differed significantly from what was laid out in the original inception report and MEL framework. Therefore, it is not possible to make any direct comparisons to baseline in this report.

Conducted in 2021, midline 1 (ML1) did not track individual girls assessed at baseline in either treatment or comparison groups because of school closures resulting from the pandemic. ML1 also did not include learning assessments; rather, it collected contextual data from girls, teachers, and government officials about general changes in learning, attitudes, and behaviours since the baseline. As schools reopened, ML2 sampled grade 8 and grade 10 girls who had also been assessed at baseline, as well as grade 6 girls for the first time.

At endline, girls who had been assessed as grade 8 (Cohort 1) students at ML2 were assessed as grade 10 (Cohort 1) students through learning assessments, surveys and FGDs. The data and outcomes from girls in grade 10 at endline (Cohort 1) were compared to those of girls enrolled in that same grade level at ML2 (Cohort 2). This cross-sectional grade-to-grade analysis guided the research findings and recommendations. Similarly, grade 6 students (no cohort) at ML2 were assessed as grade 8 students (no cohort) at endline through the learning assessments and survey. The data and outcomes of girls in grade 8 at endline were compared to the results of grade 6 girls at ML2, providing another cross-sectional comparison.

Sampled grade 10 girls were assessed on mathematics and literacy outcomes using the Secondary Grade Mathematics Assessment (SeGMA) and Secondary Grade Reading Assessment (SeGRA) tools developed at baseline, which were administered to a different cohort of grade 8 girls at that time point. Sampled grade 8 girls were assessed on mathematics and literacy outcomes using the Early Grade Mathematics Assessment (EGMA) and Early Grade Reading Assessment (EGRA), which were also developed at baseline and administered to a different cohort of grade 4 and 6 girls. All endline learning assessments were conducted in English, which is used as the language of instruction starting in grade 7.

The rationale for using grades 8 and 10 in the endline assessment is to ensure a meaningful comparison with midline for grade 10 as well as to relate to the baseline cohort. We also included grade 8 due to the significant emphasis placed on EGRA/EGMA in Ethiopia because SEGRA/SEGMA are less commonly utilized. Qualitative data was also collected from girls who had been assessed as grade 10 students at ML2 and were in grade 12 at endline through select FGDs, as well as a select group of grade 10 students. This was done to understand the transition outcomes for girls in grade 12. An overview of the cohorts sampled at each evaluation timepoint is presented in Table 1. Data was collected in May 2024.

Table 1. Cohorts Sampled by Evaluation Timepoint

	Baseline (2018)	Midline 1 (2021)	Midline 2 (2022)	Endline (2024)
Cohort 1	Grade 4	Grade 7	Grade 8	Grade 10
Cohort 2	Grade 6	Grade 9	Grade 10	Grade 12 ¹
Cohort 3	Grade 8	-	-	-
No Cohort²	-	-	Grade 6	Grade 8

The endline sample included 15 secondary schools and 8 primary schools, with a target sample of 20 grade 10 girls per secondary school and 20 grade 8 girls per primary school. The 15 secondary schools were randomly sampled. The eight primary schools were selected based on their proximity to the

¹ Data from Grade 12 girls was collected through select FGDs only. No quantitative data was collected from this grade.

² This is the extra cohort sampled at endline.

secondary school, with the nearest ones chosen due to logistical and budgetary constraints.. One implication of this decision is that the findings might indicate an easier transition for girls in these primary schools to secondary schools due to geographical proximity, potentially overlooking variations in educational outcomes and accessibility experienced by more remote schools. This selection criterion might not fully capture the diverse educational landscape and could limit the generalizability of the results to all primary schools in the region. The total number of girls sampled was 453. Other surveys were conducted with boys, classroom teachers, head teachers and woreda officials. Classroom teachers were also observed delivering a lesson. In addition to the 4 focus group discussions (FGDs) with girls, qualitative data was also collected with one FGD apiece with zonal and woreda education officials, 2 FGDs with teachers and 4 key informant interviews (KIs) with head teachers.

Past Reading Outcomes in Ethiopia

In addition to presenting results of the STAGES endline evaluation, this report considers those results within the wider language learning context in Ethiopia, including the implications of key policy decisions and past performance on reading assessments over recent years. Overall, the endline evaluation highlights key successes in the project objectives targeting learning, transition and sustainability outcomes. In reading the endline findings, it is important to consider 2 major contextual factors. First, the endline evaluation was conducted in May 2024 during a turbulent time for secondary schools across the Wolaita Zone. Government delays in paying teachers during the school year led to school closures and decreases in learner attendance, according to respondents in FGDs and KIs. Overall, each of the project-supported 17 secondary schools were closed for at least 20 days. Second, girls in the Wolaita Zone learn in a complex language environment in which the language of instruction switches to English in grade 7. However, girls supported by STAGES, who live mainly in remote and rural areas and are categorised as educationally marginalised, are not exposed to spoken or written English in their home or broader community environment. Findings from ML1 highlighted language of instruction to be a factor in girls' thoughts of dropping out of school and not transitioning to the next grade.

Data from other studies and projects shows the environment for literacy is 'hostile', with reading comprehension being a challenging skill for students. Early-grade reading performance has been assessed using the EGRA tool at least 6 times between 2010 and 2018. In 2018, the USAID-funded Reading for Ethiopia's Achievement Developed Monitoring and Evaluation (READ M&E) project carried out the endline administration of the EGRA in June 2018 to evaluate the reading fluency and comprehension levels of students in grades 2 and 3 within 5 regions of Ethiopia in 7 languages: Afaan Oromo, Af Somali, Amharic, Hadiysa, Sidamu Affo, Wolaytta and Tigrigna.³ The 2018 EGRA endline administration followed the 2014 baseline and 2016 midline administrations.

Two key measures of early-grade reading include oral reading fluency (ORF) and reading comprehension. The 2018 EGRA results show that 6.2% of Ethiopian students in grades 2 and 3 and across all languages combined achieved the targeted reading benchmark—reading fluently with full or almost full comprehension. The overall Ethiopian percentage of learners 'who demonstrate reading fluency and comprehension of grade level text at the end of grade 2 with USG assistance' was 25.0%, with Tigrigna and Amharic learners at the top (41.3% and 40.7%, respectively) and Hadiysa learners at the bottom of the range (9.1%).

³ American Institutes of Research. (2019). USAID Reading for Ethiopia's Achievement Developed Monitoring and Evaluation (READ M&E): Early grade reading assessment (EGRA): 2018 endline report. Retrieved on 12/1/2022 from https://pdf.usaid.gov/pdf_docs/PA00X5JW.pdf

In June 2014, RTI conducted a baseline EGRA for the Haddiysa and Wolaytatto languages. The results revealed that some Haddiysa- and Wolaytatto-speaking students were only beginning to learn to read in their respective languages by grade 3. For the READ M&E EGRA 2018 Endline, grade 2 and 3 Wolaytatto-speaking students were assessed. It was found that the mean ORF score for grade 2 students was 8.9 correct words per minute (CWPM), and for grade 3 students was 18.9 CWPM. Among the same students, boys' mean fluency scores were higher than girls but those differences, although statistically significant, were small and reported as practically insignificant.

The mean reading comprehension score among Wolaytatto speakers in 2018 at the READ M&E Endline evaluation was 13.11 for grade 2 students and 26.91 for grade 3 students. As seen in ORF outcomes, boys consistently outperformed girls in reading comprehension in all Ethiopian languages tested, including Wolaytatto, except for Amharic. This was true for grades 2 and 3. The differences grew between grades 2 and 3. As the results indicate, in 2018, Wolaytatto-speaking children in grades 2 and 3 were performing at low levels on the key reading indicators of ORF and reading comprehension. Girls performed even lower than boys on both measures.

Overall, the EGRA results suggest that a small percentage of Ethiopian grade 2 and 3 students tested were reading with comprehension. This percentage was even smaller among Wolaytatto-speaking students and girls. The project could not find any past projects that utilised the early grade numeracy assessment to provide context for the proceeding discussion of findings.

Endline Findings

Overall, the vast majority of girls at endline were considered improved learners and reported wanting to stay enrolled in school. In terms of learning outcomes, 85.4% of grade 8 girls were improved learners at endline in literacy and 73.3% of grade 8 girls were improved learners in numeracy. Considering transition outcomes, 80.2% of grade 8 girls at endline said they wanted to transition to secondary schools, while 93.1% of grade 10 girls at endline reported they EL girls want to transition to grade 11.⁴

The project saw significant gains in average literacy and numeracy scores among grade 8 girls. Specifically, girls in grade 8 at endline had statistically significantly higher aggregate literacy scores in English compared to those in grade 6 at ML2, indicating a positive trend, as shown in Table 2, as well as in numeracy, as shown in Table 3. The gains were statistically significant from ML2 to endline on all literacy and numeracy subtasks. After ML2, STAGES introduced new literacy interventions t

However, this increase cannot be solely attributed to the intervention due to the absence of a true longitudinal design. Additionally, the 2 years of schooling and natural development may have influenced scores, though students faced challenging external circumstances, which STAGES sought to mitigate due to political developments in Ethiopia and the impact of inflation.

Table 2. Change in Literacy Mean Scores from Grade 6 ML2 to Grade 8 Endline

	G6 ML2 Literacy Treatment (n = 269)	G8 Literacy Treatment (n = 157)	Endline	Difference to Endline	ML2	Statistical Significance – Robust Regression	OLS
--	-------------------------------------	---------------------------------	---------	-----------------------	-----	--	-----

⁴ Since there was no comparison group, and the comparison was made between grade 6 at midline and grade 8 at endline, it is difficult to attribute the gains solely to the intervention. Potential confounding factors include the additional years of schooling and natural development over time.

	G6 ML2 Literacy Treatment (n = 269)	G8 Endline Literacy Treatment (n = 157)	Difference to Endline	ML2	Statistical Significance – Robust Regression	OLS
Score	36.3	68.7	+32.46***		P = 0.000	

Note: Three asterisks (***) denote statistically significant at $p < 0.05$.

Table 3. Change in Numeracy Mean Scores from Grade 6 ML2 to Grade 8 Endline

Grade	G6 Midline 2 Numeracy Treatment	G8 Endline Numeracy Treatment	Difference	P value
Percentage Score	52.69	63.53	+10.83***	P = 0.000

Note: Three asterisks (***) denote statistically significant at $p < 0.05$.

Grade 10 girls' performance in literacy and numeracy remained unchanged from ML2 to endline despite difficult external factors that girls faced this school year. When comparing ML2 and endline literacy and numeracy scores of grade 10 girls, as shown in Table 3 and Table 4, respectively, no statistically significant differences were found, indicating similar scores. Students at secondary schools in the Wolaita Zone faced notable obstacles this school year in the months preceding endline, with considerable loss of instructional time due to teacher strikes, a shortage of textbooks related to adoption of a new curriculum and inflationary pressures affecting student motivation. When analysing these results, these challenges, namely the teacher strikes, must be considered, especially that SeGRA and SeGMA scores did not decline despite the challenging conditions.

Table 4. Grade 10 Aggregate Literacy Outcomes at ML2 and Endline (%)

Timepoint	Reading Comprehension	Vocabulary	Grammar	Overall
Midline	55.2	31.37	40.3	41.1
Endline	53.1	35.5	38.29	40.8
Change from Midline to Endline	-2.04	+2.92	-1.95	-0.22

Table 5. Grade 10 ML2 and Endline Numeracy Outcomes (%)

Timepoint	Geometry	Multiplication	Fraction	Overall
Midline	44.34	50.55	29.58	41.49
Endline	41.78	47.05	31.06	39.96
Change	-2.6	-3.5	+1.48	-1.52

In subgroup analysis by woreda, overall, Damot Sore and Damot Pulasa showed strong performance in both literacy and numeracy outcomes for grade 8, with Damot Pulasa maintaining this performance into grade 10. Kindo Koisha consistently had the lowest mean scores in both literacy and numeracy across grades 8 and 10, raising questions about what factors may be associated with those low scores compared with the other 3 project woredas. As for other subgroups, findings suggest targeted interventions may have positively impacted grade 8 girls, particularly those facing specific disabilities or socioeconomic challenges.

Girls who were surveyed reported having a clear desire to continue their education. At endline, the overwhelming majority of grade 10 girls—93.1%—indicated that they want to move on to grade 11 the following school year, which was a slight increase from ML2 (91.97%).

Early marriage, lack of parental support, early pregnancy and illness were the 4 reasons that girls cited most frequently at endline for preventing their peers from transitioning. These factors are notably related to girls' lives at home. Importantly, these home-related factors all decreased from ML2 to endline, indicating that the project interventions might have addressed these home-based barriers.

The proportion of grade 10 girls who stated they had thought about dropping out of school either rarely, sometimes or never decreased from more than a quarter at ML2 (25.56%) to about a tenth at endline (9.52%). The reasons for dropping out that girls cited most frequently were the lack of parental support and the greater importance of work and earning income.

As for sustainability, overall, it was clear from FGDs with zonal and woreda officials that sustaining key project interventions has been at the top of stakeholders' minds with the project nearing its end. In its long-term sustainability plan, STAGES identified 3 primary interventions it deemed to have the most potential for sustainability—pedagogical leadership and supervision (including school improvement planning), gender and inclusive responsive pedagogy (GIRP) and safeguarding mechanisms and practices (including through community-based mothers' and fathers' groups). Zonal and woreda officials reported that these specifically targeted interventions have been successfully incorporated and scaled at varying levels into the Wolaita Zone's educational system, with pedagogical leadership and supervision furthest along in being integrated, followed by GIRP and then safeguarding. Further, zonal officials said they are using their own budget and partnerships to fund the GIRP expansion, but funding constraints have prevented it from scaling it across the entire zone.

Intermediate Outcome Findings

Attendance among grade 10 girls was similar from ML2 to endline. About three-quarters of girls reported that they did not miss any days of school in the past week of school at ML2 and endline.

Teachers generally said they believed key education actors were responsive to the needs of girls in school. Nearly all 15 teachers surveyed said that cluster supervisors were mostly (8 of 15) or always (5 of 15) responsive to girls' needs, as well as school leadership (7 each responded mostly and always). By contrast, teachers said they believed woreda officials were not as responsive as cluster supervisors or school leadership, with only 3 of 15 saying woreda officials were mostly responsive and only 1 of 15 saying they were always responsive.

The proportion of teachers who reported receiving training in their subject matter, classroom management and GIRP increased statistically significantly from ML2 to endline. These increases indicate that interventions focused on teacher professional development have been implemented effectively.

Teachers and boys had high scores on their gender perception composite scores at endline, suggesting all these groups have supportive, positive views of girls and girls' education. The teachers' score increased slightly from ML2 to endline. The gain was not statistically significant, but there was not much room for growth due to a high ML2 score.

Girls' life skills⁵ and gender perceptions⁶ composite scores increased statistically significantly from ML2 to endline, suggesting that there were notable changes in girls' life skills and gender perceptions from ML2 to endline.

Qualitative data corroborates the overall gains in girls' gender perceptions and life skills, with head teachers, teachers and girls themselves describing the positive change in well-being. Respondents said the gains had occurred for multiple reasons, including improved gender-inclusive teaching practices, greater community acceptance of girls' education and numerous project activities focused on girls' self-esteem and well-being. In FGDs, girls noted how the social-emotional learning (SEL) training had benefited them in multiple ways by building their confidence, improving their interpersonal skills and helping them manage their emotions better.

Conclusions

The statistically significant gains in grade 8 girls' reading performance at endline from grade 6 girls' outcomes at ML2 revealed stronger performance in familiar word reading, invented word reading and ORF than reading comprehension, although gains were also statistically significant on this subtask from ML2 to endline. Grade 8 girls' percentage scores at endline were highest on ORF (83.81%), invented word reading (81.45%) and familiar word reading (71.12%) and lowest on reading comprehension (45.98%). Although the endline percentage score on reading comprehension for grade 8 girls was a statistically significant increase from the grade 6 percentage score of 16.5% at ML2, it underscores the continued gap between girls being able to read aloud the words in the reading passage and their ability to make sense of what they are reading.

It is also important to note that the statistically significant gains from grade 6 at ML2 to grade 8 at endline cannot be solely attributed to the intervention for several reasons. First, the evaluation design lacked a true longitudinal design, as well as a counterfactual. Additionally, 2 years of schooling may have influenced scores. Still, the gains in reading outcomes do underscore the importance of the literacy interventions that STAGES introduced after ML2 to improve the literacy environment in schools, and the two additional years of schooling were delivered within the timeframe of project support.

Grade 8 girls also performed better on numeracy subtasks, with statistically significant gains on all subtasks. Grade 8 girls' numeracy scores improved at endline on all 3 subtasks that grade 6 girls also were administered at ML2, with the opportunity for even further improvement and growth in numeracy outcomes.

