

# Project Evaluation Report

<b>Report title:</b>	Midline Evaluation
<b>Evaluator:</b>	Consilient
<b>GEC Project:</b>	Adolescent Girls' Education in Somalia (AGES)
<b>Country:</b>	Somalia
<b>GEC window:</b>	Leave No Girl Behind
<b>Evaluation point:</b>	Midline 1
<b>Report date:</b>	August 2022

## Notes:

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

# Adolescent Girls' Education in Somalia (AGES)

LEAVE NO GIRL BEHIND

Midline Evaluation

Prepared By:



Authors:

Payce Madden

Brenton D. Peterson

Anastasia Vasilyeva

Nicolas Wicaksono

August 3, 2022

Version 3

# Table of Contents

<b>TABLE OF CONTENTS.....</b>	<b>I</b>
<b>LIST OF FIGURES.....</b>	<b>III</b>
<b>LIST OF TABLES .....</b>	<b>III</b>
<b>LIST OF ABBREVIATIONS .....</b>	<b>VI</b>
<b>EXECUTIVE SUMMARY .....</b>	<b>1</b>
LEARNING.....	1
TRANSITION .....	2
TEACHING QUALITY .....	3
LEADERSHIP, LIFE SKILLS, AND SELF-EFFICACY.....	4
SCHOOL MANAGEMENT AND GOVERNANCE.....	5
ACCESS TO PROTECTIVE SERVICES .....	6
STRENGTHENED ECONOMIC CIRCUMSTANCES.....	6
ENHANCED SOCIAL SUPPORT FOR FEMALE YOUTH .....	6
<b>1. INTRODUCTION .....</b>	<b>8</b>
<b>2. METHODOLOGY.....</b>	<b>11</b>
2.1 OVERALL EVALUATION DESIGN .....	11
2.2 DATA COLLECTION TOOLS .....	11
2.3 NEW NFE (COHORT 4) SAMPLING DESIGN .....	14
2.4 MIDLINE RE-CONTACT RATES.....	15
2.5 CHALLENGES AND LIMITATIONS.....	20
<b>3. SAMPLE CHARACTERISTICS .....</b>	<b>22</b>
3.1 NEW NFE (COHORT 4) SAMPLE DEMOGRAPHICS.....	22
3.2 COHORT PROFILES.....	25
<b>4. LEARNING .....</b>	<b>33</b>
4.1 AGGREGATE LEARNING OUTCOMES .....	34
4.2 SUBTASK-SPECIFIC GAINS IN LEARNING .....	43
4.3 SUBGROUP PROGRAMME IMPACT .....	47
4.4 TESTING THE THEORY OF CHANGE .....	57
4.5 DISCUSSION .....	67
<b>5. TRANSITION.....</b>	<b>69</b>
5.1 AGGREGATE TRANSITION OUTCOMES .....	69
5.2 SUBGROUP TRANSITION RATES .....	73
5.3 TESTING THE THEORY OF CHANGE .....	75
<b>6. SUSTAINABILITY .....</b>	<b>78</b>
<b>7. INTERMEDIATE OUTCOMES .....</b>	<b>88</b>
7.1 ATTENDANCE .....	88
7.2 TEACHING QUALITY .....	89
7.3 LEADERSHIP AND LIFE SKILLS .....	101
7.4 SCHOOL MANAGEMENT AND GOVERNANCE .....	113
7.5 COMMUNITY ATTITUDES .....	122
7.6 INCREASED SELF-EFFICACY .....	131
7.7 STRENGTHENED ECONOMIC CIRCUMSTANCES .....	136
<b>8. LEARNING .....</b>	<b>142</b>

8.1 AGGREGATE LEARNING OUTCOMES .....	142
8.2 SUBGROUP LEARNING OUTCOMES.....	144
8.3 FOUNDATIONAL SKILL GAPS .....	146
8.4 GRADE-LEVEL COMPETENCY.....	150
<b>9. INTERMEDIATE OUTCOMES .....</b>	<b>151</b>
9.1 ATTENDANCE .....	151
9.2 TEACHING QUALITY .....	152
9.3 LEADERSHIP AND LIFE SKILLS .....	156
9.4 SCHOOL MANAGEMENT AND GOVERNANCE .....	160
9.5 INCREASED SELF-EFFICACY .....	164
9.6 STRENGTHENED ECONOMIC CIRCUMSTANCES .....	169
9.7 ENHANCED SOCIAL SUPPORT FOR FEMALE YOUTH.....	172
<b>10. ANCILLARY ANALYSIS .....</b>	<b>177</b>
10.1 MENSTRUAL HEALTH AND HYGIENE .....	177
10.2 COVID-19 IMPACTS IN AGES COMMUNITIES/SCHOOLS.....	180
<b>11. RECOMMENDATIONS.....</b>	<b>184</b>
<b>ANNEXES .....</b>	<b>187</b>
ANNEX 1 - EVALUATION METHODOLOGY.....	187
ANNEX 2 – LEARNING TABLES – MIDLINE COHORTS .....	198
ANNEX 3 – LEARNING TABLES – BASELINE NEW NFE COHORT .....	220
ANNEX 4 – LOGFRAME .....	226
ANNEX 5 – DATA COLLECTION TOOLS.....	226
ANNEX 6 – DATASETS AND REPLICATION FILES .....	226
ANNEX 7 – PROJECT MANAGEMENT RESPONSE.....	226

## List of Figures

FIGURE 1: IMPACT OF GIRL'S CHARACTERISTICS ON RE-CONTACT OUTCOMES .....	19
FIGURE 2: CHANGES IN AGGREGATE NUMERACY AND LITERACY SCORES .....	35
FIGURE 3: PERCENT OF GIRLS SUPPORTING FAMILY'S AGRICULTURAL OR OTHER ECONOMIC ACTIVITIES.....	43
FIGURE 4: DISTRIBUTION OF SCORES FOR NUMERACY SUBTASKS 9 THROUGH 11.....	45
FIGURE 5: CHANGES IN ASSESSMENT SCORES BY DISABILITY STATUS .....	49
FIGURE 6: CHANGES IN ASSESSMENT SCORES BY MENTAL HEALTH STATUS .....	50
FIGURE 7: LEARNING OUTCOME CHANGES BY GIRLS' ACCESS TO WATER .....	52
FIGURE 9: LEARNING OUTCOME CHANGES BY AVAILABILITY OF TAKEHOME MATERIALS.....	56
FIGURE 9: ML NUMERACY AND LITERACY SCORES BASED ON NUMBER OF HOURS STUDIED AT HOME .....	59
FIGURE 10: IMPACT OF SELECT YLI QUESTIONS ON ML NUMERACY AND LITERACY SCORES .....	61
FIGURE 12: CAREGIVER ATTITUDES AS PREDICTORS OF SUCCESSFUL TRANSITION.....	76
FIGURE 13: ASSESSMENTS OF TEACHING QUALITY AS A PREDICTOR OF TRANSITION OUTCOMES .....	77
FIGURE 14: PERCENT OF MALE AND FEMALE TEACHERS PARTICIPATING IN TRAININGS AT MIDLINE .....	86
FIGURE 15: CHANGE IN USE OF FORMATIVE ASSESSMENTS.....	97
FIGURE 16: CHANGE IN YLI SCORES FROM BL TO ML BY ZONE .....	106
FIGURE 17: CHANGE IN CEC ENGAGEMENT, BY ROUND .....	119
FIGURE 18: SHARE OF CAREGIVERS WHO SUPPORT INVESTMENT IN EDUCATION, BY STATE AND ROUND.....	125
FIGURE 19: INCOME BY REGION .....	137
FIGURE 20: C4 NFE GIRLS NUMERACY AND LITRACY MEAN SCORES, BY AGE .....	143
FIGURE 21: DISTRIBUTION OF LSI SCORES, BY STATE .....	156
FIGURE 22: DISTRIBUTION OF YLI SCORES, BY STATE .....	157
FIGURE 23: DOES THE CEC DO ANYTHING TO SUPPORT GIRLS' EDUCATION, BY STATE.....	162
FIGURE 24: INCOME BY REGION .....	170
FIGURE 25: SHARE OF GIRLS REPORTING ACCESS TO DEVELOPMENT OPPORTUNITIES .....	174
FIGURE 26: CHANGE IN REUSE OF PADS AND FREQUENCY OF PAD REPLACEMENT .....	179

## List of Tables

TABLE 1: SHARE OF MIDLINE RESPONDENTS COMPLETING FULL HOUSEHOLD SURVEY .....	12
TABLE 2: SAMPLE TARGETS AND ACHIEVED SAMPLE FOR SCHOOL-BASED DATA COLLECTION TOOLS.....	13
TABLE 3: POPULATION AND SAMPLE ALLOCATION, BY STATE, FOR NEW NFE COHORT.....	15
TABLE 4: RE-CONTACT RATE AND ACHIEVED SAMPLE, BY COHORT AND STATE.....	17
TABLE 5: INDIVIDUAL AND HOUSEHOLD CHARACTERISTICS OF NEW NFE (COHORT 4) SAMPLE .....	22
TABLE 6: MATCH BETWEEN POPULATION AND SAMPLE CHARACTERISTICS, NEW NFE COHORT.....	24
TABLE 7: CHARACTERISTICS OF FE COHORT, BEFORE AND AFTER PANEL ATTRITION.....	25
TABLE 8: CHARACTERISTICS OF ABE COHORT, BEFORE AND AFTER PANEL ATTRITION .....	27
TABLE 9: CHARACTERISTICS OF NFE COHORT, BEFORE AND AFTER PANEL ATTRITION .....	29
TABLE 10: CHANGES IN INDIVIDUAL AND HOUSEHOLD CHARACTERISTICS, BL TO ML.....	30
TABLE 11: NUMBER OF INSTITUTIONS & GIRLS IN THE PANEL SAMPLE .....	33
TABLE 12: BENCHMARK CHANGE FOR EACH GRADE LEVEL AT BASELINE.....	34
TABLE 13: DIFFERENCE IN OUTCOMES BETWEEN PANEL AND NON-PANEL SAMPLE GIRLS AT BASELINE .....	36
TABLE 14: EFFECT OF ML ATTENDANCE, AGE, COHORT, AND LOCATION ON LEARNING OUTCOME CHANGES.....	38
TABLE 15: AGGREGATE NUMERACY SCORE CHANGES FOR THE PANEL SAMPLE .....	40
TABLE 16: AGGREGATE LITERACY SCORE CHANGES FOR THE PANEL SAMPLE .....	41
TABLE 17: CHANGES IN NUMERACY SUBTASK SCORES FOR THE PANEL SAMPLE (P-VALUES IN PARENTHESES).....	43
TABLE 18: PANEL SAMPLE SCORES FOR NEW NUMERACY SUBTASKS AT ML .....	44
TABLE 19: CHANGES IN LITERACY SUBTASK SCORES FOR THE PANEL SAMPLE (BASELINE VALUES IN PARENTHESES).....	46
TABLE 20: LEARNING OUTCOME DIFFERENCES AMONG SELECTED SUB-GROUPS .....	51
TABLE 21: LEARNING OUTCOME CHANGES BY MARGINALITY.....	53
TABLE 22: LEARNING OUTCOME DIFFERENCES AMONG GIRLS WITH DIFFERENT SCHOOL INFRASTRUCTURE.....	55

TABLE 23: EFFECT OF 1 POINT INCREASE IN BL ATTENDANCE RATE ON CHANGES IN LEARNING OUTCOMES.....	58
TABLE 24: EFFECT OF 1 POINT INCREASE IN BL YLI SCORE AND CHANGES IN LEARNING OUTCOMES .....	60
TABLE 25: EFFECT OF GEF PARTICIPATION ON LEARNING SCORES AT MIDLINE .....	62
TABLE 26: EFFECT OF TEACHING PRACTICES ON GAINS IN NUMERACY AND LITERACY SCORES SINCE BASELINE .....	64
TABLE 27: EFFECT OF CAREGIVER ATTITUDES ON NUMERACY AND LITERACY SCORES .....	66
TABLE 28: TRANSITION PATHWAYS, ACCORDING TO STARTING POINT OR COHORT.....	69
TABLE 29: TRANSITION OUTCOMES, AMONG DISTINCT COHORTS.....	72
TABLE 30: SUBGROUP TRANSITION RATES, ACROSS COHORTS .....	73
TABLE 31: PARTICIPATION RATES AND ACTIVITIES OF GEFS, BY COHORT .....	81
TABLE 32: NUMBER OF TEACHERS TRAINED AT BASELINE AND MIDLINE.....	85
TABLE 33: ATTENDANCE RATES, BASED ON HEADCOUNTS, IN FORMAL SCHOOLS .....	88
TABLE 34: ATTENDANCE RATES, BASED ON HEADCOUNTS, IN ABE AND NFE CENTRES.....	89
TABLE 35: CHANGE IN TEACHER PROFESSIONALISM BY TRACK AND REGION.....	91
TABLE 36: CHANGE IN MEASURES OF GENDER EQUALITY.....	93
TABLE 37: CHANGE IN MEASURES OF GENDER EQUITY .....	94
TABLE 38: CHANGE IN USE OF DISCIPLINARY PRACTICES BY EDUCATION TRACK .....	95
TABLE 39: CHANGE IN USE OF DISCIPLINARY PRACTICES BY REGION .....	95
TABLE 40: USE OF PARTICIPATORY TEACHING PRACTICES IN CLASSES, FE SCHOOLS.....	97
TABLE 41: USE OF PARTICIPATORY TEACHING PRACTICES IN CLASSES, C1 NFE AND ABE.....	98
TABLE 42: USE OF POSITIVE TEACHING PRACTICES REPORTED BY GIRL STUDENTS, BY EDUCATION TRACK .....	99
TABLE 43: USE OF POSITIVE TEACHING PRACTICES REPORTED BY GIRL STUDENTS, BY REGION.....	100
TABLE 44: LIST OF YOUTH LEADERSHIP INDEX (YLI) QUESTIONS.....	102
TABLE 45: LIST OF LIFE SKILLS INDEX (LSI) QUESTIONS.....	103
TABLE 46: CHANGE IN YLI SCORES FROM BL TO ML BY SCHOOL TYPE .....	104
TABLE 47: CHANGE IN YLI SCORES FROM BL TO ML BY AREA .....	105
TABLE 48: CHANGE IN YLI SCORES FROM BL TO ML BY GIRLS' CHARACTERISTIC .....	107
TABLE 49: CHANGE IN YLI SCORES FROM BL TO ML BY HOUSEHOLD CHARACTERISTICS .....	108
TABLE 50: CHANGE IN YLI SCORES FROM BL TO ML BY CAREGIVER ATTITUDE.....	109
TABLE 51: CHANGE IN YLI SCORES FROM BL TO ML BY LEARNING ENVIRONMENT .....	111
TABLE 52: CHANGE IN YLI SCORES FROM BL TO ML BY PROGRAMME PARTICIPATION .....	112
TABLE 53: REASONS CITED BY CAREGIVERS FOR A GIRL TO BE OUT-OF-SCHOOL, BY ROUND .....	117
TABLE 54: SHARE OF CAREGIVERS WHO FEEL COST IS AN ACCEPTABLE REASON FOR NOT ENROLLING GIRL .....	128
TABLE 55: SUMMARY OF C4 NFE GIRLS SCORES AT BL .....	142
TABLE 56: SUB-GROUP IMPACTS ON BL NUMERACY AND LITERACY SCORES OF C4 NFE GIRLS .....	144
TABLE 57: GAPS IN FOUNDATIONAL NUMERACY SKILLS AMONG NEW (COHORT 4) NFE GIRLS.....	147
TABLE 58: GAPS IN FOUNDATIONAL LITERACY SKILLS AMONG NEW (COHORT 4) NFE GIRLS.....	149
TABLE 59: ATTENDANCE HEADCOUNTS IN C4 NFE CENTRES, BY STATE.....	152
TABLE 60: TEACHER PROFESSIONALISM, C4 NFE.....	152
TABLE 61: TEACHERS' INTERACTIONS WITH GIRL STUDENTS, C4 NFE .....	153
TABLE 62: USE OF DISCIPLINARY PRACTICES, C4 NFE .....	153
TABLE 63: USE OF PARTICIPATORY TEACHING PRACTICES IN CLASSES, C4 NFE .....	154
TABLE 64: USE OF POSITIVE TEACHING PRACTICES REPORTED BY GIRL STUDENTS, C4 NFE.....	155
TABLE 65: MEAN LSI SCORES, BY STATE.....	156
TABLE 66: MEAN YLI SCORES, BY STATE.....	157
TABLE 67: MEAN YLI SCORES, BY GIRL AND HOUSEHOLD CHARACTERISTICS .....	158
TABLE 68: MEAN LSI SCORES, BY GIRL AND HOUSEHOLD CHARACTERISTICS .....	159
TABLE 69: SHARE OF CAREGIVERS CITING REASONS THEY DID NOT ENROL THEIR DAUGHTER IN FORMAL SCHOOL .....	161
TABLE 70: PARTICIPATION IN DEVELOPMENT OPPORTUNITIES, AMONG SUBGROUPS.....	174
TABLE 71: CHANGE IN GIRLS REPORTING THAT MENSTRUATION PREVENTS ATTENDANCE .....	178
TABLE 72: CHANGE IN MENSTRUAL PRODUCT CLEANING PRACTICES .....	179
TABLE 73: COVID-RELATED OUTCOMES ACROSS REGIONS.....	181
TABLE 74: SAMPLE TARGETS BEFORE AND AFTER SCHOOL-LEVEL ATTRITION .....	189
TABLE 75: ACHIEVED SAMPLE, RELATIVE TO TARGET .....	189
TABLE 77: SOMALI LITERACY AMONG SUBGROUPS OF FE, ABE, AND NFE GIRLS (FULL PANEL SAMPLE).....	199
TABLE 78: SOMALI LITERACY AMONG SUBGROUPS OF FE, ABE, AND NFE GIRLS (SUBGROUPS DERIVED FROM HOUSEHOLD SURVEY) .....	202
TABLE 79: NUMERACY (ABBREVIATED VERSION) AMONG SUBGROUPS OF FE, ABE, AND NFE GIRLS (FULL PANEL SAMPLE) .....	203

TABLE 80: NUMERACY (ABBREVIATED VERSION) AMONG SUBGROUPS OF FE, ABE, AND NFE GIRLS (SUBGROUPS DERIVED FROM HOUSEHOLD SURVEY) .....	206
TABLE 81: NUMERACY (FULL VERSION) AMONG SUBGROUPS OF FE, ABE, AND NFE GIRLS (FULL PANEL SAMPLE).....	207
TABLE 82: NUMERACY (FULL VERSION) AMONG SUBGROUPS OF FE, ABE, AND NFE GIRLS (SUBGROUPS DERIVED FROM HOUSEHOLD SURVEY) .....	211
TABLE 83: FOUNDATIONAL GAPS IN NUMERACY AT MIDLINE, FE GIRLS .....	213
TABLE 84: FOUNDATIONAL GAPS IN NUMERACY AT MIDLINE, ABE GIRLS.....	214
TABLE 85: FOUNDATIONAL GAPS IN NUMERACY AT MIDLINE, OLD NFE (COHORT 1) GIRLS.....	215
TABLE 86: FOUNDATIONAL GAPS IN SOMALI LITERACY AT MIDLINE, FE GIRLS .....	216
TABLE 87: FOUNDATIONAL GAPS IN SOMALI LITERACY AT MIDLINE, ABE GIRLS.....	217
TABLE 88: FOUNDATIONAL GAPS IN SOMALI LITERACY AT MIDLINE, OLD NFE (COHORT 1) GIRLS.....	218
TABLE 89: LITERACY COMPETENCY (PARTIAL) AT LEVEL OF LOWER PRIMARY COMPLETION, BY COHORT .....	219
TABLE 90: NUMERACY COMPETENCY (PARTIAL) AT LEVEL OF LOWER PRIMARY COMPLETION, BY COHORT .....	219
TABLE 91: SOMALI LITERACY AMONG SUBGROUPS OF NEW NFE GIRLS COHORT .....	220
TABLE 92: NUMERACY AMONG SUBGROUPS OF NEW NFE GIRLS COHORT.....	223

## List of Abbreviations

ABE	Accelerated Basic Education
AGES	Adolescent Girls' Education in Somalia
ALP	Alternative Learning Programme
al Shabaab	Harakat al-Shabaab al-Mujahideen
AMISOM	African Union Mission to Somalia
BDR	Banadir (region)
BL	Baseline
C1 NFE	Cohort 1 Non-Formal Education
C4 NFE	Cohort 4 Non-Formal Education
CEC	Community Education Committee
CPYDS	Chinese Positive Youth Development Scale
DEO	District Education Officer
DiD	Difference-in-Differences
EGMA	Early-Grade Math Assessment
EGRA	Early-Grade Reading Assessment
FCDO	Foreign and Commonwealth Development Office
FE	Formal Education
FGD	Focus Group Discussion
FGS	Federal Government of Somalia
FMS	Federal Member State
GEF	Girls' Empowerment Forum
IDP	Internally-Displaced Person
JSS	Jubaland State
KII	Key Informant Interview
LSI	Life Skills Index
ML	Midline
MOE	Ministry of Education
NFE	Non-Formal Education
OOS	Out-of-School
PSU	Primary Sampling Unit
REO	Regional Education Officer
SOMGEP-T	Somali Girls' Education Promotion Programme – Transition
SOS	Somali Shilling
STI	Sexually-transmitted infection
SWS	South West State
USAID	United States Agency for International Development
USD	United States Dollar
VSL/VSLA	Village Savings and Loan Association
YLI	Youth Leadership Index



# Executive Summary

---

The Adolescent Girls' Education in Somalia (AGES) programme is a large-scale education initiative funded by both FCDO and USAID under the umbrella of the Girls' Education Challenge. The programme is implemented in three of Somalia's Federal Member States – Hirshabelle, Jubaland, and South West State (SWS) – along with the country's capital, Mogadishu. By working with communities, private education providers, and government ministries, the programme seeks to improve numeracy, literacy, financial literacy, knowledge of sexual and reproductive health, and other key outcomes for a set of over 40,000 adolescent girls who are among the most marginalised in Somalia.<sup>1</sup>

Following a baseline evaluation in November 2019, the programme has tracked three cohorts of girls through to the midline: formal school (FE) girls, who were enrolled in grades 1-2 at the time; Accelerated Basic Education (ABE) girls; and Cohort 1 Non-Formal Education (C1 NFE) girls, whose learning courses focus on functional literacy and life skills.<sup>2</sup> These three cohorts of girls were re-contacted as part of the midline alongside a new cohort of NFE girls (C4 NFE) who were enrolled in NFE courses immediately prior to the midline evaluation. Below, we discuss outcomes for the midline cohorts (FE, ABE, C1 NFE), followed by the baseline cohort of C4 NFE girls.

## Learning

---

### Midline

Learning outcomes – numeracy and Somali literacy – are central to the programme's goals. Since baseline, FE girls were the only group for which an improvement in numeracy scores from baseline to midline was observed, with an improvement greater than the expected improvement, per benchmarks developed at

---

<sup>1</sup> Financial literacy includes knowledge of financial concepts (e.g., profit and interest) that are essential to running a small business and the application of arithmetic to answer financial questions (such as the calculation of a total price or discounts). Financial literacy programming also promotes positive attitudes and behaviours regarding finances, such as encouraging savings and emphasising the importance of record-keeping. Sexual and reproductive health programming encompasses primarily knowledge of menstrual hygiene in this context.

<sup>2</sup> Each intervention pathway targets a different age group of girls and lasts for different time periods: FE interventions targeted the youngest girls (10-13 years at baseline), lasting four years; ABE programmes targeted girls aged 13-16 years and last two years; NFE or life skills interventions targeted the oldest girls (17-20 at baseline) and were comparatively short, with 11 months of planned programming. As we note elsewhere in this summary, the first cohort (C1) of NFE girls tracked since baseline completed the NFE programme in late 2020.

baseline.<sup>3</sup> For C1 NFE girls, a decline of nearly 5 points was observed, and while ABE girls' numeracy scores increased from baseline to midline, the increase was still 8.5 points below the benchmark.<sup>4</sup> In contrast, there was an observed increase in literacy scores from baseline to midline for FE girls, ABE girls, and C1 NFE girls. The increase for FE girls occurred at a rate faster than expected compared to the benchmark, whereas ABE girls are still 3 points below the benchmark. For C1 NFE girls, there was no benchmark for comparison, but the observed increase was very slight.

Analysis into why ABE and – especially – C1 NFE girls' gains in learning were smaller relative to those of FE girls suggests that the shorter learning courses the former girls were exposed to played a role. C1 NFE girls exited their courses in late 2020, which likely resulted in learning loss between the end of the course and the midline evaluation.

## Baseline

C4 NFE girls, newly enrolled just before this evaluation, naturally scored very low on both numeracy and literacy assessments, with scores of just 27.8 in numeracy and 16.9 in Somali literacy. Using comparable assessments, C4 NFE girls scored significantly lower in both subjects than did C1 NFE girls at their own baseline, though this may be due to the fact that C1 NFE girls had been exposed to multiple months of learning interventions prior to their baseline. C4 NFE girls are also older, at the start of their programme, than were C1 NFE girls.

By region, girls in Hirshabelle scored the lowest in all assessments and girls in South West State scored the highest. Age, state, and participation in household economic activities appears to be correlated, and older girls from Hirshabelle more often reported contributing to household tasks (e.g. agriculture, running a small business) that might detract from their learning.

## Transition

### Midline

Owing to the multiple forms of formal and non-formal educational programming implemented as part of AGES, the programme has a fairly complicated set of possible outcomes for transition. Continuation in formal school, transition from ABE courses into formal school, and transition from a shorter NFE course into the workforce or vocational training are the predominant pathways the programme expects to achieve. Thus far, this has been true: thus far, 84.2 percent of girls enrolled in formal school at baseline remain enrolled, and the majority have progressed at least two grade levels in that time. ABE girls have begun shifting into formal school in moderate numbers, while many others (35.2 percent) remain in their programmes. NFE girls, meanwhile, are entering the labour market and self-employment, though the most common outcome for NFE girls is idleness and unemployment following the completion of their course.

<sup>3</sup> For girls enrolled in formal school, the benchmarks for girls advancing from grade 1 to grade 2 are a 15.1 point gain in numeracy and a 23.1 point gain in Somali literacy; for girls advancing from grade 2 to 3, the equivalent benchmarks are 10.0 and 14.7 points.

<sup>4</sup> ABE girls are compared to the benchmarked differences between grade 1 and grade 2 FE girls, as ABE programming is intended to bring girls up to a basic level of numeracy and literacy. The relevant benchmarks are a 15.1 point increase in numeracy and a 23.1 point increase in literacy.

Overall, the transition rate at midline was 70.8 percent among FE girls, 46.8 percent among ABE girls, and just 25.6 percent among NFE girls. Among FE girls, the most common negative transition outcome is to drop out (14.9 percent), followed by being held back a grade (13.4 percent). Among ABE girls, outcomes are more mixed – while 38.6 percent are no longer enrolled in their ABE programme, a significant share remain enrolled (35.2 percent) or have transitioned into formal school (23.4 percent).

## Teaching Quality

### Midline

The evaluation examined four components of teaching quality and practices: overall professionalism, gender equity, the use of physical punishment, and pedagogical practices. For teacher professionalism, we find that the vast majority of girls – over 95 percent, on average – stated that their teachers make them feel generally or broadly welcome in the classroom at both baseline and midline.<sup>5</sup> At the same time, there were large shifts in teacher absenteeism, depending on the region, falling sharply and significantly in Banadir and SWS, but increasing significantly – to more than 50 percent – in Jubaland. Qualitative data suggests that teacher absenteeism may be driven in part by insecurity (including violence and economic insecurity) and low salaries. At midline, there has also been a decrease in some aspects of in-class gender equity. For instance, just 50 percent of teachers were observed to direct questions to girls and boys equally, a significant decrease from baseline. However, teachers were significantly more likely to be observed providing girls and boys with equal amounts of time to answer questions, and provided girls and boys with similar levels of positive (and negative) feedback.

The most compelling positive trend in teaching quality is the large decline in the use of corporal punishment and discipline for wrong answers. Similarly, we find significant increases in the use of several positive teaching practices, such as formative assessments, group work, and teaching lessons at the right speed. However, we note that for both negative and positive teaching practices, findings are by no means universal. Corporal punishment and, even more so, some teachers still discipline students who provide incorrect answers, and many positive teaching practices were observed to be used by only around half of teachers.

### Baseline

For C4 NFE girls, we find high average rates of girls reporting that the teacher makes them feel welcome, but significantly lower rates in Hirshabelle than in Banadir or SWS. Teacher absenteeism is, as in the other learning centres and schools, a problem: around 18 percent of girls agreed that teachers are often absent. Overall, although rates of reported teacher absenteeism are relatively low, these findings still suggest a need to strengthen teacher attendance, as absenteeism can have a severe impact on students' learning outcomes. Compared to the NFE centres evaluated at baseline, the use of corporal punishment is less common in C4 NFE centres – under ten percent of students report that their teachers use corporal punishment, with similar results obtained from classroom observations. Reported use of negative

<sup>5</sup> A “welcoming teaching” is, admittedly, a broad concept, and girls will inevitably have differing ideas of what constitutes “feeling welcome.” While this fact prevents us from saying how teachers behave more specifically, it provides a useful overall measure of how girls *feel* about their teachers and being in class, aggregating across many aspects of teacher behaviour. Additional questions unpack whether the teacher engages in gender-based favouritism, conducts the lesson at an appropriate speed, encourages participation, and uses corporal punishment, among other more specific behaviours.

disciplinary practices was highest in Hirshabelle, where teachers are also more likely to discipline students for providing incorrect answers.

Gender equity within classrooms is not relevant in NFE programmes, as they include only girls. Direct observation of classrooms did reveal that the most common participatory methods teachers use focus on encouraging participation among students who were not participating in class and using open-ended questions. In contrast, the use of group work and student-centred activities or games was very infrequent. Based on feedback from girls, we find that teachers often explain the use of lessons for their lives and provide students with ideas to study at home. In contrast, girls were much less likely to state that lessons move at a speed that is “just right”.

## Leadership, Life Skills, and Self-Efficacy

### Midline

Leadership skills, self-confidence, and life skills were primarily measured through the Youth Leadership Index (YLI). Scores on this index rose significantly from 49.8 at baseline to 53.5 at midline, and the proportion of respondents meeting the target of 70 on the YLI also increased significantly, from 12.6% at baseline for the panel sample to 17.4% at the midline, though this is still far below the endline target of 80%. Improvements in YLI scores were shared fairly evenly across C1 NFE, ABE, and FE girls, while impact varied geographically – the largest increase in scores occurred in Banadir (from 51.4 to 61.0), with a smaller (1.7 point) increase in Jubaland, and a decrease in SWS.

Leadership and self-esteem are especially critical, because AGES targets marginalised girls, many of whom face discrimination in their communities. Our analysis suggests that particular subgroups of girls experienced differential changes in YLI scores over time; of particular note is the fact that linguistic minorities – those who speak the af-Maay dialect – experienced a 5.8-point decline in scores since baseline, while those who speak the dominant dialect experienced an improvement of 7.2 points. There is some evidence that girls in households that were poorer at baseline have seen greater gains, thus far, and clear evidence that girls in worse learning environments at baseline have also seen larger gains, suggesting the programme has had success in improving the leadership skills and self-confidence of the most marginalized girls.

This evaluation round also saw the incorporation of a new measure of self-efficacy for NFE girls, based on the Chinese Positive Youth Development Scale (CPYDS), which captures girls’ views on whether they have control over their lives, can solve their own problems, and can make positive changes in their lives. A significant minority of girls (39.9 percent) report that they feel helpless when facing life’s difficulties, and 55.6 percent feel the course of their lives is mostly out of their own hands. This scale will be tracked for both C1 and C4 NFE girls going forward.

### Baseline

This round of the assessment set baseline youth leadership index (YLI) and life skills index (LSI) scores for C4 NFE girls. Across all regions, girls performed relatively well on the LSI, with most scores falling between 65-80 out of 100. Girls from SWS had the highest LSI scores, followed by girls from Banadir and then Hirshabelle. YLI scores followed a similar pattern at the broadest level, with the lowest scores emerging from Hirshabelle. A more nuanced accounting on a state-by-state basis revealed that – in comparison to Banadir and SWS – scores in Hirshabelle were bimodal, with one very high-scoring group and one very low-scoring group. In general, the spread of YLI scores was very high in both Banadir and Hirshabelle. LSI scores were – as expected – associated with a girl’s relative marginalization: for instance, girls with disabilities scored worse, as did linguistic minorities. This was also true for YLI scores: scores were significantly lower for IDPs, girls with disabilities, af-Maay speakers, and girls from poorer households.

## School Management and Governance

### Midline

To improve school governance, AGES targets the activities of Community Education Committees (CECs) and Ministry of Education representatives that work with programme schools. AGES tracks the changes in indicators that capture the progress CECs and MOE have made to promote inclusivity and improve quality assurance procedures. The sole school without a CEC at baseline has since established one, and the proportion of schools reporting engagement of the CEC increased across all measured activities from baseline to midline, with CECs increasingly active in the hiring of teachers, fundraising, and following up on drop-outs. Additionally, the proportion of schools visited by their CEC in the past year rose by 14 percentage points from baseline to midline, and the types of topics/activities monitored by CECs has also expanded over time.

Further, qualitative data suggests that CECs have employed a variety of strategies – e.g., fundraising and awareness campaigns – in an attempt to offset the negative effects of school fees, which have increased since baseline, on enrolment. CECs perceive their awareness campaigns as having significantly reduced discrimination against girls from minority groups, but the analysis also found that CEC members tended to downplay the impact of discrimination on educational opportunities compared to material and accessibility barriers and were reluctant to acknowledge the unique barriers faced by, for instance, girls from minority groups. Additionally, CECs report feeling unable to solve accessibility issues that prevent girls with disabilities or girls who live far away from attending school.

MoE departments are expected to support schools by developing a national curriculum, helping to ensure security at schools and surrounding communities, and conducting school visits to monitor teaching practices, check on administrative records, and oversee the inclusion of marginalised students, including by engaging with traditionally underserved sections of communities, such as IDPs in the area, to ensure they are included. At baseline, the engagement of state MoE and MoECHE officials was found to be limited in most areas. At midline, MoE officials reported MoE support to schools had increased significantly since baseline. In particular, officials report increased efforts to train teachers, CEC members, District Education Officers (DEOs) and arrange DEO visits to schools. There were also reports of mobilization campaigns, hiring of teachers from minority groups, and the opening of schools for students with disabilities.

### Baseline

Much of our analysis of school governance for the C4 NFE centres focuses on establishing the extent of barriers to enrolment and retention related to discrimination, as the goal of the programme's work with CECs in these areas will be to promote inclusivity. Unsurprisingly, cost was cited as a major reason why girls were not enrolled in school by 52 percent of caregivers in this cohort, with higher rates among disadvantaged groups, such as IDPs. Minority girls were subject to somewhat greater levels of mistreatment in the form of discriminatory attitudes, and caregivers of girls with disabilities were particularly likely to cite mistreatment as a reason for their daughter not having been enrolled in the past.

In general, girls took a fairly apathetic view of CECs and their tangible support for girls' education, though this may stem from the fact that the girls in question are newly enrolled in a learning programme. Just 46 percent of C4 NFE girls felt CECs take any action to support girls' education, while over one-third reported that they did not know whether CECs supported girls' education. This likely indicates an absence of CEC activities or – at best – a lack of visibility of CEC efforts. Particularly disadvantaged girls – those most likely to benefit from financial or “moral” support from CECs – were significantly more likely to say that the CECs support girls' education. According to girls, CECs mostly support enrolment of out-of-school girls (86 percent), followed by supporting dropouts to return to school (59 percent). Head teachers supported this notion, but also believed CECs were involved in a more expansive set of activities, including monitoring teacher absenteeism.