The impact of lengthy secondary school closures on grade 10 girls' endline learning outcomes must be considered when evaluating the unchanged SeGMA and SeGRA results from ML2 to endline. Overall, grade 10 girls' mean SeGRA and SeGMA aggregate scores slightly declined from ML2 to endline, but the decreases were not statistically significant. These results, however, should be interpreted with a major caveat—the closure of secondary schools in the Wolaita Zone in February and March 2024 before the endline evaluation was administered in May. All 17 secondary schools supported by STAGES were

⁵ The gender perceptions composite score was an aggregate of the following questions: It is important for girls to go to school; Education is more important for boys than for girls; Girls learn the same at school as boys; A woman's role is to do household jobs and raise children; Men should share household duties; Boys are more naturally skilled than girls at reading and writing; Boys are more naturally skilled than girls at mathematics. Scale 0-7

⁶ The life skills composite score was an aggregate of the following questions: I get nervous when I have to read in front of others; I get nervous when I have to do maths in front of others; I feel confident answering questions in class; I would like to continue studying/attending school after this year; I often feel lonely at school. Scale 0-5

closed for at least 20 days, due to teacher protests about delays in receiving their salaries. Additionally, the extra tutorial sessions at schools that multiple respondents credited as having an important impact on girls' learning also stopped because schools were closed. Girls shared in FGDs how the closures affected their motivation. Teachers reported that the closures affected attendance and the quality of teaching and learning, with project staff also detailing how classes were rushed to make up for lost instructional time. Overall, these substandard conditions very likely had a negative impact on learning outcomes in secondary schools.

Girls in Kindo Koisha had the lowest literacy and numeracy scores out of the 4 woredas in both grade 8 and grade 10. In the subgroup analysis of learning outcomes, the comparatively poor performance of girls in Kindo Koisha was notable at endline. By contrast, at ML2, grade 10 girls in Kindo Koisha, which has deteriorated politically since being stable in 2022, had the highest SeGRA and SeGMA scores out of the 4 woredas. This downward trend underscores the likely impact of external factors on learning outcomes.

The multiple positive trends found in transition outcome data suggest that STAGES has had success in addressing the multiple barriers girls face in transitioning to secondary school:

- First, the home-based factors that girls cited most frequently for preventing their peers from transitioning all declined from ML2 to endline.
- Second, girls' socio-emotional composites measuring gender and life skills increased statistically significantly from ML2 to endline.
- Third, the proportion of grade 10 girls who sometimes or often considered dropping out decreased from 18.08% at ML2 to 6.80% at endline.
- Finally, in FGDs, girls and teachers spoke positively about the changes they had observed in families' attitudes, as well as in communities' beliefs as a whole, with respect to girls' education.

Overall, all these findings suggest that project interventions have successfully mitigated the factors that prevent girls from transitioning to secondary school. Findings also suggest that families are better equipped to support their children's educational needs, as a statistically significant difference was noted between ML2 and endline in the proportion of girls who said their parents or caregivers pay for everything they need to go to school—from 56.46% to 79.25%, respectively.

As for sustainability, stakeholders at all levels seem prepared to carry out key interventions once the project closes. The efforts that STAGES made during the life of the project to target all levels of the education system—from the zonal education office to communities and classrooms—appear to have fostered a sense of ownership in project interventions in all stakeholders and set up all education actors to sustain them once the project closes. It is especially notable that the zonal office said it has already scaled the pedagogical supervision and leadership activity to all woredas in the zone and trained other regional education actors beyond the zone on the intervention, as well as reported using its own funds to expand the GIRP intervention to non-project woredas in the zone. Zonal and woreda officials seem poised to sustain STAGES activities, with interventions incorporated into 3-year strategic planning, certain project resources such as teacher mentoring and coaching checklists aligned with zonal monitoring tools and school improvement planning and mothers' and fathers' groups now considering the importance of gender and inclusion in their efforts.

Recommendations

The findings from endline suggest 9 main recommendations:

1. Focus on improving students' reading comprehension skills in both English and Wolaytatto in future iterations of literacy interventions in the Wolaita Zone, primarily by training teachers

intensively on classroom activities and instructional strategies focused on reading comprehension.

2. Expand tutorials to focus on improving knowledge in other subjects and/or honing English skills.
3. Modify the SeGRA and SeGMA if needed for future iterations of education interventions, with evaluators, implementing partners and education officials exploring the most reliable and valid way to measure higher-level learning skills.
4. Research why literacy and numeracy results in Kindo Koisha were lower than other woredas' scores.
5. Look into avenues for the zone to continue to scale up project interventions, not only including pedagogical leadership and supervision, GIRP and safeguarding mechanisms but also effective teaching and learning materials that produced gains in numeracy.
6. Connect zonal and woreda education officials to subject-based experts and other resources so they can update materials as needed after the project closes and promote the institutionalisation of materials beyond the zonal level.
7. Host a zone-wide meeting of fathers' and mothers' groups to validate their importance, promote best practices and encourage continued community involvement in education.
8. Explore ways to enable girls who have benefited from the project to apply what they have learnt, possibly by becoming mentors in their communities, to promote the success of SEL interventions.
9. Identify communities that most effectively implemented safeguarding mechanisms and fostered successful coordination between stakeholders so education officials can promote these practices across the Wolaita Zone.
10. Further research teachers' and head teachers' beliefs about support from middle management at the woreda level and intensify support to woreda education offices as needed to support the transformation of attitudes and support for girls and leadership for girls' education.

1. Background to Project

The Supporting Transition of Adolescent Girls through Enhancing Systems (STAGES) project was part of the United Kingdom's Foreign, Commonwealth and Development Office's Girls' Education Challenge-Transition (GEC-T). It was implemented by Link Education International (Link) in Ethiopia's Wolaita Zone of the South Ethiopia Regional State where the primary local language is Wolaytatto. STAGES targeted 144 primary and secondary schools in 4 of the 22 zone's woredas—Damot Pulasa, Damot Sore, Damot Woide and Kindo Koisha—from 2017 to 2024. Over the life of the project, STAGES aimed to improve the literacy and numeracy skills of 63,644 girls.

1.1 Project Theory of Change

The project's theory of change identifies 5 intermediate outcomes (IOs) that the project needs to achieve to reach its longer-term outcomes and goal. IOs are organised around 5 core themes:

- **IO 1—Attendance and retention:** Key activities include providing girls with materials and bursaries, as well as feminine hygiene products in schools, and holding additional tutorial sessions at schools to reinforce learning. To measure improved attendance and retention, the IO's key indicators included the percentage improvement in attendance rates and the percentage of girls with improved perceptions of access.
- **IO 2—School leadership, management and governance:** Key activities include training linked to existing continuous professional development (CPD) for head teachers and support of community-school structures, including PTAs, School Improvement Committees and Kebele Education and Training Boards. To measure improved school leadership, management and governance, the IO's key indicator measured the number of schools developing and implementing high quality school improvement plans which are gender, inclusion and safeguarding responsive.
- **IO 3—Quality of teaching:** Key activities include training, regular mentoring and coaching support for teachers, encouragement on communities of practice at school and cluster level. To measure improved quality of teaching, the IO's key indicator was the percentage of teachers who demonstrate improved methodology and practice.
- **IO 4—Community-based attitudes and behaviours:** Key activities include establishing gender clubs and strengthening community-school structures, as mentioned above, as well as mother and father groups. To measure improved community-based attitudes and behaviours, this IO's key indicators included the percentage of teachers and head teachers who report positive changes in gender perceptions and gender-sensitive teaching, the percentage of parents reporting greater support for secondary education, especially for girls, the percentage of girls reporting positive changes in gender perceptions and gender sensitive teaching and the percentage of boys' reporting positive perceptions of the value of girls' education.
- **IO 5—Girls' wellbeing and self-esteem:** Key activities include training teachers on gender and inclusive responsive pedagogy (GIRP), social and emotional learning approaches and school-related gender-based violence (SRGBV); providing direct socioemotional and counselling support to girls; embedding mechanisms to safeguard children in schools; and establishing gender-inclusion school improvement and action plans. To measure improvements in girls' wellbeing and self-esteem, the IO's key indicators were the percentage of girls reporting improved wellbeing and the percentage of girls reporting improved self-esteem.

The project has also introduced new literacy and numeracy interventions as a response to prior evaluations. Specifically, in numeracy, the project underwent major adaptations since baseline,

incorporating numeracy teacher training, assessing gaps in numeracy teaching, observing numeracy classes, intensifying follow-up coaching and mentoring of teachers for regular feedback and support and developing high-quality numeracy training materials to improve the teaching of numeracy in supported schools. For literacy, in response to reading comprehension outcomes at ML2, STAGES established and supported reading corners in schools, provided schools with reading books and conducted additional teacher training on teaching reading.

1.2 Key Evaluation Questions and Role of the Endline Evaluation

The purpose of the endline evaluation was to collect empirical, mixed-methods information that enables STAGES to complete two different tasks:

- evaluate progress or changes since midline 2 (ML2) via the project's evaluation questions; and
- to assess the efficiency, effectiveness, impact, sustainability and value for money (VfM) of the project.

The evaluation design for the STAGES project is intended to facilitate the measurement of primary and intermediate objectives (IOs) needed to answer 4 primary research questions:

- Evaluation Question 1: Was the STAGES project successfully designed and implemented? Was the STAGES project good value for money?
- Evaluation Question 2: What has been the change in terms of learning outcomes among girls enrolled in schools served by the STAGES project?
- Evaluation Question 3: What has been the change in transition of marginalised girls enrolled in schools served by the STAGES project?
- Evaluation Question 4: What parts of the intervention work to facilitate the transition of marginalised girls through education stages and increase their learning?

1.3 Project Context

The endline evaluation was conducted in the spring of 2024 during a turbulent time for secondary schools across Ethiopia. Government delays in paying teachers during the school year led to school closures and decreases in learner attendance, according to respondents in FGDs and KIIs. Overall, each of the project-supported 17 secondary schools were closed for at least 20 days in February and March. In addition to teachers not getting paid, STAGES stopped paying tutors to lead tutorial classes in school because teachers were absent.

A group of secondary school teachers at one school explained how their school had closed multiple times during the school year, including at the start of the school year and the beginning of the second semester. In the FGD conducted in mid-May, a teacher said they were worried the school would close again because they still had not received their salaries from April. 'Teachers are neglected, especially those in rural schools', another teacher said during the FGD. 'We have not been paid for an extended period. This has affected attendance, continuation and quality of teaching and learning'.

A head teacher at a secondary school in a different woreda also reported the issue was affecting his school. 'Teachers' salary delay inconveniences the project activities currently', he said. 'Especially, since 2 months ago, the teachers are not coming to school, and the students are also not attending their classes'. In an FGD at the same school, grade 12 girls confirmed that teachers had stopped coming to class. 'In grade 11, I was happy to go to school. ... But some things are not pleasant in the current situation', one of the girls said, 'and teachers don't teach because they don't get paid. I don't like going to school because of this'.

Students in secondary schools also faced other obstacles. Due to the adoption of a new curriculum, schools faced a shortage of textbooks, which then led to minimal opportunity for developing key reading skills such as reading comprehension. Inflationary pressures also affected student motivation because students were compelled to focus less on their educational pursuits than economic ones.

It is also important to note the COVID-19 pandemic started 2 years after STAGES began in 2018. While the endline evaluation occurred 4 years after the height of the pandemic when schools were closed in Ethiopia and around the world, the lingering effects of those closures is likely still impacting learning. As reported in Table 6, 63.6% of girls surveyed indicated that they learned less during the pandemic than during a normal school year.

Table 6. Effects of Pandemic on Learning

During the pandemic, did you learn the same, more, or less than a normal year?	Student's Grade		Total
	Grade 8	Grade 10	
Less than a normal year	101	186	287
The same as in a normal year	8	9	17
More than in a normal year	45	82	127
I do not know	2	16	18
No response	1	1	2
Total	157	294	451

2. Research Design

2.1 Evaluation Approach and Methodology

The evaluation utilised a mixed-methods, repeated measure, cross-sectional and non-experimental design. This design is most appropriate because the STAGES interventions were designed to reach all schools and girls within the selected grades in target woredas. No comparison schools were available at ML2 and were likewise not visited at endline. Furthermore, since the current project works with both primary and secondary girls—as opposed to primary-grade girls only in Link's Girls' Education Challenge 1 project—the design originally followed 3 cohorts of girls as they advanced through higher grades. This was done instead of adding new cohorts of girls as they entered primary school. Monitoring data complemented the data collected at the 4 evaluation points. Throughout all data collections, an effort has been made to minimise the amount of data collection conducted outside of STAGES-focused schools.

Due to the COVID-19 pandemic and a lack of access to comparison schools, the evaluation design differed significantly from what was laid out in the original inception report and MEL framework. While the project originally aimed to recontact previously sampled girls at baseline, logically, it was impossible to do so. Attrition rates at baseline were much higher than expected for the project. Therefore, it is not possible to make any direct comparisons to baseline results in this endline report.

Following the protocol that was established at ML2, a random sample was drawn from girls in grades 8 and 10 at EL. Analyses were conducted at the cohort level rather than at the individual level. Specifically, for grade 10, School-to-School International (STS) conducted grade-by-grade comparisons using a random sample of girls. For example, the performance of grade 10 girls at endline was compared to the performance of grade 10 girls at ML2. For grade 8, STS similarly conducted a grade-by-grade comparison using a sample of random girls—this time comparing the performance of grade 8 girls at

endline to the performance of grade 6 girls at ML2. In addition, STS identified subgroups of girls—for example, by woreda or marginalisation level—and examined growth trends within each subgroup. Doing so enabled STS to understand whether the intervention has been equally effective for different girls.

2.2 Quantitative Data Collection Tools

The quantitative data collection tools were designed to answer the aforementioned evaluation questions. The focus of the endline evaluation was primarily quantitative, with qualitative tools focused on sustainability. Qualitative data also informed a light-touch approach to VfM by highlighting areas perceived to be the most cost-effective and sustainable.

Sampled grade 10 girls were assessed on mathematics and literacy outcomes using the Secondary Grade Mathematics Assessment (SeGMA) and Secondary Grade Reading Assessment (SeGRA) tools developed at baseline, which were administered to a different cohort of grade 8 girls at that time point. Sampled grade 8 girls were assessed on mathematics and literacy outcomes using the Early Grade Mathematics Assessment (EGMA) and Early Grade Reading Assessment (EGRA), which were also developed at baseline and administered to a different cohort of grade 4 and 6 girls. In addition, a sample of grade 10 boys and woreda officials was used to gather information on their beliefs around gender and other relevant information related to the research questions. In each of the secondary schools visited, the head teacher was surveyed, and one grade 10 classroom was observed. The teacher was interviewed. Endline data collection occurred from 13–17 May 2024.

As shown in Table 7, close to 158 grade 8 and 295 grade 10 girls were assessed in both literacy and numeracy. In total, 453 girls had their literacy and numeracy skills measured.

Table 7. Learning Assessment Tools

Grade	Tool	Sample
8	Learning Assessment	158
10	Learning Assessment	295

The survey instruments were designed to connect with learning assessment data, offering greater insight into the ways that STAGES IOs— i) attendance, ii) quality in teaching, iii) school management and governance, iv) community attitudes and v) girls' well-being—were linked to the outcomes of learning and transition. Data from the survey instruments were considered alongside qualitative data to better understand the socioeconomic dynamics surrounding girls' transition and the factors contributing to the program's sustainability.

During the endline evaluation, the project also surveyed numerous other community actors to gain a clearer picture of the learning environment for girls. As a comparison group and as a measure of learning environment, the project surveyed 45 boys in grade 10. This survey did not include learning assessments for boys in literacy or numeracy but did include measures on their gender perceptions, school environment and learning experiences.

Teachers were also surveyed and observed. In grade 10, 15 teachers were surveyed, including 14 men and one woman. Enumerators additionally observed one grade 10 teacher at each secondary school. Classroom observations were broken down by grade and by teacher gender. In grade 10, 15 classrooms were observed, including 14 led by men and one by a woman. This gender imbalance reflected the reality of school employment.

Table 8. Quantitative Data Collection Tools

Grade	Girls student survey	Boys student survey	Classroom teacher survey	Classroom observation	Head teacher survey	Woreda official survey
Grade 8	159	n/a	n/a	n/a		
Grade 10	294	45	T:15	T:15	15	30
			M:14	M:14		
			W:1	W:1		
Total Sample	453					

2.3 Qualitative Data Collection Tools

The purpose of the qualitative instruments was to provide a deeper understanding of sustainability, as well as enrich the quantitative component of the endline study. Qualitative data collection helped evaluators understand the context of Link interventions and student outcomes by describing the conditions that facilitate and mitigate the intended outcomes. This process also provided data for the sustainability and VfM outcomes, as well as case studies to highlight in the report.

Eight focus group discussions (FGDs) and 4 key informant interviews (KIs) were conducted. The qualitative sample included all 4 woredas involved in STAGES. Two data collectors—one interviewer and one notetaker—conducted FGDs with woreda-level and zone-level officials. They also visited one school in each woreda and conducted KIs with the head teacher, as well as FGDs with teachers, grade 10 girls and grade 12 girls. Eighteen teachers overall participated in 2 FGDs. Data collection occurred in May 2024.

Table 9 and Table 10 provide detailed information on the qualitative and quantitative data collection tools, along with the corresponding number and type of respondents, as well as the outcomes and IOs.

Table 9. Qualitative Tools, Sample and Corresponding Outcomes and Intermediate Outcomes

Qualitative Instrument	Number and type of participants targeted	Total number of focus group discussions or key informant interviews	Total number of participants	Outcome and Intermediate Outcomes
Girls' FGDs	Up to 8 Grade 10	2	16	O2
	Up to 8 Grade 12	2	16	
Teacher FGDs	Up to 8 participants each (preferably from targeted grade levels, possibly divided into male and female teachers)	2	15	O3, IO2, IO3
Zone-level official	Up to 8 zone-level	1	8	O3

Qualitative Instrument	Number and type of participants targeted	Total number of focus group discussions or key informant interviews	Total number of participants	Outcome and Intermediate Outcomes
FGD	officials			
Woreda-level official FGD	Up to 8 Woreda-level officials	1	8	O3
Total Focus Group Discussions		8	63	
School management KIIs	4 participants (one head teacher per woreda)	4	4	O3, IO2
Total KIIs		4	4	

Table 10. Quantitative Tools, Sample and Corresponding Outcomes and Intermediate Outcomes

Quantitative Instrument	Type of participants targeted	Number per school	Total number	IO
SeGRA, SeGMA and girl survey	Grade 10 girls	20	300 (15 secondary schools)	O1, O2, IO1, IO5
EGRA, EGMA and girls survey	Grade 8 girls	20	160 (8 primary schools)	O1, O2
Head teacher survey	Secondary school head teachers	1	15	O3, IO2, IO4
Classroom teacher survey	Grade 10 teacher	1	15	O3, IO3, IO4
Classroom observations	Grade 10 classroom	1	15	IO2, IO3
Boys survey	Grade 10 boys	3	45	IO4
Woreda staff survey	Woreda staff	N/A	30	O3

2.4 Sample

The endline sample includes 15 secondary schools and 8 primary schools, with a target sample of 20 grade 10 girls per secondary school and 20 grade 8 girls per primary schools. The total number of girls sampled was 453. More information on the final quantitative sample is shown in Table 11 to Table 16.

Sampling for the grade 10 qualitative component was done from within the sample used for the quantitative data collection. Grade 12 girls, who were interviewed through the qualitative component but did not take part in the quantitative assessment, were sampled from the schools selected for the quantitative assessment. The sampling strategy was purposive, selecting participants that would provide the most insight into the research questions. This strategy aimed to select participants that would

represent a stratification across key characteristics of interest that might have intersectional relationships with the outcomes of interest.

Table 11. Quantitative Sample of Girls

Woreda	Number of primary schools sampled	Number of secondary schools	Grade 8	Grade 10
Damot Pulasa	2	4	40	75
Damot Sore	2	3	40	80
Damot Woide	2	5	38	80
Kindo Koisha	2	3	40	60
Total girls sampled			158	295
Target Sample			160	300

Table 12. Quantitative Sample of Boys

Woreda	Number of secondary schools sampled	Grade 10
Damot Pulasa	4	12
Damot Sore	3	9
Damot Woide	5	15
Kindo Koisha	3	9
Total boys		45

Table 13. Quantitative Sample by Age of Girls

Age bands	ML2	Endline
Aged 6–8 (percentage of overall sample)	0 (0%)	0 (0%)
Aged 9–11 (percentage of overall sample)	5 (.56%) ⁷	0 (0%)
Aged 12–13 (percentage of overall sample)	126 (14.2%)	7 (1.55%)
Aged 14–15 (percentage of overall sample)	337 (38.2%)	125 (27.71%)
Aged 16–17 (percentage of overall sample)	258 (29.2%)	211 (46.79%)
Aged 18–19 (percentage of overall sample)	87 (9.86%)	94 (20.84%)
Aged 20+ (percentage of overall sample)	32 (3.62%)	13 (2.87%)
Age missing	0	1 (0.22%)
Girls	882	451
Underage for grade (percentage of overall sample)	43 (5.26%)	42 (9.3%)
On-age for grade (percentage of overall sample)	492 (55.78%)	265 (58.7%)

Age bands	ML2	Endline
Overage for grade (percentage of overall sample)	347 (39.34%)	144 (31.9%)

Table 14. Quantitative Sample by Age of Boys

Age Bands	ML2	Endline
Aged 6–8 (percentage of overall sample)	0	0
Aged 9–11 (percentage of overall sample)	1 (.43%)	0
Aged 12–13 (percentage of overall sample)	33 (14.35%)	0
Aged 14–15 (percentage of overall sample)	72 (31.30%)	1 (2.22%)
Aged 16–17 (percentage of overall sample)	57 (24.78%)	19 (42.23%)
Aged 18–19 (percentage of overall sample)	41 (17.83%)	16 (35.55%)
Aged 20+ (percentage of overall sample)	25 (10.87%)	9 (19.99%)
Boys	229	45
Underage for grade (percentage of overall sample)	11 (4.80%)	1 (2.22%)
On-age for grade (percentage of overall sample)	83 (36.24%)	19 (42.22%)
Overage for grade (percentage of overall sample)	135 (58.95%)	25 (55.56%)

Table 15. Endline Sample by Disability

Sample breakdown (girls)	ML2	Endline
Girls with disability (percentage overall)⁸	Grade 6: 60 (20.55%) Grade 8: 58 (19.93%) Grade 10: 74 (24.75)	Grade 8: 39 (24.8%) Grade 10: 54 (18.37%)
Vision impairment (percentage overall)	Grade 6: 18 (6.16%) Grade 8: 11 (3.78%) Grade 10: 10	Grade 8: 18 (11.47%) Grade 10: 20 (6.8%)
Hearing impairment (percentage overall)	Grade 6: 10 (3.42%) Grade 8: 7 (2.41%) Grade 10: 1 (.34%)	Grade 8: 18 (11.47%) Grade 10: 16 (5.44%)
Mobility impairment (percentage overall)	Grade 6: 4 (1.37%) Grade 8: 9 (3.09%) Grade 10: 3 (1.00%)	Grade 8: 19 (12.10%) Grade 10: 17 (5.78%)
Cognitive impairment (percentage overall)	Grade 6: 30 (10.27%) Grade 8: 32 (11.0%) Grade 10: 53 (17.73%)	Grade 8: 23 (14.65%) Grade 10: 26 (8.84%)

⁸ Girls were coded as having a disability if they stated they had any of the impairment (cannot do at all to yes, some difficulty) included in the table: vision impairment, hearing impairment, mobility impairment, cognitive impairment.

Sample breakdown (girls)	ML2	Endline
Self-care impairment (percentage overall)	Grade 6: 8 (2.74%) Grade 8: 6 (2.07%) Grade 10: 8 (2.68%)	Grade 8: 7 (10.83%) Grade 10: 17 (5.78%)
Communication impairment (percentage overall)	Grade 6: 10 (3.42%) Grade 8: 17 (5.86%) Grade 10: 41 (13.71%)	Grade 8: 19 (12.1%) Grade 10: 20 (6.8%)

Table 16. Girls' and Boys' Characteristics

Construct of interest	ML2 ⁹	Endline
Girls' Characteristics		
Mothers (percentage of the total sample) - Under 18 - Under 16	227 (25.74 %) of the sample answered that a girl in their class was pregnant.	174 (38.58%) of girls answered that a girl in their class was a mother. 105 (23.28%) girls said that a girl in their class became pregnant in the past year
Poor households (percentage of the total sample)	220 (24.94%) girls report living in a household unable to meet basic needs.	85 (18.85%) girls report living in a household not enough to meet basic needs
Girls report a high level of well-being (percentage of the total sampled) ¹⁰	690 (78.23%) of girls report having learned or doing something interesting when they were at school.	405 (89.80%) of girls report having learned or doing something interesting when they were at school.
Overage for grade	Grade 6: 43 (5.26%) Grade 8: 492 (55.78%) Grade 10: 347 (39.34%)	Grade 8: 43 (27.39%) Grade 10: 101 (34.35%)
Intends to enrol in secondary school after completing grade 8 (percentage of the total sample)		139 (86.34%)
Intends to enrol in secondary school after completing grade 10 (percentage of the total sample)	818 (92.74%)	270 (93.10%)
Boys' characteristics		
Living without both parents (percentage of the total sample)	2 (.87%) boys report living without both their mother and their father.	Three (6.66%) boys report living without both their mother and their father.

⁹ The data reported in this column are the result of single question observations. Composite measures are reported in further on in the report.

¹⁰ Three items on well-being were asked on the boys' student survey. Item 1: Were you happy the last time you were at school? Item 2: Did you learn or do something interesting the last time you were at school? Item 3: Did you have enough energy to get things done the last time you were at school?

Construct of interest	ML2 ⁹	Endline
Boys from poor households (percentage of the total sample) ¹¹	100 (43.67%) boys report living in a household unable to meet basic needs.	Ten (22.22%) boys report living in a household unable to meet basic needs
Boy does not report a high level of well-being (percentage of the total sample) ¹²	46 (20.08%) boys are below the average level of well-being.	None of the boys reported below the average level of well-being

3. Outcomes

Outcomes related to learning, transition, sustainability and value for money are detailed in this section. Due to the original MEL framework changing and the initial longitudinal design being dropped, it is not possible to make any direct comparisons to baseline results in this endline report.

3.1 Learning Outcomes

Literacy

Literacy outcomes for grade 8 and 10 girls are presented in Table 17. The aggregate EGRA literacy mean score for girls in grade 8 was 68.7 out of 113.5, while the aggregate SeGRA literacy mean score for grade 10 girls was 10.2 out of 25. Grade 8 girls performed better in familiar words, invented words and oral reading fluency (ORF) than reading comprehension, as the percentage scores for each EGRA subtask at endline show in figure 1 and figure 2, and had statistically significant gains from ML2 to endline in all four subtasks.

Table 17. Literacy from EGRA/SeGRA Scores

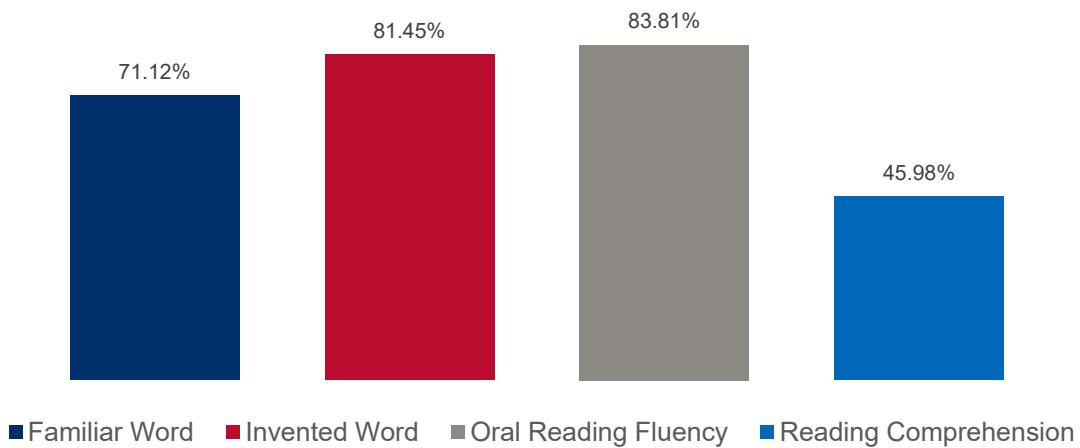
Grade	Aggregate Literacy Mean	Standard Deviation
Grade 8 EGRA: (N=157)	68.7 of 113.5	27.5
Grade 10 SeGRA: (N=294)	10.2 of 25	4.13

Note: The scale of the aggregate literacy mean for grade 8 is 0 to 113.5, while the scale of the aggregate literacy mean for grade 10 is 0 to 25.

¹¹ This was defined as boys responding that their households are unable to meet basic needs on the boys' student survey.

¹² This was indicated by boys' responding 'no' to at least one of 3 well-being questions. Three items on well-being were asked on the boys' student survey. Item1: Did you smile or laugh the last time you were at school? Item 2: Did you learn or do something interesting the last time you were at school? Item 3: Did you have enough energy to get things done the last time you were at school?

Figure 1. Grade 8 Girls Endline Subtask Literacy Outcomes



Comparison of ML2 and Endline Literacy Scores

A comparing of grade 6 ML2 and grade 8 endline literacy scores reveals an improvement, as shown in Table 18. Specifically, girls in grade 8 at endline had statistically significantly higher aggregate literacy scores in English compared to those in grade 6 at ML2, indicating a positive trend. However, this increase cannot be solely attributed to the intervention due to the absence of a true longitudinal design, which was initially planned at baseline but later scrapped due to the pandemic. Additionally, the 2 years of schooling and natural development may have influenced scores. Still, the gains in literacy outcomes do underscore the importance of the literacy interventions that STAGES introduced after ML2 to improve the literacy environment in schools, and the two additional years of schooling were delivered within the timeframe of project support.

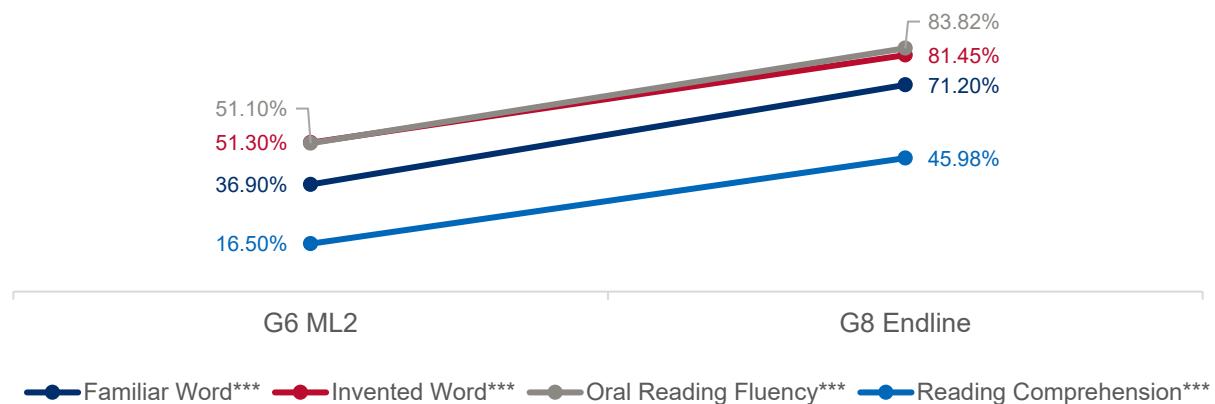
Table 18. Change in Literacy Mean Scores from Grade 6 ML2 to Grade 8 Endline

	G6 ML2 Literacy Treatment (n = 269)	G8 Endline Literacy Treatment (n = 157)	Difference to Endline	ML2	Statistical Significance – Robust Regression	OLS
Score	36.3	68.7	+32.46***		P = 0.000	

Note: Three asterisks (***) denote statistically significant at $p < 0.05$.

In each of the 4 subtasks, as displayed in Figure 2**Error! Reference source not found.**, girls' scores had a statistically significant increase from ML2 at grade 6 to endline at grade 8. Both lower-level skills, including familiar word reading and invented word reading, and higher-level literacy skills improved, including ORF and reading comprehension. According to Gough's Simple View of Reading, reading comprehension comprises word recognition and language comprehension. Measures of decoding (familiar word reading, invented word reading and ORF) showed higher performance in grade 8 girls at endline, with statistically significant gains.

Figure 2. Changes in Subtask Scores from Grade 6 ML2 to 8 Endline (%)



Grade 10 girls were assessed using the SeGRA literacy assessment in English, which includes 3 subtasks aimed at higher-level literacy skills appropriate for secondary students—reading comprehension (narrative text and questions), vocabulary (cloze exercise) and grammar (sentence revisions). On average, grade 10 girls' highest aggregate score was on the reading comprehension subtask, followed by grammar and vocabulary, as detailed in Table 19. The overall aggregate literacy score in English for grade 10 girls at endline was 40.8.

Table 19. Grade 10 Aggregate Literacy Outcomes at Endline (%)

Reading Subskill	Reading Comprehension	Vocabulary	Grammar	Overall Aggregate
Percentage Score	53.1	35.5	38.3	40.8

When comparing ML2 and endline literacy scores of grade 10 girls, no statistically significant differences were found, indicating similar scores, as displayed in Table 20. Students at secondary schools in the Wolaita Zone faced notable obstacles this school year in the months preceding endline, with considerable loss of instructional time due to teacher strikes, a shortage of textbooks related to adoption of a new curriculum and inflationary pressures affecting student motivation. When analysing these results, these challenges, namely the teacher strikes detailed in section 1.3, must be considered, especially that SeGRA scores did not decline despite the challenging conditions.