## Access to Protective Services

### Midline

Overall, access to protective services was high both within schools and at the community level. The vast majority of girls (93.9 percent) reported that they could report abuse, harassment, or other issues in school to *someone*, most commonly a teacher or the head teacher. It is important to distinguish between the ability to report abuse and a specific mechanism for doing so, but most girls expressed confidence they could report abuse. According to qualitative accounts, girls typically report incidents to teachers or head teachers and in some rare cases to CECs, to their school's Girls' Empowerment Forum (GEF), or to the police. A high proportion of girls also feel they could report abuse or harassment occurring in their community but outside their school – with many citing their teacher or head teacher, but also relying on parents, police, and clan elders or community leaders. The ability to engage protective services appears to be lowest in Banadir and for girls whose head of household or caregiver has no education of any kind.

### Baseline

Access to protective services for C4 NFE girls mirrors those outlined above for the C1 NFE girls. An overwhelming majority (93.0 percent) of girls feel they could report abuse or harassment occurring at their school, and a smaller majority (81.8 percent) believe they could report harassment, abuse, or exploitation taking place within their community.

## Strengthened Economic Circumstances

### Midline

AGES programming for NFE girls is largely concerned with increasing their long-term economic opportunities. This evaluation round included new data collection on economic empowerment, including average monthly income and types of income generation activities. Self-reported monthly income among C1 NFE girls averaged 368,128 SOS (~\$14 USD at the time of data collection). Girls in Jubaland appear to be the highest earners, while those in SWS and Banadir are reportedly the lowest earners, on average. Other factors, including lack of food and water, being divorced or orphaned, not living with one's parents, having children, and having a disability, are all negatively correlated with income. The low mean income figures reflect the fact that most girls are unemployed – 59.2 percent do not report an occupation, while a further 23.6 percent are occupied by completing domestic chores within the home. Almost none of the girls in the sample reported having their own business (8.7 percent). Barriers mentioned in the qualitative data to starting a business include lacking financial support or parental guidance, lacking knowledge, or already working for a parent's business.

### Baseline

On average, C4 NFE girls reported that they made about 509,741 SOS a month (\$19.60 USD at the time of data collection), with 51.4 percent of girls reporting no income at all. This average income is higher than the amount observed for C1 NFE girls; this cohort also differs from C1 NFE girls in the geographic distribution of earning – C4 NFE girls in Banadir appear to be the highest earners, followed by girls from SWS. Girls from Hirshabelle were the lowest earners, on average. In line with the C1 NFE cohort, most girls reported not having an occupation (52.4 percent) or doing domestic chores inside the home (17.3 percent), and almost none of the girls reported having their own business (8.2 percent).

## Enhanced social support for female youth

## Baseline

As C4 NFE girls have only just joined the program, the assessment established their baseline levels of engaging with various development opportunities that will be tracked in future evaluations. A total of 21.4 percent of girls reported they have received some form of humanitarian assistance during the past year, 14.8 percent reported participating in youth groups, 14.0 percent reported participating in service delivery activities, and 8.3 percent reported participating in local governance activities. There was substantial covariation between the first three outcomes, meaning that many of the same girls engaged in two or more of the activities.

Girls from IDP households were the least likely group to be involved in youth networks and political fora. Girls with disabilities and from minority groups had higher average rates of participation in activities or consultations related to local governance. However, girls with disabilities and minority girls were no more likely to receive humanitarian assistance, unlike IDP girls, who were significantly more likely to benefit from it. Overall, girls in Hirshabelle were much more involved in public life, whereas Banadir lagged behind in participation in all three types of opportunities. Lastly, more traditional caregiver attitudes tended to be associated with lower participation in the development opportunities for girls, and there was a positive relationship observed between certain indicators associated with girls' leadership skills and an increased likelihood of accessing development opportunities.



# 1. Introduction

Somalia is one of the most challenging possible environments for girls' education. Widespread poverty has combined with a long-running military conflict, a series of frequent and ongoing environmental shocks, and emerging public institutions, severely affecting access to education and learning outcomes in general, and for girls especially. The environment in which AGES is being implemented faces all of these challenges in their most extreme forms: conflict; attacks against schools and education specifically; droughts that cause displacement, loss of livelihoods and economic fluctuations; floods that compound the effect of droughts on food prices and food security; and an education system composed primarily of community and privately owned schools.

Beginning in the late 1980s with the rise of a resistance movement to the Siad Barre regime, which eventually overthrew the regime in 1991, Somalia fell into a long period of conflict and statelessness. The conflict that has defined the period from 1991 to the present has often been intensely localised and fluctuating, with internecine fighting the norm, punctuated by periods of intense and outright civil war. In general, fighting has been most intense in south-central Somalia, where AGES is being implemented, with significant and often repeated rounds of fighting occurring in AGES areas like Mogadishu, Kismayo, and Baidoa, among others.

The second overarching contextual factor is environmental: ongoing drought and intermittent flooding, which have devastated riverine agricultural areas. According to recent reports, the drought that began in 2020 is exacerbating food insecurity in the Southern and Central parts of the country, including AGES locations such as Gedo, Bakool, and Bay regions, deteriorating the nutritional situation and driving a surge in population displacement from rural areas to IDP camps near major urban centres.<sup>6</sup> May 2021 saw the latest iteration of heavy flooding afflicting populations settled on the banks of the Juba and Shabelle rivers, with especially heavy hit areas including in Jowhar, Hiran, Doolow, and several other project districts<sup>7</sup>. Current levels of humanitarian assistance are outpaced by the rapidly increasing needs of affected populations in and outside the IDP camps. As it stands, cereal production is expected to be lower than average this season, while the trajectory of the drought and rainfall in the future has the potential to significantly influence project outcomes.

Environmental shocks have several effects on schools and education. Drought and flooding can have direct impacts on schools and the families of students, damaging schools, reducing the income necessary to pay for schooling among families, or prompting their migration. These exogenous shocks can also raise food prices, weakening families' purchasing power. Given that the vast majority of schools in the evaluation sample are urban, the more likely scenario is that schools will be influenced by an influx of economic or environmental internally displaced people from affected areas. This has been a common occurrence in urban schools elsewhere in Somalia, where drought prompts movement of pastoralists from the countryside into cities, increasing enrolment rates dramatically, while burdening already overcrowded schools and informal social safety nets.

It is also important to note the demographic composition of the areas relevant to this study. While Somalia is often portrayed as ethnically and culturally homogeneous relative to other countries in the region,

<sup>6</sup> Food Security and Nutrition Analysis Unit, Somalia. 2022. "Somalia Food Security Outlook - February to September 2022" Available at: <https://www.fsna.org/downloads/Somalia-Updated-IPC-and-Famine-Risk-Analysis-Technical-Release-8-Apr-2022.pdf>

<sup>7</sup> <https://reliefweb.int/disaster/fl-2021-000051-som>



linguistic and cultural differences do exist and can be particularly relevant for life outcomes. The regions in which the project is being implemented include large swathes of territory – especially in South West State – associated with the traditionally agricultural clan family, the Digil-Mirifle. Less numerous in the sample – but still comprising approximately 17 percent of households – are less locally numerous marginalised groups, such as the Bantu, an all-encompassing term for agricultural peoples of broad Bantu lineage who occupy farmland in riverine areas.<sup>8</sup>

In Somalia, two primary dialects of the Somali language are spoken: af-Mahatiri is, generally, the language associated with pastoralist clans, while af-Maay is associated with sedentary agricultural groups. The former is the codified version of Somali and the language in which instructional materials and textbooks, as part of the newly established federal curriculum, are written. Many of the girls in AGES schools speak af-Maay, which places them at potential linguistic disadvantage vis-à-vis students who speak the official language of instruction, af-Mahatiri. Speakers of af-Maay face additional hurdles to learning, insofar as learning in a different dialect than one speaks at home is more difficult.

On the other hand, most of the schools targeted by AGES are situated in the oldest urban areas of Somalia, with traditions of governance, educational attainment, and trading cultures in the coastal cities of Mogadishu and Kismayo. Prior to the outbreak of civil war, many of the cities in question were centres of significant power, which may – speculatively – be reflected in more positive attitudes toward education or greater access to financial resources for education. Unfortunately, little is known about cross-regional levels of income, and even less about attitudinal outcomes such as support for education; nonetheless, the proliferation of private schools can be used as a proxy to estimate strong support for education among urban populations, despite potential biases towards boys and majority groups.

As the discussion above makes clear, education in Jubaland, South West State, Banadir, and Hirshabelle face a number of structural impediments that will make sustained improvements in educational outcomes difficult. Unfortunately, this problem is exacerbated, broadly, by the fractured institutional environment in which education in this region operates. Following the collapse of Somalia's central government in 1991, many private educational institutions were established, and formed into "umbrella" associations of schools, which were also privately run. As of 2016, 14 umbrella organisations provide education to a quarter-million students through over 1,000 affiliated schools. Despite the re-establishment of central governance, umbrella associations are prominent actors in the education sector, with only 7.4 percent of students in South Central Somalia attended government-run primary schools in 2017<sup>9</sup>.

In 2018, the FGS introduced an official curriculum for primary education nationwide, which is the first revision of Somalia's curriculum since 1996. The new curriculum established Somali as the primary language of instruction and specified official instructional materials in the core seven subjects that were being distributed to schools at the initial stages of the curriculum rollout<sup>10</sup>.

The impact of COVID-19 on education and the economy exacerbated the challenges faced by the project. The economic effects of the pandemic included the disruption of international trade, shortage of key consumer goods, price fluctuations, and declining remittances. These factors have increased strain on

<sup>8</sup> South West State is the core traditional homeland of the Digil-Mirifle, although our sample includes a large number of respondents from this clan in Mogadishu as well. Minority group respondents tend to be clustered in Jubaland and, to a lesser extent, Mogadishu.

<sup>9</sup> <https://education-profiles.org/sub-saharan-africa/somalia/~non-state-actors-in-education>

<sup>10</sup> <https://reliefweb.int/report/somalia/new-way-educate-somali-schools-starts-august-enso>

families, making it harder to justify education costs. School closures mandated across the country from March to August 2020 were compensated by a shift towards online education that nonetheless excluded a substantial number of students who could not access the technology required for digital learning<sup>11</sup>. During previous major crises that affected children and their ability to access education, such as the 2011 famine, approximately 15% of students did not return to school, leading experts to project substantial losses in enrolment even after the end of the pandemic<sup>12</sup>.

The brief overview of south-central Somalia provided here highlights several contextual factors key to project implementation and success. First, the region has been wracked by three full decades of conflict, a conflict which is far from over. Second, both dryland pastoralist and riverine agricultural economic zones in the target regions face serious environmental threats, in the form of drought and flooding. Third, the project's geographic range includes the capital city of Somalia as well as regional centres of trading and commerce that are populated partially by ethnic minorities and other marginalised groups, who may face a different set of barriers to educational attainment than those from Somalia's largest clan families. And finally, COVID-19 and the associated school closures created major challenges for the project that had to adapt to remote learning and growing economic pressures experienced by many households in the target regions. In this context, CARE and its implementing partners will face a number of challenges to both implementing the project and generating positive improvements in learning, transition, and other project outcomes.

---

<sup>11</sup> <https://reliefweb.int/report/somalia/child-protection-and-covid-19-somalia-case-study>

<sup>12</sup> <https://www.savethechildren.net/blog/covid-19-school-closures-put-decades-gains-somali-children-risk>

## 2. Methodology

### 2.1 Overall Evaluation Design

The AGES evaluation employs a pre-post, longitudinal research design, tracking girls of varied cohorts – formal school (FE), accelerated basic education (ABE), and non-formal education (NFE) – across evaluation rounds to analyse changes in learning, transition, and intermediate outcomes over time. It is important to note that the present evaluation round consists of two mostly distinct evaluations, which cover a set of educational centres and girls supported by FCDO beginning in late 2019, and a second set of centres and girls supported by USAID beginning in early 2022. This report serves as a midline evaluation for the former group and a baseline for the latter group. In this section, we provide a brief overview of the evaluation design – especially sampling and tracking across rounds – with only the necessary details to understand the inferences drawn. We refer readers to the AGES baseline evaluation report and to Annex 1 of this report for further details.

As noted above, the AGES evaluation utilises a pre-post design without an in-built comparison group. Sampling occurred within centres, meaning that the baseline consisted exclusively of girls who were enrolled at the centre in question. Household surveys were completed with the girls' caregiver and head of household, such that measures of community attitudes are representative of caregivers and heads of household of girls enrolled in AGES centres, rather than the overall adult population in the areas studied.

Due to the lack of a comparison group, the results of the evaluation are sensitive to both maturation effects (improvements that accrue naturally with age) and exogenous shocks that impact programme outcomes both among the programme sample and in the wider population. To guard against maturation effects, learning outcomes were benchmarked at baseline, with grade-level differences established based on performance of different grade levels in 2019. For instance, girls in Grade 1 at baseline – who should have advanced to Grade 3 by the time of the midline – will have their respective improvements in learning over that period compared to the difference – at baseline – between Grade 1 and Grade 3 girls. The latter difference represents the maturation effect from Grade 1 to Grade 3 that should be expected in the absence of intervention.

Unfortunately, not all outcomes can be benchmarked in this manner, and the evaluation is sensitive to the impact of exogenous events and secular trends in outcomes. Throughout the evaluation, we discuss changes from baseline to midline and our relative confidence in attributing these changes to the impact of the programme itself.

As a baseline evaluation of the second (USAID-supported) cohort of NFE centres and girls, the evaluation mimics most features of the evaluation that started in 2019. This evaluation employs an identical pre-post design, with sampling of girls occurring within the population of girls enrolled in the targeted NFE centres. The set of girls recruited in the current (baseline) round will be tracked in future rounds to provide a panel of respondents for longitudinal analysis.

### 2.2 Data Collection Tools

The AGES evaluation is mixed-methods, employing a variety of qualitative and quantitative tools that target several different groups of respondents. Quantitative tools consist of:

- Learning assessments testing numeracy and Somali literacy among cohort girls
- Household surveys with the caregivers and heads of household of most cohort girls
- School survey with head teachers of formal (FE) schools

- Attendance headcounts of ABE, NFE, and FE centres
- Classroom observations, consisting of direct observation of teaching practices and teacher behaviour

During the inception period, revisions were made to the tools to ensure they captured all necessary information. Specifically, new questions capture information regarding the impact of COVID-19 on girls and their households, and the participation of girls in Girls' Empowerment Forums, among others.

Three broader changes were made to the quantitative tools that are noteworthy. First, the sample of classroom observations was expanded to include ABE and NFE centres (both NFE centres from the original cohort and newer NFE centres supported by USAID). At baseline, the sample target was to complete two classroom observations in each formal school; at midline, the equivalent target was two observations in each school or centre visited. Note that many ABE and NFE centres are situated in the same facility as a formal school; when we refer to "centres", we are distinguishing between centres that are co-located, such that there may be multiple centres located within a single facility. Second, the survey with teachers, undertaken at the baseline, was not included in the midline evaluation.

Third, changes were made to targeting of household members within the household survey. At baseline, every cohort girl completed a survey and learning assessments, and members of her household – her caregiver and head of household – also completed questions related to the girl, her school, and details of their household. At midline, the household survey was completed only for girls who were under 18 years of age; girls 18 and over completed a shorter set of questions – capturing some of the same information – themselves. This approach was taken because a significant share of cohort girls are over 18, married, and are, in many ways, independent. However, this decision also has implications for our analysis, because the shortened set of questions applied to girls themselves does not capture all of the household characteristics normally included in the household survey. As a result, when we conduct subgroup analysis of programme outcomes or programme impacts, the sample size available for analysis is, at times, smaller.<sup>13</sup> The table below documents the number of girls in each cohort for whom a household survey was completed and, alternatively, who completed the new module targeted at girls directly.

**TABLE 1: SHARE OF MIDLINE RESPONDENTS COMPLETING FULL HOUSEHOLD SURVEY**

Cohort	Completed Full Household Survey	Completed Abbreviated Module with Girl Only
<b>FE Girls</b>	410 (98.3%)	7 (1.7%)
<b>ABE Girls</b>	214 (58.8%)	150 (41.2%)

<sup>13</sup> Wherever possible, we mitigated this effect by utilising data from the baseline to determine whether a girl fit into a subgroup of interest, though this was primarily possible with relatively fixed characteristics, i.e. those that do not change or change only slowly over time. For instance, if a girl was 18 years or older and, therefore, completed only the shorter household module, we did not capture data at midline on whether her head of household or caregiver had completed any formal education. However, as this characteristic of her parent or caregiver should not change over time, we attributed her status on this outcome from baseline to her midline data. Again, this is only possible in the context of relatively fixed outcomes; for outcomes that can change – such as whether the girl is married or has given birth – we cannot utilise information collected two years prior, and we are forced to exclude girls 18 years or older from subgroup analyses.

<b>Old (Cohort 1) NFE Girls</b>	12 (3.2%)	369 (96.9%)
<b>New (Cohort 4) NFE Girls</b>	280 (30.6%)	636 (69.4%)

**Note from the project:** Out-of-school girls were enrolled in different learning tracks (formal school, ABE, or NFE) based on their age range. Girls enrolled in formal school were reportedly 10-12 years old at the time of enrolment in 2019, thus making it impossible for them to be 18 or above in 2022. In practice, however, the lack of birth certificates and the fact that many girls do not know their exact age contribute to potential errors, despite the project's practice of requesting additional information to establish age ranges, such as events that took place in the year of the girl's birth. As a result, there are some cases where older girls have been enrolled in formal education or ABE, as the seven girls enrolled in formal education who self-declared to be 18 or older in the current evaluation.

The table below reports the sample targets for each of the quantitative tools, aside from the learning assessments and household surveys, the sample for which is discussed in more detail in a subsequent section of this report. The table indicates the number of observations completed at baseline for each tool, the target per centre or school, the overall target across all centres, and the achieved sample. While documenting the extent of data collection in this round, the expansion of data collection primarily reflects the expansion of the sample in this round to new, USAID-supported, NFE centres. This expansion accounts for the larger number of classroom observations and attendance headcounts completed relative to the baseline.

**TABLE 2: SAMPLE TARGETS AND ACHIEVED SAMPLE FOR SCHOOL-BASED DATA COLLECTION TOOLS**

<b>Data Collection Tool</b>	<b>Total Baseline Sample</b>	<b>Target per Centre</b>	<b>Overall Sample Target</b>	<b>Achieved Sample</b>
<b>Head Teacher Survey</b>	38	1 per formal school	38	37
<b>Classroom Observations</b>	70	2 per centre	302	219
<b>Attendance Headcounts</b>	373	N/A (all classes or streams in the centre)	N/A	463

At the same time, the evaluation failed to meet its targets regarding classroom observations – and completed fewer attendance headcounts than the evaluation team expected – due to the fact that many of the original NFE centres had completed their programming with the original cohort of girls.<sup>14</sup> This was not unexpected, as the length of both NFE and ABE curricula indicated that many beneficiaries who enrolled in 2019 would now have completed the programme. As a result, classroom observations and

<sup>14</sup> To illustrate, consider that the sample included 151 distinct centres at midline, though this includes significant overlap between new and old NFE centres. A simple accounting indicates that 3.1 headcounts were completed in each centre. However, this is misleading, as we conducted an average of 6.9 headcounts in each of the 38 formal schools visited, but many fewer in the typical ABE and NFE centre, due to their smaller number of distinct classes or streams and the aforementioned closure or winding down of centre activities.

headcounts were not completed in some ABE and NFE centres; where the new (Cohort 4) and old (Cohort 1) NFE centres overlapped, we completed observations in classrooms associated with the new cohort of USAID-supported students.

In addition to quantitative tools, the present evaluation includes focus group discussions (FGDs) and key informant interviews (KIIs) with six distinct sets of respondents:

- FGDs with Community Education Committee (CEC) members
- FGDs with mothers of girls enrolled in AGES centres or schools
- FGDs with teachers
- KIIs with religious leaders
- Participatory FGDs, employing risk mapping, with girls
- Participatory FGDs, employing vignette story-telling exercises, with girls

The goal of these qualitative interviews was to provide nuance to the quantitative findings, to explain patterns in the quantitative data, and to provide a more complete picture the circumstances and support for education in programme communities. To supplement the data collected, especially with regard to Ministry of Education (MOE) actions to promote inclusive education, the evaluation team completed short, remote interviews with MOE officials in several districts.

The target established during the inception period was to complete 12 qualitative interviews in each of the categories above, with the exception of religious leaders, for whom the target was to complete 8 KIIs. During the inception phase, CARE's monitoring and evaluation team and the evaluation team agreed to drop FGDs with fathers from the fieldwork plan. In total, we sought to complete 68 qualitative interviews. This target was met for the risk mapping exercises (12), vignette FGDs (13), teacher FGDs (12), and KIIs with religious leaders (8). The evaluation fell short of the target for mothers (10) and CEC members (11). The final qualitative data includes 67 interviews, as well as 9 short remote interviews with MOE officials.

## 2.3 New NFE (Cohort 4) Sampling Design

As described above, this evaluation round serves as a baseline for a second batch of NFE centres and girls, who are part of an expansion of AGES programming supported by USAID. The new cohort of NFE girls are enrolled at both existing NFE centres that were part of the earlier intervention and new NFE centres. In this section, we outline the sampling strategy for this new cohort, which will be tracked into future evaluation rounds.

The sample of New NFE girls was a multi-stage, stratified sample, clustered at the level of individual centres. Stratification took place at the state level, proportional to the population of beneficiaries (NFE girls) in each state. A sample frame of girls was provided by CARE following an enrolment and listing exercise undertaken in each NFE centre participating in the new cohort. The sample frame consisted of 11,426 girls, spread across 82 distinct centres in three states or regions – Banadir, Hirshabelle, and South West State. Importantly, the sample frame was still being constructed during the inception phase and was incomplete at the time of our initial sample draw. At that time, our best estimate indicated that Banadir would comprise 47.5 percent of the beneficiary population, while Hirshabelle and South West State would comprise 15.8 and 36.7 percent, respectively. These population proportions were used for stratification at the state level. As shown in the table below, this resulted in assignment of 22 clusters (out of 46) to Banadir, 7 to Hirshabelle, and 17 to South West State.

**TABLE 3: POPULATION AND SAMPLE ALLOCATION, BY STATE, FOR NEW NFE COHORT**

State	Population Proportion, Early Sample Frame	Population Proportion, Final Sample Frame	Clusters Assigned (Total Interviews)	Achieved Sample Size	Proportion of Total Sample
Banadir	47.5%	47.4%	22 (440)	440	48.0%
Hirshabelle	15.8%	13.5%	7 (140)	139	15.2%
South West	36.7%	39.1%	17 (340)	337	36.8%
<b>Total</b>	100%	100%	46 (920)	916	100%

The table also shows that the composition of the sample frame changed slightly between the time of our sample draw and the finalisation of the sample frame. The result is that our sample slightly overrepresents Hirshabelle and underrepresents South West State, because the sample draw was completed on the basis of the earlier, incomplete sample frame.

Within the state-level strata, centres were selected with probability proportional to their population (number of girls) size, with replacement. The result is a self-weighting sample at the level of centres; combined with proportional stratification at the state level, weights are generally not required for analysis of the new NFE sample.<sup>15</sup> With schools as the Primary Sampling Unit (PSU), the second stage of selection occurred within schools. CARE's sample frame included a complete listing of girls enrolled in each centre; thus, within centres, we selected 20 girls using simple random sampling. At the same time, we selected 20 replacement girls with simple random sampling and randomly ordered them, ensuring that any replacements that occurred in the field were selected randomly. In each NFE centre, we planned to complete 20 interviews, for a total sample size of 46 clusters and 920 interviews, exceeding the required target, based on CARE's power calculations, of 906 interviews. The final, achieved sample consists of 916 interviews. The next section describes the demographic characteristics and representativeness of the sample.

## 2.4 Midline Re-Contact Rates

The cohorts of girls recruited during the 2019 AGES baseline were re-contacted during this evaluation, producing a panel sample of girls who were successfully contacted in both rounds. In this section, we document our approach to re-contacting girls, including fieldwork procedures to maximise re-contact rates, and analyse patterns in successful re-contact.

### Re-Contact Procedures

Our approach to re-contacting girls is multifaceted and takes advantage of our enumerators' local knowledge, the assistance of teachers and other community members, and the data collected from girls and their families at baseline. Re-contact procedures were specified precisely during the training and consist of the following:

- Ask the head teacher and other teachers in the centre whether the girl is present at the time of visiting.

<sup>15</sup> Weights are needed only to correct for the very minor deviation from precise proportionality at the state level that results from the clustered sample.



- Call every phone number on file for the girl and her family (from baseline) three times, separated by six hours between call attempts. Call attempts were required to be made on a minimum of two separate days.<sup>16</sup>
- Visit the girl's household a minimum of two times, using location information from the baseline data, from teachers, and from other students in the school who know the girl or her family.
- Ask the head teacher, teachers, other students, and community leaders for contact information for the girl or her family.

An important decision concerned the extent to which teams would travel to find girls who had moved – temporarily or permanently – away from the area. Our approach sought to balance logistical considerations while maximising successful re-contact rates: if the girl lived within approximately 30 minutes of the school – or a reasonable distance, based on the team leader's judgment – an enumerator would travel to her home to interview her. Team leaders also considered whether a girl who lived outside that range could be visited later in fieldwork, when the team would be closer to her home. Because AGES is an urban and peri-urban programme, teams moved around the same cities several days or weeks and could arrange to visit a girl at home when they were “in the area” later.

Finally, if a girl could not be contacted at all, refused to participate, lived outside the areas where fieldwork teams were operating, or could not be reached safely, team leaders made the final determination of whether to drop the girl from the sample. Replacement girls were selected in the case of FE girls, who – at baseline – had been enrolled in a formal school. For such girls, replacements were selected from within the same school; replacements were pre-selected by the evaluation team from the original sample frame provided by CARE in 2019, to guard against improper sampling procedures employed by team leaders. Pre-selection ensured that replacement girls were selected randomly.

ABE and NFE (Cohort 1) girls who could not be located were not replaced. ABE and NFE girls recruited at baseline have, by the time of the midline, typically completed their participation in the educational centre where they were recruited. As such, it would not be possible to select a comparable replacement girl from within the centre, as any girls enrolled in the centre would be in a different level or stream than the original girl.

Finally, in an attempt to maximise the sample available for analysis of transition outcomes, teams recorded as much information as possible about girls who fell out of the sample, whether they were replaced or not. Enumerators indicated where the girl now lived, whether she was enrolled in school, and her grade level, if these details could be ascertained. Panel attrition is typically a major threat to inferences around transition outcomes, because attrition is generally higher among dropouts and is – more broadly – correlated with transition outcomes. Girls who have married and moved away from the area are more likely to fall out of the sample and are, simultaneously, more likely to be out-of-school. By recording information regarding the enrolment status of girls who have fallen out of the sample, we are better able to analyse transition rates from baseline to midline. The information allows us to analyse “rough-coded” transition outcomes as a robustness check of our primary findings.

## Re-Contact Rates

---

<sup>16</sup> Our typical approach involved placing a first phone call in the morning or mid-day on the first day of fieldwork at a particular school. If unsuccessful, a second series of calls (to every phone number on file) was completed that evening. A third series of calls would be completed on the following day.



Before turning to a discussion of re-contact rates, it is important to note that school-level attrition is responsible for a small decline in the sample of FE, ABE, and NFE girls from baseline to midline.

The table below reports re-contact rates by state, in separate panels, broken down by cohort or respondent type; the bottom panel reports the cross-state aggregate. As the table shows, re-contact rates were highest among FE girls, and lowest among NFE girls. This trend is not surprising, because FE girls were the youngest cohort and many remain enrolled in the same schools as at baseline, making it easier to locate them, thanks to the knowledge of their teachers and classmates. Younger girls are also less likely to have migrated to pursue educational, work, or marital opportunities – a result based on prior studies of girls’ education programmes we have conducted in Somalia.<sup>17</sup> This helps explain why NFE girls were less likely to be re-contacted successfully than ABE girls, as the former are older, on average, and more likely to marry and move away from their previous area.

**TABLE 4: RE-CONTACT RATE AND ACHIEVED SAMPLE, BY COHORT AND STATE**

State	Total Possible Sample	Re-Contact rate	Achieved Panel Sample
<b>Banadir</b>			
FE Girls	192	78.6	151
ABE Girls	218	80.3	175
NFE Girls	269	75.1	202
<b>Total</b>	<b>679</b>	<b>77.8</b>	<b>528</b>
<b>Jubaland</b>			
FE Girls	132	90.2	119
ABE Girls	108	76.9	83
NFE Girls	107	77.6	83
<b>Total</b>	<b>347</b>	<b>82.1</b>	<b>285</b>
<b>South West State</b>			
FE Girls	81	87.7	71
ABE Girls	142	74.6	106
NFE Girls	131	73.3	96
<b>Total</b>	<b>354</b>	<b>77.1</b>	<b>273</b>
<b>Aggregate</b>			
FE Girls	405	84.2	341
ABE Girls	468	77.8	364
NFE Girls	507	75.1	381
<b>Total</b>	<b>1380</b>	<b>78.7</b>	<b>1086</b>

<sup>17</sup> In this sample, though, age and re-contact rates are not unambiguously correlated. In fact, in a model that controls for the girl’s cohort and region, age has a negative but very small – and statistically insignificant – relationship with successful re-contact.

There is no clear pattern in the re-contact rates reported by state – while Banadir had the lowest success rate for FE girls, it had the highest success rate for ABE girls, suggesting that respondent type and respondent age are more important factors than geographic region. Investigating further, there *are* important differences across regions *within* states: within Jubaland, successful re-contact was much more common in Gedo than in Lower Juba; indeed, Gedo had a higher re-contact rate (91.7 percent) than any other region (the next highest occurred in Bay, at 80.3 percent). However, the sample size in Gedo is relatively small and had a larger share of FE girls – whose re-contact rates tended to be higher – than any other region.

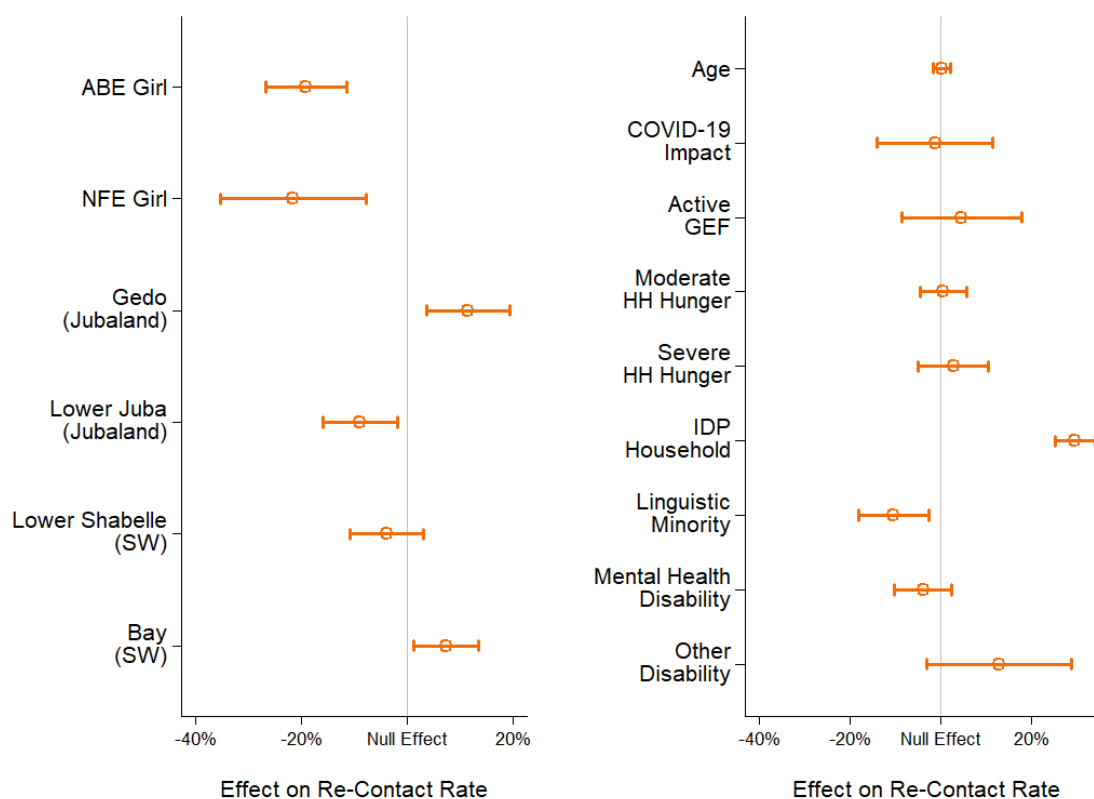
## Predictors of Successful Re-Contact

The results in the previous section do not point to any systematic relationship between successful re-contact and age, while cohort and region are more strongly predictive of successful re-contact. However, because many of the characteristics of interest – such as age and household economic status – are correlated with one another, it is not possible to disentangle their relationships with re-contact status in bivariate models, such as a difference-of-means or cross-tabulations. To understand the predictors of successful re-contact more accurately, we estimated a series of regression models predicting successful re-contact as a function of a girl’s baseline characteristics.

Our basic approach is to predict the binary outcome of re-contact as a function of cohort and region – both incorporated into the model as a series of dummy variables – and household or girl characteristics, such as household hunger at baseline, IDP status, enrolment in a school with an active GEF, and so forth. The figure below reports the regression coefficients (orange dots) and the 95 percent confidence interval (orange horizontal bars) around those coefficient estimates, for a series of predictors. This figure is drawn from a single regression, meaning that all of the variables reported were included in a single model. The regression coefficients can be interpreted as the impact of a variable, measured in percentage points, on successful re-contact.<sup>18</sup> For instance, attending a school with an active GEF is associated with a 4.7-point increase in the likelihood that a girl will be re-contacted successfully, holding all else equal.

<sup>18</sup> For cohort and region, the coefficients should be interpreted relative to the omitted category. For instance, ABE girls have a 19.2-point lower re-contact rate than FE girls (the omitted category), holding all else equal.

FIGURE 1: IMPACT OF GIRL'S CHARACTERISTICS ON RE-CONTACT OUTCOMES



The regression results reinforce at least two of the findings from less rigorous analysis in the previous section. First, ABE and NFE girls have substantially lower re-contact rates than FE girls – 19.2 and 21.7 points lower, respectively, after accounting for other factors that predict successful re-contact. Second, some of the regional patterns are reinforced in this analysis – clearly, re-contact rates are higher in Gedo and Bay and lower, on average, in Lower Juba.

At an individual and household level, there are fewer compelling findings. We measured COVID-19 impact at the school level using midline evaluation questions that asked girls or their caregivers how their family was affected by COVID-19; we constructed a measure from 0 to 1 that is the share of households at a given school who reported at least one COVID-19 impact.<sup>19</sup> Neither this, nor a similarly-constructed measure of whether a school had an active GEF, had a substantial impact on re-contact rates.

The only individual- or household-level factors with large effects on re-contact rates were IDP status and language use relative to the dominant local Somali dialect. Girls from IDP households 29.6 percentage points more likely to be re-contacted successfully, potentially because IDP households are less likely to move to a new location within the same area – as they tend to lack rights to land ownership. At the same time, linguistic minorities – which is almost perfectly correlated with speaking the af-Maay dialect of Somali – had much lower re-contact rates. We do not have convincing explanations for either the role of IDP or linguistic minority status, as the opposite effects actually seem to contradict one another – to the extent that IDP households are easier to re-contact due to their fixed location (and the importance of working with NGOs in IDP camps, prompting greater cooperation in such settings), this would

<sup>19</sup> Girls who were not successfully re-contacted lack data on direct COVID-19 impacts

simultaneously suggest that Maay-speaking households should have higher re-contact rates, at least insofar as Maay-speakers living outside of their homelands (e.g., in Banadir or Kismayo) tend to also be IDPs.

At an individual and household level, our analysis does not suggest any major concern regarding patterns of attrition that would produce bias or significantly alter the sample's composition. However, at a regional level, the sample – especially among the ABE and NFE cohorts, where girls who could not be re-contacted were not replaced – is becoming more heavily focused on Gedo and Bay and less representative of Lower Juba and Lower Shabelle. Fortunately, this impact is relatively small: at baseline, respondents from Gedo made up 7.0 percent of the sample, while Gedo made up 8.1 percent of the midline sample that was successfully re-contacted. Given this relatively small shift, we do not think any special steps need to be taken to address differential attrition in the analysis, though we urge the endline evaluation team to minimize attrition to the extent possible and consider re-weighting schemes to recover the underlying population distribution where necessary.