Figure 3. Grade 10 Girls Endline Aggregate Literacy Outcomes

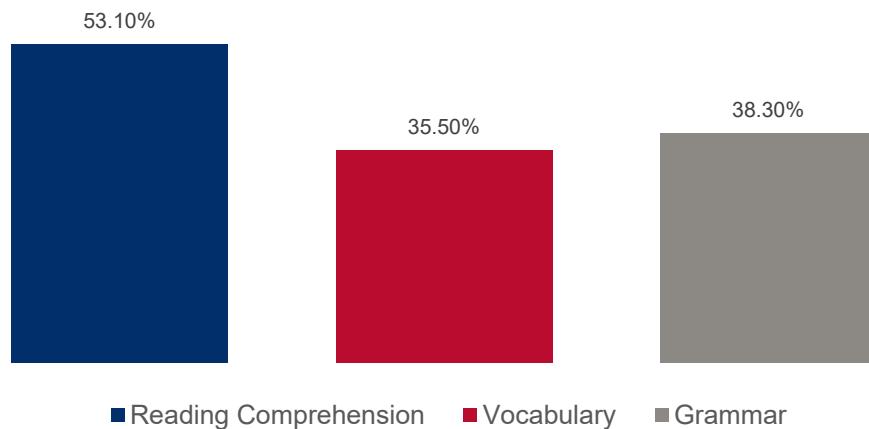
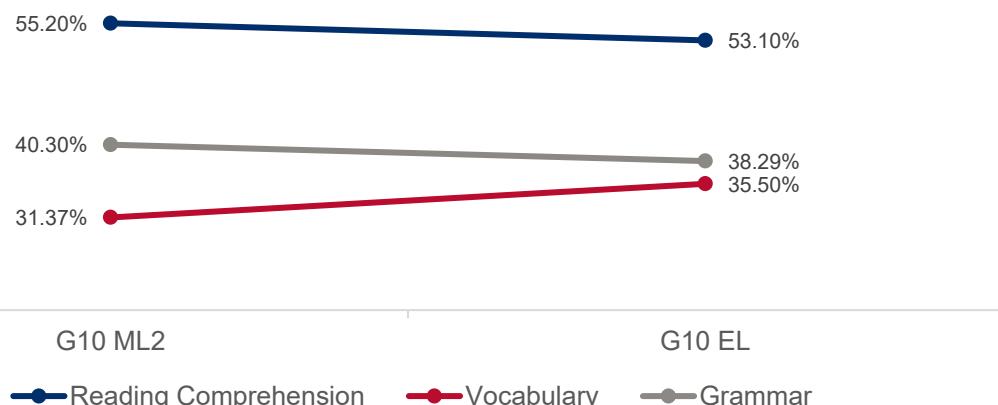


Table 20. Grade 10 Aggregate Literacy Outcomes at Endline (%)

Timepoint	Reading Comprehension	Vocabulary	Grammar	Overall
ML2	55.2	31.37	40.3	41.1
Endline	53.1	35.5	38.29	40.8
Change from ML2 to Endline	-2.04	+2.92	-1.95	-0.22

Figure 4. Changes in Subtask Scores from Grade 10 ML2 to Endline (%)



Numeracy Outcomes

Numeracy outcomes for grade 8 and 10 girls are presented in Table 21. The aggregate EGMA score for girls was 63.53 out of 100 for grade 8, while the aggregate SeGMA score was 7.19 out of 18 for grade 10. Overall, the improvements in numeracy outcomes for girls in Grade 8 from ML2 to endline were statistically significant, while scores remained similar in Grade 10 from ML2 to endline, highlighting the possibility for gains in higher grades.

Table 21. Numeracy based on EGMA/SeGMA Scores

Grade	Endline Group Mean	Standard Deviation
Grade 8 EGMA (n = 157)	63.53 of 100	18.84
Grade 10 SeGMA (n = 294)	7.19 of 18	2.91

Comparison of ML2 and Endline Numeracy Scores

For grade 8, numeracy performance statistically significantly improved from ML2 to endline, as shown in Table 22. Specifically, the aggregate numeracy score improved from 52.69 at ML2 for grade 6 girls to 63.53 for grade 8 girls at endline. This notable improvement could be attributed to the enhanced intensity or quality of numeracy instruction for grade 6 and grade 8 girls, as detailed in section 1.1.

Table 22. Numeracy Scores from ML2 to Endline

Grade	G6 Numeracy Treatment	ML2	G8 Numeracy Treatment	Endline	Difference	P value
Percentage Score	52.69		63.53		10.83***	P = 0.000

*Note: Three asterisks (***) denote statistically significant at p < 0.05.*

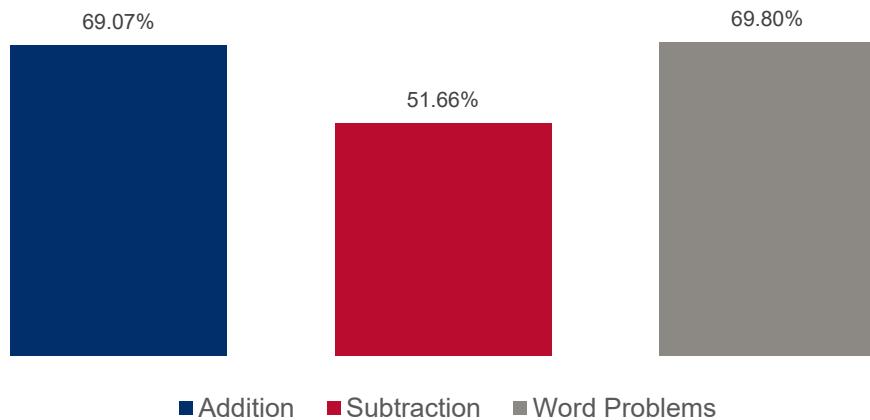
Performance on numeracy subtasks varied. As shown in Table 23, grade 6 girls at ML2 performed better in quantitative reasoning (67.28%), word problems (56.62%) and addition (55.09%) than in subtraction (40.69%) and missing numbers (31.04%). These disparities could potentially result from differences in the quality or frequency of instruction in certain numeracy skills.

Table 23. Grade 6 ML2 Subtask Numeracy Outcomes (%)

Quantitative Reasoning	Addition	Subtraction	Missing Number	Word Problem
67.28	55.09	40.69	31.03	56.62

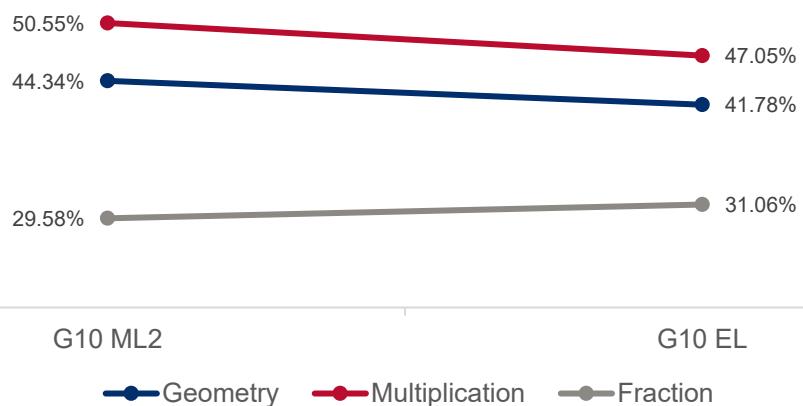
Numeracy outcomes for grade 8 girls at endline show similar trends, as shown in Figure 5, with scores in addition (69.07%) and word problems (69.80%) higher than subtraction (51.66%). On average, grade 8 girls at endline scored statistically significantly higher than grade 6 girls at ML2 in addition (69.07% to 55.09%, respectively), word problems (69.85% to 56.62%) and subtraction (51.66% to 40.69%). The grade 8 EGMA did not include the quantitative reasoning or missing number subtasks.

Figure 5. Grade 8 Girls Endline Subtask Numeracy Outcomes



On average, grade 10 girls' numeracy scores remained low at endline. Grade 10 girls were assessed with the SeGMA, which tests higher-level numeracy skills than the EGMA used for grades 6 and 8. The SeGMA included 3 subtasks—geometry and measurement, multiplication and fractions. As shown in Figure 6, grade 10 girls performed best on multiplication at both ML2 and endline, scoring 50.5% and 47.05%, respectively, on average, followed by geometry and measurement. Scores were notably lower on the fractions subtask than the other 2 numeracy subtasks. There was no statistically significant difference between the scores at ML2 and endline. As noted with the SeGRA results, the challenges that grade 10 girls faced this school year should be considered when analysing the endline results.

Figure 6. Changes in Subtask Scores from Grade 10 ML2 to Endline (%)



Subgroup Analysis of the Learning Outcome

In this section, trends in learning are explored by woreda, demographic subgroups and barriers to learning. Average literacy and numeracy aggregate scores are presented by grade for girls in each woreda and subgroup, as well as those affected by each barrier.

Learning Outcomes by Woreda at Endline

Girls' performance in literacy and numeracy was compared across woredas (Table 24). Among the grade 8 girls, Damot Sore had the highest mean literacy score (74.11) with a relatively low standard deviation,

indicating more consistent performance. Damot Pulasa followed with a mean score of 70.09 and a higher standard deviation, while Damot Woide had a similar mean score (69.62) but a lower standard deviation, suggesting more variability than Damot Sore but less than Damot Pulasa. Kindo Koisha had the lowest mean literacy score (60.63) and the smallest standard deviation, indicating the most consistent but lowest performance among the 4 project woredas.

Girls in Damot Pulasa achieved the highest mean numeracy score (68.18). Damot Sore followed closely with a mean score of 65.46 and a standard deviation of 16.66, showing more consistency. Damot Woide had a mean score of 60.79, while Kindo Koisha had the lowest mean numeracy score (59.65) and a standard deviation of 16.64.

The scale for the grade 10 literacy and numeracy scores is different than that of Grade 8 due to the use of different literacy and numeracy assessments (SeGRA and SeGMA). The literacy assessment was scored on a scale of 0 to 25, and the numeracy assessment was scored on a scale of 0 to 18.

In grade 10, for numeracy, Damot Pulasa had the highest mean score (7.64) with a standard deviation of 3.36, indicating higher performance but some variability. Damot Woide followed closely with a mean score of 7.62 and a lower standard deviation, showing more consistency. Damot Sore and Kindo Koisha had similarly lower scores with the same standard deviation.

In literacy, Damot Pulasa again led with the highest mean literacy score (10.69). Damot Sore and Damot Woide had similar mean scores (10.36 and 10.22, respectively), with Damot Sore showing slightly more variability (SD = 4.50) compared to Damot Woide (SD = 3.42). Similar to grade 8, grade 10 learners in Kindo Koisha had the lowest mean literacy score (9.45).

Overall, Damot Sore and Damot Pulasa showed strong performance in both literacy and numeracy outcomes for grade 8, with Damot Pulasa maintaining this performance into grade 10. Kindo Koisha consistently had the lowest mean scores in both literacy and numeracy across grades 8 and 10, raising questions about what factors may be associated with those low scores compared with the other 3 project woredas.

Table 24. Literacy and Numeracy Outcomes at Endline by Woreda, Grades 8 and 10

Woreda	Literacy		Numeracy	
	Mean	Standard Deviation	Mean	Standard Deviation
Grade 8	EGRA		EGMA	
Damot Pulasa	70.09 of 113.5	33.57	68.18 of 100	23.75
Damot Sore	74.11 of 113.5	24.41	65.46 of 100	16.66
Damot Woide	69.62 of 113.5	26.76	60.79 of 100	16.80
Kindo Koisha	60.63 of 113.5	23.62	59.65 of 100	16.64
Grade 10	SeGRA		SeGMA	
Damot Pulasa	10.69 of 25	4.91	7.64 of 18	3.36
Damot Sore	10.36 of 25	4.50	6.6 of 18	2.94
Damot Woide	10.22 of 25	3.42	7.62 of 18	2.36
Kindo Koisha	9.45 of 25	3.711	6.51 of 18	2.954

Note: The scale of the aggregate literacy mean for grade 8 is 0 to 113.5, while the scale of the aggregate literacy mean for grade 10 is 0 to 25, and the aggregate numeracy mean for grade 10 is 0 to 18.

Figure 7. Grade 8 Literacy and Numeracy Outcomes at Endline by Woreda

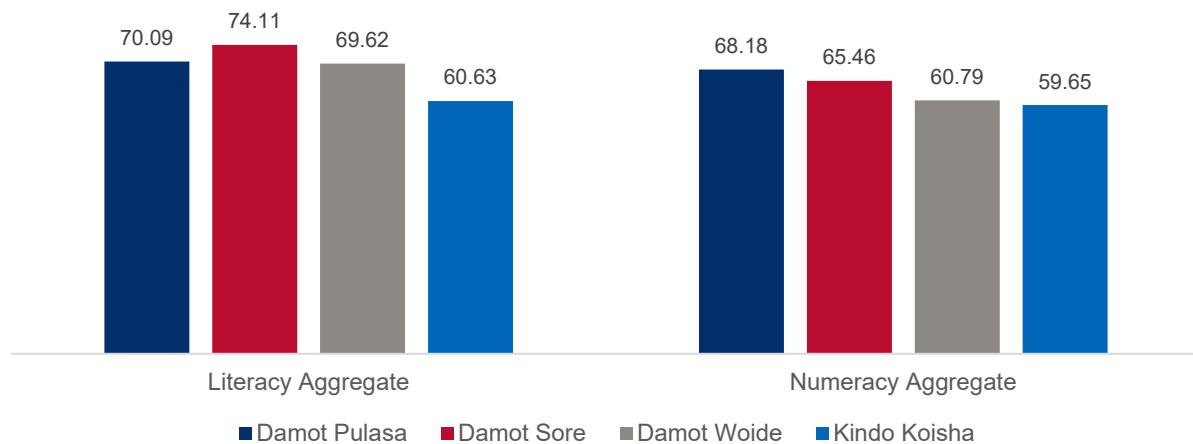
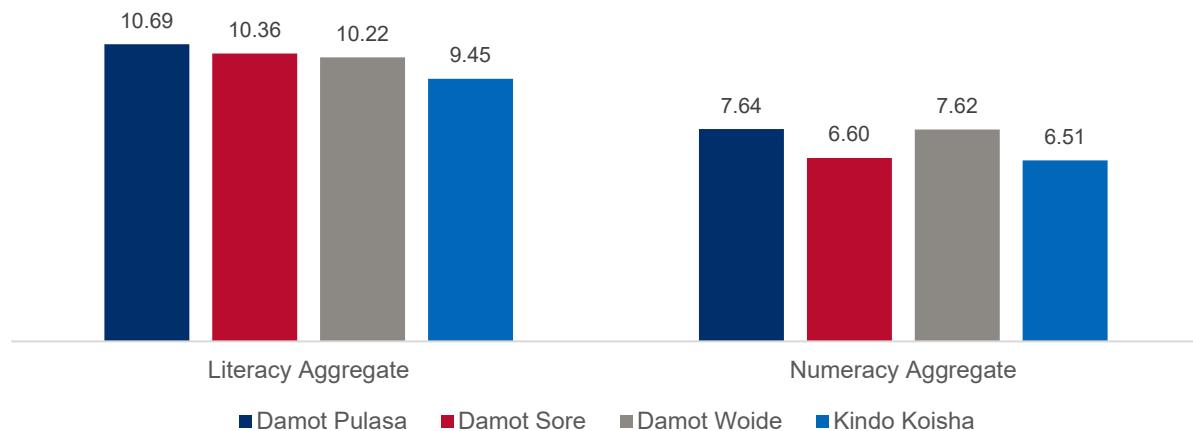


Figure 8. Grade 10 Literacy and Numeracy Outcomes at Endline by Woreda



Learning Outcomes by Subgroups

Learning outcomes by subgroups are presented in Table 25 for girls in grades 8 and 10. At endline, grade 8 girls showed improvements from ML2 in literacy scores across various subgroups, notably girls with difficulty seeing (+32.64), difficulty hearing (+32.12) and difficulty walking or climbing stairs (+41.75). Similarly, numeracy scores improved from ML2 to endline for grade 8 girls across various subgroups, particularly for those with difficulty seeing (+12.16) and remembering/concentrating (+13.85).

For grade 10 girls, the overall trends were less consistent. While girls with difficulty hearing showed modest gains in both literacy (+1.60 points) and numeracy (+2.43 points) from ML2 to endline, other subgroups experienced declines. Girls with difficulty walking or climbing stairs and those with serious illness saw decreases in both literacy and numeracy scores, indicating potential challenges that need addressing.