## 2.5 Challenges and Limitations

The baseline evaluation of AGES documented a number of limitations related to the evaluation design, which would impact our ability to draw strong inferences regarding programme impact and causal attribution. Generally, these limitations continue to apply at midline and apply equally to the new NFE (USAID-supported) cohort that has been introduced in the latest round. The most important methodological limitations of the evaluation are outlined below, though we refer readers to the baseline evaluation report for additional details.

### Lack of a Comparison Group

By far the most serious threat to drawing valid inferences regarding the impact of AGES is the lack of a comparison group and the consequent lack of a counterfactual that would allow us to attribute changes over time to the programme's impact. In the absence of an explicit comparison group, the AGES evaluation employs a pre-post design, which compares outcomes between baseline, midline, and endline, testing for changes over time. Any temporal changes in outcomes over time cannot be attributed to the intervention, except by assuming that external factors – such as an outbreak of conflict, drought, improvements in stability, or significant advances in nationwide educational attainment – were not responsible for the changes in outcomes. The weakness of such comparisons stems from the possibility of positive (or negative) trends over time in the broader program environment.

This issue is mitigated, in part, by the proactive benchmarking of learning and transition outcomes that occurred at baseline, but benchmarked outcomes come with comparatively strong assumptions regarding the comparability of “benchmark girls” to cohort girls. For instance, to draw conclusions in the context of learning, we must assume that benchmark girls – e.g., those who were in grade 4 at baseline – are equivalent to girls who were in grade 2 at baseline (grade 4 now) in all ways except in their exposure to AGES interventions. This may or may not be a valid assumption, as girls who reached grade 4 in the absence of AGES interventions may be more motivated, or different in other, fundamental ways, from the girls tracked as a cohort since baseline.

The lack of a comparison group causes greater problems in the case of ABE and NFE girls, for whom no valid benchmark sample exists. This limitation applies equally to the New NFE cohort recruited in this evaluation round, who will be longitudinally tracked going forward.

### Sample Size and Statistical Power

As discussed in the baseline evaluation report, the baseline sample design took into account explicit power analysis performed by CARE. However, the target samples at baseline were not met, due to a loss of sampling points (both schools and centres) and improper survey administration by two of Consilient's enumerators. The baseline report analysed the changes in statistical power that would result; now that

midline data collection is complete, our assumptions regarding panel attrition – which we expected to be between 75 and 80 percent, overall – have largely been validated. The current panel sample remains sufficient for nearly all of the desired analysis, though disaggregation of changes over time into relevant subgroups often results in such small samples that we cannot draw reliable inferences. More concerning is the expectation of further attrition between midline and endline, particularly among the ABE and NFE cohorts, for whom replacement girls are not available. The panel sample for these groups will become dangerously small at endline and make it difficult to assess changes in learning scores and other outcomes over time.

On the other hand, the New NFE cohort recruited in this round, with a total baseline sample of 916 girls, is well-powered for a greater number of comparisons over time.

### Additional Limitations

- **Comparability of learning assessments across time** – comparability of learning assessments across time remains a concern, as any changes in overall difficulty will obscure and bias analysis of changes in learning outcomes over time. However, updated learning assessments at midline very closely mirrored those employed at baseline, with CARE’s Monitoring & Evaluation team carefully adjusting the questions (to avoid recall effects) without changing the overall difficulty.
- **Ceiling and floor effects** – ceiling and floor effects remain a concern, though our focus is on ceiling effects, in light of the relatively strong performance of all three cohorts at baseline. CARE sought to actively mitigate against ceiling effects by incorporating additional, more difficult, subtasks into the learning assessments at midline. While these more difficult subtasks cannot be utilised for baseline-to-midline comparisons (because the two learning assessments are of significantly different difficulties), the more difficult assessments will facilitate comparisons from midline to endline, the point at which ceiling effects are most likely to be observed.
- **Panel attrition** – panel attrition remains a concern, in line with our discussion at baseline. The midline evaluation has provided empirical insights regarding the extent of attrition that should be expected across rounds, with aggregate attrition over the two years separating baseline and endline of 21.3 percent. We would expect a similar level of attrition going forward. Attrition could potentially increase, owing to the older age of girls and their greater likelihood of marrying and/or migrating; on the other hand, it could also decrease, because the set of girls who were successfully re-contacted at midline are a non-random subsample of the baseline – these girls have “self-selected” into staying in the sample and we would expect lower attrition among this subsample going forward.<sup>20</sup>
- **Previous exposure to educational interventions** – the baseline evaluation revealed that a number of ABE and NFE girls – who were generally expected to have never had previous exposure to schooling – had been enrolled in school or other educational interventions in the past. As part of the standard household survey, caregivers were asked whether their girl had

<sup>20</sup> This logic is supported by other girls’ education projects we have evaluated in Somalia, where the highest attrition rates tend to occur from baseline to midline. Girls who were successfully re-contacted at midline in, e.g., the SOMGEP-T evaluation, had higher re-contact rates at endline, because the sample’s average “type” of girl had shifted from one likely to migrate to one more likely to remain in-situ and able to be re-contacted.

ever attended school in the past, and 54.5 percent of respondents at baseline indicated that they had. In practice, though, previous exposure to education does not present a specific methodological threat to the evaluation; it simply raised the starting point for many girls and may have made it more difficult for the programme to show positive impact, as there was less “low-hanging fruit” of which the programme could take advantage. Nonetheless, there is little reason to believe that prior exposure to schooling or educational programming should present inferential concerns in this evaluation or the endline that will follow.

## 3. Sample Characteristics

### 3.1 New NFE (Cohort 4) Sample Demographics

The baseline sample for new NFE centres was stratified by state and selected with probability proportional to size. While sampling weights are technically required to adjust for slight variation in cluster size and the slight deviation from strict proportionality at the state level, the weights in question are almost uniform; specifically, frequency weights range from just 45.85 to 48.26, a very small range for weights. Throughout this discussion, we report unweighted sample demographics as the incorporation of weights does not affect the conclusions drawn.

The table below reports the composition of the new (Cohort 4) NFE sample at baseline, focusing on household socio-economic status, disability status, and vectors of marginalisation, such as language spoken and minority status. In most ways, the sample mirrors that of the cohorts recruited in the original baseline in 2019, with similar rates of household hunger and land ownership, and a similar proportion of IDP households. In reality, the new NFE girls are significantly more disadvantaged than the original cohorts were *in 2019*, but this reflects the worsening economic situation in Somalia since that time. At *present*, the new NFE girls mirror their counterparts from the earlier sample, who have also experienced a significant decline in household economic security.

TABLE 5: INDIVIDUAL AND HOUSEHOLD CHARACTERISTICS OF NEW NFE (COHORT 4) SAMPLE

Characteristic	Share of Sample
HH has poor roof (mud/thatch/cardboard/plastic)	22.0%
Went to sleep hungry most/all nights, last 12 months	6.7%
Went without water for home use most/all days, last 12 months	8.1%
HH owns land, either solely or jointly	23.1%
Girl has physical disability	3.2%
Girl has physical disability, alternative coding	3.4%
Girl has cognitive, behavioral, or communicative disability	4.4%
Girl has cognitive, behavioral, or communicative disability, alternative coding	5.6%
Girl has mental health disability	15.5%

Girl has mental health disability, alternative coding	17.0%
IDP household	40.9%
Household speaks af-Maay	33.3%
Linguistic minority	5.8%
Member of minority group, non-Somali, or occupational caste	13.8%

**Note from the project:** The proportion of new NFE girls (Cohort 4) who reported going to sleep at night feeling hungry many/most days is 23.8%, while the proportion who reported experiencing any level of food insecurity is 68.3%. Similarly, 27.7% of the new NFE girls reported going without enough water at home many/most days, while 66.9% reported any level of limitation in the access to water. Table 5 only provides information on the proportion reporting extreme food insecurity and lack of access to water (most days/ always).

Overall, however, there are no large, systematic differences between the new NFE and previous NFE samples. The new NFE girls are younger than the Cohort 1 NFE girls, by 0.8 years, on average. This age difference is *at present*, not at the time of their enrolment; at the time of recruitment, the new NFE girls were significantly older than Cohort 1 NFE girls – 19.3 years old, compared to 17.9 years among Cohort 1 NFE girls at baseline.<sup>21</sup> Their households are slightly (1.5 percentage points) less likely to own land, and they are slightly more likely to have a physical or cognitive disability; at the same time, they are somewhat less likely to have a mental health disability related to anxiety or depression.

The sample's composition was not affected drastically by attrition between selection and the completion of fieldwork. This is not because few girls were replaced; indeed, attrition and non-response were frequent, and the final sample included 104 replacements that were pre-selected prior to fieldwork. These girls were interviewed when, upon arriving at the school, the original girl drawn for the sample could not be located, indicated they were not enrolled in the NFE centre in question, or refused to participate. The most frequent reason for replacement, which accounted for 34.3 percent of all replacements of New NFE girls, was that the girl had permanently moved away from the location in question and would not be a participant in that NFE centre going forward.

We investigated patterns of attrition and replacement by region, district, and age. The lowest contact rates were found in Lower Shabelle (SW State), where we located and interviewed 84.8 percent of girls from the original sample draw, while the highest contact rate was in Bay (SW State), at 94.5 percent. The lowest contact rates, at a district level, tended to be clustered in Banadir, though the sample size of individual districts within Banadir was often very small. We expected contact rates to be linearly related to age, because age is often a predictor of successful re-contact in our panel studies. However, this was not the case – the lowest contact rates were among girls 18-22 years of age, with higher contact rates among both younger (14-17) and older (23+) girls.

<sup>21</sup> To clarify, Cohort 1 NFE girls were 17.9 years at baseline and now average 20.1 years; at the “baseline” for Cohort 4 NFE girls, they are 19.3 years, on average.

Our takeaway from this analysis is that contact rates were relatively low but successful contact was not systematically related to a girl’s characteristics.<sup>22</sup> However, a low success rate when sampling for a baseline is of comparatively less importance than the re-contact rates we analyse within the tracked panels at midline, which we discuss in a subsequent section. Attrition from subsequent rounds in a tracked panel reduces the statistical power of our tests, because – even if girls can be replaced – replacements are never perfectly comparable to the girls they replace. In contrast, attrition, refusal, or unsuccessful contact at baseline may alter the composition of the sample slightly but, do not fundamentally undermine our ability to draw inferences about changes over time when the baseline sample is re-contacted in future rounds.<sup>23</sup>

How representative is the eventual New NFE sample of the underlying population of New NFE beneficiaries? The table below reports the few characteristics for which we can make this comparison, reporting the characteristics of the population (sample frame), the original sample draw (assuming 100 percent successful contact), and the eventual, achieved sample, factoring in replacements of girls from the original sample who could not be located.

**TABLE 6: MATCH BETWEEN POPULATION AND SAMPLE CHARACTERISTICS, NEW NFE COHORT**

Characteristic	Sample Frame	Drawn Sample	Achieved Sample
<b>Age Group</b>			
<b>Age 15 years and under</b>	21.8%	23.9%	22.5%
<b>16-17 years</b>	22.9%	21.2%	21.2%
<b>18-19 years</b>	17.2%	17.1%	17.0%
<b>20-22 years</b>	16.5%	15.1%	16.7%
<b>23+ years</b>	21.7%	22.6%	22.6%
<b>State</b>			
<b>Banadir</b>	58.4%	47.2%	48.0%
<b>South West</b>	27.8%	36.5%	36.8%
<b>Hirshabelle</b>	13.9%	16.3%	15.2%

As the table shows, the achieved sample deviates substantially from the population, primarily in terms of regional representation. With regard to age, the mean age shifted from 18.9 years in the population, to

<sup>22</sup> The sample frame provided by CARE did not record a large number of characteristics of the girls, which limits our ability to analyse patterns of successful contact and replacement.

<sup>23</sup> In other words, attrition from a tracked sample at midline constitutes a very real threat to valid inferences. Unsuccessful contact and refusals from the initial sample draw are a comparatively minor issue.



19.0 years in the final sample, with minor shifting across age groups. However, these differences were relatively small and do not represent systematic bias.<sup>24</sup>

Regional representation differs more substantially, due to the expansion of the sample frame after the initial sample was drawn. As noted above, the complete sample frame had a larger number of girls from Banadir than the partial sample frame that was used for state-level stratification; the result is that regions in South West State and Hirshabelle are overrepresented in the final sample, and Banadir is underrepresented. To guard against the possibility that imbalance in state- or region-level representativeness could alter our results, we calculate sampling weights that adjust for this imbalance – upweighting respondents in Banadir – and report topline findings for learning outcomes using these weights as a robustness check.

## 3.2 Cohort Profiles

In this section, we document the characteristics of the FE, ABE, and NFE samples for whom we analyse baseline-to-midline changes in programme outcomes. As discussed in the previous section, panel attrition has affected each of the three distinct samples, with attrition as high as 24.9 percent among the NFE cohort. While re-contact and attrition are not systematically correlated with demographic or household characteristics, it is important to understand the nature of the different cohorts, in part because this allows us to explain some of the trends in programme outcomes we discuss later in this report.<sup>25</sup>

The table below describes the original baseline sample of FE girls – those who were sampled from within formal schools in 2019 – and the sample of girls who were successfully re-contacted at midline. The table allows us to judge how the sample’s characteristics have changed as a result of overall attrition. Both columns report characteristics, as measured at baseline; thus the difference between the two columns are a function *exclusively* of attrition – both due to inaccessibility and due to individual-level attrition.

**TABLE 7: CHARACTERISTICS OF FE COHORT, BEFORE AND AFTER PANEL ATTRITION**

Sample Characteristics	Original Baseline Sample (n = 421)	Post-Attrition Baseline Sample (n = 341)
------------------------	------------------------------------------	------------------------------------------------

<sup>24</sup> Self-reported age, among girls interviewed for the baseline, also differed somewhat from the age recorded in the sample frame: mean self-reported age was 19.3 years, compared to 19.0 years reported in CARE’s listing exercise. In total, 38.0 percent of girls reported an age different from that provided by CARE. However, two facts reduce our concern about this discrepancy: first, a significant number of girls may have had birthdays between the time of the listing exercise and our data collection, accounting for changes of one year; second, differences between the ages reported are not, in any way, systematic. Age does not appear to have been consistently under- or over-reported in order to ensure eligibility for the programme, for instance, and appears to be random measurement error.

<sup>25</sup> Note that the analysis in this section does not touch on the relative representativeness of the cohorts, relative to the underlying populations that they represent. Readers interested in a discussion of representativeness in the context of the original baseline sample should refer to the baseline evaluation report; representativeness of the New NFE cohort (for whom this report serves as a baseline) is discussed above under “Baseline Sample Demographics.”

<b>Family Structure and Characteristics</b>		
Girl has only one living parent	10.7	11.4
Girl has no living parents	0.7	0.9
Girl does not live with either parent in her HH	3.6	3.5
Female-headed household	34.9	33.1
HoH has no education of any kind (no Quranic)	22.1	22.0
HoH has no formal education (may have Quranic)	80.8	79.8
Caregiver has no education of any kind (no Quranic)	29.0	29.9
Caregiver has no formal education (may have Quranic)	85.7	85.3
<b>Household Wealth and Socio-Economics</b>		
HH has poor roof (mud/thatch/cardboard/plastic)	18.1	19.1
Went to sleep hungry most/all nights, last 12 months	6.9	6.3
Went without water for home use most/all days, last 12 months	4.8	4.4
HH owns land, either solely or jointly	38.0	38.4
Household owns a phone	84.8	84.5
Household owns a smartphone	20.7	21.1
HoH does not have an occupation or does not earn a wage	37.1	36.7
Household is engaged in pastoralism	1.4	0.6
<b>Disability Status</b>		
Girl has physical disability	0.0	0.0
Girl has physical disability, alternative coding	0.7	0.6
Girl has cognitive, behavioral, or communicative disability	0.0	0.0
Girl has cognitive, behavioral, or communicative disability, alternative coding	0.5	0.3
Girl has mental health disability	10.5	9.4
Girl has mental health disability, alternative coding	11.6	10.9
<b>Indicators of Household Marginalization</b>		
Household speaks af-Maay	27.8	26.7
Linguistic minority	5.0	6.2
Member of Bantu or occupational minority groups	8.6	9.4

As the table shows, FE girls at baseline were widely disadvantaged. A majority of girls hailed from homes where neither their caregiver nor their head of household had any formal education. Despite their location in urban and peri-urban areas, 18.1 percent of FE girls lived in households with poor-quality roofs,

constructed of mud, thatch, plastic, or other poor-quality building materials. Household hunger was relatively rare, but landlessness was common and over one-third (38.1 percent) of girls were IDPs.<sup>26</sup>

Physical and cognitive disabilities were relatively rare within the sample, at least as reported by girls themselves or their caregivers. A moderately-liberal coding of disability indicated that just 0.7 percent of girls had a physical disability, while another 0.5 percent had cognitive, behavioural, or communicative disability. In contrast, mental health disabilities – defined as relatively frequent anxiety or depression – was much more common.

As the table also shows, there have not been major shifts in sample composition from baseline to midline. Despite aggregate attrition of 14.3 percent, the sample is largely unchanged in terms of household composition or socioeconomic status. There have been minor shifts in the share of respondents whose parents have received no formal education – a slight drop in that share – and a slight drop in the number of girls with mental health difficulties, but few other changes.

Moving to the ABE sample, the table below reports identical information regarding sample composition for ABE girls, who were aged 13-17 years at baseline, but were concentrated within the 13-15 year old range at that time.<sup>27</sup> While the education of caregivers and heads of household improved slightly among the FE sample as a function of attrition, the opposite has occurred among ABE girls – the girls who were successfully re-contacted at midline are slightly more likely to come from households without a formally-educated adult member.

**TABLE 8: CHARACTERISTICS OF ABE COHORT, BEFORE AND AFTER PANEL ATTRITION**

Sample Characteristics	Original Baseline Sample (n = 484)	Post-Attrition Baseline Sample (n = 364)
<b>Family Structure and Characteristics</b>		
Girl has only one living parent	10.1	9.9
Girl has no living parents	1.2	1.1
Girl does not live with either parent in her HH	5.0	4.9
Female-headed household	33.7	34.3
HoH has no education of any kind (no Quranic)	22.1	23.4
HoH has no formal education (may have Quranic)	82.9	84.1
Caregiver has no education of any kind (no Quranic)	27.9	30.5
Caregiver has no formal education (may have Quranic)	86.0	86.3
<b>Household Wealth and Socio-Economics</b>		
HH has poor roof (mud/thatch/cardboard/plastic)	22.1	21.2

<sup>26</sup> We use the term “household hunger” loosely. At baseline, 6.3 percent of FE girls came from households that had gone to bed hungry most days or always over the previous year. This is a fairly severe version of hunger. A slightly more liberal definition – the share of households that went to sleep hungry many (10 or more) nights or more often was 23.2 percent at baseline.

<sup>27</sup> The mean age of ABE girls at baseline was 14.1 years in the full sample as well as in the share of the baseline sample used for our panel analysis. The mean age of these same girls at midline is 16.7 years.

Went to sleep hungry most/all nights, last 12 months	7.9	7.7
Went without water for home use most/all days, last 12 months	6.2	6.9
HH owns land, either solely or jointly	31.4	32.1
Household owns a phone	86.4	86.8
Household owns a smartphone	13.2	13.5
HoH does not have an occupation or does not earn a wage	36.6	34.9
Household is engaged in pastoralism	2.7	1.1
<b>Disability Status</b>		
Girl has physical disability	0.2	0.3
Girl has physical disability, alternative coding	0.2	0.3
Girl has cognitive, behavioral, or communicative disability	0.6	0.8
Girl has cognitive, behavioral, or communicative disability, alternative coding	1.0	1.4
Girl has mental health disability	12.8	12.1
Girl has mental health disability, alternative coding	16.9	16.5
<b>Indicators of Household Marginalization</b>		
Household speaks af-Maay	31.6	27.2
Linguistic minority	7.2	4.7
Member of Bantu or occupational minority groups	11.6	11.0

Other measures of household marginalisation or privilege – at least in the context of socioeconomic status and wealth – are, essentially, unchanged. For instance, the share of households who own land or a smartphone is virtually identical between the full baseline sample (484) and the smaller sample (364) of girls successfully re-contacted at midline.

On the other hand, the ABE sample experienced a significant change in the number of IDP households included at midline, increasing from 33.5 to 43.1 percent. Again, this shift is not a function of interviewees changing their responses across rounds, as all results are based on responses at baseline; the only explanation for the shift is via differential attrition – in this case, higher attrition among non-IDP households that increased the share of IDP households remaining in the panel sample. Meanwhile, although we cannot assess changes in the inclusion of minority groups from baseline to midline, it is clear that the ABE sample is particularly representative of minority groups. Indeed, more so than either the FE or NFE cohorts, the ABE cohort is comprised of a majority of marginalised groups along this axis, with 53.6 percent of girls hailing from households of ethnic minorities and occupational castes.

Finally, the table below reports the sample characteristics of NFE girls. The sample of NFE girls fell from 507 at baseline to 381 at midline, as a function of school-level attrition ( $n = 8$ ) and individual-level attrition ( $n = 118$ ). However, as with ABE girls, the overall contours of the sample have not changed appreciably in response to attrition. The girls who remain in the sample are slightly less likely to have a mental health disability, and are slightly more likely to live in a household where the household head has a paid, wage-earning occupation. However, minor shifts in terms of two or three characteristics are expected, even as a function of random sampling variation – there is no evidence from this table that the sample has changed in a systematic manner as a function of attrition.

TABLE 9: CHARACTERISTICS OF NFE COHORT, BEFORE AND AFTER PANEL ATTRITION

Sample Characteristics	Original Baseline Sample (n = 515)	Post-Attrition Baseline Sample (n = 381)
<b>Family Structure and Characteristics</b>		
Girl has only one living parent	17.5	17.6
Girl has no living parents	1.0	1.3
Girl does not live with either parent in her HH	8.5	7.1
Female-headed household	42.7	43.0
HoH has no education of any kind (no Quranic)	19.8	21.3
HoH has no formal education (may have Quranic)	81.7	80.6
Caregiver has no education of any kind (no Quranic)	24.5	26.0
Caregiver has no formal education (may have Quranic)	83.9	83.7
<b>Household Wealth and Socio-Economics</b>		
HH has poor roof (mud/thatch/cardboard/plastic)	19.6	19.4
Went to sleep hungry most/all nights, last 12 months	6.2	7.1
Went without water for home use most/all days, last 12 months	4.7	5.2
HH owns land, either solely or jointly	29.1	28.6
Household owns a phone	88.2	88.2
Household owns a smartphone	20.2	20.2
HoH does not have an occupation or does not earn a wage	43.7	39.9
Household is engaged in pastoralism	1.2	0.8
<b>Disability Status</b>		
Girl has physical disability	0.2	0.3
Girl has physical disability, alternative coding	0.8	1.0
Girl has cognitive, behavioral, or communicative disability	0.6	0.3
Girl has cognitive, behavioral, or communicative disability, alternative coding	1.2	1.0
Girl has mental health disability	10.9	9.4
Girl has mental health disability, alternative coding	16.9	15.5
<b>Indicators of Household Marginalization</b>		
Household speaks af-Maay	25.4	24.7
Linguistic minority	6.8	7.6
Member of Bantu or occupational minority groups	9.9	11.5

In the discussion thus far, we have focused on changes in the three cohort samples due to panel attrition. However, changes in sample characteristics can also arise due to changes over time, even where the set of respondents remains fixed. These shifts are also important, not because they represent the effects of

panel attrition, but because they tell us something important about how girls and their households have changed over time in response to maturation and exogenous events.

In the table below, we report the baseline and midline characteristics of the panel sample for each cohort. To be clear, the set of girls is fixed from baseline to midline – this discussion includes only those girls who appeared in both the baseline and midline evaluation rounds. Our interest is in how they and their households have changed since the baseline was conducted in late 2019.<sup>28</sup>

The most important trends shown in the table concern shifts household economics. Surprisingly, household hunger is not dramatically more common at midline than it was at baseline, despite the ongoing drought in much of southern and central Somalia. Gedo, for instance, is in the grip of a severe drought that has worsened significantly in recent months. While the overall share of households reporting moderate-to-severe hunger has increased, this increase has not been as stark as one might expect – increasing from 6.7 to 7.8 percent in the aggregate. Over the same period, the share of households participating in a VSL or savings group has also increased, from 4.5 percent of caregivers at baseline to 9.6 percent at midline, further complicating notions of drought-induced vulnerability and hinting at the possibility that the programme has significantly improved resilience.

On other measures of household economics, the sample has become more vulnerable. For instance, a larger share of households are headed by an individual who does not have an occupation and does not earn a wage. This outcome may have shifted due to changes in *who* constitutes the head of household for a given girl (e.g., if she marries and her head of household is now her husband). However, major changes in head of household occupation among FE and ABE girls cannot be explained in this way, since marriage rates among both these groups are still very low, owing to their young age.

**TABLE 10: CHANGES IN INDIVIDUAL AND HOUSEHOLD CHARACTERISTICS, BL TO ML**

Characteristic	FE Girls		ABE Girls		NFE Girls	
	Baseline	Midline	Baseline	Midline	Baseline	Midline
Girl is currently married	0.3	1.2	0.0	4.7	N/A	N/A
Girl has ever been married	0.3	1.6	0.5	5.2	N/A	N/A
Girl has a child	0.0	0.6	0.0	1.9	N/A	N/A

<sup>28</sup> Where data is missing for NFE girls, this is due to the fact that many NFE girls were – at midline – over the age of 18. Girls 18 and above did not complete a full household survey at midline, which prevents us from measuring many of the outcomes of interest in this discussion. Of the 364 ABE girls who were successfully re-contacted at midline, 214 completed a full household survey; the analysis in the table always relies on a comparable sample from baseline to midline, consisting of the exact same respondents across rounds. This means that, for some outcomes, the ABE girl sample is just 211 respondents, while in others it includes the full 364 girls who were re-contacted. This issue is more problematic in the context of NFE girls, more of whom fell above the age cut-off for completing a household survey – just 12 NFE girls completed a full household survey. Therefore, for outcomes recorded only through the full household survey, the sample size of NFE girls is too small to analyse changes from baseline to midline.

Went to sleep hungry most/all nights, last 12 months	6.4	8.8	7.8	8.6	7.4	6.6
Went without water for home use most/all days, last 12 months	4.6	9.5	6.9	8.0	5.5	6.8
HH owns land, either solely or jointly	39.0	38.1	32.1	27.7	29.6	29.6
Household owns a phone	83.9	90.1	82.9	89.1	N/A	N/A
Household owns a smartphone	21.4	30.4	7.1	11.4	N/A	N/A
HoH does not have an occupation or does not earn a wage	36.6	48	25.2	48.6	N/A	N/A
Journey to school is unsafe, per girl	1.8	2.1	1.1	6.3	2.5	7.4
Journey to school is unsafe, per caregiver	1.9	1.9	0.0	3.3	N/A	N/A

Girls in ABE centres – and their caregivers – are almost six times more likely to report that the journey to their school is unsafe now, compared to baseline. Although 6.3 percent of girls reporting an unsafe journey may not seem like a major increase, it is significant, given how rarely girls generally report insecurity of this kind. The trend seems to be driven by increases in Banadir (from 1.7 to 9.1 percent of girls reporting insecurity) and Lower Juba (from 1.5 to 7.4 percent). Importantly, this trend is not driven by changes in sample composition, as our analysis is restricted to girls who appeared in both the baseline and midline evaluation rounds.

# Midline Results

---

Formal Education, Accelerated Basic Education, and Non-Formal Education Cohorts



## 4. Learning

The AGES programme focuses on two primary outcomes. The first is girls' numeracy and literacy, specifically, girls' command over reading in Somali; the second is "transition": the advancement of girls enrolled in schools from one grade to the next, the transition of girls not enrolled in schools into formal education institutions or alternative learning institutions, and the transition of older girls into gainful employment. This section presents findings regarding the first outcome – numeracy and Somali literacy. It specifically focuses on first identifying aggregate changes for each type of girls in the sample, followed by changes in the specific subtasks in the numeracy and literacy assessments. Next, we present findings on the trends between disability and other demographic sub-groups. The final sub-section analyses the extent to which the intermediate outcomes, as identified in the project theory of change, serve as predictors of changes in learning outcomes.<sup>29</sup>

Before proceeding, a few methodological notes are warranted. First, the baseline evaluation report had ran a number of tests on the scores results, including running discrimination analyses to determine how well each subtask correlated with students' overall performance in the EGRA or EGMA. As the fundamental structure of the assessments have not changed, we expect the findings of the baseline to largely hold true. Second, the analysis in this section only utilizes the panel sample of girls – meaning the girls who were in the sample at both baseline and midline. This included girls from the FE, ABE, and C1 NFE cohorts. The below table highlights our panel sample, in terms of the number of institutions of each type that were in the sample, and the number of girls. The distribution of learners is even across the learning cohorts, ranging from 332 in the FE cohort to 366 in the C1 NFE cohort.

TABLE 11: NUMBER OF INSTITUTIONS & GIRLS IN THE PANEL SAMPLE

	BDR	JSS	SWS	Total
<b>Institutions</b>				
<b>Formal Education</b>	14	8	10	32
<b>Accelerated Basic Education</b>	16	9	9	34
<b>C1 Non-Formal Education</b>	15	13	7	35
<b>Total</b>	45	30	26	101
<b>Girls</b>				
<b>Formal Education</b>	142	119	71	332
<b>Accelerated Basic Education</b>	175	83	106	364
<b>C1 Non-Formal Education</b>	187	83	96	366
<b>Total</b>	504	285	273	1062

<sup>29</sup> Additional tables of results are provided in Annex 2. Specifically, Annex 2 documents subgroup learning outcomes at midline (as opposed to subgroup-specific *gains* in learning, presented below), analysis of foundational gaps in literacy and numeracy, and documentation of the share of students achieving grade-level competency in literacy and numeracy.

Third, as articulated in the baseline methodology, our analysis of changes in learning outcomes over time uses benchmarks obtained during baseline from a separate sample of 454 girls, drawn separately from the same schools which were surveyed at baseline. As described, the purpose of the benchmark is to help account for maturation effects – the fact that girls on average are likely to develop their skills as they become older, not necessarily due to their education or any intervention – and to attempt to separate the programme’s effects on girls’ learning from learning that would have occurred regardless. While not a perfect substitute for a comparison group in an experimental or quasi-experimental setting, benchmarks provide an additional robustness check to assess programme impact. The table below outlines expected changes from one grade level to the next due to maturation effects, and the midline girls to which we compare to benchmark changes. Fourth, by default, we run the analyses that follow without using weights. However, we also measure results using weights as robustness checks, and will report the weighted results should they be substantially different from the initial results.

**TABLE 12: BENCHMARK CHANGE FOR EACH GRADE LEVEL AT BASELINE**

Girls' Grade at BL	Benchmark	Change, Numeracy	Change, Literacy
<b>Primary 1</b>	Primary 1 to Primary 2	15.05	23.06
<b>Primary 2</b>	Primary 2 to Primary 3	10.03	14.70
<b>ABE</b>	Primary 1 to Primary 2	15.05	23.06
<b>C1 NFE</b>	N/A	N/A	N/A

And finally, during the midline assessment, the EGMA had been expanded to include 11 subtasks, as compared to 8 in the baseline. This was due to analysis of the BL results demonstrating that girls’ scores on the numeracy tests skewed positive, and such ceiling effects may mean that there is relatively little room for improvement for the numeracy assessment, which in turn might obfuscate the programme’s true effect. To grade girls’ numeracy skills at ML, we use only the 8 subtasks that were also administered at BL; including the three additional (and more challenging) subtasks lowers ML scores compared to ML scores calculated solely on the basis of the original 8 subtasks, and thus would obscure girls’ improved learning between BL and ML, or make decreases in learning appear more steep than they otherwise are. To maintain symmetry in the assessment tool between BL and ML, the 8 subtask grading scheme is preferred. Moreover, as the benchmarks calculated to estimate maturation effects as girls age were based on data collected at BL, benchmarks are also based on the original set of 8 subtasks. To enable a comparison of BL and ML changes in the panel sample’s scores relative to the benchmark, it is important that the same grading criteria are used.<sup>30</sup>

## 4.1 Aggregate Learning Outcomes

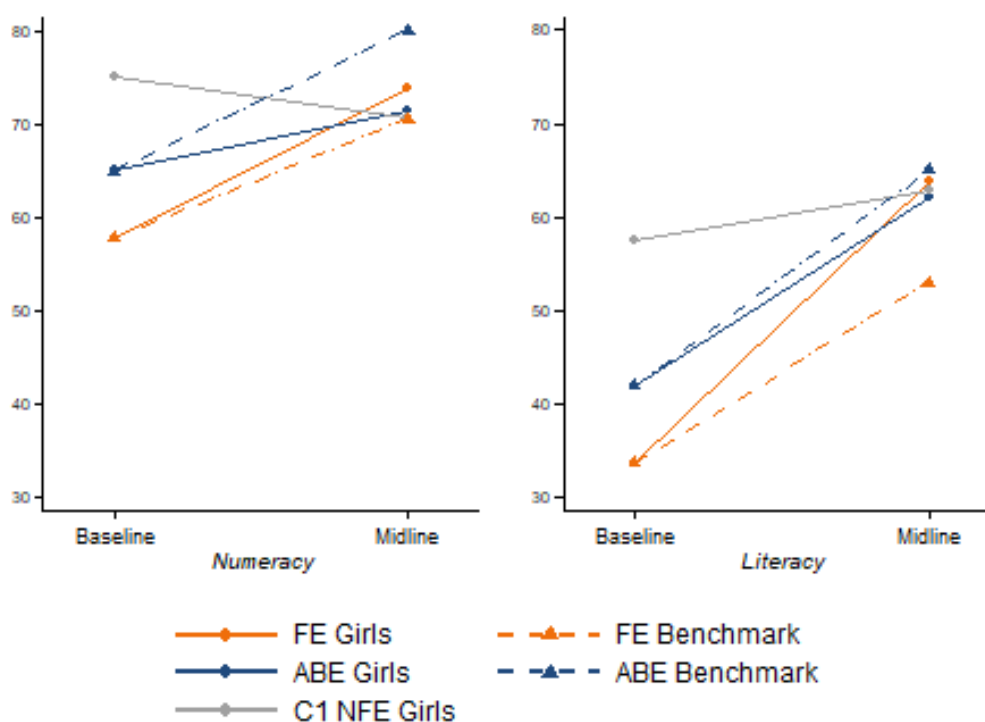
This sub-section presents key findings on learning outcomes for girls in each learning cohort and geography, at the level of aggregate scores for both numeracy and literacy tests. We first focus on outcomes across cohort types, and then across states and present the key findings of each analysis.

<sup>30</sup> Having said that, the fact that girls at ML have now completed the three additional numeracy subtasks mean that we will be able to make ML to endline comparisons based on the expanded set of numeracy subtasks, in addition to the BL to endline comparison using the original set of 8 subtasks. At endline, the research team will use the ML to endline comparisons based on the expanded numeracy assessment as a robustness check, and to supplement the BL to endline comparison, where needed.

## Cohort Type

The graph below presents a high-level cursory view of the changes since baseline, and compared to benchmark expected changes for ABE and FE girls. A cursory review of the graphs reveals a number of noteworthy trends. Overall, the numeracy scores paint a mixed picture of numeracy learning between BL and ML. C1 NFE girls, who in the BL were noted to have considerably higher scores compared to FE girls (most likely due to the former two categories of girls being older on average), registered a nearly 5 point decline in their scores.<sup>31</sup> While ABE girls overall had an increase in their numeracy scores, this increase was still well below the benchmark (8.5 points lower). FE girls were the only category of girls whose learning improved above the benchmark expected improvement. In contrast, literacy scores paint a more optimistic picture overall, as all categories of girls experienced an increase in their scores. For FE girls, the increase happened at a rate that was faster than expected when compared to the benchmark, which suggests that the programme may be having a very large “true effect” on FE girls’ learning in particular. Similar to numeracy scores, ABE girls experienced an increase in literacy, but are still nearly 3 points below the benchmark. C1 NFE girls also experienced an increase in their scores; while we lack a benchmark for this particular cohort of girls, the BL to ML increase was very slight, and considerably smaller than the increases demonstrated by ABE and FE girls.