Interestingly, girls living in poverty in grade 8 showed significant improvement in literacy (+19.29 points) and numeracy (+10.68 points), although grade 10 girls in this subgroup had minimal changes. These findings suggest targeted interventions, especially bursaries that the project provided for vulnerable girls, may have positively impacted grade 8 girls, particularly those facing specific disabilities or socioeconomic challenges.

Table 25. Learning Scores of Key Subgroups by Grade ¹³

Characteristics	Average literacy score at Endline (aggregate)	Change in average literacy score since ML2	Average numeracy score at endline (aggregate)	Change in average numeracy score since ML2 ¹⁴
All girls ¹⁵	Grade 8: 68.77 Grade 10: 10.21	Grade 8: +32.46**	Grade 8: 63.53 Grade 10: 7.19	Grade 8: 10.83***
Difficulty seeing	Grade 8: 67.05 Grade 10: 10.05	Grade 8: +32.64 Grade 10: -0.07	Grade 8: 63.75 Grade 10: 7.6	Grade 8: +12.16 Grade 10: -0.9
Difficulty hearing	Grade 8: 67.05 Grade 10: 8.68	Grade 8: +32.12 Grade 10: +1.6	Grade 8: 63.75 Grade 10: 6.43	Grade 8: +9.14 Grade 10: +2.43
Difficulty walking or climbing stairs	Grade 8: 68.35 Grade 10: 9	Grade 8: +41.75 Grade 10: -3.5	Grade 8: 63.46 Grade 10: 6.82	Grade 8: +9.39 Grade 10: -1.68
Difficulty remembering or concentrating	Grade 8: 69.57 Grade 10: 9.62	Grade 8: +40.65 Grade 10: -0.59	Grade 8: 64.48 Grade 10: 7.12	Grade 8: +13.85 Grade 10: -1.12
Difficulty with self-care	Grade 8: 64.52 Grade 10: 9.47	Grade 8: +29.68 Grade 10: +0.97	Grade 8: 62.16 Grade 10: 7.23	Grade 8: +10.85 Grade 10: -0.86
Difficulty with communication	Grade 8: 64.23 Grade 10: 9.75	Grade 8: +40.08 Grade 10: +0.35	Grade 8: 62.18 Grade 10: 7.6	Grade 8: +10.87 Grade 10: -0.1
Serious illness	Grade 8: 64.76 Grade 10: 8.68	Grade 8: +23.96 Grade 10: -1.67	Grade 8: 60.01 Grade 10: 6.89	Grade 8: +2.98 Grade 10: -0.65
Poverty	Grade 8: 52.37 Grade 10: 10.88	Grade 8: +19.29 Grade 10: +0.3	Grade 8: 62.11 Grade 10: 7.27	Grade 8: +10.68 Grade 10: +0.46

Note: Three asterisks (***) denote statistically significant at $p < 0.05$.

Learning Outcomes by Barriers

Note of 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 was only a small difference. Specifically, differences in learning scores for all girls did not vary from the learning scores of some subgroups—including girls who reported that teachers treat boys and girls differently or girls who reported not using drinking water facilities. However, it is important to keep in mind the small sample sizes of subgroups. As such, the relationships between overall learning assessment scores and single items from a survey—such as teachers' treatment of boys and girls or teacher absenteeism—should be interpreted with caution. Relationships with constructs generated from a set of items—such as household support, life skills or

¹³ Testing on the difference of significance was not conducted for disability subgroups due to small sample size.

¹⁴ All girls here reflect the total number girls whose learning assessment could be paired with the quantitative tool that contains the answer to the indicators used to identify these subgroups.

¹⁵ All girls here reflect the total number girls whose learning assessment could be paired with the quantitative tool that contains the answer to the indicators used to identify these subgroups.

¹⁶ This change is G6 ML2 to G8 endline. Since the age group of students is different, the development and years of schooling should be taken into account.

decision-making—can be interpreted with greater confidence because of higher reliability in the observed data to assess the construct of interest. It is more reliable to interpret the aggregate of multiple questions rather than a single question.

Learning outcomes

Learning scores aggregated by barrier are presented in Table 26. It is important to remember that the total number of girls in each of the barrier categories varies and may be quite small. The following barriers are highlighted in the report:

- Of all barrier groups in grade 8, girls who reported not feeling safe travelling to and from school had the lowest average scores in literacy and numeracy.
- Of all barrier groups in grade 10, girls who reported not feeling safe at school posted the lowest average scores in literacy and numeracy.

Table 26. Learning Scores, Given of Key Barriers, by Grade

Barrier	Average literacy score (aggregate)	Average numeracy score (aggregate)
All girls ¹⁷	Grade 8: 68.77 Grade 10: 10.21	Grade 8: 63.53 Grade 10: 7.19
Difficult to move around the school	Grade 8: 35.2 Grade 10: 13	Grade 8: 58.66 Grade 10: 5.5
Doesn't use drinking water facilities	Grade 8: 67.51 Grade 10: 10.11	Grade 8: 62.69 Grade 10: 7.19
Doesn't feel safe travelling to/from school	Grade 8: 33.08 Grade 10: 9.46	Grade 8: 48 Grade 10: 7.2
Doesn't feel safe at school	Grade 8: 47.75 Grade 10: 8.16	Grade 8: 69.6 Grade 10: 5.83
Agrees teachers treat boys and girls differently in the classroom	Grade 8: 68.84 Grade 10: 10.13	Grade 8: 63.04 Grade 10: 7.14
Agrees teachers are often absent from class	Grade 8: 65.86 Grade 10: 10.48	Grade 8: 61.85 Grade 10: 7.65

3.2 Transition Outcomes

This section presents findings from the girls' survey related to transition outcomes. Grade 10 girls were asked about their intention to stay in school, as well as common barriers that keep their peers from staying in school. At the start of STAGES in 2017, students were only required to complete their education through grade 10 in secondary school, but now they must finish grades 11 and 12 due to the Ethiopian government changing the requirements for secondary school.

Girls who were surveyed reported having a clear desire to continue their education. At endline, the overwhelming majority of grade 10 girls—93.1%—indicated that they want to transition to grade 11 after

¹⁷ All girls here reflect the total number of girls whose learning assessment could be paired with the quantitative tool that contains the answer to the indicators used to identify these subgroups.

they have completed grade 10 (as reported in Table 27). Similar rates were recorded at ML2, underscoring that the project's emphasis on transition matches the desire among girls.

Table 27. Girls Survey, Percentage of Grade 10 Girls Who Want to Continue to Grade 11

In the first year after you have completed Grade 10, what do you really wish to do?		
	ML2	Endline
Continue to senior secondary	254	270
	93.73%	93.1%

At endline, early marriage, lack of parental support, early pregnancy and illness were the 4 reasons that girls cited most frequently for preventing their peers from transitioning, as shown in Table 28. These factors are notably related to girls' lives at home. Importantly, these home-related factors all decreased from ML2 to endline, indicating that the project interventions might have addressed these home-based barriers, notably home chores. The proportion of girls who said chores at home were a factor preventing girls from transitioning to secondary school declined from 74.20% at ML2 to 55.40% at endline.

Qualitative data suggest as well that STAGES interventions may have addressed behaviour at home, especially chores at home, that prevented girls from transitioning. In FGDs, girls shared how attitudes had changed within their own homes. 'The training from STAGES made our families understand the importance of girls' education', a grade 10 girl said, 'so we don't face pressure from obligations at home anymore'. A grade 12 girl added, 'I don't have to prioritise domestic work over attending school as both genders are taught to share responsibilities'.

By contrast, school-based reasons such as poor or unsafe infrastructure, gender-specific latrines and access to water were the factors that girls reported least frequently as barriers to transition. It is difficult to tell whether these factors were cited infrequently because they were not a common experience for the girls surveyed—for instance, their schools have these things—or if they are not impactful when it comes to transition. STAGES contributed to the construction of 4 secondary schools and repaired and constructed separate latrines for girls at all secondary schools.

Table 28. Factors Preventing Girls from Transitioning to Secondary School, Grade 10 Girls

Factor	ML2	Endline
Early marriage	79.30%	69.10%
Lack of parental support	77.50%	62.90%
Early pregnancy	77.10%	62.60%
Illness	77.49%	61.56%
Home chores	74.20%	55.40%
Migration	58.70%	51.70%
Lack of confidence	65.70%	48.90%
Lack of community support for education	54.90%	43.50%
Low expectations	67.53%	41.50%
Work/income earning is more important	57.20%	36.70%
School is too far away	52.40%	35.00%

Factor	ML2	Endline
Low performance/grade	40.60%	31.63%
SRGBV	38.70%	30.61%
Quality of education is poor	28.04%	29.59%
Infrastructure is poor/unsafe	36.90%	27.20%
No water supply	29.20%	26.50%
Latrines are there, but not separate latrines for girls	13.30%	26.50%
School fees	33.58%	24.83%
No latrines	19.60%	24.15%
Language barriers	22.50%	12.90%
No food at school	18.82%	11.90%
None	1.48%	1.02%
I do not know	0.00%	1.02%
No response	0.74%	0.00%

Understanding the socio-emotional well-being of the girls who have successfully transitioned may help potentially uncover what attributes have helped sustain their enrolment. A notable finding was that grade 10 girls had high scores on 2 composites—gender perceptions and life skills composites, as shown in Table 29.¹⁸ (These measures will be discussed later in the report.) Grade 10 girls' mean scores on both the gender composite and the life skills composite increased statistically significantly from ML2 to endline. These findings suggest that high levels of socio-emotional well-being, as well as a positive perception of gender equality in and out of the classroom, may help girls surmount the barriers that might potentially prevent them from continuing their education. In other words, these girls not only believe they deserve equal access to education, but they also have the life skills necessary to help them attain it. Additionally, grade 10 girls at endline showed higher mean scores (19.1) than at ML2 (18.6) decision-making composite,¹⁹ indicating that girls are confident in making important decisions about their lives.

Table 29. Grade 10 Girls Socio-emotional Composites

Composite	Scale	ML2		Endline	
		N	Mean Score	N	Mean Score
Gender Composite	0-7	271	5.7	294	6.5
Life Skills Composite	0-5	271	3.6	294	4.3

¹⁸ Correction notice: The ML2 report included incorrect gender and life skills composites due to a coding error in the variables. This error has been corrected in the endline analysis. This rectification standardizes the data interpretation across the ML2 and endline evaluation timepoints and ensures consistency in our analysis.

¹⁹ The decision making composite included items: 'Whether or not you will go to school', 'Whether or not you will continue in school past this grade', 'When/ at what age you will get married', 'If you will work after you finish your studies', 'What type of work you will do after you finish your studies', 'How you spend your free time', 'How often you spend time with your friends'. Individual items were scored from my family decides for me (1) to I decide (3).

Further supporting the project findings that support at home is critical to transition, one of the top factors that girls reported for keeping them in school was encouragement from family and friends. As shown in Table 30, the top 3 factors that girls reported as keeping them in school were support for basic school needs, encouragement from family and friends and teacher encouragement. Notably, over 75% of girls in grade 10 indicated that support for basic school needs helped keep them in school. The results, and those following, suggest that parental support is likely a prerequisite for a successful transition.

Table 30. Factors that Have Helped Grade 10 Girls Stay in School at ML2 and Endline

Factor	ML2	Endline
Support for basic school needs	69.70%	77.90%
Family/friend encouragement	64.60%	75.80%
Teacher encouragement	57.90%	61.90%
Not being married	57.90%	54.10%
School is girl-friendly	51.30%	52.00%
Lack of home chores	44.60%	51.00%
School is a safe place	35.40%	49.30%
Separate latrines for girls exist	35.10%	47.90%
Good grades	54.20%	46.60%
Community support	37.50%	44.60%
Infrastructure is safe	32.50%	39.80%
Latrines exist	27.70%	38.10%
I do not know	0.74%	34.00%
High quality of education	34.70%	32.90%
Bursaries	28.40%	27.20%
No need to work/earn income	33.90%	25.50%
High English proficiency	23.60%	21.80%
School is close	2.60%	5.40%
None	0.74%	1.70%
No response	10.70%	1.70%

The proportion of grade 10 girls who stated they had thought about dropping out of school either rarely, sometimes or never decreased from more than a quarter at ML2 (25.56%) to about a tenth at endline (9.52%). As shown in Table 31, the reasons for dropping out that girls cited most frequently were the lack of parental support and the greater importance of work and earning income. This highlights the importance of parent support in keeping the girls in school.

Table 31. Reported Reasons for Grade 10 Girls Considering Dropping Out at Endline

Factor	ML2	EL
Work/income earning more important	15.94%	50.00%

Factor	ML2	EL
Lack of parental support	81.16%	46.43%
School is too far away	23.19%	17.86%
Home Chores	28.99%	17.86%
School Fee	31.88%	14.29%
Infrastructure is poor/unsafe	8.70%	14.29%
Low expectations	11.59%	10.71%
Illness	5.80%	7.14%
Quality of education is poor	4.35%	3.57%
No latrines	0.0%	3.57%
Early marriage	0%	3.57%
Low performance/grades	4.35%	3.57%
None	0%	3.57%
No food at school	1.45%	1.45%
No water supply	4.35%	0%
Latrines are there, but not separate latrines for girls	0%	0%
Early pregnancy	0%	0%
SRGBV	1.45%	0%
Language barriers	2.9%	0%
Migration	0%	0%
Lack of confidence	15.94%	0%
I do not know	0%	0%
No response	4.35%	0%

3.3 Sustainability Outcome

In its long-term sustainability plan, STAGES identified 3 primary interventions it deemed to have the most potential for sustainability—pedagogical leadership and supervision (including school improvement planning), GIRP and safeguarding mechanisms and practices (including through community-based mothers' and fathers' groups). Education officials are also determined to continue other key project activities. Overall, it was clear from FGDs with zonal and woreda officials that sustaining key project interventions has been a key priority for stakeholders, with the project nearing its end.

Overall, the project's design, which included close collaboration with the government in initial planning and implementation, including joint monitoring and activity adaptation, has paid off as the project closes, with the government seemingly poised to sustain STAGES interventions through long-term planning. Zonal and woreda officials reported that the three specifically targeted interventions have been successfully incorporated and scaled at varying degrees into the Wolaita Zone's educational system, with

pedagogical leadership and supervision furthest along in being integrated, followed by GIRP and then safeguarding. A zonal official said:

'Our involvement with STAGES has been transformative, aligning seamlessly with the government's priorities and policies on gender equality, inclusive education and safeguarding children in educational settings. The project's interventions have become deeply embedded within our regular programs, reflecting our commitment to institutionalizing its successful models and practices'.

Zonal and woreda education officials detailed how they have integrated project interventions into their long-term planning and regular programs. The zone has embedded the project's key interventions by incorporating them into their three-year strategic plans. Further, it is funding this scale-up for all the key project interventions with its own financial resources after gaining the technical expertise needed for implementation from STAGES. '[This planning] clearly shows diffusion of the project's experience and resources throughout the zone', a zonal official said. Officials said they have also adapted project training materials and manuals for GIRP into their CPD programming, and according to STAGES personnel, materials from the project's subject-specific interventions in literacy, numeracy and socio-emotional learning (SEL) have been as well.

Pedagogical Leadership and Supervision

The scale-up of the pedagogical leadership and supervision intervention exemplifies the zonal office's commitment to sustaining and institutionalizing project practices. 'We will collaborate closely with schools, communities and other stakeholders to amplify the impact of these efforts and ensure sustainable implementation', a zonal official said about the office's dedication to sustaining the pedagogical leadership and supervision intervention.

This zonal office has implemented the intervention not only in all 18 non-project woredas in the Wolaita Zone but also conducted trainings on the intervention outside of the zone. Zonal officials detailed how they have trained education officials in the South Ethiopia Regional State, formerly known as the Southern National, Nationalities and Peoples' Region. 'Generally, we are being seen as a centre of excellence in the sector, in the region', a zonal official said.

The success that the zonal office observed with the intervention in the woredas that STAGES targeted has motivated it to deploy STAGES-trained experts to lead a cascade training across the zone. 'The initial results from the target areas were highly encouraging, demonstrating positive impacts on pedagogical approaches, classroom environments and overall student learning outcomes', a zonal official said.

Zonal officials said they have primarily utilised the project's resources for scale-up but also used government funding despite it currently being in short supply. The scale-up has included a comprehensive training package, including conducting initial training and follow-up coaching, establishing communities of practice and completing sustainability checklists as part of regular monitoring, according to zonal officials. In addition, the intervention has been incorporated into CPD. 'We understood the long-term value of investing in the professional development of our educational leaders and supervisors, as they play a pivotal role in driving systemic change', a zonal official said.

GIRP

While the pedagogical supervision and training intervention has already been scaled across the entire zone, the zonal office said GIRP is not far behind, with the approach being implemented in 18 non-project woredas already. In an FGD, zonal officials said they plan to scale the intervention to all woredas in the zone starting in September 2024. 'Following the initial success of the GIRP intervention in the target woredas', a zonal official said, 'we recognised the immense value it holds and decided to scale it up across the entire Wolaita Zone'.