FIGURE 2: CHANGES IN AGGREGATE NUMERACY AND LITERACY SCORES



<sup>31</sup> One explanation for this outcome is that C1 NFE girls had already been enrolled in NFE programmes for as long as four months at the time of the baseline and their programmes ended in September 2020. The latter fact means that C1 NFE girls had been out of schooling – unless they shifted into formal education, which few did – for almost a year and a half before the midline evaluation. Combined, these disjunctures in timing can explain the high starting point of C1 NFE numeracy scores and their deterioration since the baseline.

However, despite the substantial average increase in literacy, it should be noted that reading skills at BL were considerably lower than numeracy skills for all categories of girls; current levels of literacy are around the same levels for numeracy, hovering somewhere in the range of 60 to 65. This suggests that the large increase in literacy may be a “catch-up” effect experienced when girls are initially introduced to structured education programmes after spending substantial time being out of school or irregularly attending school. Additionally, aggregate scores appear to also be converging, both among girls in different cohorts for the same assessment.

In light of the above, we first investigate whether differential attrition led to ABE and C1 NFE girls in the panel sample to register smaller changes in learning relative to FE girls. In other words, if the ABE and C1 NFE girls who were not reached for the learning assessments during ML were systematically different from those who were – namely, that they were higher performers – then the ML results might be biased downwards because those who would have performed well had left the sample. By definition, because these girls have left, we cannot compute changes in their scores from BL to ML. A comparison of the average BL numeracy scores of ABE and C1 NFE girls between those who remained in the panel sample and those who did not showed no significant difference. For example, for the ABE girls, those remaining in the sample had a BL average score of 64.99, compared to 68.11 for girls who would no longer be in the sample, a difference of 3.12 points, though this difference was not statistically significant. The full list of comparisons are outlined below.

**TABLE 13: DIFFERENCE IN OUTCOMES BETWEEN PANEL AND NON-PANEL SAMPLE GIRLS AT BASELINE**

Cohort	Numeracy Diff at BL	P-Value	Literacy Diff at BL	P-Value	Age Diff at BL	P-Value
<b>FE Girls</b>	-2.61	0.59	0.51	0.92	0.04	0.76
<b>ABE Girls</b>	-3.12	0.36	7.69	0.03	0.15	0.30
<b>C1 NFE Girls</b>	-1.41	0.43	2.13	0.53	-0.04	0.68

The table indicates that, for numeracy scores, girls who remained in the panel registered lower scores compared to girls who left the panel, but none of these differences were significant. On the literacy measure, girls in the panel sample scored higher in BL than girls who wound up leaving, while there is almost no discernible age difference between girls who stayed and who left the panel sample. Taken as whole, this comparison does not suggest that differential attrition might bias learning outcomes downwards for some cohort of girls, meaning that we can have some confidence that C1 NFE and ABE girls’ lower performance compared to FE girls was not caused by a sample that had changed in important ways between BL and ML. In fact, the opposite concern might be more salient: ABE girls who remained in the panel sample on average scored 7.69 points higher on the literacy test than their ABE peers who left the sample, and this difference is significant. This might suggest that the ABE girls’ increase in learning in the literacy measure, already somewhat underwhelming compared to the FE girls, is in fact overestimated, as ABE girls who might otherwise register low scores at ML may have dropped out of the sample for various reasons.

The above possibility means that caution is warranted when examining ABE girls’ score changes from BL and ML. Nonetheless, even with the potential assistance of differential attrition buoying ABE girls’ scores, ABE girls still appear to perform relatively poorly compared to FE girls, thus necessitating analysis on why FE girls are outperforming, by a substantial margin, the other two cohorts.

Three further characteristics we investigate are:

- Whether the girl was still attending school at the time she was interviewed for ML;

- Her age at baseline;
- Her scores at baseline.

In the first case, girls who are no longer in school at ML may experience a slower rate of learning outcome increases compared to her peers and/or in fact registering a decline in their scores from BL to ML for several possible reasons. This is because such girls have had a shorter period of time receiving the programme's intervention, and therefore experienced a lower level of treatment intensity compared to girls who are still in the programme after the two years since the BL. Naturally, the former girls are more likely to underperform compared to the latter. Indeed, in our ML panel sample, 84.15% of C1 NFE girls are no longer attending school or learning centre, as compared to 39.84% of ABE girls, and 15.14% of FE girls.

The analysis thus far is limited insofar as we have not actively controlled for alternative explanations when analysing the comparative performance of different cohorts. At the most basic level, systematic and intentional differences in the ages of the three cohorts at recruitment suggest the need for regression models that can account for these additional factors. In the table below, we report results from three linear regression models predicting numeracy scores as a function of age, cohort type, state, and enrolment status at midline. These models are reported as Models 1-3; equivalent models for Somali literacy are reported as Models 4-6. We provide three versions of the analysis of each outcome because it is useful to see how – for instance – the reported effect of age on learning outcomes changes when we take into account the cohort of each girl. In each column, we report the regression coefficient – the average change in learning scores associated with a one-unit change in the predictor variable – and its standard error; results noted with asterisks are statistically significant.<sup>32</sup>

As seen in the multivariate linear regression table below under models 2 and 5, having stopped attending school/learning centre is a significant determinant of changes in numeracy and literacy score: no longer being in school at ML is associated with a lower rate of change compared to the reference group. However, both attendance status at ML and cohort type are still significant in these models, suggesting that while no longer attending school/learning centre is a determinant in score changes, it does not fully explain why FE girls are outperforming ABE and C1 NFE girls.<sup>33</sup> The most straightforward explanation for the comparatively better performance of FE girls lies in their more consistent exposure to programming. By comparison, NFE girls' exposure to structured learning interventions ended in late 2020, leaving a gap of over one year between exposure and their participation in this assessment; in comparison, FE girls – the majority of whom remain in school at midline – have not experienced a substantial gap in learning, aside from COVID-driven disruptions.

<sup>32</sup> Many of the predictors are binary (yes/no) variables, which are interpreted relative to a reference category. For instance, when Model 2 shows that being a member of the ABE cohort is associated with an 8.1 point drop in numeracy scores, this indicates ABE girls score 8.1 points lower than FE girls, specifically, after accounting for their state of residence and their current enrolment status. The results for Jubaland are similarly interpreted relative to a reference category – in this case, relative to the performance of girls from Banadir.

<sup>33</sup> The subsequent section on transition outcomes will provide a more in-depth analysis on the determinants of dropping out.

TABLE 14: EFFECT OF ML ATTENDANCE, AGE, COHORT, AND LOCATION ON LEARNING OUTCOME CHANGES

	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
	Change in Numeracy Scores	Change in Numeracy Scores	Change in Numeracy Scores	Change in Literacy Scores	Change in Literacy Scores	Change in Literacy Scores
<b>ABE Girl</b>		-8.13**	-4.22		-6.81	1.79
		(3.48)	(4.61)		(4.68)	(6.62)
<b>C1 NFE Girl</b>		14.29***	-5.52		-16.51***	2.77
		(3.70)	(7.41)		(4.72)	(9.69)
<b>Jubaland</b>		-9.73***	-9.63***		-2.44	-2.24
		(3.48)	(3.50)		(4.57)	(4.58)
<b>South West State</b>		0.35	0.32		-4.27	-4.33
		(2.92)	(2.88)		(3.69)	(3.65)
<b>Not in School at ML</b>		10.76***	-10.55***		-11.89***	-11.42***
		(2.92)	(2.94)		(3.61)	(3.67)
<b>Age</b>	-2.89***		-1.32	-3.63***		-2.91**
	(0.41)		(0.96)	(0.54)		(1.34)
<b>Constant (Ref. Group)</b>	47.50***	21.03***	35.69***	70.65***	33.52***	65.74***
	(6.04)	(2.61)	(10.84)	(8.10)	(3.43)	(14.92)
<b>Observations</b>	1,062	1,062	1,062	1,062	1,062	1,062
<b>R-squared</b>	0.08	0.12	0.13	0.07	0.08	0.09

Robust standard errors in parentheses

\*\*\* p < 0.01, \*\* p < 0.05, \* p < 0.1

**The second characteristic we investigate is age.** There are several reasons why an older age might lead to lower learning outcome changes. Older girls might be expected to contribute more to household tasks, thus taking away time and energy that could otherwise be spent on learning. The social expectations placed upon older girls, such as the expectation to marry or have children, may induce mental stress that can detract from their ability to study as effectively. For instance, the caregiver data shows that C1 NFE girls were more often than both ABE and FE girls to appear sad, anxious, or very nervous. An additional potential source of disruption to learning for older girls is the stigma of being an older learner in the same grade level as younger girls, especially if they are in the same school. Qualitative data from FGDs provide some accounts of this, including one anecdote of boys telling girls in the classroom that they looked too old to be students and should be married instead.<sup>34</sup> When asked what the biggest challenges girls faced in school was, another group of teachers were more emphatic:

*The biggest challenges are the class and their age, because they are ashamed to attend elementary school at an older age and sit in the classroom with younger students and not fully have school supplies.*

-FGD with Teachers, Banadir, Int. 508

<sup>34</sup> FGD with Teachers, Banadir, Int. 506

A girl's age at baseline is positively related to her numeracy and literacy scores at ML,<sup>35</sup> simply because older girls are more likely to have had more opportunities to learn and develop their numeracy and literacy skills even if outside the setting of a school, as already noted in the baseline report. In other words, older girls at ML naturally tend to score higher because they started from comparatively high scores. However, we find that age at baseline is in fact *negatively correlated* with changes in learning outcomes from BL to ML: the older a girl is, the likelier she is to register slower rates, or even negative rates, of learning changes over time. This can be seen in models 2 and 5 in the regression table, where the coefficient for age is negative for both assessments and significant for reading.

However, when introducing the third characteristic, **BL test scores**, as an additional variable in models 3 and 6, the negative effect of age fades, and in fact age becomes a positive predictor of learning outcome changes (significant for the literacy assessment). As expected, higher BL test scores lead to slower rates of growth from BL to ML because higher baseline scores mean that girls have comparatively little room for improvement, as they have exhausted more of the “easier” sub-tasks and the remaining opportunities for growth rest in the sub-tasks that are more challenging. Indeed, as indicated in the BL, it is the older girls who tend to score better, hence for the data up until the current ML assessment, age serves as a good proxy for BL performance and the effects thereof on future changes in learning outcomes.

The effect of BL score on future changes is significant for both assessments, as seen in the regression table above. This can be discerned graphically in the figure at the top of this sub-section, which shows that while the different cohort types started with substantially BL test scores, their ML test scores have largely converged, with lower-performing girls at BL (who tend to be younger FE girls) making a substantial leap in test scores, while other categories of girls registered more muted increases, or even decreases, in the case of C1 NFE girls in the numeracy assessment.

Taken together, the findings on the three characteristics investigated through the multivariate model suggest a number of key takeaways. First, dropping out of school/learning centre, unsurprisingly, leads to lower levels of learning outcome changes over time, and this finding is robust across all models, with and without control variables. Second, despite the negative relationship between age and improvements in learning – after controlling for the girl's cohort – the programme might still be helping older girls learn more effectively than they might otherwise have. The qualitative data strongly suggests that placing older learners in the same learning environment with younger girls might lead to the former facing some level of stigma that hinders learning, while more generally, the social expectations imposed by communities on older girls might also be a burden for their learning. Given this, we would expect smaller learning improvements among older girls, all else equal; while the regression models above suggest that this is the case, we cannot say whether the programme is having an outsized impact on older girls, nor the direction of the effect. In other words, we do not take the negative relationship between age and improvements in learning to be an indictment of the programme or its approach, as we cannot say how older girls' scores would have changed in the absence of the programme.

And finally, we find that higher performers at BL will eventually leave less room for growth up until ML, and that the “catch-up” growth of girls who were lower performers at BL has led girls from all three cohort types to have, on average, test scores that converge at ML. On one hand, this is a positive indicator of the programme's ability thus far to support the lowest-performing girls and put them on a trajectory that can help them catch up to where they need to be in terms of grade-level numeracy and literacy skills.

On the other hand, that higher-performing girls at BL register much lower levels of learning increases, or even demonstrate decreases, is concerning, as it suggests that the programme may be reaching a plateau

---

<sup>35</sup> We ran a regression model controlling for cohort type and state of interview, and find that age at BL is a positive predictor of scores at ML.



in terms of how much more the current panel sample of girls can learn, or even that the programme is devoting too few resources to them compared to lower performers. Although ceiling effects might be partially responsible for why there is a slower rate of growth for ABE and C1 NFE girls, which we have already discussed extensively at BL and in the methodology in the present report's annex, the fact that ABE girls are still below the benchmark, which were calculated based on an additional sample of girls for whom ceiling effects are also a concern, suggests that these effects are to an extent real and not simply a result of our measurement tool being inadequately challenging.

## Geography

Examining trends in learning outcome changes across different states revealed a few trends on geographic patterns in the observed results. As the table below highlights, ABE and C1 NFE girls in Jubaland experienced the most dramatic decreases in their numeracy scores compared to girls in other districts, though this was not observed for FE girls in Jubaland. In fact, Jubaland ABE girls were the only ABE girls who registered a decline in their scores. Similarly, Jubaland ABE girls had the smallest increases compared to their cohort peers in other states for the literacy assessment and Jubaland C1 NFE girls were the only category of girls who showed a decrease in their literacy scores at ML, though Jubaland girls in the FE cohort showed the single biggest increase in scores for any category, at 34.40 points. Collectively, these findings suggest that Jubaland in particular is a difficult learning environment for girls outside the FE cohort. Although the p-values associated with these figures indicate that some of these changes may not be statistically significant, we should note that the sample size for non-FE Jubaland girls is small, which can have a direct bearing on the ability of statistical tests to detect significant results.

TABLE 15: AGGREGATE NUMERACY SCORE CHANGES FOR THE PANEL SAMPLE<sup>36</sup>

Region	ML Mean Score	ML Standard Deviation	Change from BL	Change Relative to Benchmark	P-value	Number of Observations
<b>Formal School Girls</b>						
Banadir	76.56	19.40	20.24	7.87	0.00	142
Jubaland	73.42	22.15	14.75	2.44	0.00	119
South West State	69.37	25.74	9.93	-4.49	0.07	71
Aggregate	73.90	21.97	16.07	3.28	0.00	332
<b>ABE Girls</b>						
Banadir	73.62	21.85	8.11	-6.94	0.04	175
Jubaland	61.23	32.50	-4.22	-19.28	0.56	83
South West State	76.02	22.65	12.23	-2.82	0.03	106
Aggregate	71.49	25.46	6.50	-8.56	0.02	364
<b>C1 NFE Girls</b>						
Banadir	75.63	17.92	-1.00	N/A	0.63	187
Jubaland	53.17	38.15	-19.31	N/A	0.01	83

<sup>36</sup> \* indicates that the p-value changes when running the analysis with weights, such that the coefficient (Change relative to BL) becomes significant/insignificant at the 90% confidence level.



South West State	75.91	21.29	1.78	N/A	0.46	96
Aggregate	70.61	26.42	-4.42	N/A	0.06	366

TABLE 16: AGGREGATE LITERACY SCORE CHANGES FOR THE PANEL SAMPLE<sup>37</sup>

Region	ML Mean Score	ML Standard Deviation	Change from BL	Change Relative to Benchmark	P-value	Number of Observations
<b>Formal School Girls</b>						
Banadir	66.42	31.89	32.56	13.98	0.00	142
Jubaland	70.06	30.92	34.40	15.90	0.00	119
South West State	48.18	33.83	17.56	-4.44	0.00	71
Aggregate	63.83	32.95	30.01	10.73	0.00	332
<b>ABE Girls</b>						
Banadir	64.60	33.52	21.67	-1.39	0.00	175
Jubaland	57.06	36.37	15.18	-7.88	0.17	83
South West State	62.09	31.33	21.61	-1.45	0.00	106
Aggregate	62.15	33.62	20.18	-2.88	0.00	364
<b>C1 NFE Girls</b>						
Banadir	69.98	30.19	8.27	N/A	0.10*	187
Jubaland	52.78	39.33	-1.88	N/A	0.76	83
South West State	57.69	31.99	5.83	N/A	0.14	96
Aggregate	62.85	33.68	5.33	N/A	0.07	366

Two possible covariates that make logical sense, where Jubaland ranks highest among the states, is the extent to which girls spend time working, either in agriculture to support the household, or in another setting, such as the household's small/micro business. Spending more time assisting the household with economic activities takes time and energy – both of which girls could otherwise be investing into their learning. In particular, at ML, Jubaland had the highest proportion of panel girls who claimed to contribute to these household economic activities, and had the sharpest increase in these proportions between BL and ML, as indicated in the graph below. This suggests that in the intervening period, there has been a comparatively greater need for more hands supporting household subsistence or income-generating activities, possibly due to ongoing droughts or other economic challenges faced by households, leading girls to spend less time on their studies.<sup>38</sup> Indeed, when asked specifically about how COVID-19

<sup>37</sup> See above.

<sup>38</sup> It is important to note that Jubaland respondents were not necessarily more likely to claim that these activities prevented girls from enrolling in school. However, the time spent on these other activities may have other effects on learning beyond enrolment, including time spent on homework or reviewing materials, energy levels while attending classes, and so forth.

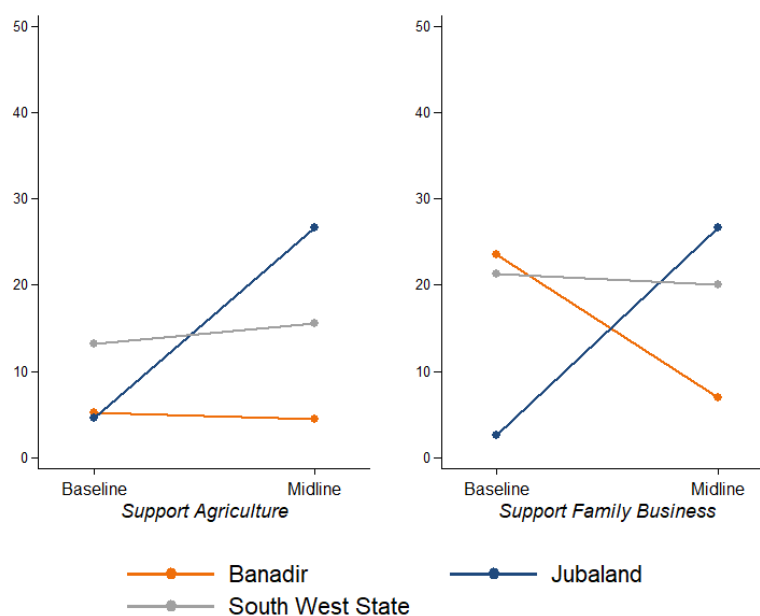
has impacted their households, girls from Jubaland were the least likely to say that the pandemic had affected their household's economic activities (such as lost jobs, an inability to do business, losing customers, etc), with only 18.2% of Jubaland girls claiming as much, compared to 30.8% of South West State girls and 41% of Banadir girls. What this suggests is that in Jubaland, closures of businesses may not have been as strictly enforced as the other states, meaning that girls who are not physically attending school may be asked to spend more time supporting the businesses/household economic activities that were still ongoing.<sup>39</sup>

In general, while Somalia in general has been confronted with myriad crises in the last two years, a confluence of other factors may also make Jubaland especially difficult for household economies and girls' learning. Like many other areas in Somalia, the Jubaland regions covered in this study are experiencing severe droughts and environmental shocks, both of which would curtail girls' ability to focus on their education in schools. At present, one of Jubaland's regions (Lower Juba) has the highest rate of severe household hunger in our sample, and protein-poor diets are also common, according to the midline data.

---

<sup>39</sup> It should also be noted that Jubaland girls were the girls who claimed that they studied the most hours a day during the COVID-19 pandemic than girls from other states, while Jubaland girls were also the least likely to claim that their overall workload has increased as a result of the pandemic. However, our questions did not adequately capture whether a girl's time spent studying had decreased during the pandemic. A big decrease in studying time, for example, might offset an increase in time spent supporting household income generating activities, such that a girl's overall workload still decreases during the pandemic. Similarly, while Jubaland girls claimed to spend the more time studying compared to girls from other states, this figure might still be a decrease from the time they spend prior to the COVID-19 pandemic; if true, than this might still result in a slower rates of learning outcome growths or even decreases between BL and ML. In fact, very high amounts of time spent studying might also be indicative of poor quality studying, as distractions or other issues might prolong the time a girl needs to spend before understanding course materials. This would possibly lead to comparatively poorer learning outcomes. More granular data would be needed to further test the above hypotheses in a more rigorous manner.

FIGURE 3: PERCENT OF GIRLS SUPPORTING FAMILY'S AGRICULTURAL OR OTHER ECONOMIC ACTIVITIES



## 4.2 Subtask-Specific Gains in Learning

Having discussed the changes in aggregate scores in the previous sub-section, here we disaggregate both numeracy and literacy scores into their constituent subtasks, and discern trends in changes in girls' mastery of subtask specific skills. Here, we focus on changes in girls' scores in each of the subtasks, while we leave an updated analysis of foundational learning gaps and grade level competencies, as defined under the FGS curriculum, in the annex.

### Numeracy Subtasks

When analyzing numeracy subtasks, the analysis necessarily follows two steps. First, we examine the changes in subtasks scores for the subtasks that were administered at both BL and ML, namely, subtasks 1 to 8. Because these were present in the assessment tools at both rounds, we are able to measure changes over time. Second, because the ML assessment included three new subtasks – multiplication, division, and word problems involving multiplication and division – no changes have occurred yet to discern skills gained (or lack thereof) in these specific areas. These additional subtasks were included to help avert potential ceiling effect problems, as BL numeracy scores were already relatively high, but will not be used in the present ML study report as part of the score calculations. However, because they will be used at the endline report, the analysis in this section also contains a brief excursion to analyze results from subtasks 9 to 11.

TABLE 17: CHANGES IN NUMERACY SUBTASK SCORES FOR THE PANEL SAMPLE (P-VALUES IN PARENTHESES)<sup>40</sup>

Subtask Number	Description	FE Girls	ABE Girls	C1 NFE Girls	All Girls
----------------	-------------	----------	-----------	--------------	-----------

<sup>40</sup> \* indicates that the coefficient changed significance when we run the analysis using weights, with significance defined at the 90% confidence level.

1	# Identification	8.67 (.00)	1.75 (.517)	-4.77 (.193)	1.67 (.342)
2	Quant. Discrimination	10.39 (.00)	4.89 (.102)	-4.7 (.216)	3.31 (.077)
3	Missing #	10.29 (.002)	10.99 (.002)	.16 (.963)	7.04 (.00)
4	Addition (1 Digit)	11.27 (.00)	4.09 (.245)	-4.75 (.155)	3.29 (.091)*
5	Addition (2 Digit)	28.13 (.00)	11.04 (.013)	-4.59 (.185)	11 (.00)
6	Subtraction (1 Digit)	18.04 (.00)	4.34 (.298)	-6.26 (.036)	4.97 (.028)
7	Subtraction (2 Digit)	26.3 (.00)	12.28 (.013)	-5.08 (.18)	10.68 (.00)
8	Word Problem	15.44 (.00)	2.61 (.535)	-5.4 (.14)	3.86 (.108)*

The above table highlights several noteworthy findings. First, and reflecting the analysis made in the previous section on the differences between the 8 subtask and 11 subtask grading schemes, both FE and ABE girls experienced increases in their numeracy scores, across all of the original subtasks. Increases in subtasks are relatively small at the earlier subtasks, which is sensible given that in general girls scored highly on the easier subtasks at baseline, thus leaving relatively little room for improvement from BL to ML. Positive changes in subtasks scores increased between subtasks 3 and 7, with FE girls' biggest gains made in subtask 5 (multi-digit addition) at 28.13 points, while ABE girls made their biggest gains in subtask 7 (multi-digit subtraction) at 12.28 points. Gains then taper off at subtask 8 (word problems): in theory, this may be because at this level of difficulty, skills gains are likely to be smaller in magnitude as the effects of question difficulty outstrip room for "catch up growth" in skills. In practice, however, we observe that subtask 8 scores are in fact higher, at both BL and ML, compared to simpler addition and subtraction questions. This might suggest room for improvement on subtask 8 has also been more exhausted compared to middle subtasks, as students already found subtask 8 to be easier. More broadly, this suggests that girls may be more familiar with these skills in the context of practical application more than in the context of abstract calculation.

Conversely, C1 NFE girls experienced decreases in all but one subtasks, with the magnitude of decrease roughly similar (within 2 points of each other), suggesting that the overall numeracy learning quality of C1 NFE of girls was comparatively poor between BL and ML. Once again, possible explanations for this include the aforementioned suggestion in the previous sub-section that C1 NFE girls, being older on average, may be subjected to additional social expectations and stigma that younger girls face less.

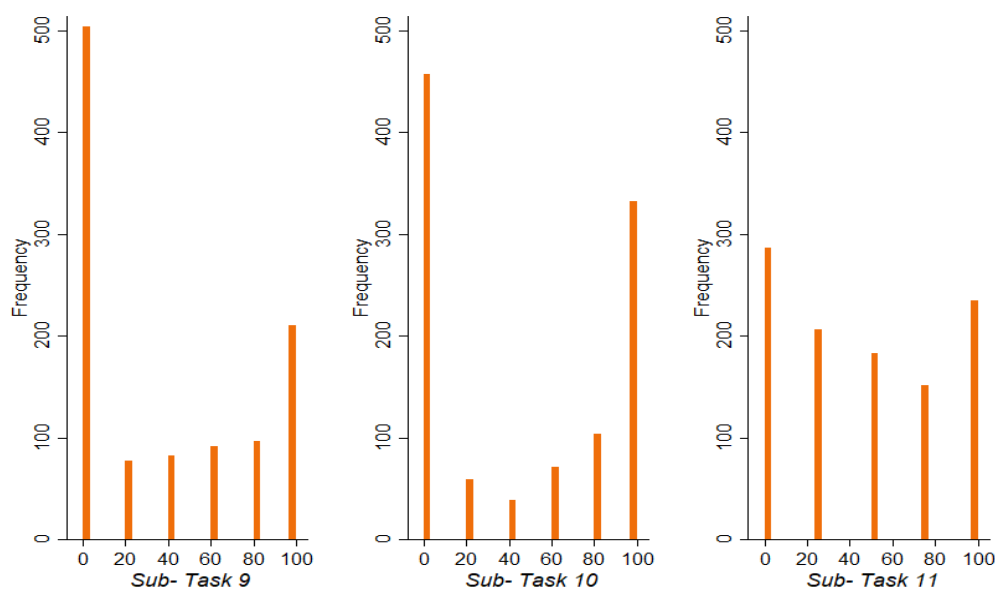
**TABLE 18: PANEL SAMPLE SCORES FOR NEW NUMERACY SUBTASKS AT ML**

Subtask Number	Description	FE Girls	ABE Girls	C1 NFE Girls	All Girls
9	Multiplication	40.42	39.12	31.26	36.82
10	Division	49.64	46.26	41.53	45.69

<b>11</b>	Word Problem (Multiplication and Division)	49.70	45.54	43.85	46.26
<b>Overall BL Score</b>	8 Subtask Grading Scheme	57.83	64.99	75.03	66.21

Interestingly, for multiplication and division problems, there is a trend where girls' scores across all cohorts are higher in subtask 11 than in subtask 9, despite the supposed increase in difficulty; for two of the three cohorts, subtask 11 was their highest score among the three new subtasks. This is also observed in the below histograms, where subtasks 9 to 10 clearly follow a bimodal distribution of scores, whereas subtask 11 is more akin to a uniform distribution. This might suggest that while, in theory, word problems involving multiplication and division are more difficult than problems where a clear procedure for multiplication and division is already provided by the assessor, in practice, girls may be more familiar with these skills in the context of practical application, rather than abstract calculation. Indeed, this finding dovetails with the finding that girls also score more highly on word problems with addition and subtraction than in the abstract calculations. Quite possibly then, these concepts are better learnt through tangible real-life examples, though we again stress that the scores for subtasks 9 – 11 remain much lower than girls' overall numeracy scores.<sup>41</sup>

**FIGURE 4: DISTRIBUTION OF SCORES FOR NUMERACY SUBTASKS 9 THROUGH 11**



## Literacy Subtasks

Similar to the numeracy subtasks, FE and ABE girls experienced increases in their literacy scores; however, the magnitude of increase is far greater than in the numeracy subtasks. As an illustrative example: where FE girls increased their numeracy subtask 1 score by 8.67 points, they increased their

literacy subtask scores by 21.14 points, which is a thrice the increase in the comparable numeracy subtask. This is likely due to the fact that girls of all cohorts had substantially lower literacy scores at baseline compared to numeracy score, which leaves ample room for “catch up” and growth even among the easiest literacy subtasks.

Where the literacy subtask scores differ from the numeracy subtasks scores is in regards to C1 NFE girls’ performance. Unlike in the numeracy assessment, we see an increase in all but one of the literacy subtask scores. This means that unlike in numeracy skills, C1 NFE girls have experienced a small, but tangible, increase in literacy skills since baseline. The fact that these girls have not experienced a sharper increase in literacy is, in part, a function of the short duration of the NFE learning programme and the fact that the programme ended in late 2020, over one year prior to this midline evaluation, potentially contributing to learning loss. In addition, the baseline was conducted after the C1 NFE cohort had begun their learning programme, potentially inflating their baseline scores and leading to the flatter-appearing trend in literacy skills documented earlier.

Perhaps more concerning for the program as a whole is if there is a limit to how much girls can develop their literacy skills in the current project locations: C1 NFE girls, who already had higher literacy skills at BL, may have reached that limit earlier, while their FE and ABE peers, who had lower scores, had room between BL and ML to catch up. If true, than it is possible that ABE and FE girls’ literacy skills gains will stagnate between ML and endline. The fact that our benchmark scores for literacy gains between grades 1 and 2 and between grades 2 and 3 (see sub-section 4.1) are still positive<sup>42</sup> and sizable suggests that there is still room for growth, and that C1 NFE girls’ slower growth could be a symptom of either their uniquely challenging status as older learners or a function of their more limited exposure to learning interventions and the fact that the interventions had wound down long before the midline was conducted. Nonetheless, from the perspective of program adaptation, it is recommended that program experts conduct more regular mini-assessments of literacy in between ML and endline in order to confirm that literacy skills development for all cohorts has not reach a plateau, and, if it has, that proper remedial actions are taken in a timely manner.

**TABLE 19: CHANGES IN LITERACY SUBTASK SCORES FOR THE PANEL SAMPLE (BASELINE VALUES IN PARENTHESES)**

Subtask Number	Description	Formal Education	Accelerated Basic Education	C1 Non-Formal Education	Total
1	Letter Sound Identification	+21.1* (61.6)	+13.9** (68.0)	+2.7 (77.7)	+12.3* (69.2)
2	Words Commonly Used	+34.2* (32.9)	+23.4** (40.5)	+8.5* (56.2)	+21.6* (43.5)
3	Reading fluency	+38.2* (21.9)	+26.4* (31.5)	+12.7* (46.8)	+25.4* (33.8)
4	Reading Comprehension 1	+31.4* (31.8)	+21.1* (40.2)	+5.3 (58.7)	+18.9* (44.0)
5	Reading Comprehension 2	+32.5* (30.2)	+20.8* (39.7)	+4.4 (57.6)	+18.8* (42.9)
6	Reading Comprehension 3	+22.7* (24.6)	+15.5* (31.9)	-1.6 (48.1)	+11.9* (35.2)

\* Change from baseline to midline is statistically significant at the 5 percent level.

<sup>42</sup> The benchmark for grade 3 to 4 is also positive, though not shown in sub-section 4.1

Like numeracy subtasks, the biggest magnitude increases for literacy were gained in the middle subtasks, specifically, subtask 3. Gains increased in magnitude from subtask 1 before peaking at subtask 3, and then tapering off as the questions become more challenging. Once again, this is likely due to easier questions in which girls scored well during BL leave comparatively less room for growth, leading to smaller gains for easier questions. However, unlike in the numeracy subtasks, girls' baseline scores for more challenging subtasks in the literacy assessments were not higher than in easier subtasks. In this case, the diminishing gains in later questions is likely due to their comparatively more difficult nature, rather due to there being little room for growth.<sup>43</sup>

### 4.3 Subgroup Programme Impact

In this sub-section, we present the key findings regarding differences in *changes* in learning outcomes from baseline to midline, reporting results for a variety of sub-groups of concern to the programme, including girls with disabilities, girls from economically disadvantaged households, and girls from schools that are lacking in basic facilities.<sup>44</sup> We analyse differences among sub-groups in terms of the aggregate numeracy and literacy scores. Due to the sheer number of sub-groups by which one could disaggregate data, this sub-section primarily focuses on sub-groups who have demonstrated noteworthy differences in learning trajectories between BL and ML, including those with large differences in learning trajectories, and those for whom differences remain substantial and/or significant after introducing control variables. Moreover, for some sub-groups such as girls with disabilities, the sample sizes are miniscule: in one case, only one girl in the entire sample met the criteria to be considered having a particular disability. Therefore, we present findings based on characteristics which at least 50 girls have, in order for our quantitative analysis to have adequate meaning. A fuller table of sub-group results is appended in the annex for the reader's consideration, but caution is warranted as statistical tests are unlikely to yield meaningful results with such small samples.

Our approach to analysis is as follows. We first use a simple regression model on all the sub-group characteristics of concern to the AGES programme to test whether those who belong in the sub-group and those who do not experienced the same level of learning outcomes changes between BL and ML for both the numeracy and literacy assessments. Specifically, we use a difference-in-difference (DiD) linear regression model to measure such changes. A DiD model reveals differences in the rate of change of observed outcomes between two groups, which would tell us, for example, whether girls with a certain disability are improving their scores at a rate equivalent to girls without said disability. We use the sub-group characteristic measured at BL to run this analysis, as that is the reference period after which these characteristics would impact students' learning. Next, for the characteristics that were statistically significant without controls,<sup>45</sup> we ran a second model with two control variables in order to check for robustness: cohort and state. As noted in previous sub-sections, there is a sizable difference in learning outcomes among girls in different cohorts, while some geographic differences were observed. These

<sup>43</sup> Refer to annex for a complete breakdown of BL and ML scores for each subtask.

<sup>44</sup> Because this report constitutes a midline for the cohorts analysed in this section, we do not discuss subgroup differences in learning outcomes, but focus our attention on subgroup differences in *trends* in learning outcomes over time. Subgroup differences were analysed extensively at baseline, because one goal of the baseline report was to understand the nature of marginalization and patterns learning outcomes across different types of girls. We report equivalent information at midline in Annex 2, allowing readers to see – in the midline cross-section – the learning performance of different demographic subgroups.

<sup>45</sup> Which we define at the 90% confidence level for the purpose of exploring possible associations.

findings suggest that cohort type and location can impact learning outcomes beyond the characteristics under study in this sub-section. Therefore, a controlled difference-in-difference model was utilized to prevent the analysis from being biased by spurious relationships or collinearity between the sub-group characteristics and the above control variables.