The zonal office said it is using its own budget and partnerships to fund the GIRP expansion, but funding constraints have prevented it from scaling it across the entire zone. The scale-up includes conducting initial trainings and follow-up coaching, establishing communities of practice and providing on-site mentoring and support, primarily through the coaching and mentoring tool developed by the project that education officials have now adapted into their own tools. 'The coaching and mentoring checklist developed through the STAGES project aligns closely with the zone's education goals and priorities', a zonal official said. Additionally, GIRP will also be integrated into the zone's CPD, according to zonal officials. 'Our vision is to institutionalise the GIRP approach within our educational system', a zonal official said.

Woreda officials confirmed this scale-up process, with one woreda education head mentioning that he had trained officials in non-project areas in the zone. This official also believed that GIRP would be sustained due to its institutionalisation as part of classroom lesson preparation and CPD. Additionally, he said the community recognised its value. 'They are witnessing when marginalised girls enrol and score good results on their education', he said.

Safeguarding

Safeguarding mechanisms and practices seem not to have been expanded as wide in the Wolaita Zone as GIRP and pedagogical training and supervision have been, based on the FGD with zonal officials. While at least 146 schools in the 4 woredas targeted by STAGES have safeguarding mechanisms in place, zonal officials said 'many' other schools in the 18 woredas not targeted by the project have adopted them, but the exact number is unknown. The schools that have safeguarding mechanisms in place typically have incorporated them into their SIPs with components such as awareness-building campaigns, capacity building for staff and students and collaboration with local stakeholders to ensure reporting and response mechanisms are functional, zonal officials said. In the FGD, zonal officials asked for technical support in implementing safeguarding practices. They mentioned that although reporting mechanisms have been robust in identifying and documenting cases, they need to strengthen the response to and support of incidents. A zonal official said:

'We are working on enhancing coordination between schools, communities and relevant authorities to ensure prompt and effective interventions. We're looking for examples of how these systems work in practice, including any challenges or successes you've observed in their implementation and response effectiveness'.

It is important to note, however, that the viability and effectiveness of safeguarding mechanisms may depend more on buy-in and participation in respective communities than other STAGES interventions, including GIRP and pedagogical training and supervision, due to the need to collaborate locally with authorities and stakeholders. As one zonal official said, 'The involvement of parents and community members in the implementation of safeguarding interventions has been crucial in ensuring a holistic and sustainable approach'.

Other Sustainability Possibilities and Potential Challenges

Although STAGES primarily focused on 3 key interventions in its long-term sustainability plans, educational officials said they plan to continue other project activities as well. 'We are committed to maintaining and expanding the girls' clubs, tutorial support programs and safeguarding mechanisms across all schools in the targeted woredas after the project and continue expanding to other non-targeted woredas', a zonal official said.

Zonal officials were realistic about the challenges they would face in sustaining project activities. The challenges that they mentioned included limited funding, staff turnover and lack of specialised expertise in monitoring and evaluation as well as technical areas such as gender, inclusive education and child protection in order 'to align with evolving contexts'. 'By proactively addressing these challenges and

leveraging partnership', a zonal official said, 'we are confident in our ability to sustain the progress made through STAGES'.

Other respondents, including woreda officials and head teachers, shared views similar to those of zonal officials about the prospects of sustaining STAGES interventions. Woreda officials said activities that involve their strengthened capacity can continue, such as GIRP, because 'we have trained manpower and the documents are in our hands', but those activities that relied on the project's financial contributions, would be harder to continue, including bursaries and the provision of teaching and learning materials.

In a KII, a head teacher echoed the sentiments about the project's training. 'We got skills on how we [can] continue the projects' activities after they phase out', the head teacher said. He was also optimistic about what the school could accomplish on its own based on current achievements. 'We have trained manpower', he said, 'and as a beginning we are doing some activities through community mobilization without LINK support such as building toilet facilities for boys, classroom buildings'.

Education officials also underscored the importance of mobilising community resources organised by the project to continue interventions. Mothers' and fathers' groups were singled out as a key community partner, as they have played a crucial role in ensuring vulnerable children stay in school and safeguarding systems function. 'Mother and father meeting groups are one of the continuation mechanisms of project activities', a woreda official said.

3.4 Value for Money

This endline study includes a 'light touch' Value for Money analysis as described by GEC-T guidance. Four of the OECD DAC criteria are included in light touch analyses—relevance, effectiveness, efficiency and sustainability. Each of the criteria is examined below. Sustainability is treated separately in this report.

Findings from quantitative and qualitative data collection are considered in the following results. Cost data was not analysed as part of the light touch analysis.

Overall, the evaluation found that STAGES met these criteria, mirroring the findings from the Ethiopian government endline that said that the project was 'relevant, responsive and efficient'.

Relevance

The relevance of STAGES is understood as the degree to which the project invested in the right activities to address the needs and barriers of beneficiaries, specifically marginalised girls.

Resources appear to be well-allocated to meet the needs of the target beneficiary populations, especially marginalised girls. The relevance of project design is evident in the improved outcomes demonstrated by marginalised girls after participating in STAGES.

Zonal and woreda officials shared how the project's interventions, including GIRP and child protection and safeguarding activities, are connected to their priorities, as well as the government's. For instance, a zonal official shared how GIRP was 'very well-aligned with the priorities, goals and policies of our office and the Ministry of Education at the zonal, regional and national levels'. The zonal official added, 'Promoting gender equality, inclusive education practices and creating equitable learning opportunities for all students, regardless of gender, disability, or background, are central to our educational vision and strategies'.

Efficiency

No wastage of resources was detected in qualitative data collection. It was clear from FGDs with zonal and woreda officials and head teachers that stakeholders are actively participating in school performance

appraisal meetings and developing school improvement plans. A zonal official said, 'Ensuring that school resources are mobilised and utilised effectively to support the learning needs of marginalised girls has been a key priority during these processes'.

Effectiveness

The findings of this endline study suggest that STAGES met expectations for effectiveness. Results, as measured chiefly by learning outcomes and a range of IOs, indicate a mix of positive changes and unchanged learning outcomes from ML2 to endline.

In terms of learning outcomes, 85.4% grade 8 girls were improved learners at endline in literacy and 73.3% grade 8 girls were improved learners in numeracy. Considering transition outcomes, 80.2% of G8 EL girls want to transition to secondary schools 93.1% of EL girls want to transition to senior secondary.

All government officials and school officials stated that they believe the project is making a difference in its schools and communities. A zonal official said, 'The STAGES project has provided us with a comprehensive framework and resources to develop long-term strategies and action plans. We have collaboratively crafted zone-wide checklists and guidelines to sustain the project's interventions beyond its culmination in 2024'.

When asked if the project had any unexpected positive or negative outcomes, head teachers said they could not think of anything negative to share. One head teacher said:

'There are no negative outcomes. The project provides more positive, unexpected things. They highly support girls' education by providing different supports. That makes girls free from embarrassment, confident and [they] equally participate in the classroom and education...Also, their parents and families are aware of the positive impact of the project, and I am witnessing the project's significant impact'.

Teachers also shared the significant impact that STAGES had had on their classrooms. 'Absolutely, STAGES has been transformative for how we understand and support our vulnerable girls', a secondary school teacher said. 'I think the most significant aspect was the training on gender-responsive and inclusive pedagogy'.

Sustainability

As noted in greater detail in Section 3.3 of this report, it was clear from FGDs with zonal and woreda officials that sustaining key project interventions has been at the forefront of stakeholders' minds with the project nearing its end. In its long-term sustainability plan, STAGES identified 3 primary interventions it deemed to have the most potential for sustainability—pedagogical leadership and supervision (including school improvement planning), GIRP and safeguarding mechanisms and practices (including through community-based mothers' and fathers' groups). Zonal and woreda officials reported that these interventions that STAGES specifically targeted have been successfully incorporated and scaled at varying levels into the Wolaita Zone's educational system, with pedagogical leadership and supervision furthest along in being integrated, followed by GIRP and then safeguarding. Although STAGES focused on 3 key interventions in its long-term sustainability plans, educational officials said they plan to continue other project activities as well.

4. Key Intermediate Outcome Findings

This section presents the key IOs and ML2 and endline results for each of the project's 5 IOs. The results are presented only for grade 10 girls at ML2 and endline.

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 marginalised girls are prerequisites for better learning, transition and sustainability outcomes for these girls. Therefore, it is important to measure the realization of these IOs to confirm that STAGES made progress in the areas that the project believes will improve learning, transitioning and sustainability outcomes.

4.1 Attendance

Under the IO on attendance, the project aimed to reduce the following barriers to education—poor access to secondary schools in extreme and remote areas; and limited opportunities and access to academic, social-emotional, basic needs, life skills and hygiene inputs in school.²⁰ One of the key activities that STAGES implemented was the provision of bursaries to vulnerable girls to encourage them to stay in school.

Attendance among grade 10 girls was similar from ML2 to endline. About three-quarters of girls reported that they did not miss any days of school in the past week at ML2 and endline, as displayed in Table 32. The slight increase from 74.54% of girls reporting they did not miss any days at ML2 to 76.53% at endline was not statistically significant. Additionally, although the average number of days that grade 10 girls missed at endline (0.47) was slightly greater than at ML2 (0.34), the increase was not statistically significant. The proportion of girls who missed 5 days, however, increased from 0.74% at ML2 to 2.72% at endline.

Table 32. Grade 10 Girls' Average Number of Absences at ML2 and Endline

In the past school week, how many days were you absent?	ML2 (n=271)	Endline (n = 294)
0 days	74.54%	76.53%
1 day	19.19%	10.88%
2 days	4.8%	7.14%
3 days	0.74%	2.38%
4 days	0.0%	0.34%
5 days	0.74%	2.72%

Perceptions of Access

Grade 10 girls were also asked about their perceptions of school access and safety. Overall, girls' perceptions seem to have favourably shifted, based on statistically significant differences on multiple questions from ML2 to endline, as shown in Table 33.

First, the proportion of girls who said it was reasonably easy for them to get to school and back increased from 77.12% at ML2 to 86.39% at endline. A notable barrier for girls transitioning to secondary school is

²⁰ Refer to the STAGES' theory of change for specific activities.

the distance they must travel from home to school. This statistically significant increase suggests that there were notable changes in students' ease of getting to and from school between the 2 time points, with more students finding it reasonably easy at endline compared to ML2. Further, this favourable gain mirrors the decline from ML2 to endline in the proportion of grade 10 girls who cited that school was too far away as a barrier to girls transitioning to secondary school—52.40% to 35.00%, respectively—as detailed earlier in section 3.2.

Statistically significant increases were also measured with respect to several questions about school safety. The proportion of girls who reported feeling safe travelling to and from school increased from 85.20% at ML2 to 94.56% at endline, while the proportion of girls who said the school infrastructure was safe increased from 73.80 at ML2 to 91.50 at endline. Both of these gains were statistically significant, suggesting that there was a notable improvement in students' perceptions of safety travelling to school over the observed period, as well as their perceived safety of school infrastructure.

A statistically significant difference was also noted between ML2 and endline in the proportion of girls who said their parents or caregivers pay for everything they need to go to school—from 56.46% to 79.25%, respectively. This increase suggests notable changes in the financial support students received from their parents or caregivers over the observed period.

In addition, while only nearly 30% of girls at ML2 said they faced fewer challenges in attending school compared to the previous school year, more than half of girls said so at endline. This statistically significant increase suggests notable changes in the challenges students faced in attending school over the observed period, with more students experiencing fewer challenges at endline compared to ML2.

Qualitative data from girls mirrors their responses to the girls' survey. In FGDs, girls shared that nearly everyone feels safer at school thanks to the project's efforts in bolstering safeguarding systems. 'I absolutely feel safe because the school has taken the right steps to make us feel that way', a grade 12 student said. Still, not all instances of SRGBV have been eliminated. 'While most of us feel safe, some girls still face harassment incidents at times, which we can report to school', a grade 10 student said.

Table 33. Grade 10 Girls' Perceptions of School Access and Safety

Survey Question	Response	ML2	Endline
Transportation			
Is it reasonably easy for you to get to school and back?	Yes	77.12%	86.39%
	No	22.5%	13.61%
	No response	0.37%	0.0%
Safety			
Do you feel safe travelling to and from school?	Yes	85.2%	94.56%
	No	14.76%	5.1%
	No response	0%	0.34%
Do you feel safe at school?	Yes	94.46%	97.96%
	No	5.54%	2.04%
Do you think that the school infrastructure is safe for	Yes	73.8%	91.5%
	No	24.35%	8.16%

Survey Question	Response	ML2	Endline
students like yours?	I do not know	1.48%	0.34%
	No response	0.37%	0.0%
Financial barriers			
Do your parents / caregivers pay for everything that you need so that you can go to school?	Yes	56.46%	79.25%
	No	43.17%	19.73%
	No response	0.37%	1.02%
Challenges			
Compared to last year, are you facing greater challenges in attending school?	Fewer	29.15%	51.7%
	The same	28.78%	19.39%
	Greater	25.09%	19.73%
	Did not answer	16.97%	9.18%

4.2 School Governance and Management

STAGES highlights school governance and management as a key input to improving learning, transition and sustainability outcomes for girls. The project identifies 3 ways in which school governance and management can form a barrier to girls—through lack of community awareness and support, gaps in perceptions of girls' education and gaps in support structures for girls to persist and learn.

Teachers generally said they believed key education actors were responsive to the needs of girls in school. Nearly all 15 teachers surveyed said that cluster supervisors were mostly (8 of 15) or always (5 of 15) responsive to girls' needs, as well as school leadership (7 each responded mostly and always). By contrast, teachers said they believed woreda officials were not as responsive as cluster supervisors or school leadership to the needs of girls in school, with only 3 of 15 saying woreda officials were mostly responsive and only 1 of 15 saying they were always responsive.

Head teachers and teachers were also asked about their schools' safeguarding plans and policies. Fourteen of the 15 head teachers surveyed reported the presence of a gender policy and guidelines, while one reported the school had partial plans. All 15 head teachers said their school had strategies to prevent SRGBV. Ten of 15 teachers said their school had developed and integrated safeguarding action plans into their SIPs, while 4 said they did not know and one said their school had not. Of those 10 teachers who said their school had integrated safeguarding action plans, 2 said they were very comprehensive, with 5 saying they were only slightly comprehensive.

In FGDs and KIIs, multiple respondents—including those at the zonal and woreda officials, head teachers and teachers—explained how STAGES had improved school management and governance at all levels. 'School is much improved now, due to the leadership and governance intervention by the project that has targeted school, cluster, woreda and zone level leaders', said a teacher, noting how the participation of community stakeholders had increased. Other teachers underscored how their direct supervision had improved due to trainings conducted by the project. 'Supervisors approach supervision in a different paradigm compared to the previous times', a teacher said, 'because now they are not just [evaluating] but supporting, approaching this responsibility more positively'.

Head teachers also spoke about how their knowledge, attitudes and practices had changed due to STAGES. 'The training has positive impacts on me. It helps me to encourage and advise girl students to

join and return to school, actively participate in the class and study well', said a head teacher, who highlighted how the top 9 ranked students at the school were all girls. Another head teacher noted how the project has intervened at all levels of education, including those in his position. 'As a school leader, I have got enough knowledge on how to manage the school, how to supervise and how to give feedback for teachers', he said.

Zonal and woreda officials corroborated what teachers and head teachers reported about improvements in supervision and the creation of more inclusive environments at schools. A woreda official said that no women were head teachers or education office heads in the targeted woredas before STAGES had conducted the pedagogical leadership and supervision training, but now 4 women were head teachers, and 2 were education office heads. 'This is the outcome of the intervention and CPD training', the woreda official said. 'Preparation of [lesson planning] is also changed and because of that supervision of the classroom has become effective'.

A zonal official noted how head teachers' practices had improved. 'Head teachers have become more proactive in fostering inclusive and conducive learning environments', the official said. 'They have implemented strategies to encourage equal participation, address gender-related barriers and create safe spaces for all students to learn and thrive'.

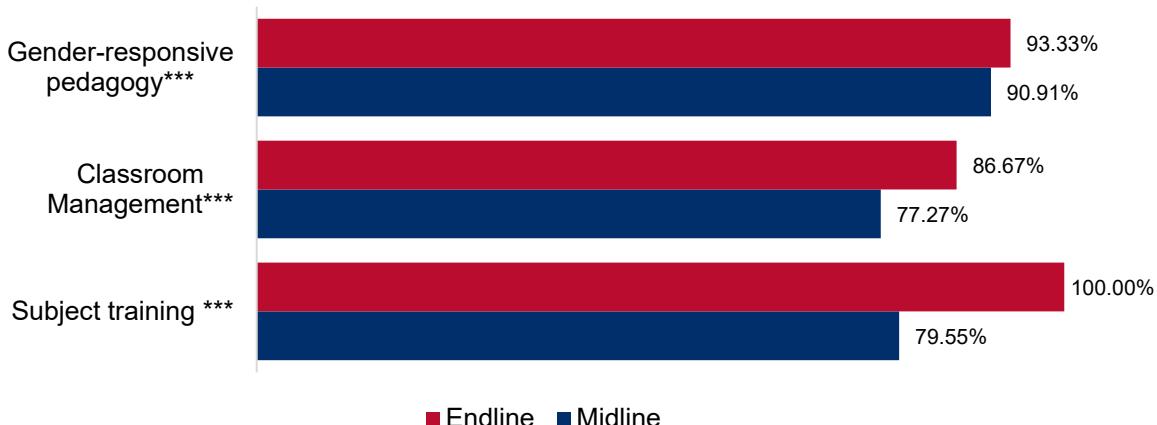
Many respondents also discussed how the project's safeguarding systems had helped to develop safer schools. One woreda official reported that his woreda, which once had the highest rate of child trafficking in the zone, had observed a 'significant' drop in child trafficking due to different focal points for reporting SRGBV incidents, including school and community members, the kebele representative and the anonymous project letter boxes.

Other respondents recounted specific cases of SRGBV that had been thwarted. A woreda official described what happened when someone threatened to abduct a girl. 'By following the reporting mechanism, the perpetrator was accused, and the court punished [him]', the official said. 'Previously even when that kind of violence happened, it was treated by elder negotiation'.

4.3 Quality of Teaching

Quality of teaching is an IO that STAGES has identified as key to increasing learning, transition and sustainability outcomes among girls. The ToC theorises 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.

Figure 9. Proportion of Teachers Receiving Training, Over Time



The proportion of teachers who reported receiving training in their subject matter, classroom management and GIRP increased statistically significantly from ML2 to endline, as illustrated in Figure 9. These increases indicate that interventions focused on teacher professional development have been implemented effectively. Additionally, all 15 teachers surveyed at endline reported that the GIRP training was very useful; 12 of 15 said the classroom management training was very useful, with the other 3 believing it was somewhat useful.

In FGDs and KIIs, respondents detailed how the project's GIRP training had improved teaching methodology and encouraged teachers to be more conscious of gender. 'We have witnessed significant changes in teaching practices as a result of this intervention', a zonal official said. 'Teachers are now more adept at employing learner-centred, gender-responsive and disability-inclusive teaching techniques'.

In the 2 FGDs conducted with teachers, the impact that GIRP had made in classrooms was evident from what different teachers said:

- 'The GIRP training from STAGES was excellent and really transformed my classroom practices'.
- 'Yes, it opened our eyes to unconscious gender biases in things like the examples and language we were using'.
- 'Yes, I'm much more intentional now about making sure I call on girls equitably, use examples relevant to females, depict women in non-traditional roles, etc.'.
- 'I feel much better prepared now than I was before thanks to the training and resources STAGES provided. The inclusive pedagogy methods helped me learn how to structure lessons and activities so all students can participate effectively regardless of disability or circumstance'.

In FGDs, teachers also said their relationship with supervisors had improved thanks to project interventions. 'In terms of supervisory support from leadership', a teacher said, 'we've actually seen quite an improvement thanks to the training and coaching STAGES provided our directors and mentors'. Another teacher added, 'Yes, they now do a good job of following up to ensure we're implementing what we've learned, whether it's inclusive pedagogy, gender-responsive practices, positive disciplining, or other key skills'. The teachers agreed, however, that it was difficult to find time for self-reflection and peer coaching due to all their other responsibilities.

A woreda official noted how some teachers were initially uncomfortable with the training and considered it too time-consuming, but they became 'highly interested' in it after seeing the results in their classroom. 'Now all teachers have the willingness to proceed with the project's modality', the official said.

One respondent drew a direct link between improved teaching practices to improved transition outcomes in secondary school. 'When I was a secondary school deputy head, I witnessed how much change was registered in the school regarding the transition of girls to secondary school and their participation in the class', a woreda official said. 'That comes through the teacher's encouragement of girls based on the training which they took'.

In addition, to understand changes in teachers' attitudes and classroom practices towards girls, a teacher gender perception composite was formulated that measured teachers' attitudes towards boys' and girls' social roles and academic abilities. The scale and mean scores from ML2 and endline are listed in Table 34. **While there was not a statistically significant increase between gender perception scores of teachers from ML2 to endline**, the positive upward trend and the relatively high score suggest that the project interventions to improve gender perceptions among teachers may have had an effect.

Table 34. Gender Perception Composite Scores Among Teachers

Composite	Scale	Mean Score (ML2)	Mean Score (Endline)	P value
Gender Perception Teacher	0-15	13.59 (N= 44)	13.8 (N=15)	0.251

During the endline evaluation, enumerators also observed lessons to measure teacher's behaviours in the classroom. The proportion of teachers observed exhibiting certain behaviours either never, sometimes, frequently or always during a lesson at ML2 and endline is shown in Table 35. The proportion of teachers observed performing 2 types of actions increased statistically significantly from ML2 to endline—asking questions to monitor students' progress and recognising effort and persistence with all students. The data suggests that the frequency with which teachers asked questions to monitor students' progress varied significantly between the two time points, with endline showing a higher incidence of frequency and always asking questions. There was also a significant difference in the frequency with which teachers recognized effort and persistence between the two time points. Specifically, at endline, teachers showed a greater tendency to recognize these qualities more frequently, including a notable increase in the "always" category.

Teachers' lesson plans were also examined as part of the classroom observation at endline. All but one of the teachers had a lesson plan, and of those 14 that did, half specifically addressed the needs of girls, boys and children with disabilities, while the other half addressed some of the 3 groups but not all of them. Additionally, all but 2 classrooms had students seated in an arrangement that considered the different needs of students. All but one classroom provided students equal access to textbooks, learning materials and other relevant facilities in the school.

Table 35. Classroom Observation of Teacher Behaviour at ML2 and Endline

Teacher Behaviour	Frequency	ML2	Endline	P value
Teacher interacts equally with girls and boys	Never	7.1%	0%	0.23
	Sometimes	33%	13%	
	Frequently	38%	47%	
	Always	21%	40%	
Teacher encourages all students to engage in discussions	Never	21%	0%	0.2
	Sometimes	29%	27%	
	Frequently	36%	47%	
	Always	14%	27%	
Teacher maintains all student's attention	Never	4.8%	0%	0.7
	Sometimes	33%	27%	
	Frequently	33%	47%	
	Always	29%	27%	
Teacher asks questions to monitor students' progress	Never	7.1%	0%	0.04*
	Sometimes	33%	13.3%	
	Frequently	52%	53.3%	

Teacher Behaviour	Frequency	ML2	Endline	P value
Teacher recognises effort and persistence with all students	Always	7.1%	33.3%	0.0***
	Never	12%	0%	
	Sometimes	62%	13.3%	
	Frequently	21%	53.3%	
	Always	4.85%	33.3%	
Teacher set high expectations for all students' participation	Never	4.8%	0%	0.3
	Sometimes	33.3%	13.3%	
	Frequently	45.24%	60%	
	Always	16.7%	26.7%	
Teacher talking time is equal to the student talking time	Never	12%	6.7%	0.52
	Sometimes	50%	40%	
	Frequently	26%	46.7%	
	Always	12%	6.7%	
Teacher asks 'how', 'why' questions or asks students to explain a concept.	Never	14%	0%	0.08
	Sometimes	38%	42.9%	
	Frequently	38%	35.7%	
	Always	9.5%	21.4%	
Teacher's attention is on students' learning	Never	2.4%	0%	0.09
	Sometimes	24%	0%	
	Frequently	38%	33.3%	
	Always	36%	66.7%	

Note: Three asterisks (***) denote statistically significant at $p < 0.001$.

Teaching Students with Disabilities

In FGDs, teachers discussed how they felt more prepared to teach students with disabilities thanks to STAGES trainings and adaptive equipment that the project distributed, including slanted writing surfaces and large pencil grips. One teacher said he taught students with mobility-related and sight-related disabilities. 'We were trained and translated the trainings to [our] regular classrooms, and we were able to keep them in school', the teacher said. 'Otherwise, they could withdraw like others had done before the project'.

Teachers explained the specific strategies they use to teach students with disabilities. 'We have improved our capacity on how to use more visual aids, hands-on materials and techniques like peer support to reinforce lessons for students with hearing or cognitive impairments', a teacher said. Still, they said they do not have the capacity to teach students who are blind, have low vision, are deaf, or are hard of hearing. 'While we are very motivated to support these vulnerable girls', another teacher said, 'we do face

challenges in terms of lacking specialised skills and knowledge, like sign language for deaf students or Braille for the visually impaired'.

4.4 Positive Community Attitudinal Change (IO4)

As part of its objective to improve the learning and transition outcomes for girls, STAGES aimed to change communities' attitudes toward girls' education by creating girls' clubs, working with mothers' and fathers' clubs and strengthening other community-based school organizations, as noted earlier in IO4. To report on this IO, this section details several quantitative measures as well as qualitative data.

Teachers and head teachers were asked questions to gauge their gender perceptions and practice of gender-responsive teaching practices. According to a gender perception composite, teachers' gender perceptions slightly increased from 13.6 at ML2 to 13.8 at endline, as detailed in Table 36.²¹ The gain was not statistically significant, however, indicating no meaningful change over time.

Table 36. Teacher Gender Perception Composite at ML2 and Endline

Composite	ML2	Endline	P-value
Gender Perception Composite	13.6	13.8	0.74

Additionally, head teachers at endline were asked about gender-responsive training and practices. Of the 15 head teachers surveyed at endline, 11 reported that most or all teachers had received training in gender sensitivity and mainstreaming. All but one said that the school system promoted gender-responsive teaching and learning methods. Further, all 15 head teachers reported that all teachers had received training on gender-responsive teaching methodology in the past year.

Boys were also asked about their gender perceptions at endline. A set of 7 survey items examining their attitudes towards the abilities of girls compared to boys, as well as gender norms, was used to calculate a composite score ranging from 0 to 7.²² The mean score at endline was 5.94, as shown in Table 37, with 18 of 45 respondents scoring a maximum of 7. Additionally, 43 of 45 boys agreed a lot when asked if girls should go to secondary school, with the remaining 2 agreeing a little.

Table 37. Boys' Gender Perception Composite Score at Endline

Timepoint	Frequency	Mean score	Standard Deviation
Midline2	230	5.66	1.60
Endline	45	5.94	1.18

Qualitative data supports the high measures of gender perceptions from quantitative data. All respondents who participated in FGDs and KIIs reported that STAGES had positively impacted communities' views on education. They said that communities' attitudes toward girls' education have improved, and community members have also become more engaged in supporting education overall. 'The issues are taken as their own agenda in the community', a woreda official said, 'and [community members] can say, "If Link does that much work, why wouldn't we work for our own community?"'

²¹ Teachers' gender perceptions were assessed using 5 items: 'Education is more important for boys than girls', 'Girls who get pregnant while still at school should be allowed back in school', 'Boys' education should get preference when money is scarce', 'I think boys are naturally more skilled than girls at reading and writing' and 'I think boys are naturally more skilled than girls at mathematics'. Each item was coded from 'disagree a lot' (0) to 'agree a lot' (3).

²² These items included: 'It is important for girls to go to school', 'Education is more important for boys than for girls', 'Girls learn the same at school as boys', 'A woman's role is to do household jobs and raise children', 'Men should share household duties', 'Boys are more naturally skilled than girls at reading and writing' and 'Boys are more naturally skilled than girls at mathematics'. Each item was coded from 'disagree a lot' (0) to 'agree a lot' (1).

In FGDs, respondents shared how attitudes had changed in communities as a whole. 'Going to school is a positive journey for me to realise my aspirations in life', a grade 10 girl said. 'I'm grateful that societal attitudes towards female education have changed, which allows me to enjoy attending school'.

Teachers echoed girls with their responses about the change in people's attitudes. 'We've actually seen positive shifts in parental support, especially from mothers, since STAGES did substantial community engagement work', a teacher said. Another teacher praised the efforts of mothers' and fathers' groups in mobilizing the community and resolving any issues immediately that arise. 'They are true role models in empowering girls protecting them', the teacher said, 'for example by not marrying them off, by keeping them in school, by overseeing the overall environment to make sure that no issues [related to SRGBV] are seen in their community'. Another teacher added that the collaboration of the school improvement committee and local education board committee in championing gender equity in education has also been vital. Their oversight has helped create a truly supportive environment'.

Teachers noted, however, that awareness still needs to be raised in the community. 'Even with the new student-centred, participatory methods we've learned', a teacher said, 'there are still some parents and community members who resist or don't understand why we're using these "modern" techniques'. Another teacher added that some families are still prioritizing girls' domestic labour or practising early marriage, but 'we are working to change this'.

In addition to changing attitudes about girls' education, communities are also more engaged in supporting their schools overall, especially through the implementation of school improvement plans. The need to raise funds locally to offset funding shortfalls from the government is a factor in driving increased community engagement, according to respondents. 'Previously this school got enough budget from the government', a head teacher said, 'and we did not need support from the community. But now we have faced budget deficit from the government side, and we understand the [how essential] community mobilization is'. Therefore, the school improvement committee is sharing strategies and plans with families of current and former students, as well as wealthy members of the community who can fund projects, the head teacher said.

In addition to raising money directly from community members, schools also use their improvement plans to organise income-generating activities. A woreda official reported that when STAGES started, 9 schools generated approximately 250,000 birr (£3,400), but now that total has increased twentyfold to 5,000,000 birr (£68,000). Selling agricultural products from the school compound—including teff and firewood—is one of the primary ways that schools generate income, according to zonal and woreda officials. 'This resource is creating a venue for the schools to improve their environment in support of girls and other marginalised groups', said a zonal official, who noted the funds support girls' clubs, tutorial classes, provision of sanitary pads and other activities.

4.5 Gender Perceptions²³ and Life Skills²⁴

Grade 10 girls' gender perceptions, which explore their attitudes about the abilities of girls compared to boys and gender norms and their life skills, which examine their comfort answering questions in the classroom and interest in continuing to study, were measured at ML2 and endline. Following STAGES

²³ Gender perceptions were an aggregate of the following questions: It is important for girls to go to school; Education is more important for boys than for girls; Girls learn the same at school as boys; A woman's role is to do household jobs and raise children; Men should share household duties; Boys are more naturally skilled than girls at reading and writing; Boys are more naturally skilled than girls at mathematics. Scale 0-7

²⁴ Life skills composite was an aggregate of the following questions: I get nervous when I have to read in front of others; I get nervous when I have to do maths in front of others; I feel confident answering questions in class; I would like to continue studying/attending school after this year; I often feel lonely at school. Scale 0-5

ToC, it is argued that both high gender perceptions and high life skills are vital in increasing outcomes of learning, transition and sustainability.

STAGES supported girls' SEL through 2 avenues—by incorporating SEL into its teaching training and specifically targeting girls with SEL support in schools with girls' clubs, mentoring, and other activities. Therefore, teachers, head teachers and education officials spearheaded efforts to make the school system and individual schools more responsive to students' emotional and social needs, while girls were also targeted to their specific, individual needs.

Grade 10 girls were asked a series of questions to measure their gender perceptions and life skills, and the items were analysed to create a composite score, as displayed in Table 38. These statistically significant gains mirror other data points related to SEL, including the decline in the proportion of girls from ML2 to endline who said that lack of confidence was a barrier to transitioning to secondary school—65.70% to 48.90%, respectively.

Table 38. For both composites, the increase in scores from ML2 to endline was statistically significant, suggesting that there were notable changes in girls' life skills and gender perceptions from ML2 to endline. These statistically significant gains mirror other data points related to SEL, including the decline in the proportion of girls from ML2 to endline who said that lack of confidence was a barrier to transitioning to secondary school—65.70% to 48.90%, respectively.

Table 38. Learning Outcomes: Life Skills Composite and Gender Perception for Grade 10 girls

Composite	ML2	Endline	P-value
Life Skills Composite ²⁵	3.7	4.3	0.000***
Gender Perception Composite ²⁶	5.7	6.6	0.000***

*Note: Three asterisks (***) denote a statistically significant at $p < 0.05$.*

Gender Perceptions by Subgroup

Gender perception scores were high across subgroups. As shown in Table 39, multiple subgroups of interest, including those with learning or physical difficulties, had high gender perception scores. Regionally, Damot Sore had the highest mean gender perception score at 6.75, with Kindo Koisha having the lowest at 6.13. Among subgroups of girls with learning or physical difficulties, girls with difficulty hearing had the lowest mean at 6.3, whereas those with difficulty communicating scored highest at 6.8. Despite the variation between these subgroups, the mean scores were still quite high.

Table 39. Gender Perception Composite Grade 10 by Subgroup

	Gender Composite Mean Endline	Freq	Gender Composite Standard Deviation Endline
Grade 10	6.6	294	0.87
Woreda			
Damot Pulasa	6.56	75	0.86

²⁵ To compute the life-skills scale mean score 5 items were used in in grades 10. To compute the decision-making scale, used 7 items. To compute the gender-perceptions scale, Grade 10, used 7 items.