## Disability Status

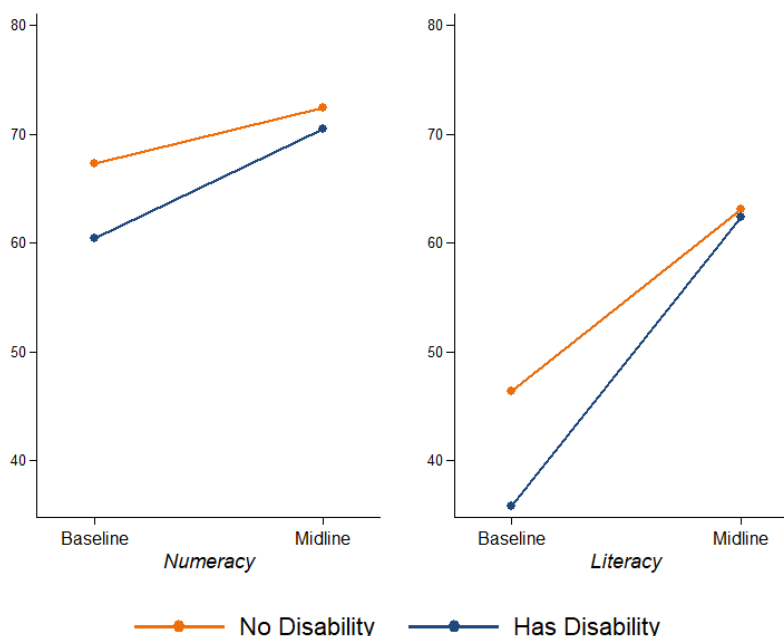
Having a disability can be one of the most serious barriers preventing girls from enrolling in school and participating actively in their programme of study. It is therefore imperative that education empowerment programmes specifically address girls with disabilities' unique needs and concerns. In the panel sample, specific categories of disabilities are rare, leading to some sub-groups having a sample size of 1; therefore, it is difficult to pinpoint the programme's effect on exact types of disabilities. However, at a more general level of categorization of disabilities, the data suggests that the programme holds promise in levelling the playing field, as girls noted to have a mental health disability specifically, or any disability in general, had a better learning trajectory in numeracy and literacy scores compared to girls who were not in those sub-groups. Specifically, girls with a disability increased their numeracy scores, on average, by 10.01 points, which is 4.82 higher than girls without any disabilities, while they also increased their literacy scores by 26.54 points, which is 9.85 points higher than the increase experienced by girls without disabilities. The literacy figure is significant at the 95% confidence level. Similarly, girls with mental health disabilities experienced a 11.43 point increase in their numeracy scores - 6.65 higher than the change experienced by girls without mental health disabilities. They also increased their literacy scores by 27.87 - 11.52 higher than girls without mental health disabilities.<sup>46</sup> These figures are significant at the 95% confidence level.

---

<sup>46</sup> These numbers were derived using a more flexible criterion for when a girl is considered to have disabilities: whether the girl herself or her caretaker claimed that she had a (mental health) disability. These numbers were also derived using the DiD model without controls, though as can be seen in the Annex, the difference in differences is largely the same even after control variables were added.



FIGURE 5: CHANGES IN ASSESSMENT SCORES BY DISABILITY STATUS



The increase in scores between BL and ML for girls with mental health disabilities or girls with any disabilities was smaller for the numeracy assessment than for the literacy assessment, which is again in line with the earlier findings demonstrating the far larger increase in literacy skills compared to numeracy skills. Nonetheless, that girls with disabilities are experiencing a greater rate of positive change for both numeracy and literacy skills, compared to their peers without disabilities, is a positive indication. Indeed, it is not entirely surprising given that the programme has emphasised improving girls with disabilities' learning.<sup>47</sup> Methods which teachers used when teaching these sub-populations of girls include moving them to the front row if they have hearing or vision impairments<sup>48</sup>, providing private specialized instruction,<sup>49</sup> and general awareness raising to parents and other students to de-stigmatize disabilities.<sup>50</sup>

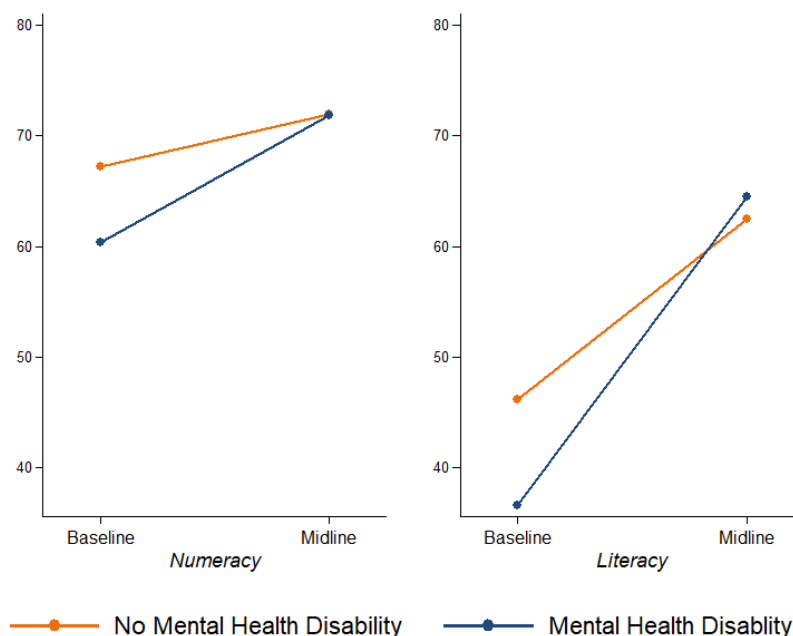
<sup>47</sup> We find that there is a statistically significant difference in BL numeracy scores between girls with a mental health or any disability who did and who did not stay in the sample, which might suggest that there is differential attrition. However, these girls who dropped from the sample had higher BL numeracy scores than those who did not stay, possibly implying that girls with disabilities who dropped from the sample might be, *ceteris paribus*, on a better learning track than girls who stayed in the sample. This suggests that the direction of bias is downwards, and that our current results might in fact underestimate the change in the learning of girls with disabilities.

<sup>48</sup> FGD with Teachers in Banadir, Int. 507

<sup>49</sup> FGD with Teachers in Jubaland, Int. 509

<sup>50</sup> FGD with Teachers in South West, Int. 501

FIGURE 6: CHANGES IN ASSESSMENT SCORES BY MENTAL HEALTH STATUS



This finding is in line with our analysis of aggregate learning outcomes, where we found that girls who scored lower at BL had much steeper increases in learning outcomes between BL and ML, to the point of catching up with, or even surpassing, their higher-performing peers. The programme's interventions likely helped removed some of the most pressing challenges limiting girls' learning; once removed or mitigated via other measures, the lowest performing girls' most affected by these barriers experience rapid growths. In addition, it is also possible that disability status, or low academic performance at BL, signalled to teachers and programme implementers that particular girls need additional attention or resources. In turn, these factors allowed them to catch up to their peers, who were initially less-affected by barriers such as disability, and therefore enjoyed comparatively fewer marginal benefits from the programme's various initiatives.

Even so, several teachers claimed that nothing is being done to accommodate students with disabilities, clearly highlighting the need for continued interventions and disability-sensitive programming, despite the positive progress.

### Household and Demographic Characteristics

A number of household and demographic characteristics were found to have significant effects on changes in girls' learning outcomes. In general, these characteristics have coefficients that are expected. Several demographic characteristics that are expected to hinder a girl's ability to learn are shown in the data to have led programme girls to underperform relative to other girls. For instance, in the case of girls who are or were married, the change in their numeracy scores from BL to ML was negative; moreover, it was also more negative than their peers who have never been married. While girls who have been married do demonstrate an increase in their literacy scores of 5.51 points, this is considerably lower than their unmarried peers whose rate of improvement was 13.53 points greater. The discrepancy highlights the comparative inability of girls who have been married to take advantage of the learning resources around them through the programme, compared to their unmarried peers. As indicated in the table below, a similar trend is observed for girls who are currently mothers. Indeed, these results are unsurprising, as

childcare or filial responsibilities can exact a toll on girls' energy and time, leading to poorer learning.<sup>51</sup> In addition, per the previous sub-sections, there appears to be stigma inside of schools against older girls learning, with one teacher claiming that married girls (who skew older) will need separate classes:

*There are girls who are already married and uneducated who come to our school to study, so we make special classes for them so that they do not interfere with the young students in the class because [married girls] are shy to study with them in the same classes.*

-FGD with Teachers, Banadir, Int.508

**TABLE 20: LEARNING OUTCOME DIFFERENCES AMONG SELECTED SUB-GROUPS<sup>52</sup>**

Subgroup	Numeracy			Literacy			Sample Size
	Change from BL	DiD	P-Value	Change from BL	DiD	P-Value	#
Full Sample	5.73	N/A	N/A	18.1	N/A	N/A	1062
Girl is or was married	-5.62	-12.16	0.00	5.51	-13.53	0.01	71
Gone without clean water most days	16.70	11.63	0.02	27.14	9.55	0.15	60
Caregiver has no formal education	4.95	-5.35	0.09	16.90	-8.57	0.00	909
HoH has no formal education	4.90	-4.48	0.07	16.44	-9.23	0.00	867
Girl is a mother	-4.72	-11.01	0.03	5.48	-13.33*	0.04	54
Girl has only one living parent	0.97	-5.46	0.07	15.32	-3.24	0.37	138
Girl owns phone	6.52	5.77*	0.11	19.52	10.09	0.02	916

<sup>51</sup> FGD with Teachers, Middle Shabelle, Int. 512

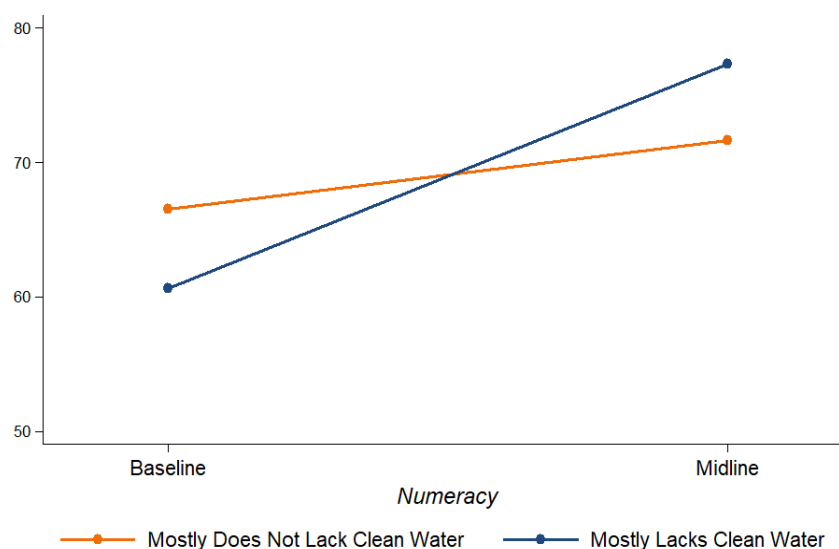
<sup>52</sup> All figures presented are from the model without controls. However, for these variables, if they are found to be significant without controls, the difference-in-difference coefficient remained very similar, if not the same, even with the introduction of controls, suggesting that these variables are good predictors of girls' learning trajectories. The coefficients for the main effects of these variables, however, were different in the model with controls. A \* indicates that the DiD coefficient changes from significance to insignificance, or vice versa, when we re-run the model with weights, with the cutoff at the 90% confidence level.

Despite this, it is worth noting that qualitative respondents indicate that, under normal circumstances, married girls will simply stop learning and drop from school.<sup>53</sup> Without a proper comparison group consisting of married girls who did not enrol in the AGES programme, it is not possible to draw broader conclusions about how well married girls would normally fare. That these girls remain in the programme and are still making some, albeit small, progress on literacy may be one achievement of the programme itself.

In addition to girls' marital and parental status, the education of the people in close proximity to them also has an impact on their learning. Understandably, girls who lack adult figures who are able to mentor or tutor them on their studies miss out on a potentially important source of support for learning. For both the literacy and numeracy measures, girls whose head of household or caregiver did not receive any formal education experienced still experienced a positive change in their scores; however, the magnitude of these increases is considerably smaller than that of girls with formally educated adults as caretakers or household heads. A less direct proxy for having adult figures who could mentor girls and provide the necessary resources to learn is whether a girl has lost one or both biological parents. While we find no significant effects for when a girl loses both parents, girls who have lost a single parent only experienced a very small growth in their numeracy skills, thus lagging behind their peers.

Finally, we also find that girls who own a phone experienced a greater increase in their literacy scores compared to their peers, though the direction of causality is ambiguous. Girls who own phones are likely exposed to a greater volume of media which require reading, thus potentially improving their numeracy skills; concurrently, it is also possible that girls who are already good at reading and who are comparatively quick learners are likelier to own phones. Regardless, the data suggests that phone ownership is a reasonable proxy for potential success in literacy learning.

**FIGURE 7: LEARNING OUTCOME CHANGES BY GIRLS' ACCESS TO WATER**



One counterintuitive finding, however, is that going without clean water in the household is associated with a positive increase in numeracy scores, and that this increase is greater in magnitude compared to peers who do not experience shortages of clean water as frequently. One possibility might be differential treatment intensity, in that schools and teachers will focus more attention in helping girls whom they

<sup>53</sup> FGD with Teachers, Lower Shabelle, Int.511; FGD with Teachers, Banadir, Int.504;

know are vulnerable to economic and other shocks.<sup>54</sup> The extra assistance may help these disadvantaged girls quickly catch-up with their peers. Indeed, though our qualitative data do not indicate any special initiatives around providing girls with water, it does suggest that some schools and education institutions are concerned more generally about staff and student health, with local CECs paying for water and utilities for school staff, especially in light of recent droughts.<sup>55</sup> Even in the absence of differential treatment intensity, particularly disadvantaged girls at BL might enjoy greater marginal benefits from the same program compared to their peers, which would explain their steep upward trajectory in learning changes. This is suggested in the above graph: consistent with conventional wisdom, girls who had at baseline claimed that their households lack clean water had lower numeracy scores compared to their peers, but were eventually able to catch up and surpass them.

While this may generally be a positive indication of the programme's ability to empower the learning of especially disadvantaged girls, cautioned should be exercised, as this is a relatively small sample of girls who were without clean water. With ongoing droughts in Somalia, a higher number of girls might potentially experience a lack of clean water; the programme's capacity to help these additional girls learn in spite of difficult household economy circumstances, either through directly providing basic WASH support at school or by using alternative teaching methods, might be stretched thinly in the coming periods.

**TABLE 21: LEARNING OUTCOME CHANGES BY MARGINALITY<sup>56</sup>**

Subgroup	Numeracy			Literacy			Sample Size
	Change from BL	DiD	P-Value	Change from BL	DiD	P-Value	#

<sup>54</sup> Another possible explanation is differential attrition among girls whose households lack water for most days in the past 12 months. If the girls who fit this category and who in turn dropped from the panel sample were systematically different from the girls who fit this category and who remained in the sample, then the results from the sample might be biased. For example, if the girls who went without clean water most days are still in the sample/in school, it might suggest that they are among the most resilient girls who are eager to learn in spite of the challenges their households face. If true, then the positive coefficient in this model has less to do with the effect of having no water on learning outcomes changes, but is more a proxy for a girl's own resilience and determination to learn. However, we find that at BL, girls who mostly lack water and dropped from the panel and girls who mostly lack water and remained in the panel had no statistically significant differences in test scores at BL. This suggests that both groups of girls were equally likely, *ceteris paribus*, to be on similar learning trajectories, though this additional robustness check still has its limits, and caution is still warranted in interpreting the relationship between lacking water and learning outcome changes.

<sup>55</sup> FGD with Teachers, Lower Juba, Int.510

<sup>56</sup> The values in this table are calculated using a regression model without controls. A \* indicates that the DiD coefficient changes from significance to insignificance, or vice-versa, when we run the model with weights. The cutoff is defined at the 90% confidence level.

<b>Full Sample</b>	5.73	N/A	N/A	18.1	N/A	N/A	1062
Bantu and minorities	1.84	-4.76	0.13	14.84	-4.04	0.26	196
Girl speaks minority language relative to home area	-7.27	-13.87	0.07	23.76	6.01	0.37	67
Girl speaks af-Maay	3.61	-2.89	0.35	16.00	-2.92	0.40	284
English is the language of instruction	19.33	3.90	0.39	22.24	-7.90	0.17	52

That lower performers at BL, do not see a steeper rise in their learning outcomes is indicative of the programme's various initiatives to improve learning not adequately compensating for the major and specific challenges which a lower status creates for girls' learning, such as poverty or lack of access to social services outside the scope of the programme's foci. Teachers and school administrators do not generally highlight minority status as a major consideration when attempting to help disadvantaged girls. As one teacher described it:

*There is no discrimination on minority groups here, the school doesn't separate students by their clans, even as teachers we don't know their tribes, and the students are really friends and we do not see anyone discriminated against.*

-FGD with Teachers, Banadir, Int. 504

Another teacher similarly emphasised that "the school does not differentiate students in terms of clans, and all students are the same."<sup>57</sup>

Though the data suggests that the programme may have helped prevent girls from secondary/minority groups from lagging too far behind their peers, they have not as yet caught up with their peers from more dominant groups.

In addition, we find that girls who speak a language that is different than the dominant local language experience a decline in their numeracy scores, whilst on average those who spoke the local language experienced an increase. English as the language of instruction is not a significant predictor of numeracy learning, but the coefficients are large and positive for numeracy skills. However, this is likely because only FE girls were instructed in English the sample, and FE girls in general have been noted to outperform ABE and C1 NFE girls. English as the language of instruction is also not significantly related to literacy scores, though the coefficient being negative is hardly surprising, as the literacy tests were administered in Somali, which these specific girls may have had less practice within a formal education setting.

## School Resources and Infrastructure

Finally, we turn to a discussion of girls' differential access to school resources can impact their learning. Most of the measures highlighted in this sub-section are from surveys done in FE schools for girls, meaning

<sup>57</sup> FGD with Teachers, Banadir, Int. 505.

that caution is warranted in making broader claims, as the sub-sample may not adequately cover ABE and NFE schools. Nonetheless, for most of these measures, the effect on learning is as expected.

**TABLE 22: LEARNING OUTCOME DIFFERENCES AMONG GIRLS WITH DIFFERENT SCHOOL INFRASTRUCTURE<sup>58</sup>**

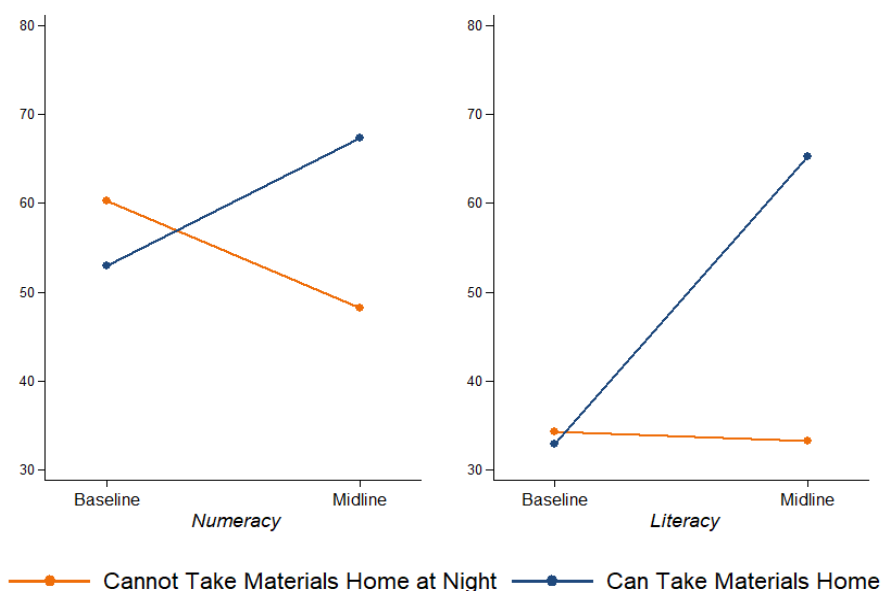
Subgroup	Numeracy			Literacy			Sample Size
	Change from BL	DiD	P-Value	Change from BL	DiD	P-Value	#
Full Sample	5.73	N/A	N/A	18.1	N/A	N/A	1062
Materials can be taken home at night	21.65	23.68	0.00	32.42	33.44	0.00	111
Not enough seats in class for student	-3.58	-10.84	0.04	13.90	-4.93	0.31	151
School has water source within 1km	20.26	9.88	0.05	38.81	21.53	0.01	181
School only has cement floor (no dirt)	18.77	8.56*	0.07	32.77	8.25	0.25	227
Fewer than 5 hours of instruction a day	4.05	-16.52	0.00	14.79	-21.02	0.02	95
Teachers often absent	10.34	6.38	0.02	22.51	6.05*	0.07	293
Textbooks are shared between students	16.53	3.18	0.49	34.62	12.56	0.09	148

It is immediately clear that being able to take home school materials at night is a substantial and significant predictor of girls' better performance on both the numeracy and literacy assessments. The table below demonstrates the magnitude of these differences. For each assessment type, the difference-in-difference coefficient is the largest among the school infrastructure measures in the table. In fact, among all subgroup characteristics in this study that have a sample size of greater than 1, being able to take home materials at night is the single largest DiD coefficient, indicating that this ability is one of the strongest ways to give girls an advantage in learning, as girls who are able to take materials home at night were

<sup>58</sup> Numbers presented are from the DiD model without controls. A \* indicates that the DiD coefficient changes from significance to insignificance, or vice-versa, when we run the model with weights. The cutoff is defined at the 90% confidence level. The use of weights adds two new characteristics that negatively influence a girl's numeracy skills: girls who do not use drinking water facilities at school, and girls who do not use toilets at school. Girls who were in the former group had a 0.94 point decline in their numeracy scores from BL to ML, which is change that is 8.52 lower than their peers who do use drinking water facilities. Girls who were in the latter group also had a decline in numeracy scores (1.77) points, which is a change that is 8.94 lower than their peers.

experience much larger increases in learning outcome scores between BL and ML than girls who were not able to do so. This is logical: not every person finds the traditional classroom setting most conducive for learning, perhaps doubly so for Somali girls who may have household responsibilities or, as indicated in previous sub-sections, may face stigma for being older learners in an early grade level.<sup>59</sup> For these girls in particular, the relative flexibility in managing their time, learning environment, and energy afforded by being able to take home school materials may be a significant catalyst for improved learning.

**FIGURE 8: LEARNING OUTCOME CHANGES BY AVAILABILITY OF TAKEHOME MATERIALS**



The physical infrastructure of the school and its surroundings also impacts learning. Having water sources within 1km is also associated with better learning, as schools that are better equipped to handle basic student needs are clearly better suited to creating a conducive learning environment. Schools with only cement floors that do not contain dirt is also associated with large increases in both numeracy and literacy learning, and that these gains were over 8 points greater than the gains made by girls in schools with mixed flooring. On a related note, having an inadequate number of seats in the classroom is associated with poorer learning outcomes, as students may need to stand, share seats with other students, or not show up to class altogether. Taken together, these factors point to the importance of students' physical comfort in shaping their learning outcomes.

And finally, it comes as little surprise that having fewer hours per day of instruction disadvantages learning; though their literacy still improved, the rate of development is still considerably lower than peers who had at least 5 hours of instruction a day. Teacher absenteeism should be expected to hurt girls' learning by depriving girls of learning time, yet the data indicates that girls who said their teachers were frequently absent had a more positive learning trajectory than their peers whose teachers were not as frequently absent. One plausible explanation, like many supposedly negative characteristics that should

<sup>59</sup> Of the variables presented in the table, taking materials home at night is the only one that captures the experiences of ABE and C1 NFE girls in addition to FE girls. Indeed, for older girls, the ability to be flexible in their learning environment might be even more critical.



be expected to hurt learning, is the aforementioned differential treatment intensity, whereby the programme focuses more resources in improving schools which face the biggest challenges, such as teacher absenteeism.<sup>60</sup> To the extent that this is true, the coefficient of teacher absenteeism may be positively associated with stronger learning precisely because teacher absenteeism invites stronger interventions by programme implementers, thereby putting students previously affected on a steeper upwards trajectory.

Finally, and perhaps somewhat counter-intuitively, sharing textbooks with other students is also positively associated with greater improvements in learning. In theory, we might expect that a mode of studying where girls are reliant on other students for study materials to decrease learning. However, if the default is no access to books whatsoever due to an inability to afford them or a lack of them at schools, than sharing textbooks may be an improvement in access to learning materials. Though not as impactful as being able to take home materials at night to study on their own time, girls' access to shared books may provide increased opportunities to learn, compared to the alternative. The other alternative possible explanation, in line with the above suggestion of differential treatment intensity, is that having to share books at BL is a signal of being especially disadvantaged, thus inviting programme implementers to pay additional attention to girls who fall under this category.

## 4.4 Testing the Theory of Change

According to the theory of change, the outputs of the programme are hypothesized to have a positive impact on learning outcomes via a series of intermediate outcomes. In regards to numeracy and literacy, four intermediate outcomes connect programme outputs with expected increases in learning outcomes: increased attendance in class, acquisition of life skills (namely leadership and negotiation skills), improved quality of teaching, and finally, more pro-education community attitudes, namely among girls' caregivers. A later section in this report will highlight changes in the intermediate outcomes from BL to ML, and what the intermediate outcomes are at ML. This sub-section instead focused on assessing the extent to which the intermediate outcomes can be used to anticipate girls' learning trajectories by examining the relationship between each of the intermediate outcomes at BL and their relationship to the panel sample's changes in learning outcomes through to ML.

### Attendance

We first examine whether attending classes more often at BL, which is an important prerequisite for learning, leads to positive changes and increased rates of change in learning outcomes between BL and ML. To measure attendance, we use data from the BL headcount survey, where the research team visited several classes or streams in all school types, counted the number of students physically present in class, and matched those students against the class' enrolment record, where enrolment records are available. The attendance rate for each school and each cohort type is the percentage of girls counted in the headcount who were also listed in the enrolment records, averaged across the classes/streams in each

---

<sup>60</sup> Indeed, our measure for teacher absenteeism, as with all variables in this analysis, is lagged, meaning that we measure associations between student outcomes by midline to whether or not they belonged in certain sub-groups at baseline.

school-cohort type. We use the attendance rate for each school-cohort type as a proxy for the attendance rate of the girls who attended that school-cohort type.<sup>61</sup>

As noted in the below table, attendance rates were largely did not have a statistically significant effect on either the scores at ML or the change in scores from BL to ML when we controlled for cohort type, age, and state. In one case, when measuring the effect which attendance rates have on the BL – ML scores, we find that an additional 1 percentage point increase in attendance rate leads to a 1.22 point decrease in the change in literacy scores from BL to ML, meaning that girls with higher attendance rates experienced slower growths in learning outcome from BL to ML. However, when we in turn control for BL scores, the effect of attendance was no longer significant, and BL scores were a stronger predictor of levels of change over time. This is likely because, as with many other characteristics discussed in previous subsections, attendance rates at BL are good predictors of scores at BL, which in turn is a good predictor of levels of change. As noted, girls scoring higher at BL tend to register lower rates of increases in scores from BL to ML, likely because the effect of the programme has a larger marginal effect on BL low performers, which in turn makes the former's their learning outcome increases sharper.

**TABLE 23: EFFECT OF 1 POINT INCREASE IN BL ATTENDANCE RATE ON CHANGES IN LEARNING OUTCOMES<sup>62</sup>**

	Effect on Score Change, Numeracy	P-Value	Effect on Score Change, Literacy	P-Value
<b>BL score not controlled</b>	-0.09	0.90	-1.22	0.03
<b>BL score controlled</b>	0.50	0.36	-0.09	0.87

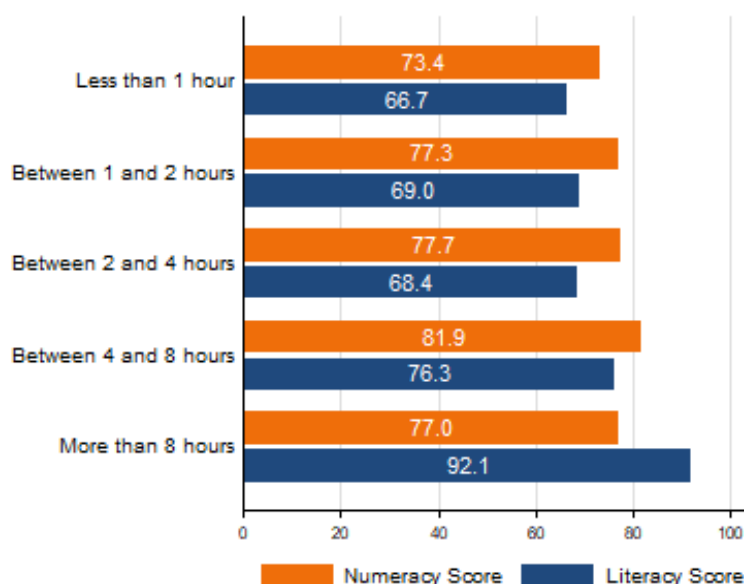
It is critical to note, however, that the BL to ML period coincided with the COVID-19 pandemic in Somalia, which resulted in considerable school closures throughout the country. As such, BL attendance data as described above are unlikely to be the best measure of learning, as many girls would not have had classes in schools. To account for this, we used a second set of measures to examine learning: whether a girl was able to study at home during the COVID-19 pandemic, and among those who could, the average number of hours they spend per day to study. Overall, after controlling for a number of covariates, the findings are consistent with the theory. For the first measure, we find that girls who claimed they were able to study at home during school closures experienced a larger positive increase in their numeracy and literacy scores compared to girls who could not study at home, with the former's growth advantage at

<sup>61</sup> In several instances, the attendance rate far exceeds 100%, as there were many more students physically present in class than enrolled according to the records. In one case, there were 12 students attending a class that had 1 student officially registered. To prevent the regression results from being overly determined by the extreme outliers, we dropped three school-cohort types with attendance rates over 300% from the data, and the panel sample girls who are from those schools have missing attendance rates, and therefore were not included in the regression models.

<sup>62</sup> Results from regression models with additional controls for cohort type, age, and state

4.24 and 11.15 points, respectively.<sup>63</sup> Next, among those who were able to study, more hours spent studying led to progressively higher test scores, though for the numeracy assessment, studying more than 8 hours a day appeared to lead to lower scores compared to studying for 4 – 8 hours a day, either because of fatigue or because who need to study for 8 hours a day may be the girls who are academically not as strong as those who need fewer hours.<sup>64</sup>

**FIGURE 9: ML NUMERACY AND LITERACY SCORES BASED ON NUMBER OF HOURS STUDIED AT HOME**



Though the results are as expected, caution should be exercised in interpreting these results on their own. Physical attendance in class may not necessarily equate to actual learning. Much depends on other factors, such as the student's (mental) health, which might affect their ability to focus in class, or the teaching quality, which we discuss later in this sub-section. It is best to address these factors together holistically, rather than through standalone interventions. In addition, in our survey, the question about whether girls were able to study at home was only asked to FE and ABE girls, and not to C1 NFE girls; the inclusion of C1 NFE girls may lead to different results. Thus, these results, on their own, should be seen as suggestive but not conclusive evidence of an accurate theory of change.

## Youth Leadership Index

The Youth Leadership Index (YLI) is a composite index on a scale of 0 to 100, measuring girls' self-perceived leadership skills in the context of school. It is scored based on 21 questions asking girls whether they felt they exhibit a number of characteristics that are indicative of higher leadership skills, including

<sup>63</sup> Significant at the 99% confidence interval for the literacy assessment, but not significant for the numeracy assessment.

<sup>64</sup> In the regression model, those who studied more than 8 hours a day scored 20.45 points higher on the literacy test than those who studied less than an hour, while those who studied between 4 and 8 hours a day on average scored 6.6 points higher in the numeracy test than those who studied for less than an hour. These were the increments that were significant at at least the 90% confidence interval. However, almost all studying coefficients were positive, meaning that studying more than an hour a day in general led to higher scores at ML.

thinking about the consequences of one's actions, being able to express one's thoughts clearly to others, and being able to organize their peers to do an activity in pursuit of common goals. To determine whether the YLI can be a predictor of larger increases in learning outcome scores, we examine the relationship between BL YLI scores of the girls in the panel sample and the change in their learning outcomes.

**TABLE 24: EFFECT OF 1 POINT INCREASE IN BL YLI SCORE AND CHANGES IN LEARNING OUTCOMES<sup>65</sup>**

	Effect on Score Change, Numeracy	P-Value	Effect on Score Change, Literacy	P-Value
<b>BL score not controlled</b>	-0.16	0.02	-0.03	0.71
<b>BL score controlled</b>	0.02	0.63	0.18	0.00

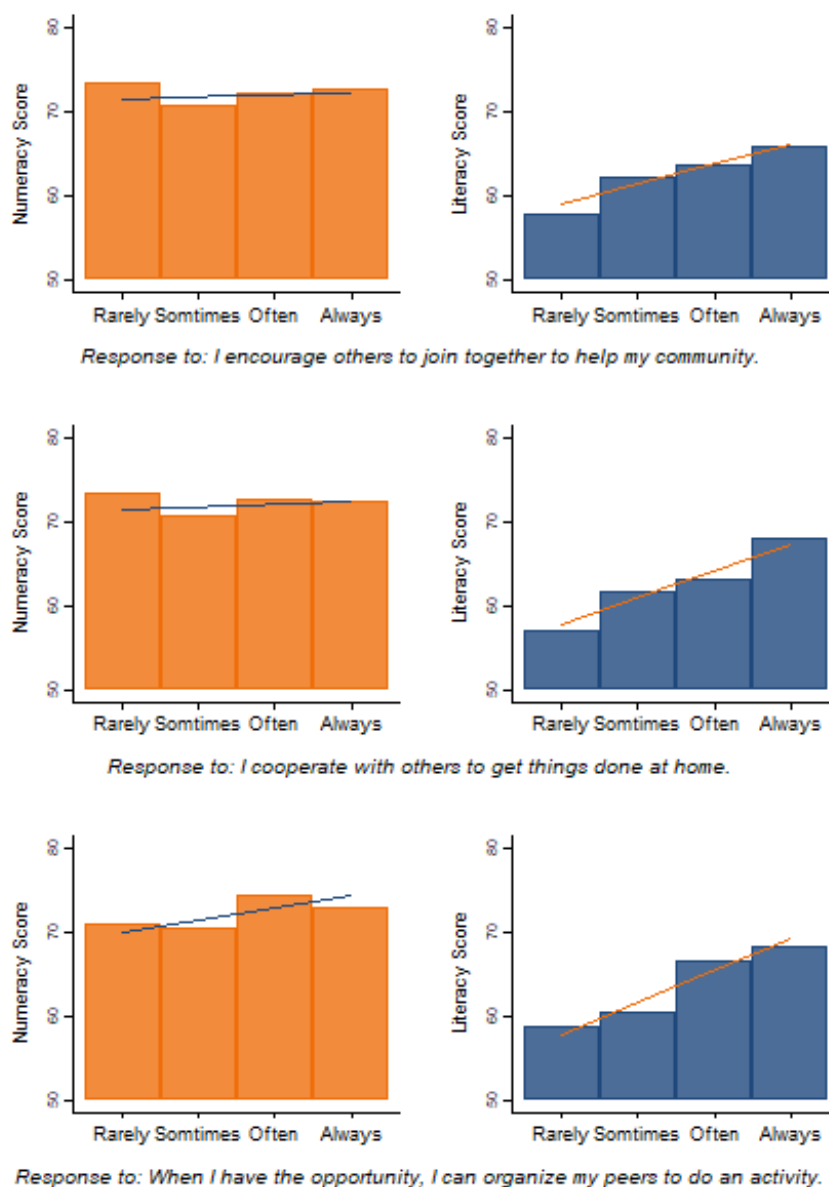
Using a regression model with girls' age, cohort type, and state as control variables, we find that an additional point increased in the YLI has a negative effect on the numeracy learning outcome change from BL to ML, and that this result is statistically significant. For the literacy outcome score, we also find that a 1 point increase in the YLI leads to a 0.03 point decrease in literacy learning outcome change, though this coefficient is not statistically significant.

However, when we add an additional control variable – the BL score of each respective assessment, we find that the effect of YLI increases in learning outcome changes becomes slightly positive (0.02), and that this coefficient was not significant. This is likely because the YLI is a good predictor of BL numeracy scores, and as has been discussed on many occasions throughout the previous sub-sections, higher scores at BL are associated with slower learning outcome changes compared to lower performers at BL, simply because the program likely had a larger marginal effect on the most disadvantaged girls, causing their learning outcome growth to outstrip that of girls who were higher-performers at BL. In other words, the YLI is still a predictor of both higher levels of learning outcomes at the time the YLI is measured, and of higher learning outcomes in the future, but also a predictor of slower learning growth rates.