²⁶ The gender perception construct is constructed from responses to 7 statements about girls' and boys' education.

	Gender Composite Mean Endline	Freq	Gender Composite Standard Deviation Endline
Damot Sore	6.75	60	0.6
Damot Woide	6.68	99	0.73
Kindo Koisha	6.13	60	1.17
Key Subgroups			
Difficulty Seeing	6.4	20	0.99
Difficulty Hearing	6.3	16	1.13
Difficulty Walking or Climbing	6.41	17	1.18
Difficulty Remembering	6.35	26	1.13
Difficulty with Self Care	6.35	17	1.06
Difficulty Communicating	6.8	20	0.41

Life Skills by Subgroup

Mean life skill scores were also higher across subgroups than gender perception scores, as displayed in Table 40. Regionally, Damot Sore had the highest mean life skills score at 6.75, with Damot Pulasa having the lowest at 4.18. Among subgroups of girls with learning or physical difficulties, girls with difficulty remembering had the lowest mean at 3.96, whereas those with difficulty communicating scored highest at 4.6.

Table 40. Life Skills Composite Grade 10 by Subgroup

	Life Composite Skills Mean Endline	Freq	Life Composite Standard Skills Deviation Endline
Grade 10	4.3	294	0.98
Woreda			
Damot Pulasa	4.18	75	1.09
Damot Sore	4.45	60	0.94
Damot Woide	4.37	99	0.89
Kindo Koisha	4.35	60	1.02
Key Subgroups			
Difficulty Seeing	4.55	20	0.68
Difficulty Hearing	4.18	16	1.27
Difficulty Walking or Climbing	4.41	17	1.12
Difficulty Remembering	3.96	26	1.24
Difficulty with Self Care	4.47	17	0.79

	Life Composite Endline	Skills Mean	Freq	Life Composite Standard Deviation Endline	Skills Endline
Difficulty Communicating	4.6		20	0.75	

Qualitative data corroborates the overall gains in girls' gender perceptions and life skills, with head teachers, teachers and girls themselves describing the positive change in well-being. Respondents said the gains had occurred for multiple reasons, including improved gender-inclusive teaching practices, greater community acceptance for girls' education and numerous project activities focused on girls' self-esteem and well-being, including the focus on menstrual health and hygiene and SEL and the creation of girls' clubs.

In FGDs, girls noted how the SEL training had benefited them in multiple ways by building their confidence, improving their interpersonal skills and helping them manage their emotions better. 'The SEL training was very impactful. It taught me techniques to manage difficult emotions like stress, anger and shyness in a better way', a grade 12 girl said. Another grade 12 girl said the female role models who shared their stories during SEL training changed the course of her life. 'I have experienced from different people that I can achieve a good level', she said. 'I was going to go abroad, but taking their experience made me pay attention to education'. Another grade 12 girl said, 'The training provided valuable lessons on building self-trust, being confident and communicating assertively—skills that have transformed my life'.

Head teachers and teachers said they noticed how girls' well-being and SEL skills had improved. In an FGD, teachers explained how the formation of girls' clubs was transformative. 'The club has really helped build girls' confidence, leadership skills and awareness of important issues like gender-based violence, sexual health and future opportunities', the teacher said. A head teacher explained how girls at his school had changed:

'In addition, SEL training increases the emotional intelligence of girl students and builds their confidence. Because of this training a lot of shy students become confident and introduce themselves and present class activities in front of others with confidence. Even their families were surprised by their changes, and last year some of them joined university'.

5. Conclusion and Recommendations

5.1 Conclusions

Overall, the endline evaluation highlights key successes in the project objectives targeting learning, transition and sustainability outcomes. First, although grade 8 girls had statistically significant gains in literacy and numeracy from ML2 to endline compared with grade 10 girls' unchanged outcomes between the 2-time points, the negative effect of secondary school closures on grade 10 girls' performance on the SeGRA and SeGMA must be considered. Second, multiple positive findings from endline, including significant gains in socio-emotional composites and a decrease in the proportion of girls considering dropping out, suggest that the project has had success in addressing the multiple barriers girls face in transitioning to secondary school. Finally, with the project closing in the fall of 2024, qualitative data from FGDs and KIIs with project stakeholders suggests that education officials at all levels, as well as the communities they serve, are eager and well-positioned to carry on project interventions, especially those that the project targeted for sustainability—pedagogical training and leadership, GIRP and safeguarding.

The statistically significant gains in grade 8 girls' reading performance at endline from grade 6 girls' outcomes at ML2 revealed stronger performance in familiar word reading, invented word reading and ORF than reading comprehension, although gains were also statistically significant on this subtask from ML2 to endline. Grade 8 girls' percentage scores at endline were highest on ORF (83.81%), invented word reading (81.45%) and familiar word reading (71.12 percent) and lowest on reading comprehension (45.98%). Although the endline percentage score on reading comprehension for grade 8 girls was a statistically significant increase from the grade 6 percentage score of 16.5% at ML2, it underscores the continued gap between girls being able to read aloud the words in the reading passage and their ability to make sense of what they are reading.

It is also important to note that the statistically significant gains from grade 6 at ML2 to grade 8 at endline cannot be solely attributed to the intervention for several reasons. First, the evaluation design lacked a true longitudinal design, as well as a counterfactual. Additionally, 2 years of schooling may have influenced scores. Still, the gains in reading outcomes do underscore the importance of the literacy interventions that STAGES introduced since ML2, including supporting reading corners in schools, providing them reading books and conducting additional teacher training on teaching reading. Further, the two additional years of schooling were delivered within the timeframe of project support.

Grade 8 girls also performed better on numeracy subtasks, with statistically significant gains on all subtasks. Grade 8 girls' numeracy scores improved at endline on all 3 subtasks that grade 6 girls also were administered at ML2—addition (55.09% at ML2 to 69.07% at endline), subtraction (40.69% to 51.66%) and word problems (56.62% to 69.80%). The opportunity for even further improvement and growth in numeracy outcomes is possible with percentage scores ranging from 51.66% to 69.80%.

The impact of lengthy secondary school closures on grade 10 girls' endline learning outcomes must be considered when evaluating the unchanged SeGMA and SeGRA results from ML2 to endline. Overall, grade 10 girls' mean SeGRA and SeGMA aggregate scores slightly declined from ML2 to endline, but the decreases were not statistically significant. These results, however, should be interpreted with a major caveat—the closure of secondary schools in the Wolaita Zone in February and March 2024 before the endline evaluation was administered in May. All 17 secondary schools supported by STAGES were closed for at least 20 days due to teacher protests about delays in receiving their salaries. Additionally, the tutorial supported that multiple respondents credited as having an important impact on girls' learning also stopped because schools were closed. Girls shared in FGDs how the closures affected their motivation. 'In grade 11, I was happy to go to school. ... But some things are not pleasant in the current situation', a grade 12 girl said, 'and teachers don't teach because they don't get paid. I don't like going to school because of this'. Teachers reported that the closures affected attendance and the quality of teaching and learning, with project staff also detailing how classes were rushed to make up for lost instructional time. Overall, these substandard conditions very likely had a negative impact on learning outcomes in secondary schools.

After ranking highest out of the 4 project woredas at ML2 on SeGRA and SeGMA, girls in Kindo Koisha, which dealt with labour unrest at schools this year, had the lowest literacy and numeracy scores out of the 4 woredas in both grade 8 and grade 10. In the subgroup analysis of learning outcomes, the comparatively poor performance of girls in Kindo Koisha was notable at endline. The mean aggregate scores for both grade 8 girls on EGRA and EGMA and grade 10 girls on SeGRA and SeGMA in Kindo Koisha were the lowest of the 4 woredas. By contrast, grade 10 girls in Kindo Koisha had the highest SeGRA and SeGMA scores out of the 4 woredas at ML2. This downward trend underscores the likely impact of external factors on learning outcomes. Kindo Koisha has deteriorated politically in the past 2 years since ML2 was administered. School personnel and teachers protested educational leadership in the woreda in late 2023 into the summer of 2024, prompting administrative changes at the woreda level. These external factors outside of the project's control likely had an impact on learning outcomes in the woreda.

The multiple positive trends found in transition outcome data suggest that STAGES has had success in addressing the multiple barriers girls face in transitioning to secondary school.

- First, the home-based factors that girls cited most frequently for preventing their peers from transitioning—including early marriage, lack of parental support, early pregnancy and illness—all declined from ML2 to endline.
- Second, girls' socio-emotional composites measuring gender and life skills increased statistically significantly from ML2 to endline.
- Third, the proportion of grade 10 girls who sometimes or often considered dropping out decreased from 18.08% at ML2 to 6.80% at endline.
- Finally, in FGDs, girls and teachers spoke positively about the changes they had observed in families' attitudes, as well as in communities' beliefs as a whole, with respect to girls' education.

Overall, all these findings suggest that project interventions have successfully mitigated the factors that prevent girls from transitioning to secondary school. Findings also suggest that families are better equipped to support their children's educational needs, as a statistically significant difference was noted between ML2 and endline in the proportion of girls who said their parents or caregivers pay for everything they need to go to school—from 56.46% to 79.25%, respectively.

As for sustainability, stakeholders at all levels seem prepared to carry out key interventions once the project closes. The efforts that STAGES made during the life of the project to target all levels of the education system—from the zonal education office to communities and classrooms—appear to have fostered a sense of ownership in project interventions in all stakeholders and set up all education actors to sustain them once the project closes. It is especially notable that the zonal office said it has already scaled the pedagogical supervision and leadership activity to all woredas in the zone and trained other education actors beyond the zone on the intervention, as well as reported using its own funds to expand the GIRP intervention to non-project woredas in the zone. In KIIs and FGDs, all types of respondents—from zonal and woreda education officials to head teachers, teachers and girls—shared positive details about STAGES and seemed motivated to sustain activities. Overall, zonal and woreda officials seem poised to sustain STAGES activities, with interventions incorporated into 3-year strategic planning, certain project resources such as teacher mentoring and coaching checklists aligned with zonal monitoring tools and school improvement planning and mothers' and fathers' groups now considering the importance of gender and inclusion in their efforts.

5.2 Recommendations

Focus on improving reading comprehension skills and strategies in both English and Wolaytatto in future iterations of literacy interventions in the Wolaita Zone, primarily by training teachers intensively on classroom activities and instructional strategies focused on reading comprehension. Girls in grade 8 and grade 10 had room for improvement in their reading comprehension outcomes on the EGRA and SeGRA, respectively. There was especially a disconnect between grade 8 girls' performance on the ORF and reading comprehension subtasks; this gap suggests that while girls read the majority of the words fluently in the short reading passage, they could not make meaning of it, based on their weaker performance in answering reading comprehension questions about the story. Therefore, future literacy interventions should focus on improving teachers' abilities to improve students' reading comprehension skills for both narrative and expository texts in both Wolaytatto and English. Through classroom activities and instructional strategies, teachers can encourage their students to adopt certain skills related to reading comprehension, including making connections, visualising, predicting, inferring, asking questions, summarising, synthesising and determining important ideas and maintaining meaning. Published in 2021, The Reading Comprehension Interest Group's *Reading Comprehension Strategies and Tools for*

Teachers of Mother Tongue Based Multilingual Education (MTB-MLE) Classrooms is a useful resource for developing these skills.

Expand tutorials to improve knowledge in other subjects and/or hone English skills. In FGDs and KIIs, girls in grades 10 and 12, teachers, and other respondents said that tutorials were one of the most impactful activities in improving learning outcomes. Quotes from FGDs with grade 10 and 12 girls especially spoke to the essential value of this intervention in helping girls grasp skills they could not fully understand during regular class sessions and develop their fluency in English. Future projects should use the tutorial activity as a best practice to ensure girls maximise their learning opportunities. Additionally, future iterations of the tutorial activity may consider incorporating elements of bolstering girls' English language skills because increasing exposure to English outside the classroom is key to developing fluency.

Modify the SeGRA and SeGMA if needed for future iterations of education interventions, with evaluators, implementing partners and education officials exploring the most reliable and valid way to measure higher-level learning skills. Per the request of project funders as part of the GEC evaluation design, researchers developed and adapted the SeGRA and SeGMA specifically for the STAGES evaluations because EGRA and EGMA cannot measure competency in secondary school students' higher-level literacy and numeracy skills. Whether the SeGRA and SeGMA, as they are currently constructed, are appropriate for future evaluations should be considered. In conjunction with implementing partners and education officials, evaluators should consult and review tools used in other Girls' Education Challenge projects, as applicable, if alternatives are deemed necessary.

Research why literacy and numeracy results in Kindo Koisha were lower than other woredas' scores. The literacy and numeracy results for both grade 8 and 10 girls were lowest in Kindo Koisha at endline out of the 4 project woredas. In contrast, at ML2, grade 10 girls in Kindo Koisha had the highest SeGRA and SeGMA aggregate scores. This change may warrant further investigation to determine if it was anomalous or related to specific factors affecting student outcomes in the woreda.

Look into avenues for zonal educational officials to continue to scale up project interventions, including pedagogical leadership and supervision, GIRP and safeguarding mechanisms, and effective teaching and learning materials that produced gains in numeracy. In an FGD with zonal officials, they reported how other regional education stakeholders viewed them as 'a centre of excellence' in the sector for the pedagogical supervision and training intervention. They also had trained regional counterparts from the South Ethiopia Regional State in the practice. Garnering such respect from education officials presents opportunities to scale up even more interventions from the project, not only those STAGES targeted for sustainability but also teaching and learning materials that brought about statistically significant gains in numeracy.

Connect zonal and woreda education officials to subject-based experts and other resources so they can update materials as needed after the project closes and promote the institutionalisation of materials beyond the zonal level. In FGDs, zonal and woreda education officials noted that one of the obstacles to sustaining interventions such as GIRP and pedagogical leadership and supervision was the need to remain knowledgeable and up to date on subject matter such as gender, inclusive education and safeguarding. Before STAGES closes, implementing partners should attempt to help education actors build knowledge-sharing partnerships with organisations and community-based groups to ensure they can access specialised knowledge and best practices. Additionally, as noted in an FGD, zonal officials should continue to promote pedagogical leadership and supervision, GIRP and other interventions beyond the zonal level and train officials at the regional level in these best practices, with the ultimate goal of scaling them up beyond the Wolaita Zone.

Host a zone-wide meeting of fathers' and mothers' groups to validate their importance, promote best practices and encourage continued community involvement in education. In FGDs and KIIs, respondents

described how instrumental community-based groups such as the mothers' and fathers' groups were in supporting education in their communities by encouraging children to remain enrolled in schools, raising awareness about the importance of education, ensuring safeguarding mechanisms are actively applied and organising income-generating activities. With the project closing, a zone-wide meeting of these groups' representatives to recognise exemplary results and promote the best practices of these groups, some of whom may have developed practical solutions to address localised problems, may be an effective way to validate their importance, motivate them to continue their work and expand the use of best practices.

To promote the success of SEL interventions, explore ways to enable girls who have benefited from the project to apply what they have learnt, possibly by becoming mentors in their communities. In FGDs and KIIs, respondents credited the project's SEL interventions in helping girls develop emotional intelligence, better manage their feelings, become more confident communicators and improve their knowledge of topics such as SRGBV and sexual and reproductive health. Although STAGES targeted teachers with SEL trainings, it may be helpful, with the project closing and some trained teachers likely relocating to other schools, to identify women, especially those who have transitioned to university, who could apply what they have learnt in their respective communities. One effective strategy could be to train them to be mentors or establish a Big Sisters-type program.

Identify communities that most effectively implemented safeguarding mechanisms and fostered successful coordination between stakeholders and other effective parts of SIPs that promote gender and inclusion so education officials can promote these practices across the Wolaita Zone. In an FGD, zonal officials mentioned that improvements in safeguarding mechanisms could be made. 'We are working on enhancing coordination between schools, communities and relevant authorities to ensure prompt and effective interventions', a zonal official said. Other effective practices targeting gender and inclusion involve multiple community actors, so it would be prudent to promote them as well. STAGES could help officials by identifying communities that are implementing the most effective safeguarding interventions and promoting these best practices. This would lead to other communities replicating these successful models and ensuring they are formally adopted into schools' improvement planning.

Further research teachers' and head teachers' beliefs about support from middle management at the woreda level and intensify support to woreda education offices as needed to support the transformation of attitudes and support for girls and leadership for girls' education. At endline, only 1 of 15 teachers said they believed that woreda officials were always responsive to girls' needs, while only 3 of 15 teachers said they believed that they were mostly responsive. Due to the small sample size of teachers at endline, further research should be conducted to determine if these results are indeed indicative of wider sentiments in schools and what specific needs these woreda officials need to be more responsive about. Based on this research, future projects can design activities to support middle management at the woreda level.