A similar interpretation can be made for the effect of the YLI scores on literacy outcome changes. Without controlling for BL literacy scores, it would appear that YLI scores are negatively correlated with learning outcomes changes. However, controlling for BL scores, higher YLI scores predict a bigger increase in literacy scores. Unlike the effect on numeracy score changes, however, the magnitude of the positive effect on score change is greater, and it is statistically significant at the 99% confidence level. The finding suggests that, for a group of girls with similar BL scores, those with higher YLI scores will have a higher rate at which their literacy score improves. When comparing girls with different BL scores, it may even be possible that for the higher performers, additional interventions on the elements covered under the YLI might somewhat “offset” the growth-slowing effects of having higher BL scores. In the figure above, it is clear that the YLI questions tend to better distinguish better and poorer performers at ML for the literacy assessment than for the numeracy assessment, as seen in the steeper increase in ML literacy test scores as girls responded more positively to the YLI questions.

<sup>65</sup> Numbers are derived from the regression model with controls.

FIGURE 10: IMPACT OF SELECT YLI QUESTIONS ON ML NUMERACY AND LITERACY SCORES



One possible reason why the YLI and its components might be a stronger positive predictor of literacy outcomes has to do with the nature of the behaviours which the YLI survey sub-module measures. Many of these are likely more conducive to literacy learning than they are to numeracy learning. The behaviours captured under the YLI include many “open-ended” tasks that require that girls learn to communicate more and better. Examples include whether a girl’s friends ask her for advice, whether a girl is able to communicate her ideas differently when someone does not understand her initial explanation, and whether she is able to contribute to discussions in the household even when other family members hold different perspectives. Being able to do these tasks requires a certain level of mastery over the use of language to not only to express ideas accurately, but also to tailor how those ideas are expressed to different audiences. While these tasks do not necessarily require communication to be done in writing, it is probable that general language skills are positively correlated with literacy skills, therefore allowing

YLI scores to affect literacy learning growth positively and somewhat counteracting the suppressive effects of high BL scores on future learning growth.

## GEF Participation

Expanding on the analysis of life skills (i.e. YLI scores) and their impact on learning outcomes, this section investigates the role of Girls' Empowerment Forum (GEF) participation in improving learning outcomes over time. GEFs are the primary mechanism through which AGES seeks to increase self-esteem, leadership ability, and life skills among its beneficiary girls. GEFs are after-school programmes that provide girls with a peer support network, a positive female role model (typically in the form of a female teacher-mentor), opportunities for tutoring and direct mentoring, and other benefits. GEFs are also a mechanism for girls to participate in school governance, and girls engage in awareness-raising around educational and health rights, encourage enrolment of out-of-school girls, and participate in their respective Community Education Committees (CECs), among other activities.

Previous research – conducted as part of the SOMGEP-T programme – has shown that participation in GEFs promotes higher learning scores. GEFs generate several intermediate benefits that lead to improved learning scores, including greater retention in school, higher attendance rates, and higher self-esteem.

Similar findings emerge from the AGES panel sample tracked since baseline. To start, we assessed the relationship between self-reported GEF participation (reported retrospectively at midline) and learning outcomes, in a regression framework that controlled for region, age, and the cohort of which the girl is part. A more robust approach also controlled for baseline learning scores, in essence analysing the change in scores from baseline to midline as a function of GEF participation. The results in the top panel of the table below show that GEF participation is associated with significantly higher numeracy (+10.5 points) and Somali literacy (+9.3 points) scores at midline.

Of course, GEF participation is not randomly assigned and, as girls self-select into participation; to the extent that more motivated and higher-performing girls may choose to participate, this biases our findings toward a positive relationship between GEF participation and learning outcomes. Indeed, girls who participate in GEFs already had numeracy and literacy scores that were 4.9 and 8.0 points higher than non-participants at baseline.<sup>66</sup> To account for this form of bias, we controlled explicitly for baseline learning scores, with few changes in the overall findings – GEF participation is still strongly predictive of higher numeracy and literacy scores.

**TABLE 25: EFFECT OF GEF PARTICIPATION ON LEARNING SCORES AT MIDLINE**

	Effect on Score Change, Numeracy	P-Value	Effect on Score Change, Literacy	P-Value
<b>GEF Participation</b>				
<b>BL score not controlled</b>	10.5	0.00	9.3	0.00
<b>BL score controlled</b>	9.1	0.00	6.6	0.00
<b>Ongoing GEF Participation</b>				
<b>BL score not controlled</b>	10.6	0.00	10.7	0.00

<sup>66</sup> This finding is reinforced by the fact that greater continuity of participation (i.e. girls who, at midline, say they continue to be in contact with the GEF) had even higher *baseline* learning scores. This suggests that motivated self-selection is at work.

<b>BL score controlled</b>	8.9	0.00	7.4	0.00
----------------------------	-----	------	-----	------

The bottom of the table builds on this analysis, utilising a measure of GEF participation that takes into account continuity of engagement. We define GEF participation in the top panel of the table based on a respondent's statement that they have participated in GEF activities in the past; in the bottom panel, we consider respondents to be participants if they participated in the past *and* continue to have contact with the GEF. This approach does not change our results, despite the stricter definition.

Consistent with evidence from SOMGEP-T and – to some extent – evidence from our previous discussion of YLI scores, GEF participation may improve learning scores in part through its effect on self-esteem and the generation of peer support networks. Specifically, GEF participation is associated with a 1.4-point increase in YLI scores from baseline to midline, though the difference between GEF participants and non-participants is not statistically significant. More telling is the impact of participation on continued enrolment, with GEF participants about 12.5 percentage points more likely to remain enrolled in education at midline, even while accounting for differential educational starting points.

**Note from the project:** The majority (78%) of the Cohort 1 girls who participated in GEFs were ABE and NFE students. Among those who continued to participate in GEFs, 50% are ABE students. ABE and NFE students had higher baseline scores due to being older and having previous exposure to shorter education in emergencies learning opportunities as well as learning through market activities (see sections above on learning outcomes for subgroups). Therefore, the higher baseline scores are not a result of the self-selection of 'better off' students, but rather from having a disproportionately higher proportion of participants from ABE and NFE cohorts in GEFs.

## Teaching Quality

A core part of the AGES Theory of Change is that improved pedagogy will produce higher learning scores, both because students learn more in class and because students are more likely to attend if their teachers are consistently present, make them feel welcome, and encourage their schooling. During the midline assessment, girls were asked a number of questions regarding their teacher's pedagogy, including whether the teacher uses corporal punishment, and the extent to which teachers can fulfil their basic responsibilities, such as attending classes. In this section, we analyse whether student-reported measures of teaching quality are associated with greater gains in learning scores from baseline to midline. Our interest in this section is not on the relationship between teaching quality and learning at a single point in time, as better schools may simply have both better teachers and better students simultaneously. Rather, we are interested in whether quality teaching generates greater *gains* in learning scores relative to a student's baseline performance.

The measures of teaching quality we choose to focus on are – in part – derived from analysis performed during the baseline evaluation. Importantly, however, we measure teaching quality at midline, as girls were asked to reflect on their teacher's performance in the previous year of their schooling. This is the period when high-quality teaching should have a direct impact on learning outcomes.<sup>67</sup> The outcome we study is the *change* in scores from baseline to midline for an individual respondent. The teaching practices we analyse are listed in the table below, each of which is coded as a binary variable. For instance, girls

<sup>67</sup> In cases where a girl is not enrolled in school or a learning programme at the time of data collection, they were asked to reflect on their teacher's performance from the period when they were in school, typically in late 2019 and 2020 (as all girls were enrolled in school or learning programmes at the time of the baseline November 2019).



who report that their teacher often encourages participation are coded in one direction, while teachers who sometimes, rarely, or never encourage participation are coded in the other.

For each of the teaching quality measures in the table below, we estimate a linear regression predicting the change in learning scores as a function of the teaching quality measure. Each regression also includes age, state, and current enrolment status as control variables. In the top panel, we limit our analysis to FE or formal school girls, as their greater exposure to schooling may mean that teaching quality has an outsized effect on them.<sup>68</sup> In the bottom panel, we employ the full sample of girls re-contacted since baseline, including ABE and C1 NFE girls, which usefully provides a larger sample; in the regressions in the full sample, we also control for the girl's cohort, as we have previously documented differential improvements in learning scores across cohorts. For each regression, we report the effect of a teaching quality measure on the *change* in learning scores (and the p-value associated with the regression coefficient). The results can be interpreted as a difference-in-differences estimate comparing trends between girls who experienced a particular teaching quality outcome (i.e. an unwelcoming teacher or one who employs corporal punishment) and girls who did not.

**TABLE 26: EFFECT OF TEACHING PRACTICES ON GAINS IN NUMERACY AND LITERACY SCORES SINCE BASELINE**

	Effect on Gain, Numeracy	P-Value	Effect on Gain, Literacy	P-Value
<b>FE Girls Only</b>				
My teacher does not make me feel welcome in classroom	2.1	0.71	-0.9	0.89
My teachers are often absent	-1.7	0.72	9.6	0.06
My teacher encourages participation	6.6	0.05	3.7	0.44
My teacher explains how things learning are useful in our lives	3.5	0.45	4.1	0.30
My teacher's lessons move too fast for me	-0.3	0.93	-5.2	0.14
My teacher treats boys/girls differently in the classroom	-4.8	0.17	1.5	0.75
Teacher punishes students who get things wrong in a lesson	-1.3	0.71	-5.1	0.14
Teacher used corporal punishment in last week	-0.3	0.94	9.2	0.08
<b>FE, ABE, and C1 NFE Combined Sampled</b>				
My teacher does not make me feel welcome in classroom	-5.6	0.08	-7.6	0.04
My teachers are often absent	-3.1	0.24	5.4	0.09

<sup>68</sup> Further controls may be useful in some cases, but are not generally necessary. By analysing “change scores”, we explicitly control for any pre-existing characteristics that influenced baseline learning scores for a given girl, reducing the importance of controlling for these characteristics.



My teacher encourages participation	4.5	0.04	4.0	0.18
My teacher explains how things learning are useful in our lives	6.5	0.01	4.5	0.14
My teacher's lessons move too fast for me	3.8	0.07	4.7	0.08
My teacher treats boys/girls differently in the classroom	-2.6	0.45	3.0	0.54
Teacher punishes students who get things wrong in a lesson	0.4	0.86	-0.2	0.94
Teacher used corporal punishment in last week	5.5	0.10	8.2	0.03

The results in the top panel of the table are broadly consistent with a view that teaching quality has shaped improvements in learning scores, but there are also a number of confusing or contradictory results. The most straightforward findings centre on encouragement of participation and emphasis of the real-life utility of lessons to students: for both numeracy and Somali literacy, FE girls who had teachers who encouraged participation and linked lessons to "real life" experienced greater gains from baseline to midline.<sup>69</sup> Absentee teachers and those who used corporal punishment within the last week both produced contradictory results across subjects, with students of these teachers experiencing large gains in Somali literacy. Frankly put, we do not have a good explanation for why teachers that are often absent would produce improved learning scores; the only speculation we can offer is that girls who described their teachers as frequently absent may hold their schools and teachers to higher standards in some way (e.g., they attend a better school or come from education-oriented families and expect greater effort from their teachers), resulting in a confounded relationship between teacher absenteeism and learning performance.

In some ways, the bottom panel of the table helps to clarify these contradictory results, in part by providing a larger sample size, which "stabilises" the findings. In this larger sample, we find that teacher demeanour, encouragement of participation, and explanations of how content links to a student's everyday life all predict greater gains in learning scores for both numeracy and Somali literacy. However, other results continue to confound our expectations: teacher absenteeism produces contradictory results, and both the use of corporal punishment and lessons that move too quickly for the student's comfort levels both predict higher gains in learning scores.<sup>70</sup> Of course, score gains are driven largely by continued

<sup>69</sup> Although only one of the four analyses was statistically significant, the consistency of both the direction and magnitude of effect gives us greater confidence in the result.

<sup>70</sup> One possible explanation for the finding regarding the pace of lessons is a form of social desirability bias: because this question in particular can be perceived as asking a girl to rate her own competency in following along with lessons (rather than the intended interpretation of whether the teacher exercised good judgement by using an appropriate speed), it may be the higher-performing girls who have had more confidence in their own abilities who are willing to admit that the teacher's lessons are done too quickly. If true, it would mean that teaching quickly is positively related to test scores (and thus negatively related to learning growth rates), when in fact the respondents who are willing to admit that teaching is done too quickly are considerably different from those who are not willing.

enrolment and underlying demographic predictors; however, the fact that these poor teaching practices predict greater improvements, despite controlling for other factors, is extremely surprising.

## Community Attitudes

Finally, we assess whether community attitudes around girls' education also impact their learning outcomes and growth trajectories. Specifically, we asked girls' caregivers about their attitudes towards education, as their beliefs regarding the worth of education can often have important effects on how much resources they avail to support their daughters'/dependent's learning. The table below presents the results of the DiD regression model controlling for age, cohort type, and state.

TABLE 27: EFFECT OF CAREGIVER ATTITUDES ON NUMERACY AND LITERACY SCORES

	Numeracy				Literacy			
	Effect on BL Score	Change from BL	DiD	P-Value of DiD	Effect on BL Score	Change from BL	DiD	P-Value of DiD
<b>Caregiver aspires to university education for girl</b>	3.78	5.48	-2.03	0.53	7.88	17.53	-5.00	0.20
<b>Caregiver believes girls' education worthwhile, even if funds are limited</b>	-0.14	5.72	-0.03	0.99	-4.22	18.64	4.24	0.32
<b>Caregiver believes work or HH chores are acceptable reason to not attend school</b>	0.30	5.35	-0.64	0.77	-2.58	17.60	-0.91	0.77
<b>Caregiver believes cost of education is acceptable reason to not attend school</b>	6.20	3.32	-6.39	0.00	5.85	17.17	-2.56	0.35

In general, we find that all but one measure had a significant effect on assessment scores, and said measure was only significant in its effects on numeracy score changes, but not on literacy score changes. One possible reason that the effects of these attitudes is more pronounced and significant is likely because these questions are attitudinal, not behavioural. In other words, attitudes will likely have a bigger impact on girls' learning when translated into concrete action, such as providing more funds for girls or explicitly asking girls to stay home and not attend school. Moreover, people may always be the best judge of their future actions, and may find that how they respond to certain situations would differ from how they believed they would respond. While some of these actions have certainly already been translated into action for some girls in our panel sample, for other girls, the corresponding actions associated with these attitudes have not been taken, which, on balance, might lead to caregivers' beliefs having less of an effect on girls' learning.

However, it is worthwhile to note that the two attitudes relating a family's fiduciary concerns with girls' education – whether education is worth it in spite of limited funds and whether the cost of education warrants keeping girls at home – take on unexpected signs in the regression model. We would expect that believing that school is worth it despite the cost would allow caregivers to better support girls' education, and conversely, that caregivers who believe it is acceptable to sacrifice education to save funds

would be less supportive. Yet, the data shows the opposite, though only the latter had a significant effect. One possible explanation might be that girls who come from households who believe that education is not as important as household finances would need to be more resolute and determined to learn, so as to convince their caregivers to still support their education and/or find other sources of support. These girls in turn are likelier to score higher in the numeracy and literacy assessments compared to girls who may not necessarily have to go through these hurdles. In theory, this might partially explain why girls from less supportive households might have a higher average score at BL (and subsequently slower rates of growth). However, measuring determination and resourcefulness is beyond the scope of this evaluation, and as such, the above hypothesis must necessarily remain speculative at this point.

## 4.5 Discussion

In the preceding sub-sections, we examined the numeracy and literacy assessment scores and how they have changed from BL to ML. We first began with an analysis of changes across regions and cohort-types, followed by changes in sub-task scores in each of the assessments. We then examined how changes differed across the major sub-groups of concern to the programme – disability status, demographics, household circumstances, and availability of school infrastructure – before finally proceeding to a final analysis of how intermediate outcomes – attendance, youth leadership, teaching quality, and caregiver (community) attitudes – predict future learning outcome changes.

Throughout these sub-sections, a consistent finding is that girls who performed more poorly at BL experienced growth in their learning outcomes that far outpaced growth experienced by BL higher performers. For girls who are members of disadvantaged groups, such as girls with disabilities, girls whose teachers were frequently absent, and girls who lacked clean water in their households, scores at BL were considerably lower than for girls not from these groups. Yet, by ML, in several measures of disadvantaged groups, in-group girls had made considerable headway in closing the gap between them and out-group girls, if not outright exceeding scores by out-group girls. In terms of sub-tasks score changes, in many cases it was in the sub-tasks where girls had low BL scores that large increases were observed.

In all, the evidence points to the AGES programme being generally effective in lifting the bottom-most performers and helping equalise their learning with comparatively less disadvantaged girls. The inclusion of various initiatives under AGES likely helped the most disadvantaged girls overcome serious challenges that might otherwise, in the absence of appropriate interventions, lead them to continue to be behind other girls, if not fall further behind. In addition, it is possible that various heuristics which correlate with poorer performance at BL, such as having a disability or having teachers that used corporal punishment in class, served as a signal to programme teachers and other implementers that a girl will need additional attention and resources to succeed. As a result, these girls in particular were able to enjoy sizable learning gains over time.

While positive indicators of the programme's progress thus far, several concerns remain. For several categories of sub-groups, in-group girls have not seen the same steeper increase in assessment scores. This is particularly true for girls who have been married, who are mothers, whose caregivers and parents have no formal education, and for girls belonging to smaller or less dominant ethnic groups, including Bantus and other minorities. These identity/group categories still exert a negative impact on girls' learning outcomes, causing girls in these groups to experience slower learning growths on average.

The logic outlined above on why lower performers at BL generally experience larger increases also means that higher-performers at BL saw considerably flatter learning growths compared to girls at lower levels at BL. In a way this is to be expected: being comparatively less disadvantaged girls who already possessed more developed numeracy and literacy skills, these girls were less affected by some of the socio-economic and pedagogical challenges faced by poorer performers. As the programme had a major focus on helping marginalized girls, these higher-performers likely experienced smaller marginal benefits from the

programme. They may have also potentially received fewer benefits overall from the programme if much of the attention paid by the teachers and implementers managing the programme's daily activities is directed to their lower performing peers.

From the perspective of the equalising learning outcomes among the girls in the AGES programme, this may not be a major concern, as the data reflects generally decreasing learning outcome disparities among top performers and low performers. However, seen from a broader perspective, the plateau seen among higher performers may be cause for alarm if it indicates that the programme is approaching the limits of how much additional learning it can help girls achieve. After all, many of these higher BL performers remain overage for their grade level, and are only high-performers relative to especially low performing girls. As the baseline report and the present report have made clear, there is a concern that measurement ceiling effects are responsible for the apparent plateau among high performers; high-performers may have developed skills at higher levels than the level captured by our current EGRA and EGMA tools, which would on average make them appear to experience a little change in their scores, when in reality our tools simply do not capture this additional progress. Mitigation measures have been put in place at ML to reduce ceiling effects distortions in the ML to endline comparisons, and more conclusive findings will be drawn then. Nevertheless, we recommend that programme leaders continue to reflect on the potential plateau among high-performers and identify potential adaptations.

Based on the data, we highlight three possible areas which the programme may be able to leverage to overcome potential plateaus as more and more of the lower-performing girls now reach learning outcome levels that are comparable with high performers. These are by no means exhaustive, but may be important considerations. First, allowing for some flexibility in the learning environment, specifically allowing girls to take study materials home at night, has been shown to have a significant and substantial impact on improving learning outcomes, though this data was only collected among FE girls. Still, being able to take home materials at night resulted in the biggest boost in learning outcome change among all the sub-groups, for both the literacy and numeracy assessments.

Second, after controlling for BL test scores, we find that age is positively correlated with learning outcome changes: among girls who were at similar BL levels of numeracy and literacy, older girls increased their learning outcome scores faster. This is in spite of the qualitative data highlighting major challenges which older learners faced, including perceived or real stigma in school for being older, overage-for-grade girls placed in the same learning environment as younger girls. In this sense, the program appears to have made some headway in removing, or overcoming, said stigma. As many of the high-performers at BL are older girls in the ABE and C1 NFE cohorts, they are already at potential risk of encountering a learning plateau; any additional assistance to help boost their learning and off-set the plateau may be beneficial.

And finally, the data also revealed that YLI scores were positively correlated with literacy and numeracy score growths over time, though this result was only significant at at least the 90% confidence level for the literacy assessment score. Particularly noteworthy is that the YLI score remained a positive predictor of higher learning growths over time even after controlling for BL numeracy and literacy scores. In other words, the skills measured under the YLI helped girls offset, to a degree, the plateau in their learning growth which was caused by already having relatively high scores to begin with. Indeed, the behaviours measured under the YLI are unique opportunities for girls to learn to communicate and solve open-ended challenges – skills that require language fluency and fluid intelligence that may have carry-over applicability in girls' reading and comprehension skills. These opportunities may not be typically used in a traditional classroom setting, but the data suggests that their inclusion in girls' development may provide novel and effective means of learning.

## 5. Transition

The second core outcome in the evaluation of AGES concerns retention and life pathways, broadly construed. “Transition” is a core outcome for all GEC, GEC-T, and LNGB projects, and it seeks to capture how the project affects girls’ pathways through life. Operationalising successful transition is always complex, because there are many possible pathways that can be considered a success, including retention in school, shifting into employment, or re-enrolment in formal school for those who were not enrolled previously. Transition is more than a measure of retention; it incorporates alternative education, vocational training, and even the possibility that an AGES beneficiary will start a small business. It is further complicated by the fact that successful transition is defined by a girl’s starting point: for a girl enrolled in formal school, continued enrolment in the same grade level is insufficient for a successful outcome; in contrast, for a girl who was out-of-school or enrolled in an NFE programme, transitioning into formal education is a success, regardless of her grade level. In these ways, transition accounts for the complex environment in which AGES is implemented.

The analysis in this section is broken into three parts, each with a distinct goal. In the first section, we define transition – as it is used by AGES – in more detail and report the overall transition rates for girls who were initially recruited into AGES programming in 2019. In the second section, we analyse subgroup-specific transition outcomes, assessing whether transition rates differ as a function of a girl’s household characteristics, marginalisation, or other factors. Finally, we analyse the relationship between the programme’s intermediate outcomes – such as teaching quality, attendance, and self-esteem – and transition rates, with an eye to providing evidence for or against the programme’s Theory of Change.

### 5.1 Aggregate Transition Outcomes

As the previous section alluded to, transition is a complex, multidimensional outcome. In much of our analysis, we will define transition in a binary fashion – success or failure – despite the fact that this obscures significant variation in how girls’ lives evolve in response to the programme. At times, we will describe and present results for more specific pathways, to provide additional depth.

AGES considers successful transition to include – depending on the girl’s “starting point” at baseline – enrolment in formal school, appropriate grade progression, enrolment in technical or vocational education, and shifting into gainful, age-appropriate employment or small enterprise ownership. The starting point for a girl is critical, as some outcomes may represent forward progress for a girl enrolled in an NFE programme at baseline but backsliding for a girl who was previously enrolled in formal school. Below, we document the potential pathways a girl can take, depending on her starting point, and how they are classified in terms of transition.

TABLE 28: TRANSITION PATHWAYS, ACCORDING TO STARTING POINT OR COHORT

Starting Point	Successful Transition	Unsuccessful Transition
<b>FE Girl Enrolled in grades 1-2 at baseline</b>	<ul style="list-style-type: none"> <li>Retention in formal school, with progression through the grades (e.g., a girl in grade 1 has reached grade 3 two years later)</li> <li>Drops out but is enrolled in a technical or vocational education programme</li> </ul>	<ul style="list-style-type: none"> <li>Drop out</li> <li>Retention in formal school without appropriate grade progression</li> <li>Transition into employment or self-employment</li> </ul>

<b>ABE Girl Enrolled in ABE at baseline</b>	<ul style="list-style-type: none"> <li>• Enrolment in formal school, at any grade level</li> <li>• Retention in ABE, with progression through levels (where available)</li> <li>• Transition into a technical or vocational education programme</li> <li>• Transition into age-appropriate, non-exploitative employment</li> <li>• Transition into self-employment</li> </ul>	<ul style="list-style-type: none"> <li>• Drop out</li> <li>• Transition into NFE</li> <li>• Retention in ABE without progression in levels</li> </ul>
<b>NFE Girl Enrolled in NFE at baseline</b>	<ul style="list-style-type: none"> <li>• Enrolment in formal school, at any grade level</li> <li>• Retention in NFE, with progression through levels (where available)</li> <li>• Transition into a technical or vocational education programme</li> <li>• Transition into age-appropriate, non-exploitative employment</li> <li>• Transition into self-employment</li> </ul>	<ul style="list-style-type: none"> <li>• Drop out</li> <li>• Transition into ABE</li> <li>• Retention in NFE without progression in levels</li> </ul>

As the table makes clear, the starting point of the girl is central to defining successful transition. A formal school girl who drops out of school to pursue employment is not considered a success because she is comparatively young and employment after completing just 2-3 years of primary schooling is not sufficient. In contrast, an NFE girl who has transitioned into employment or self-employment is considered successful because she is older and employment is a desirable outcome for older girls, and because NFE courses are shorter, such that her learning programme, begun at baseline, has certainly ended by the time of the midline evaluation.

The sample we analyse for transition outcomes consists of the set of baseline girls – FE, ABE, and NFE – who were successfully re-contacted at midline.<sup>71</sup> The total sample includes 1,062 girls, defined by their status at baseline: 332 formal school girls, 364 ABE girls, and 366 NFE girls.

Aggregating across the three groups, the overall transition rate from baseline to midline is 47.2 percent. The highest transition rates, by far, come from formal school (FE) girls, 70.8 percent of whom experienced successful transition over this period. In contrast, transition rates are much lower – 46.8 and 25.6 percent, respectively – among ABE and NFE girls.

Two patterns in these aggregate results require explication. First, transition rates appear to be quite low, overall. However, it is difficult to contextualise these rates, because transition among ABE and NFE programme participants is seldom studied. A similar programme implemented by CARE in Somaliland, Puntland, and Galmudug (SOMGEP-T) recently recorded a transition rate of 56 percent among girls

<sup>71</sup> We do not include replacement FE girls in this analysis because replacement girls may not have been enrolled in formal school at the time of the baseline. Moreover, because replacement FE girls were selected explicitly from *within* schools, their transition rates would be artificially inflated, owing to the fact that they are enrolled at the time of their selection into the sample. Replacements of this can be included in transition analysis in future rounds, however.



enrolled in Alternative Learning Programmes (ALP) and 84.0 percent among ABE girls. Transition rates for AGES ABE and NFE girls are lower, but this may reflect the fact that over two years passed between baseline and midline. Transition success is, in some ways, cumulative over time; if an NFE girl completed her programme and enrolled in formal school for one year but dropped out prior to the time of the midline, she would be classified as unsuccessful. As a result, a lag between reporting periods has the potential to depress transition rates.

Second, the very large gap between FE, ABE, and NFE girls can be explained as a function of the girl's ages and the differential options open to them in terms of transition. A girl's "type" is closely correlated with her age, with FE girls the youngest and NFE girls the oldest; in related research, we have found that age is associated with lower transition rates, holding all else constant.<sup>72</sup> The much older NFE girls (20.1 years at midline, compared to 13.7 years for FE girls) contributes to a lower transition rate.

Perhaps more important is the fact that girls enrolled in formal school face a different set of possible pathways. For a girl in formal school, remaining in school and progressing through the grades is considered a successful outcome. While we do not wish to downplay the difficulty of remaining enrolled – especially through the COVID-19 pandemic – retention and progression is a pathway with at least some degree of path dependence – after all, the girl is already enrolled, and initial enrolment tends to be a major barrier. In contrast, ABE and NFE girls whose educational programmes have concluded can also enrol in formal school, which is considered a successful outcome. But enrolment itself is a major barrier, especially for girls who would be starting their formal education at a comparatively late age. For ABE and NFE girls, there are few straightforward options that are considered a success: employment is contingent of labour market conditions and enrolment in formal school is possible, but relatively infrequent due to the aforementioned barriers. Remaining enrolled in ABE or NFE programming may be possible in some cases, but these programmes are not intended to last for multiple years, meaning that retention and advancement is not the "expected standard" that it is in formal schools. As a result, ABE and NFE girls face a different post-programme environment than formal school girls, and this is reflected in their lower transition rates.

Our discussion of structurally different transition *options* is reinforced by the results in the table below, which documents the observed transition paths of the three different types of girls. The results show that the vast majority of FE girls remain enrolled in school just over two years after the baseline, with some variation in whether the girl has progressed sufficiently (two grades) over that period. Even for girls who were held back at least once during this time, though, remaining enrolled is a natural – if costly and difficult, given the context – outcome.

In contrast, ABE girls and, especially, NFE girls are much less likely to remain enrolled. To a large degree, this is because their programmes are time- and level-limited – for most girls, an NFE programme does not last multiple years, which means that there is not a natural "default" pathway for NFE girls to follow after completion of the programme. Of course, the AGES Theory of Change implies that NFE girls will transition into gainful employment or other opportunities, which many of them do; nonetheless, as noted above, this is neither a default outcome or one over which the girls themselves exercise significant control, as employment opportunities are so limited in the Somali context.

---

<sup>72</sup> This is not surprising, given that dropout rates from formal school are closely correlated with age. It is safe to assume that a similar relationship applies to informal education programmes, with older girls more likely to drop out to get married, pursue employment opportunities, and so forth.

TABLE 29: TRANSITION OUTCOMES, AMONG DISTINCT COHORTS

Outcome/Path	Share of Girls
<b>Formal School (FE) Girls</b>	
Remain enrolled, progress two grades	70.8%
Remain enrolled, held back (failed to progress two grades)	13.4%
Transition into ABE/NFE	0.4%
Drop out	14.9%
Drop out into employment or self-employment	0.7%
<b>ABE Girls</b>	
Enrol in formal school	23.4%
Remain enrolled in ABE, progress in ABE levels	20.6%
Remain enrolled in ABE, without sufficient level progression	14.6%
Drop out or completion of programme without alternative outcome	38.6%
Drop out into employment or self-employment	2.8%
<b>NFE Girls</b>	
Enrol in formal school	6.1%
Remain enrolled in NFE, progress in NFE levels	3.1%
Remain enrolled in NFE, without sufficient level progression	3.8%
Transfer into ABE or informal education	1.8%
Drop out or completion of programme without alternative outcome	68.7%
Drop out into employment or self-employment	16.5%

Another finding that emerges from the table above concerns the post-programme pathways of ABE versus NFE girls. As expected, both types of girls have a tendency to finish their programmes and “drop out” into idleness, i.e. complete the programme without transitioning specifically into employment, formal education, or other options. However, this is much more common among NFE girls for two reasons: first, ABE girls are significantly more likely to remain enrolled in the longer ABE programmes – 35.2 percent of ABE girls remained enrolled in ABE at midline, compared to just 6.9 percent of NFE girls. In addition, ABE girls are much more likely to enrol in formal school, while NFE girls are more likely to transition into employment or self-employment. These observed pathways align with the programme’s intentions and expected outcomes – ABE is viewed, in part, as an opportunity to bring girls who have fallen behind in school back up to an acceptable level of learning and allow them to shift back into formal schooling if they desire. NFE programming was not designed to facilitate this, focusing instead on functional life skills that can enhance a girl’s economic opportunities. These different approaches and emphases are, quite naturally, reflected in the achieved outcomes of the different girls.



## 5.2 Subgroup Transition Rates

The discussion of transition thus far has highlighted the varied pathways taken by girls from different starting points, and emphasised the very different trends across FE, ABE, and NFE cohorts. The most dramatic divergence is across cohorts, though the very low transition rates of NFE girls (and moderately low among ABE girls) was not unexpected. Indeed, the programme and the evaluation team anticipated low transition rates, in part because most girls enrolled in NFE two years ago at baseline have completed their courses, and in part because of the age gap across cohorts and the role that age plays in shaping transition outcomes.

In this section we move away from the more nuanced description of transition pathways and focus on transition as a binary outcome, complicating the picture by investigating how transition rates differ across particular demographic groups in the sample. For each cohort, the table below reports transition rates for specific subgroups. For instance, the first pair of columns reports transition rates among formal school (FE) girls, beginning with those in Banadir, then Jubaland, and so on. The first column (N) refers to the number of respondents who fall into a given subgroup (e.g., pastoralist) within a cohort (e.g., FE girl). The second column (%) reports that transition rate for that subgroup, within the cohort. The two right-most columns report the transition rate for a given subgroup, aggregating across cohorts. Importantly, the top line of the table (“Overall”) documents transition rates for the overall sample of – for instance – FE girls, allowing readers to compare rates within a particular subgroup to this overall rate, to determine whether the subgroup performs better or worse than the cohort’s norm.

The first finding from the table concerns the relationship between geography and transition rates. In fact, there is a surprising lack of a relationship between geography and transition outcomes. We might expect girls in Banadir – a comparatively larger, more cosmopolitan area than the urban areas studied in Jubaland and South West State – to have greater schooling and labour market opportunities. We might also expect their households to be more financially stable, on average. However, this does not appear to translate into systematically higher transition rates – ABE girls in Banadir are more likely to successfully transition than other ABE girls, but this pattern is reversed exactly among NFE girls, and the result is an aggregate transition rate that does not differ very substantially across states (Jubaland is markedly lower, but rates in Banadir and South West State are similar). It is possible that this is because, among the AGES cohort, girls in Banadir are not systematically better-off financially, at least based on the available data. Many of the Banadir locations supported by AGES are situated in outlying districts, with a large number of IDPs, which is reflected in the fact that transition rates in the country’s capital are not higher than those in comparatively marginalised regions.

TABLE 30: SUBGROUP TRANSITION RATES, ACROSS COHORTS

Subgroup	FE Girls		ABE Girls		NFE Girls		Aggregate	
	N	%	N	%	N	%	N	%
Overall	332	70.8	364	46.8	366	25.6	1062	47.2
<b>Geography</b>								
Banadir	142	71.8	175	60.7	187	17.7	504	48.4
Jubaland	119	73.8	83	20.9	83	24.8	285	43.8
South West State	71	63	106	48	96	40.6	273	49.1
<b>Education of Parents/Adults</b>								
Head of Household has no education	75	74.4	85	54.5	78	24.4	238	51.9
Caregiver has no education	102	69.4	111	56.5	97	23.7	310	51.2

Neither HoH nor caregiver has any education	69	73.2	76	59.4	73	26	218	53.3
<b>Household Economic Characteristics</b>								
HH's main livelihood is pastoralism	2	50	4	78	3	0	9	45.5
HoH does not earn a regular wage	122	71.4	127	56.4	149	23.6	398	48.5
HH has a poor quality roof	62	62.8	77	41.5	70	28.5	209	43
HH went to sleep hungry most nights, last 12 months	21	73.7	28	54.4	27	13.1	76	44.8
HH owns land	130	71	117	47.2	109	29.3	356	49.8
<b>Girl Characteristics</b>								
Girl is married	1	0	4	17.9	32	33.8	37	31.2
Girl is a mother	0	N/A	2	0	52	36.6	54	35.2
Girl has a mental health disability	40	78	70	54.7	78	23.2	188	46.4
Girl has <i>any</i> disability	49	77.7	81	48.5	93	27.9	223	46.3
Girl has a non-mental health disability	12	61.4	15	41.8	19	43.1	46	47.1
Girl has a mental health disability, alternative coding #1	56	78	91	52.2	79	22.9	226	48.5
Girl has any disability, alternative coding #1	73	71.7	106	49.1	94	27.7	273	47.9
Girl has a non-mental health disability, alternative coding #1	24	49.6	22	48.2	19	43.1	65	47.3
<b>Marginalisation</b>								
Member of minority group	39	73.5	86	38.9	71	28.2	196	43
Af-Maay speaker	91	61.8	99	42	94	34.4	284	45.8

Additionally, there does not appear to be a strong relationship between parental education and transition rates. In fact, girls from households in which the head of household or caregiver have no education at all (neither formal nor Quranic) actually have slightly higher transition rates than the averages for their respective cohorts. This is less true for economically-marginalised households, which have slightly – but not significantly – lower transition rates. For instance, FE and ABE girls living in households with a poor-quality roof are 8.0 and 5.3 points less likely to transition successfully, compared to their cohort average. Households that have experienced chronic hunger over the last year also produce lower transition rates.

The bottom panel of the table reports transition rates among marginalised categories of girls. The patterns broadly match the trends in transition rates that we observe – namely, minority members have the lowest transition rates overall, though the relatively small sample size at this level of disaggregation leads to a degree of noise in our estimates. Speakers of af-Maay, which is the dialect of much of South West State and whose speakers are marginalised nationally, also have lower overall transition rates.

## 5.3 Testing the Theory of Change

Moving beyond the disaggregation of transition outcomes by subgroup in the previous section, we now subject the AGES Theory of Change, as it pertains to transition, to a tentative test. Our goal in this section is to assess whether the programme’s intermediate outcomes are correlated with transition rates in the manner the Theory of Change hypothesizes. For instance, we expect positive caregiver attitudes toward girls’ education to be correlated with higher transition rates, all else equal.

The approach we take utilises a linear regression model predicting binary transition success, while controlling for a number of ancillary factors that are correlated with transition outcomes. Our default control variables include: the girl’s cohort (i.e. the programme into which she enrolled at baseline), because transition rates are starkly different across cohorts; the girl’s age, because older girls tend to have lower transition rates; and region, accounting for systematic differences in transition rates across geographic space.<sup>73</sup> Wherever possible, we use baseline measures of our intermediate outcomes to predict transition, because we are interested in the impact of *prior* intermediate outcomes on transition.<sup>74</sup>

To start, consider a model of transition success that includes only age, cohort, and region. In this model, our earlier findings are buttressed: after controlling for age and region, cohort remains very strongly correlated with transition outcomes. There is a sharp and monotonic decrease in transition rates as one moves from FE, to ABE, to NFE cohorts.

Building on this model, we consider the impact of caregiver attitudes. The figure below reports two series of regressions: in the left panel, we report regressions predicting successful transition as a function of the control variables noted above and a single measure of caregiver attitudes. For instance, the orange line at the bottom represents the relationship between a caregiver’s aspirations and successful transition, after controlling for region, cohort, and age of the girl. If a caregiver reported, at baseline, that they hope their daughter will complete a university education, transition rates are 6.4 points higher, on average, though this difference is not statistically significant.

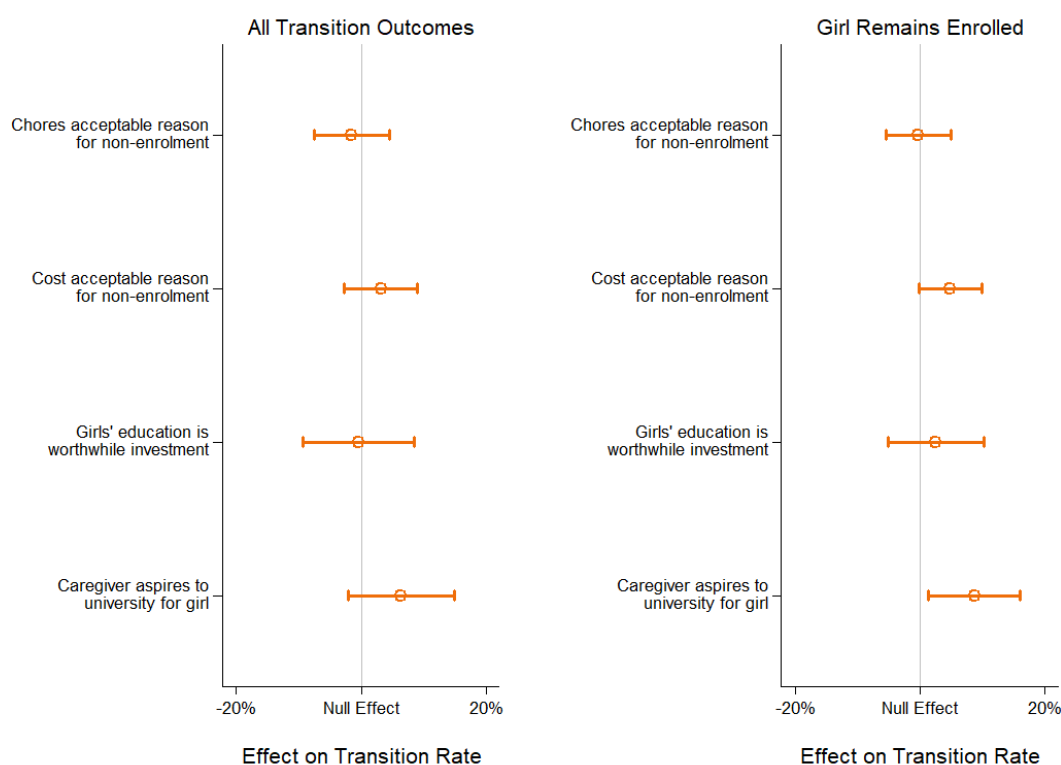
Each of the bars in the left panel represent a different regression. The next bar indicates how transition rates are affected if a caregiver agrees that “a girls’ education is a worthwhile investment, even when funds are limited.” Again, this model controls for other factors that predict transition rates. The dot indicates the relationship between this positive caregiver attitude and transition rates, while the horizontal orange bar is the 95 percent confidence interval for the estimate. When the dot is near the vertical line at “Null Effect,” it indicates that the variable has no relationship with transition rates. When the dot is to the right of the vertical line, it indicates a positive impact on transition rates; the horizontal orange bars tell us whether the effect is statistically significant (if they do not cross the vertical line at “null effect”). As the

<sup>73</sup> Note that we control explicitly for a girl’s cohort in pooled regressions that include girls from all three cohorts. In some of the regression models described in this section, we limit our analysis to formal school girls, in which case there is no need to control for cohort. The specific models are typically those that analyse the effect of school-level intermediate outcomes (e.g., teaching quality) on transition, because teaching quality was not measured in ABE and NFE centres during the baseline.

<sup>74</sup> Using measures for predictors that are derived from the midline evaluation would involve predicting a pre-midline transition outcome (girls enrol, drop out, etc. *between* baseline and midline) with an outcome that is measured at midline. To the extent that transition success or failure can influence some of the variables of interest – such as self-esteem – this would bias our results.

graph shows, when a caregiver states that education is a worthwhile investment, it has not discernible impact on transition rates, but when a caregiver aspires for their daughter to receive higher education, there is a positive – but not statistically significant – impact on transition rates.

**FIGURE 11: CAREGIVER ATTITUDES AS PREDICTORS OF SUCCESSFUL TRANSITION**



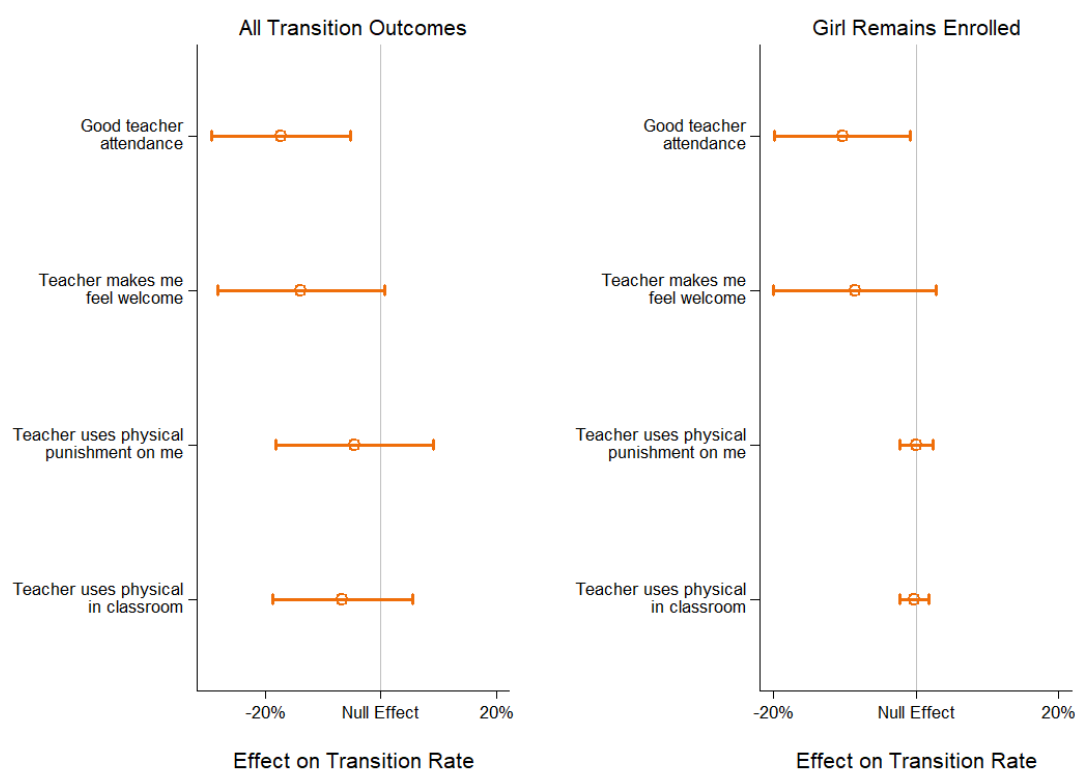
The right panel reports an identical set of models, but focuses on a slightly different outcome – whether the girl remained enrolled in their educational programme or another programme. For instance, if a girl moved from ABE to formal school, remained enrolled in formal school, or remained enrolled in NFE, they would be counted as a success under this definition. However, shifting into paid employment is not classified as a success. The logic of this second set of models is that caregiver attitudes toward girls' education should have their strongest relationship with enrolment and retention in school; there is less reason to expect a strong relationship between caregiver attitudes and a girl's employment outcomes. After all, moving into the labour market is not the same as attending university.

The results shown in the right panel vindicate this approach only slightly. For instance, there is now an even stronger relationship between caregiver aspirations and the likelihood that a girl remains enrolled. However, when caregivers profess that cost is an acceptable reason for a girl to drop out of school or remain out-of-school, this is also associated with a higher likelihood that their daughter remains enrolled, a finding that runs directly counter to our expectations. These findings are also stable when we limit our analysis to formal school girls. The null effects and contradictory results reported here suggest that caregiver attitudes may be correlated with transition outcomes but that the relationship is insufficiently direct for us to detect it among the noise of other factors, such as household income and exogenous shocks, that likely drive much of the variation in transition.

A similar story emerges when we analyse teaching quality and its impact on transition rates. As the figure below shows, the link between teacher characteristics – as reported by girls themselves – and their own transition outcomes are very weak and contradict conventional wisdom. For instance, girls who “strongly

disagree” that their teacher is often absent – in other words, indicate that their teacher attends class consistently – have lower overall transition rates (left panel) and are less likely to remain enrolled in school (right panel).

FIGURE 12: ASSESSMENTS OF TEACHING QUALITY AS A PREDICTOR OF TRANSITION OUTCOMES



It is difficult to reconcile these results with our theoretical expectations. However, it is important to emphasise that tests of this kind, which analyse the cross-sectional correlation between attitudes or attributes and an outcome, are a comparatively weak test of a particular hypothesis. We cannot eliminate alternative explanations for an observed pattern that the use of physical punishment by one’s teacher is only loosely related to lower transition rates. It is possible that girls who are willing to report a teacher’s use of corporal punishment are systematically different from those who are unwilling, which could impact their likelihood of remaining enrolled or gaining employment. The clearest conclusion is that there is not a direct and straightforward relationship between these particular aspects of teaching quality – as reported on by girls themselves – and transition outcomes.

Importantly, other findings linking intermediate outcomes and programme outputs to transition rates are clearer and more concrete. Perhaps the most robust finding in our regression models was that YLI scores – which measure self-esteem and leadership skills among girls – are correlated with higher transition rates. For instance, a 1-standard deviation change in a girl’s YLI score (approximately 18 points on a 100-point scale) is associated with a 7.1 point increase in successful transition, even after controlling for other predictors of transition. This finding is robust to alternative conceptualisations and definitions of transition, such as a focus on enrolment and grade progression – a higher YLI score is a statistically significant predictor of better transition and enrolment outcomes.<sup>75</sup>

<sup>75</sup> This is true among the formal school subsample and in the broader sample with girls from all cohorts.

Two other factors are also strong predictors of transition outcomes: a household's socioeconomic status and the impact of COVID-19 on their household. We used a simple measure of household socioeconomic status – a food security measure asking how often any member of the household had gone to bed hungry over the last year due to lack of food or resources to acquire food. We prefer this measure over others for two reasons: first, it was captured at the baseline, while many of our better measures of household economic status were added only in the midline evaluation round; second, it captures a universal need, food, that will affect any family, rather than assessing livestock or land ownership as a measure of wealth, both of which are shaped by a family's livelihood sector.

The findings regarding food security are stark: girls from households that have experienced hunger many or most days over the last year ( $n = 278$  out of a sample of 1062) are 6.1 points less likely to transition successfully. COVID-19's impact is slightly less stark and is concentrated among transition outcomes related to enrolment. Girls who report that COVID-19 had an economic impact on their household are only slightly (0.8 points) less likely to transition successfully. However, COVID-19 has a more specific impact on enrolment status – girls who report an economic impact of COVID-19 on their household are 3.4 percentage points less likely to remain enrolled. This pattern makes sense, because the economic shock of COVID-19 is unlikely to have a negative impact on a girl's labour market participation; if anything, an economic shock to the household might spur a girl's entry into the labour market, increasing this form of transition. At the same time, it will reduce enrolment by driving girls into the labour market to support their families, and by reducing the money available to support a girl's education.

We performed a number of additional analyses of transition outcomes, but found few compelling predictors at midline. For instance, the relationship between school management and attendance, on one hand, and transition rates, on the other, was null in our data. However, this is not a strong indicator that these intermediate outcomes do not impact transition rates – it is a function of the “levels of analysis” problem, in which school-level outcomes (attendance measured through classroom headcounts and school management measured at the school level) are only weakly linked to an individual-level outcome, such as transition. We would not expect aggregate attendance in a given school to be strongly related to a girl's likelihood of remaining enrolled; rather, her personal attendance may predict transition, if it were measured directly.<sup>76</sup>

## 6. Sustainability

### Self-replication rate of village savings and loans (VSL) groups

The first indicator of community-level sustainability for AGES focuses on the programme's economic interventions, which established and supported Village Savings and Loans Associations (VSLAs), with the goal of improving economic opportunities for the parents of AGES girls and the girls themselves. VSLAs provide a vehicle for household savings, making investment and saving for large household purchases –

<sup>76</sup> Of note, girls' baseline learning scores were highly predictive of successful transition. Both Somali literacy and numeracy scores were associated with a statistically significant increase in transition rates. For instance, a 1-standard deviation increase in Somali literacy scores (35.4 points) was associated with a 6.1-point increase in successful transition. This effect was even stronger when considering the relationship between baseline learning scores and successful retention and grade progression, but it predicts the broader transition outcome – including entry into the labour market – effectively as well.

including school fees or other educational expenses – more feasible. In many cases, they also provide loans, often the only viable source of finance for poorer households.

The concern, from a sustainability perspective, is whether programme-supported VSLAs will continue to operate after the end of the programme. If VSLAs have been established for a lengthy period, have an active membership, and provide real and equitable benefits to their members (e.g., they make loans or provide a safe venue for savings), they are more likely to be maintained over the long run. Savings groups, as with any organisation or institution, will have short lives if they are inactive, have weak leadership or ambivalent membership, and provide few actual benefits for participation.

To assess the performance of VSLAs in AGES communities, we asked both cohort girls – those enrolled in NFE programmes at either baseline or midline – and the caregivers of formal school and ABE girls about their participation in local savings groups. The first finding that stands out is the relatively limited penetration of VSLAs into local communities. Informal savings groups are common in Somalia and it is not uncommon for groups to form organically. However, among a total of 910 caregivers surveyed in the midline evaluation, just 6.5 percent were participating in a savings group.

At the same time, the share of caregivers participating in a savings group has increased over time. Using a sample that is comparable across rounds, the share of caregivers participating has increased from 4.6 to 8.8 percent from baseline to midline. This suggests that the programme may be increasing participation, but these improvements are coming from a low starting point. A similar outcome can be seen among NFE girls, which provides even stronger evidence of the programme’s impact on savings group participation. Among the “old cohort” of NFE girls, 27.6 percent have participated in a savings group in the past, compared to just 3.8 percent among the “new cohort” of NFE girls for whom this evaluation serves as a baseline. This suggests that the baseline rate of savings group participation in AGES communities might be quite low, as girls newly recruited into the programme – who are unlikely to be enrolled in savings groups at this early stage – have very low participation rates. In contrast, girls who have benefited from two years of programming participate at a starkly higher rate, highlighting the role of the programme in savings group participation.

Just as important as overall savings group participation is the extent to which these groups remain active. According to our sample of caregivers – among whom participation is fairly low – a significant majority (83.1 percent) report that their group remains active. NFE girls participating in savings groups paint a less optimistic picture – among this group, just 36.4 percent report that their savings group is still active, including 38.1 percent among “old NFE girls,” whose groups may have been started as early as late 2019. While it may seem compelling that over one-third of girls report that their savings group remains active, this is a fairly low share, in light of the short period over which the groups have been active and the fact that the programme is still active. We can expect the number of defunct groups to increase once the programme ends.

The current status of AGES-supported VSLAs is difficult to gauge, and our approach only provides suggestive evidence of their relative strength. We cannot say whether groups meet frequently, have effective leadership, are viewed as trustworthy, and have significant group assets saved and available for lending. These factors will be critical to the longevity of VSLAs over the long run. Our evidence suggests two shortcomings that, combined, suggest poor sustainability of the current VSLAs: first, comparatively few AGES households participate in savings groups, considering the “natural” frequency with which these groups operate. Second, among those who have participated in savings group, many report that their groups have been disbanded or are no longer active. It is difficult to say whether these findings should be linked directly to the programme, given that VSLAs are formed by other organisations and also occur outside the scope of explicit programmes. However, the fact remains that relatively few girls and women in AGES communities are in an active savings group and the savings groups that exist in these communities



have a marginal track record of longevity. The current situation suggests that much greater institutionalisation of VSLAs is necessary before they will be self-sustaining in the long run.

### Sustainability Indicator Score: 1 (Latent)

#### Proportion of parents able to support costs of girls' education

Expanding on the discussion of VSLAs in the previous section, we now consider whether parents are better situated to support the costs of girls' education at midline, compared to baseline. Participation in VSLAs may help households achieve greater financial stability, but this indicator also connects to income-generating activities, bursaries and cash support provided to girls' households, and broader economic trends in AGES communities.

At an attitudinal level, there is an increasing sense that cost and economic scarcity are not valid reasons to remove a girl from school or prevent her from being educated. For instance, the share of caregivers who feel it is acceptable for a girl to miss out on schooling because "education is too costly" has fallen from 61.3 percent at baseline to 45.3 percent at midline. The number of caregivers who cite cost or financial burden as the specific reason that *their* daughter is not enrolled in school has also fallen since the baseline, even when restricting our analysis to the same set of respondents over time. This provides reasonably strong evidence of attitudinal change regarding trade-offs between household finances and girls' education – with opinions moving in favour of greater educational access for girls, even when it is financially burdensome.

Over the same period, households have become at least somewhat more financially secure. Despite the vagaries of a drought that has worsened significantly since data collection ended, fewer households report having gone hungry in the last year than at baseline – the share of households that did not ever go to bed hungry in the last year rose from 21.4 to 29.9 percent over the last two years. The share of households that have savings of some kind has also risen; while still low (8.3 percent), it is almost four times as high as at baseline, among households that have been included in the programme over that period.

One potential cause of these improvements, despite the economic hardships brought by drought and the COVID-19 pandemic, are cash stipends and bursaries provided to families through AGES. At baseline, 71.1 percent of respondents had received financial support of some kind; among the newer NFE cohort, 67.8 percent have received financial support at midline. Unsurprisingly, financial support of this kind is seen by caregivers as critical to facilitating enrolment and attendance – in the current round, 90.6 percent of respondents who have received financial support say it has made their daughter's enrolment more likely, and 91.0 percent say it has improved her attendance. There are a number of pathways by which financial support can improve attendance – even in the absence of school fees, the labour provided by girls can be remunerative, providing an incentive for her to work inside or outside the home, in lieu of schooling. There are also direct expenses that inhibit attendance – such as the cost of sanitary pads – that can be alleviated via fairly limited financial support.

The financial support provided through AGES will not outlive the programme. It is reasonable, therefore, to assume that households will become slightly worse off as the programme winds down. At the same time, the improved financial standing since baseline is unlikely to have been solely the result of receiving a cash stipend or bursary. This provides a measure of hope that households' abilities to support girls' education – for the daughters in our cohort or younger daughters – will remain elevated after the end of direct programme support.

### Sustainability Indicator Score: 2.5 (Emerging/Becoming Established)

#### Parental support for girls' and boys' participation in GEFs and BEFs



To assess the sustainability of the GEFs, we surveyed members who have varying levels of exposure to CARE programming – those who have been beneficiaries since its inception, those who were original participants but who have graduated from the program, as well as new entrants into AGES. The GEFs aim to develop community-level social networks which will maintain and strengthen the programs and ideals championed by CARE in perpetuity. To do so they empower girls to take active roles in their community through activities which promote topics and causes important to women, including financial empowerment, education, and health.

To assess their long-term viability, we asked AGES girls whether they were GEF members and, if so, whether they were still connected with fellow members and what types of activities they have participated in under the GEF umbrella. Some high-level takeaways from our aggregated data are that a relatively small percentage of girls have participated in a GEF (14 percent) but their connection to fellow members is relatively strong, with 55 percent reporting that they are still in contact with other GEF members. They have also chosen to take on high-profile and high-impact causes. These include raising awareness in their community for sensitive women's issues and participating in CECs. Both activities are highly visible within communities, lending the GEFs a platform that may translate into recruitment opportunities beyond CARE's programmatic mandate.

Since our sample contains girls who have been participating in AGES for varying lengths of time, and as such have different levels of exposure to GEFs, we can look at how program recruitment and retention evolves as the program developed. Additionally, some girls have graduated from their treatment arm, which gives us a view into how much participation may wind down after CARE's programs expire. To parse out these time-dependent effects we disaggregated the girls in our sample into those who had been in the program since its inception in 2019 and are still (Group 1), those who were original members but whose program arms have mostly winded down (Group 2), and lastly those who were recently enrolled (Group 3).

**TABLE 31: PARTICIPATION RATES AND ACTIVITIES OF GEFs, BY COHORT**

Outcome	Aggregate	Formal (FE) Girls	ABE and "Old" NFE Girls	"New" NFE Girls
<b>Participation Rate</b>	14%	16%	29%	5%
<b>Percent Still in Contact with GEF Members</b>	55%	68%	50%	68%
<b>Girls Educational Support</b>				
<b>Enrolling OOS Girls</b>	49%	46%	46%	68%
<b>Teaching Other Girls / Study Groups</b>	14%	18%	13%	18%
<b>Participation in CECs</b>	17%	18%	16%	27%
<b>Changing Community Attitudes</b>				
<b>Awareness Raising</b>	90%	90%	97%	56%
<b>Trainings</b>	28%	11%	34%	21%
<b>Preventing Early Marriage</b>	20%	18%	19%	24%

<b>Business &amp; Finances</b>				
<b>Joint Business</b>	5%	2%	15%	21%
<b>Savings Group</b>	13%	0%	6%	6%
<b>Total Activities</b>	2.39	2.04	2.48	2.38

As the table above shows, there is considerable variation in their participation and retention, including a noticeable 18 percentage-point drop off between those who continue to receive the program (1 & 3) and those who no longer are (2). But the 50% continued-contact rate for Group 2 girls is still considerable and may suggest that there is a relatively high lower bound for continued relationships. The Group 2 girls also exemplify the strength of their engagement by having participated in a wider variety of activities, two and a half on average.

Taken together, girls who participated in GEFs take part in impactful, long-term, and high-profile activities. Within their communities these activities are high-profile and meaningful, both qualities which are likely to provide GEF members with a sense of purpose and group unity. The decline in between-member relationships seen by program graduates is not strikingly large and remains at a high level. And the participation rate's doubling between Group 1 and Group 2 girls gives some evidence that, given time, the Group 3 girls who follow them will continue that rising trend.

### **Sustainability Indicator Score: 2 (Emerging)**

#### **Proportion of GWDs who remain in school**

AGES seeks to target the most marginalised girls in south-central Somalia's urban areas, and this necessitates a particular focus on girls with disabilities (GWDs), who face particularly onerous barriers to enrolment and participation in school. Disability prevalence in Somalia is not known with precision, but is theorised to be higher than the global average, given its recent and ongoing history of conflict and the poor condition of medical infrastructure and staffing. In under-resourced schools and poor communities, GWDs face significant practical obstacles to school attendance, due to a lack of assistive equipment and teachers who are not trained to adapt teaching or school administration to the needs of GWDs. This is combined with often intense stigma regarding disabilities in Somalia, such that many teachers we interviewed noted that their efforts to promote attendance by GWDs was primarily oriented around reducing discrimination and teasing.<sup>77</sup>

AGES targeted GWDs for enrolment in formal schooling and non-formal learning programmes. However, in order for the programme's gains on this front to be sustained, it is important to assess how many GWDs have remained enrolled two years later, at midline, particularly given the dislocation of COVID-19 and the winding down of the NFE interventions in which some GWDs took part.

In the aggregate, enrolment rates – where enrolment is classified as current participation in any learning programme, not exclusively formal primary school – among the three cohorts tracked since baseline are 52.8 percent, with much higher rates among the formal school cohort and much lower rates among the NFE cohort. When we analyse GWDs, their enrolment rates are not dramatically different from those of non-GWDs: in the aggregate, 49.8 percent of GWDs are enrolled. This rate increases to 53.1 – actually higher than non-GWDs – when we use an alternative, more liberal definition of disability status.

<sup>77</sup> See, e.g.: FGD with CEC Members, MiddleShabelle, Int. 107.

Unfortunately, this aggregate trend is driven by comparatively higher enrolment rates among girls with mental health disabilities (severe anxiety and/or depression), while enrolment rates among girls with physical, cognitive, and communicative disabilities are much lower. When we study the impact of disability status more systematically, we find that girls with any disability *other* than a mental health disability (i.e. physical, cognitive, communicative) have enrolment rates 7.1 to 7.9 points lower than non-GWDs. After accounting for age and region – in addition to cohort – girls with these types of disabilities are 4.3 to 4.8 points less likely to be enrolled than non-GWDs.

On the other hand, girls with mental health disabilities have enrolment rates 4.4 to 5.4 points *higher* than non-GWDs, a finding that is robust to the inclusion of basic control variables in a linear regression. Given that mental health disabilities are much more common in our sample ( $n = 188$  versus  $n = 46$ , in a panel sample at midline of 1,062 respondents) than other forms of disability, the higher enrolment rate among girls with mental health disabilities is obscuring the much lower enrolment rates among girls with physical, cognitive, and communicative disabilities.

The finding that girls with non-mental health disabilities are particularly disadvantaged is consistent with a significant amount of qualitative evidence. Indeed, when we interviewed teachers and CEC members about how they accommodate or promote enrolment of GWDs, they focused almost entirely on those with physical disabilities. For girls with vision and hearing difficulties, few solutions are possible – at best, teachers and administrators are able to seat these students near the front of the class, but a lack of assistive devices is still a hard limitation.<sup>78</sup> Interviewees seem to view mobility disabilities – difficulty walking – as more problematic, because it is a stronger impediment to getting to school. Again, for these girls, there are few solutions, as assistive devices are rare, the distance to schools – even in the urban areas served by AGES – are significant, and transportation options that could accommodate girls with mobility challenges are too expensive for poor families and communities.<sup>79</sup>

We note the particular difficulties faced by girls with physical disabilities because this helps explain their much lower enrolment rates. Perhaps the two largest barriers to enrolment of GWDs in Somalia are the physical challenges of attending school without assistive devices or necessary equipment – an issue that affects those with physical disabilities but not mental health disabilities – and social stigma – again, an issue that is more often directed at children with physical disabilities, rather than mental health disabilities that are invisible to most.

The evidence in this section suggests three general conclusions. First, the programme has done a commendable job of enrolling GWDs – according to even a strict classification of disability status, 11.9 percent of girls recruited at baseline and 20.4 percent of girls enrolled in the newest NFE cohort have a disability of some kind. Second, their continued enrolment over the last two years, while lower than those of non-GWDs, is still reasonably high. Third, however, the rate of enrolment among girls with physical disabilities is much lower and may not be sustained in the absence of programmatic support – not in the form of assistive devices, but awareness-raising and encouragement of CECs and teachers to reach out to and promote enrolment among this population. Without the explicit encouragement and monitoring provided by the programme, it is very possible that efforts targeting this marginalised group will diminish and enrolment rates will slip further, particularly among those with physical disabilities.

<sup>78</sup> FGD with CEC Members, Bay, Int. 101; FGD with CEC Members, LowerJuba, Int. 104; FGD with CEC Members, LowerJuba, Int. 105; FGD with CEC Members, Baidoa, Int. 109.

<sup>79</sup> FGD with CEC Members, LowerJuba, Int. 105; FGD with CEC Members, MiddleShabelle, Int. 107; FGD with CEC Members, Bay, Int. 101; FGD with Mothers, Bay, Int. 203.

## Sustainability Indicator Score: 2 (Emerging)

### Proportion of GEFs and BEFs implementing community actions to support attendance and retention

As education is inherently empowerment, and girls are particularly disadvantaged in Somalia's cultural and educational system, GEF activities are particularly oriented towards the enrolment, attendance, and retention of girls, CARE's core programmatic objectives. The activities GEFs undertake in this vein can be grouped into three categories: direct support to girls' education, mitigating attitudes which indirectly prohibit their education, and female-led business promotion.

The first of these, direct support to girls' education, is likely to have the largest and longest-lasting impact on AGES communities. GEFs directly support girls' education by helping enrol out-of-school (OOS) girls, easing the learning burden on these new students as well as pre-existing ones through private tutoring and study groups, as well as by participating in their local CEC, through which meaningful institutional transformations could be made. The three groups of girls shown in the table in the previous section uniformly possess high rates of participation in activities aimed at OOS enrolment, participation that is particularly pronounced among Group 3, who may be enthusiastic in part due to their own experience without school over the past two years. In both study groups and CEC participation Group 2, for whom the program has largely ended, lags marginally behind the continuing participants in Group 1. Our analysis of this slight drop is in line with our thinking behind Group 2's lower rate of continued contact with GEF members – a drop is to be expected, and it is assuring that it is relatively small.

Societal expectations such as teenage marriage and childrearing play an outsized role in girls' school enrolment as well as dropout rates. The more senior program participants (i.e. those who have been involved in AGES the longest, Groups 1 and 2) near-universally contribute to raising awareness about issues such as these. The figure is much lower within Group 3, but that may be a function of their recent introduction to the program and having simply not had the opportunity to join those activities yet. And the final category, female-led business promotion, may improve attendance and retention by increasing the real or perceived returns to education. Savings groups increase women's access to capital within the community and are a mechanism through which women can increase the returns to their own businesses. In this Groups 2 and 3 are particularly entrepreneurial relative to Group 1, with 15 percent and 21 percent respectively engaging in joint businesses with other GEF members.

## Sustainability Indicator Score: 2 (Emerging)

### Proportion of umbrella schools adopting new methodologies

AGES programming is intended to expose teachers to new methods for teaching in order to help improve learning outcomes for students. Exposure to new teaching methods can potentially result in the long-term adoption of these methods by teachers targeted by the programme, thus improving sustainability. These teachers' peers may also observe the successful use of new methodologies and be interested in or incentivised to adopt new teaching methods themselves, further expanding the impact and sustainability of AGES programme interventions. Furthermore, CARE's coordination and involvement with the Ministry of Education may influence this ministry to help train teachers or disseminate information on new teaching practices, another factor which would strengthen programme sustainability.

To examine this aspect of sustainability, in the school survey, respondents were asked how many male and female teachers at the school had been trained in a variety of skills over the past year. The below table shows the total and average number of teachers per school receiving training by skill area at baseline and midline.

TABLE 32: NUMBER OF TEACHERS TRAINED AT BASELINE AND MIDLINE

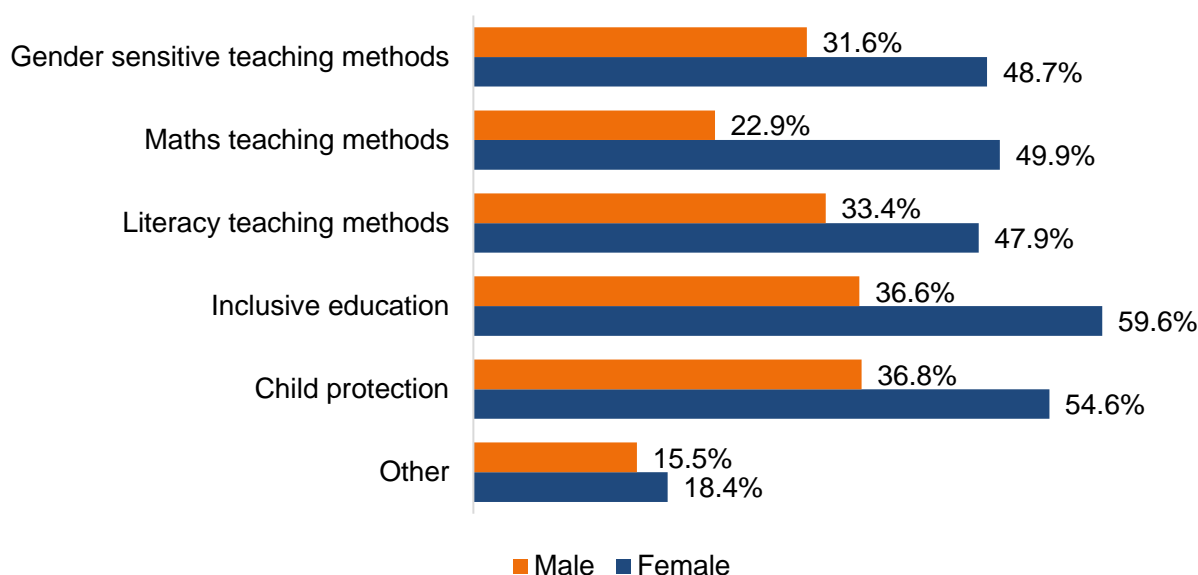
Training area	Total		Average per school	
	BL	ML	BL	ML
<b>Gender sensitive teaching methods</b>	73	296	2.8	8.0
<b>Maths teaching methods</b>	141	213	4.1	5.8
<b>Literacy teaching methods</b>	213	277	6.3	7.5
<b>Inclusive education</b>	266	340	7.8	9.4
<b>Child protection</b>	129	323	4.4	9.0
<b>Other</b>	38	130	1.4	3.5

At midline, a significantly greater average and total number of teachers had received training on gender sensitive teaching methods over the past year than at baseline. A higher—though not significantly so—average and total number of teachers also received training on maths and literacy teaching, inclusive education, child protection, and other topics. This data suggests that AGES programming has been at least moderately successful in improving teacher access to and participation in trainings (whether AGES-sponsored or otherwise), a positive sign for programme sustainability.

Additionally, the below figure shows the average percent of male and female teachers per school who received training in each of the above areas at midline. Notably, on average, female teachers were substantially more likely to have participated in trainings than male teachers.<sup>80</sup> This may occur if female teachers are prioritized by schools for trainings, or targeted by the Ministry of Education, AGES programme implementers, or other groups. This may have a positive impact on girls' education if, for example, girls are more likely to participate in classes taught by female teachers. However, it is also important for male teachers to receive trainings to ensure programme efficacy and sustainability, particularly on topics such as gender sensitivity and inclusive education.

<sup>80</sup> We note that there were fewer total female teachers than male teachers, so these numbers do not necessarily correspond to a greater absolute number of female teachers receiving trainings than male teachers.

FIGURE 13: PERCENT OF MALE AND FEMALE TEACHERS PARTICIPATING IN TRAININGS AT MIDLINE



Qualitative data from interviews with teachers suggests that teachers are not only participating in trainings, but also implementing new teaching methods taught in trainings in at least some cases. One teacher, for example, stated that he had started using visual examples and materials when teaching math because he had been taught to do so in a training:

*After we received training on how to present lessons, we provide students the lessons in audio and video format, for example in math when we teach them how subtract we bring books and visually subtract one book from another to make them understand.*

FGD with teachers, Banadir, Int. 508

However, other teachers mentioned that they had not yet received training, and as a result, had not changed their teaching practices over the past year.<sup>81</sup> Furthermore, when asked what was needed to improve their teaching, many respondents emphasized that there was a need for more training.<sup>82</sup>

### Sustainability Indicator Score: 2 (Emerging)

## Proportion of teachers implementing inclusive education strategies in class

A key component of AGES programming is not just that teachers implement improved teaching practices, but also that their teaching practices are inclusive (of male and female students, students from different clans, IDP students, students with disabilities, and other groups). Training on inclusive education—as with teaching methodologies—may help improve programme sustainability by providing teachers with strategies for inclusivity that they will continue to use throughout their careers, even after programme support is removed. Advocacy work with the Ministry of Education may also strengthen the sustainability

<sup>81</sup> For example, FGD with teachers, Banadir, Int. 506; FGD with teachers, South West State, Int. 501

<sup>82</sup> For example, FGD with teachers, Banadir, Int. 503; FGD with teachers, Banadir, Int. 504; FGD with teachers, Banadir, Int. 507

of interventions aimed at improving inclusivity by enabling the government to continue implementing inclusivity-related interventions after programme support ends.

The above section shows that at midline, 340 teachers reported receiving training on inclusive education in the past year. However, in the analysis of gender equity in the *Teaching Quality and Practices* section, we found mixed results for gender inclusivity in the classroom: Girls and boys were observed to have consistently equal access to learning materials, have similar amounts of time to answer questions, and receive similar amounts of positive (and negative feedback), but only half of teachers were observed asking questions to girls and boys equally. Given that there have been mixed changes in gender equity indicators at programme midline—i.e., while programme implementation has already occurred and is still ongoing—these findings have potentially concerning implications for sustainability. The use of inclusive education strategies is likely to be highest while interventions are actively targeting teachers to encourage implementation of these strategies, and is likely to then fall off somewhat once support (from AGES implementers, the government, or others) is removed. As such, a low level of inclusivity mid-programme suggests that sustainability of inclusive teaching practices may be a challenge.

However, while these results for gender equity suggest more needs to be done to strengthen gender-inclusive teaching practices to ensure sustainability, results from the household survey suggest a potentially more positive outlook for other dimensions of inclusivity. For example, among girls who dropped out of school, less than 1 percent (only two girls total) stated that they dropped out because the teacher mistreated them or discriminated against them. No girls who dropped out stated that they did so because “the teacher does not know how to teach someone like me;” this includes eleven girls with a mobility-related disability, one girl with a hearing disability, and four girls with a vision-related disability. Respondents to the household survey also overwhelmingly answered that children with disabilities have a right to go to school, with more than 99 percent of respondents agreeing with this statement.

IDP students stated that teachers made them feel welcome at approximately equivalent rates as non-IDP students (96 percent for both groups).

Given the limited amount of information present in the data regarding inclusivity along non-gender-related lines, although these results are positive, they are not conclusive. We also note that there has been very little change in indicators related to IDP inclusivity from baseline to midline, suggesting positive results may not be attributable to programme impact. In order to further examine sustainability of inclusive education strategies—including minority status, disability, and IDP inclusivity—it may be useful to collect additional qualitative or quantitative data at endline in order to better understand how teaching practices vary by subgroup.

### **Sustainability Indicator Score: 2 (Emerging)**



## 7. Intermediate Outcomes

### 7.1 Attendance

Attendance comprises the first intermediate outcome for the AGES programme, and one which is expected to directly drive improvements in both learning and transition outcomes. Attendance is viewed as likely to affect other programme outcomes as a result of:

- improved learning opportunities, along with the positive community and parental attitudes to girls' education will positively contribute to the attendance rates of the marginalised girls;
- improved learning environment along with the increasing attendance rates will lead to improved learning outcome in the specific subject/areas of interventions of the marginalised girls;
- improved attendance rates, often identified as the major predictor of the school dropouts, will have a positive impact on the girls' transition to formal / non-formal education or employment.

Data on attendance is drawn exclusively from attendance headcounts conducted during site visits by the external evaluator's field teams. In the context in which AGES is being implemented, direct headcounts are generally considered the most reliable approach to assessing attendance rates, as record-keeping at schools and learning centres is often poor.

The headcount surveys were administered in each grade or stream. If there was more than one class in the same grade or stream, team leaders completed a headcount in *every* class or stream/level in the centre. This provides the maximum possible sample of headcounts for analysis. As we do elsewhere in this report, we limit our analysis of changes from baseline to midline to the comparable sample of schools or centres. For instance, we do not include schools or centres which were visited at baseline but excluded – for security reasons – at midline, as this would introduce bias in our calculation of attendance rates. Instead, we construct a “comparable sample” of schools – those schools appearing in both samples – and calculate attendance rates from among the classes surveyed in those schools only.

Attendance headcounts have two shortcomings that should also be noted. The first is that they provide only a snapshot of attendance on a single day, and attendance rates can fluctuate significantly from day to day. While this does not introduce systematic bias, it does increase the variability in our results and decreases our confidence in the results slightly. Second, the measurement of attendance is based on the total enrolment numbers – which serve as the denominator in our calculation – meaning that the validity of the attendance rate calculation is dependent on the reliability of enrolment records.

TABLE 33: ATTENDANCE RATES, BASED ON HEADCOUNTS, IN FORMAL SCHOOLS

State	Baseline	Midline
<b>Banadir</b>	89.3%	89.8%
<b>Jubaland</b>	76.6%	74.7%
<b>South West State</b>	91.7%	85.6%
<b>Total</b>	89.3%	84.1%

The table above reports headcount attendance rates for formal primary schools at baseline and midline. Note that the AGES cohort girls enrolled in primary school at baseline entered grades 1 or 2 exclusively and are now concentrated predominantly – among those still enrolled – in grades 3 through 6. However, our analysis utilises attendance from *all* classrooms in each school, including those in higher grades,



because our interest is in documenting school-level performance. This can be understood in light of the advancement of cohort girls from grades 1-2 to grades 3-6 and beyond, in many cases, because attendance levels across all classrooms in a given school are relevant to a girl's learning outcomes if she moves through those grade levels over time.

As the table above shows, attendance rates have generally stayed steady in Banadir and Jubaland, with a slight decline from baseline to midline in the latter state. However, attendance rates in South West State have declined more significantly, dropping 6.1 points, a change that is statistically significant. The decline in South West State – relative to those in other states – is difficult to explain, as the region is not facing greater challenges, in terms of drought or food security than its neighbour, Jubaland, to the south. The decline also cannot be explained as a function of small sample size and variability that might stem from that, as the sample includes 68 headcounts in South West State at baseline and another 57 headcounts at midline. One possible explanation is simply reversion to the mean, as attendance rates reported at baseline from South West State were inordinately high – higher than the other states in the sample and – to the best of our knowledge – higher than those observed in any of our previous GEC or GEC-T evaluations in Somalia. Driven largely by the declines in South West State, overall attendance rates based on headcounts have fallen by 5.2 points since baseline among formal primary schools.

In addition to primary schools, the baseline and midline evaluations collected headcount data from ABE and NFE centres. As in primary schools, headcounts were completed in every eligible classroom, though most ABE and NFE centres have a more limited number of levels/classes, resulting in a much smaller sample size of headcounts. For instance, at midline, our total sample of ABE headcounts is just 25 classrooms.

As the table shows, attendance rates in ABE centres have improved over time, increasing by 6.0 points, while rates in NFE centres have stayed steady since the baseline. The comparison of attendance rates in NFE centres is slightly unwieldy and should be taken with a grain of salt, as the cohort of NFE girls tracked from baseline to the midline completed their learning courses in late 2020 and have moved out of the NFE centres as of that time period. Therefore, the attendance rates reported for NFE centres at midline are based on headcounts with later NFE classes, which do not generally include any of the cohort girls tracked for this evaluation.

**TABLE 34: ATTENDANCE RATES, BASED ON HEADCOUNTS, IN ABE AND NFE CENTRES**

Type of Centre	Baseline	Midline
<b>ABE Centre</b>	79.6%	85.6%
<b>NFE Centre</b>	84.0%	84.1%

## 7.2 Teaching Quality

We now examine four components of teaching quality and practices: overall professionalism, gender equity, the use of physical punishment, and pedagogical practices. Each of these aspects of teaching quality is linked to both learning and transition outcomes: A girl will likely learn more when teachers use high-quality and inclusive teaching practices, which may also incentivize girls to attend and stay in school. In contrast, if teachers are often absent, do not engage students, or use abusive teaching practices, girls may be more inclined to drop out.

Two data sources are used in the below analysis. First, direct, in-person classroom observations were conducted to better understand teachers' practices within lessons. At baseline, classroom observations were conducted only in formal schools; at midline, observations were also conducted in ABE and NFE schools. Note that our analysis of data from classroom observations from FE schools uses the panel sample

of schools observed at both baseline and midline unless otherwise noted. Second, girls were surveyed regarding teachers' behaviours within classes and with students.

While we generally prefer the slightly less subjective data provided by direct observation of classroom practices, we note three shortcomings of the classroom observations in this specific case. First, the fact that observations were not conducted in ABE or NFE centres at baseline means that we cannot assess change in pedagogical practices over time. Second, the sample size of ABE and NFE centres is also very small – just 27 ABE classroom observations, for instance. Finally, because the first cohort of ABE and NFE students had both completed their programmes by the time of data collection, most observations in these centres were conducted in classrooms populated by students that are not part of the tracked cohort; rather, they are populated by a later cohort of ABE and NFE students.<sup>83</sup> This disconnect between the girls being tracked and the classrooms observed is somewhat problematic, as we cannot draw direct conclusions regarding the quality of teaching to which our tracked sample of girls was exposed. With that said, given that observations were conducted in the same centres – but with later cohorts – it is likely that teaching practices observed are *indicative*, as teaching practices are likely to be similar between the previous and current classes within the same centres.

## Teacher Professionalism

To analyse teacher professionalism, we examine responses to two questions in the survey with girls in which girls were asked whether they agree or disagree with the following statements: (1) “My teachers make me feel welcome in the classroom” and (2) “My teachers are often absent for class”. We disaggregate this analysis by the education track of the respondent, the region of the respondent, and baseline/midline, and focus our analysis on the panel sample of girls contacted at both baseline and midline.

The table below shows results for these two statements. We find that girls overwhelmingly state that their teachers make them feel welcome in the classroom; over 95 percent of respondents agreed strongly or somewhat with this statement, and results were consistent across education tracks and regions with no significant change (and little absolute change) from baseline to midline. This suggests that teachers are (and consistently have been) effective at creating a welcoming learning environment.

Results for teacher absenteeism, however, reveal some challenges. At midline, on average, around one-fifth of girls agreed strongly or somewhat that their teacher was often absent. Rates of reported teacher absenteeism were fairly consistent across all three education tracks, ranging from around 19 percent for ABE girls to around 22 percent for FE girls. It is important to note, however, that there was significant variation in results by region. Reported teacher absenteeism fell sharply and significantly in Banadir and SWS, but *increased* significantly—to more than 50 percent—in Jubaland. Further disaggregating by district, we find that these results are driven by very high reported absenteeism rates in Kismayo (68 percent) and Dhobley (79 percent). The reason for these high absenteeism rates is not entirely clear;

---

<sup>83</sup> In those NFE centres in which both Cohort 1 and Cohort 4 students are being tracked, classroom observations at midline were completed with the Cohort 4 NFE classrooms. In NFE centres that only included Cohort 1 students, observations were conducted with whichever classrooms were active, though – again – the students in such classes are not part of a cohort tracked by the evaluation team. Finally, in ABE centres, most students in observed classrooms are not part of a cohort tracked by the evaluation team, except cases in which a student has re-enrolled or remained enrolled in ABE for longer than intended (see Section 5, on transition outcomes, for documentation of the share of ABE girls who remain enrolled in ABE centres at midline).

while Kismayo and Dhobley have faced insecurity over the last year, it has not been notably more severe than in many other locations also targeted by the programme.

**TABLE 35: CHANGE IN TEACHER PROFESSIONALISM BY TRACK AND REGION**

Outcome	FE			C1 NFE			ABE		
	BL	ML	Diff.	BL	ML	Diff.	BL	ML	Diff.
<b>Feels welcome</b>	96.1%	97.6%	1.5	96.7%	96.4%	-0.3	96.7%	97.3%	0.6
<b>Rarely absent</b>	28.1%	21.7%	-6.4	27.5%	20.6%	-6.9	27.9%	18.8%	-9.1
	Banadir			Jubaland			South West State		
	BL	ML	Diff.	BL	ML	Diff.	BL	ML	Diff.
<b>Feels welcome</b>	94.8%	98.0%	3.2	97.5%	95.4%	-2.1	98.5%	97.1%	-1.4
<b>Often absent</b>	25.2%	6.6%	-18.6*	22.1%	50.7%	28.6*	38.6%	13.8%	-24.8*

We note that teacher absenteeism may, at least in part, be driven by high levels of insecurity, including violence and economic insecurity, within the areas where AGES operates. When asked about the challenges faced by teachers, many respondents stated that in insecure areas, teachers might not attend school because they felt endangered:

*Teachers face a variety of challenges, including fear or insecurity in the workplace because they are unable to attend places where there is no security because they are in danger.*

- FGD with CEC members, Jubaland, Int. 101

Respondents, however, also frequently mentioned that teachers' salaries were low, which reduced teachers' incentives to attend classes regularly.

## Classroom Gender Equity

We now analyse differential treatment of students by gender within the classroom and teacher attitudes towards gender equity. We do not assume that teachers hold positive views about gender equality or equity, given that AGES is implemented in some of the most marginalised areas of Somalia where girls are subject to a wide range of barriers, including cultural barriers. Understanding teachers' attitudes towards gender roles and girls' education is thus an important metric of long-term programme impact.

To analyse gender equity, we use data from classroom observations and surveys with girls. In the below analysis, we analyse both questions regarding gender equality and those regarding equity. Equality, in this case, refers to a situation in which girls and boys are given exactly the same opportunities to, for example, participate in class. Equity, in contrast, refers to a situation in which girls may be given additional opportunities in order to help overcome the effects of systematic discrimination or barriers facing girls' lives that may prevent them from reaching their educational potential. As one CEC member eloquently stated:

*In my opinion, girls are more vulnerable than boys; you can see girls' needs increasing as they get older; they require things that boys do not.*

## - FGD with CEC members, Jubaland, Int. 105

The AGES programme seeks not just to ensure equality of opportunities, but to be gender transformative and to allow girls to overcome a wide range of barriers. In order to achieve this goal, gender equity—not just equality—is necessary.

However, several questions employed during the evaluation assess gender equality, rather than equity. For example, during classrooms observations, enumerators observed whether teachers asked girls and boys questions that were equally demanding. In cases where the teacher was not observed to ask equally demanding questions, it is not clear whether more difficult questions were asked to girls or boys. Differences in the difficulty of questions asked may not be a negative; if, due to systematic discrimination in past learning, girls are starting from a lower level of learning than boys or are less comfortable answering questions than boys, then asking girls easier questions may help engage girls in class and build their confidence. Unfortunately, we do not measure these dynamics.

Given these methodological challenges, we first review evidence from questions measuring gender equality. We then discuss in more detail questions dealing with gender equity. In these questions, we can distinguish between outcomes that favour boys, those that favour girls, and those that treat the two equally. The analysis below focuses on FE schools, as classroom observations were not conducted in NFE or ABE programmes at baseline, and only a limited number of observations were conducted in these school types at midline.

We first analyse responses from girls who were asked whether teachers treat boys and girls differently in the classroom. At baseline, 34.9 percent of girls in FE agreed strongly or somewhat with this statement, while at midline, 40.7 percent of girls in FE agreed. This increase was not significant; regardless, these numbers still represent a large percentage of girls who experience differential treatment from teachers. While the question does not distinguish between teachers favouring girls versus boys, we expect girls to be more likely to identify differential treatment if it disadvantages them. As such, it is possible to interpret this finding as evidence—though not conclusive—that many girls face inequity in the classroom.

To further investigate this, we now review evidence from classroom observations. The below table shows mixed results for changes in measures of gender equality between baseline and midline in FE schools.<sup>84</sup> We do not analyse results for ABE or C1 NFE below, as many of these schools have no or very few boy students. At baseline, almost all classrooms were observed to provide equal access to learning materials for girls and boys; at midline, this rate declined, although not significantly. A significant decline was found, however, for teachers directing questions equally at both genders; at midline, teachers were only found to direct questions to both genders equally in most of their classes 50 percent of the time. In contrast, however, there was a large and significant improvement in teachers' provision of the same amount of time to girls and boys to answer questions, while there was a substantive, though not significant, increase in the percent of teachers asking equally-difficult questions to students of both genders. Given these mixed results, and the measurement issues regarding equality as opposed to equity discussed above, we do not wish to overstate any conclusions; instead, we now turn to a discussion of gender equity.

<sup>84</sup> we note that for the last three questions listed in the table, the behaviour was considered to be observed if it was observed by the enumerator in at least two out of three class blocks.

TABLE 36: CHANGE IN MEASURES OF GENDER EQUALITY

Outcome	BL	ML	Difference	p value
Equal access to learning materials	97.0%	89.7%	-7.3	0.31
Same amount of time to answer questions	45.2%	84.6%	39.4	0.02*
Questions directed at both genders	73.9%	50.0%	-23.9	0.04*
Questions equally difficult	25.0%	48.7%	23.7	0.15

Our main measures of gender equity in the classroom observations include the number of times teachers called on boy and girl students, whether positive feedback was provided to boys and girls, and how often harsh language was used with boys and girls. The below table reports results for these three metrics of equity. For the number of times teachers called on girl and boy students respectively, we first note that at midline, boys attempted to answer or ask a question significantly more than girls—an average of 7.6 times over three blocks for boys compared to 6.8 times for girls. There were also significantly more boys than girls within classes (an average of 27.8 boys and 23.0 girls per class). These two dynamics may explain, at least in part, why teachers were observed calling on boy students more often (though not significantly more often) than girl students at midline. Even if this is the case, however, greater participation from boys than girls may represent a concerning trend whereby girls feel less comfortable or confident within classrooms than boys.

Within the qualitative data, while many respondents stated that girls were free to participate in class and ask questions, several teachers mentioned barriers to girls' participation. One teacher in Banadir, for example, stated that girls may feel more shy when asking questions to male teachers than to female teachers.<sup>85</sup> Many teachers stated that in general, girls were more shy and more likely to feel embarrassed about asking questions than boys. This perception—while likely rooted in reality—may make teachers less likely to call on girls or support them in class, as they may believe that inherent differences between girls and boys make girls less able to meaningfully participate in lessons. As such, there is a need to both work with girls to continue improving their confidence and work with teachers to encourage them to increase girls' participation proactively.

For positive feedback, enumerators were asked only to record whether or not teachers provided positive feedback to girls and boys, not the number of times this occurred. As such, the below table shows the percent of teachers observed to provide positive feedback to girls or boys in at least two out of three observed class blocks. This data shows an increase (though not significant) in the provision of positive feedback to both girls and boys at midline, with a larger increase for girls than boys and with teachers providing positive feedback to girls and boys at the same average rate. These are positive trends, but we emphasize that given the discrimination faced by girls, it may be useful to further encourage teachers to provide additional positive feedback to girls to help overcome barriers which may make them more shy or less confident in class than boys.

Lastly, we find slightly more prevalence in the use of harsh language with boy students than girl students, but a larger increase in the use of harsh language against girls at midline than against boys. While this increase is not significant, it represents a potentially concerning trend, particularly given that it may discourage students from participating in or attending class.

<sup>85</sup> FGD with teachers, Banadir, Int. 505

TABLE 37: CHANGE IN MEASURES OF GENDER EQUITY

Outcome	BL	ML	Difference	p value
<b>Times teacher called on girl/boy</b>				
<b>Girl student</b>	5.8	6.1	0.3	0.78
<b>Boy student</b>	5.3	6.7	1.4	0.21
<b>Teacher provided positive feedback to girl/boy (at least two blocks)</b>				
<b>Girl student</b>	65.2%	80.0%	14.8	0.09
<b>Boy student</b>	75.7%	80.0%	4.3	0.74
<b>Times teacher used harsh language with girl/boy</b>				
<b>Girl student</b>	0.6	1.9	1.3	0.08
<b>Boy student</b>	1.6	2.3	0.7	0.50

## Disciplinary Practices

Our analysis of disciplinary practices examines whether teachers punish students for wrong answers and whether corporal punishment is used within the classroom. Both of these practices can have a negative impact on learning; in the former case, students may be afraid to take part in class lest they be punished for a wrong answer, while in the latter case, students may be deterred from attending school or staying in school if they face the threat of physical violence. We utilise two main data sources to understand these disciplinary practices: the survey with girls and classroom observations. We note that disciplinary practices observed during classroom observations likely represent a lower bound on their occurrence, as teachers are likely to reduce their use of negative practices while under observation.

The below table shows the change in use of these two disciplinary practices by education track. Analysing results from the survey with girls, we find that use of discipline for wrong answers and corporal punishment has significantly and substantively decreased at midline for all three cohorts of girls. This includes an approximately 50 percentage point drop in reported use of corporal punishment within classrooms; we note that reported use of discipline for wrong answers, while declining significantly, declined by relatively less than the use of corporal punishment. We also note that use of negative disciplinary practices is most prevalent among formal schools and least prevalent among non-formal schools. Overall, these are positive findings, but also suggest a need to continue to focus on reducing teachers' use of discipline for wrong answers, as this remains relatively more prevalent (especially among formal schools) than corporal punishment.

Looking at results from the classroom observation, we find a somewhat different trend. We note that enumerators only observed the presence (or absence) of physical punishment in four ABE and four C1 NFE programmes at midline; as such, we only report results for FE schools. For FE, among schools observed at both baseline and midline, we find an increase—though not a significant one—in the use of physical discipline against both boys and girls. The increase is somewhat larger for boys, at around 14 percentage points, than for girls, at around 10 percentage points. The reason for this gap in results between the survey with girls and the classroom observations is not fully clear. Given that classroom observations occur in only one classroom, which all surveyed girls may not necessarily attend, it is possible that observations occurred in classrooms where the use of physical discipline has increased abnormally. Alternatively, it is possible that at midline, girls felt pressure to report that teachers did not use corporal punishment, although we would expect this to have also affected baseline results.

TABLE 38: CHANGE IN USE OF DISCIPLINARY PRACTICES BY EDUCATION TRACK

Outcome	FE			C1 NFE			ABE		
	BL	ML	Diff.	BL	ML	Diff.	BL	ML	Diff.
Discipline for wrong answer (reported by girls)	80.4%	56.3%	-24.1*	57.9%	35.5%	-22.4*	67.6%	45.9%	-21.7*
Use of corporal punishment (reported by girls)	71.3%	20.8%	-50.5*	52.4%	10.3%	-42.1*	68.8%	14.6%	-54.2*
Boys disciplined physically (observed)	36.4%	50.0%	13.6	-	-	-	-	-	-
Girls disciplined physically (observed)	32.6%	42.2%	9.6	-	-	-	-	-	-

Further analysing by region in the below table,<sup>86</sup> we find that change in the use of negative disciplinary practices was most stark in Banadir where, notably, the reported prevalence of corporal punishment declined from almost 79 percent to a remarkably low 5 percent at midline. In contrast, we note that in South West State, there was no significant decline in the use of discipline for wrong answers (although there was a significant and substantive decline in the use of corporal punishment). As above, this suggests a need to continue focusing on reducing discipline for wrong answers, as well as to focus efforts on schools in Jubaland and South West State where negative practices are more prevalent and have declined relatively less over time.

TABLE 39: CHANGE IN USE OF DISCIPLINARY PRACTICES BY REGION

Outcome	Banadir			Jubaland			South West State		
	BL	ML	Diff.	BL	ML	Diff.	BL	ML	Diff.
<b>Discipline for wrong answer, reported by girls</b>	69.4%	33.5%	-35.9*	78.3%	59.7%	-18.6*	55.7%	53.1%	-2.6
<b>Use of corporal punishment, reported by girls</b>	78.8%	5.0%	-73.8*	57.4%	37.9%	-19.5*	79.7%	21.4%	-58.3*

<sup>86</sup> We do not analyse results from the classroom observations by region for these questions due to low sample size (below 20 observations).



Within the qualitative data, several respondents mentioned that students still faced punishment at schools. A teacher in Jubaland, for example, stated that it was difficult to manage classes if students did not fear the teacher; because of this, it was necessary to beat students. He further stated that while corporal punishment of students was forbidden by the Ministry of Education, teachers did not follow this guideline.<sup>87</sup> Another teacher in Banadir attributed the use of punishment to outdated practices experienced by teachers themselves when they were students:

*[Teachers] need training in how to deal with students. For example, teachers were punished when they were young, so they need to change the discipline so students can learn and be disciplined without punishment.*

- FGD with teachers, Banadir, Int. 506

However, many respondents also mentioned that practices had begun to change. Teachers in Banadir and Jubaland, for example, stated that teachers had previously punished students who made mistakes, but now tried to advise the students instead of punishing them.<sup>88</sup> Overall, these findings validate those from the quantitative analysis, and suggest that while use of discipline for wrong answers and corporal punishment still exists, these practices are changing over time.

## Pedagogical Practices

Our last metric for teaching quality is the pedagogical practices used by teachers. This includes the use of formative assessments, both as reported by the teacher during classroom observations and as verified by enumerators, who asked teachers to view the formative assessments. It also includes a number of teaching practices observed in classroom observations, including:

- Using student-centred activities or games
- Allowing students to instruct each other
- Asking open-ended questions that encourage thinking
- Asking questions that solicit students' opinions
- Trying to involve students that are not participating
- Allowing students to work together in groups

As in the section on gender equity, our analysis of data from classroom observations focuses on FE schools.

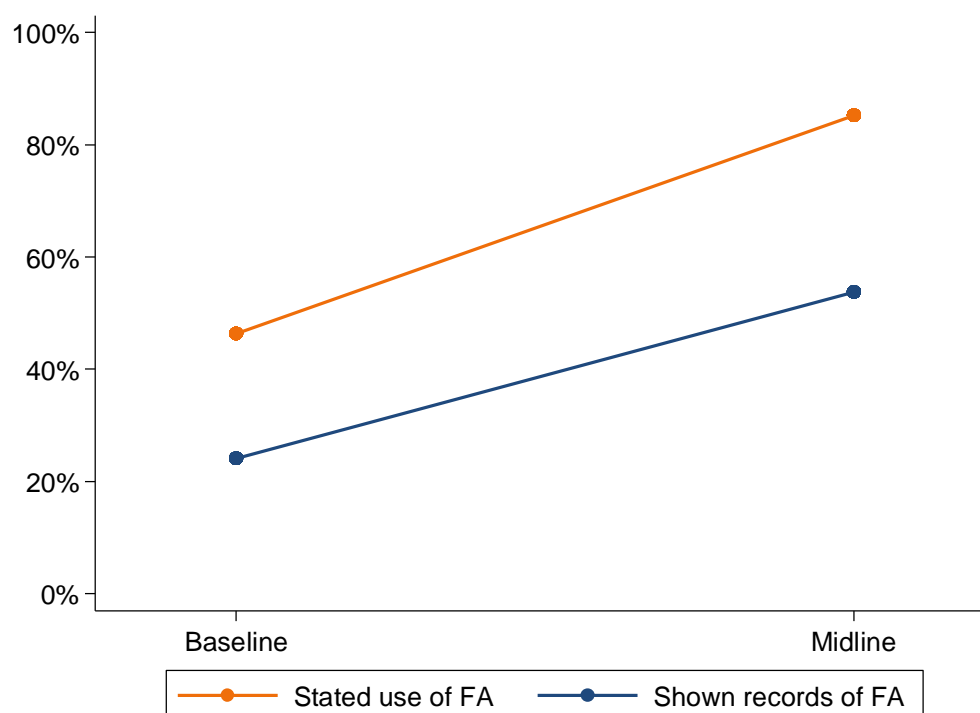
The below figure shows the change over time in purported and demonstrated use of formative assessments (labelled as "FA"). Both stated and demonstrated use of formative assessments increased substantively and significantly at midline. However, we note that teachers' abilities to produce records of formative assessments when asked to do so remains substantially below their reported use of assessments. This may occur if teachers overstate their use of formative assessments or if they do not keep proper records; either situation is problematic. We also note that, while sample size is low and results are thus not conclusive, the use of formative assessments, while increasing in every region, has dramatically increased in Jubaland at midline. At baseline, only 16.7 percent of teachers in Jubaland reported using formative assessments and records were observed in zero classrooms. At midline, in contrast, 88.2 percent of teachers in Jubaland reported using formative assessments and records were produced in 82.4 percent of classes. Indeed, records of formative assessments were observed more frequently at midline in Jubaland than in Banadir (52.2 percent) or South West State (50.0 percent), in a stark contrast from baseline.

<sup>87</sup> FGD with teachers, Jubaland, Int. 509

<sup>88</sup> FGD with teachers, Banadir, Int. 506; FGD with teachers, Jubaland, Int. 502



FIGURE 14: CHANGE IN USE OF FORMATIVE ASSESSMENTS



We now look at the use of participatory and student-centred practices in classrooms. The below table presents results from classroom observations in FE schools, and reports the percent of teachers observed to have undertaken the practice in at least two out of three observed class blocks. We find non-significant increases in the use of most participatory teaching practices at midline, with the exception of the use of activities or games, which decreased substantially though not significantly, and the use of open-ended questions, which decreased slightly. Overall, these findings show a slightly positive upward trend, though not a definitive one, and suggest that teaching practices may be improving, on average, in schools targeted by the AGES programme. Further evidence from the endline evaluation, however, will be needed to confirm this result.

Despite these positive trends, we note that overall, the use of participatory teaching practices remains fairly low. Only around half of teachers were observed to use open-ended questions, solicit student opinions, involve students who were not participating, and use group work in at least two of three class blocks, while only 15 percent were observed to use activities or games. Continuing to strengthen the use of these positive teaching practices may help improve teaching efficacy, thereby improving learning outcomes and attendance as students are incentivized to stay in classes that are interesting and engaging.

TABLE 40: USE OF PARTICIPATORY TEACHING PRACTICES IN CLASSES, FE SCHOOLS

Practice	BL	ML	Difference	p value
Student-centred activities or games	27.8%	13.0%	-14.8	0.06
Students instruct each other	59.3%	70.4%	11.1	0.28
Open-ended questions	52.8%	47.1%	-5.8	0.58
Questions solicit student opinions	55.6%	55.6%	0.0	1.00

<b>Involve students who are not participating</b>	48.2%	63.0%	14.8	0.16
<b>Group work</b>	27.8%	41.5%	13.7	0.18

The below table additionally reports results for ABE and NFE cohort 1 schools, for whom classroom observations were only conducted at midline. Sample size for these two programme types is small—27 ABE centres and 20 C1 NFE centres—so results should not be considered definitive. However, the table shows that a fairly high percentage of ABE classes—comparable to findings for FE schools at midline—reported using formative assessments, although only around 40 percent of teachers were able to show records of these assessments. Formative assessments were reported to be used, and records were produced, at a lower rate in C1 NFE schools, suggesting a need to focus on strengthening their use in these school types.

Furthermore, as in FE schools, teachers were observed to use activities or games and group work less frequently than other participatory teaching practices. Results also suggest that the use of participatory teaching practices is lower in ABE classrooms than in FE or C1 NFE classes. For example, this gap is notable for open-ended questions, which were observed to be used in at least two out of three class blocks only 30 percent of the time for ABE teachers, compared to 65 percent for C1 NFE and 47 percent for FE. Indeed, in ABE classrooms, only student instruction of their peers was observed at a rate over 50 percent. While, as with FE schools, rates also remain low in NFE cohort 1 classes, these findings may suggest a relatively more urgent need to improve teaching practices for ABE teachers.

**TABLE 41: USE OF PARTICIPATORY TEACHING PRACTICES IN CLASSES, C1 NFE AND ABE**

<b>Practice</b>	<b>C1 NFE</b>	<b>ABE</b>
<b>Formative assessments (stated use)</b>	50.0%	74.1%
<b>Records of formative assessments</b>	30.0%	40.7%
<b>Student-centred activities or games</b>	33.3%	22.2%
<b>Students instruct each other</b>	50.0%	51.9%
<b>Open-ended questions</b>	65.0%	29.6%
<b>Questions solicit student opinions</b>	50.0%	44.4%
<b>Involve students who are not participating</b>	60.0%	40.7%
<b>Group work</b>	25.0%	22.2%

While the above tables provide useful information on teaching practices, as noted above, classroom observations are vulnerable to bias, as teachers may be more likely to use positive teaching practices when they know they are being observed. As such, in the below table, we present data from the survey with girls to supplement information from classroom observations. Girls were asked the following questions:

- Does the teacher/facilitator explain how the things you are learning will be useful to you in your life?
- Does the teacher/facilitator give you ideas for how you can learn outside the learning centre/class as well as inside it?

- Does your teacher/facilitator suggest ways you can continue to study after school/at home?
- Do you find the lessons go at a good speed for you?
- If you don't understand something, do your teachers/educators use a different language to help you understand?
- Does your teacher/facilitator encourage students to participate during lessons, for example by answering questions?

This data is disaggregated by education track; we first discuss results for FE schools, followed by NFE and ABE. The table presents the percent of surveyed girls within the panel sample who answered with the answer given in italics; for example, the top line of the table shows the percentage of girls who stated that their teacher “often” explains the use of lessons for their lives.

The table shows that in FE schools, the use of most positive teaching practices increased at midline compared to baseline, with the exception of teachers providing students with ideas to study from home, which started from a very high base level of 93 percent and decreased by a statistically insignificant two percentage points at midline. The practice of explaining the use of a lesson or subject for students’ lives increased significantly, and remains the second most frequently used positive teaching practices reported by girls. In addition, girls reported that lessons went at a good speed more than 20 percentage points more at midline than at baseline, a significant increase. These positive findings confirm the results from classroom observations, and suggest that the use of positive teaching practices may be increasing in FE schools targeted by the AGES programme.

Looking now at NFE schools, we find somewhat more mixed results. While the use of most positive teaching practices increased at midline—including a significant increase in the use of different language to explain if students are confused—we also find a significant decline in the percent of girls reporting that teachers give them ideas to study at home. However, we note that as with FE schools, this question started from a very high baseline value of almost 97 percent, and remains at a relatively high level at midline. As such, while a decline may be of some concern, it does not fully negate other findings that suggest strengthened teaching practices.

Lastly, for ABE programmes, we find a significant increase in girls reporting that lessons move at the right speed, that teachers explain lessons in different ways, and that teachers often encourage participation. These are positive findings, particularly for the latter two practices, which nearly three-quarters of girls now report that their teachers practice. We note, however, that although girls were significantly more likely to report that lessons move at the right speed, the overall percentage of ABE girls reporting this remains low, at only slightly over half. While this may, in part, reflect variations in the skill levels of students—if classes have a large range of student backgrounds and abilities, it may be difficult for the teacher to teach at the right speed for every student—this finding also shows a need for continued improvement in the pace of lessons.

**TABLE 42: USE OF POSITIVE TEACHING PRACTICES REPORTED BY GIRL STUDENTS, BY EDUCATION TRACK**

Practice	FE			C1 NFE			ABE		
	BL	ML	Diff.	BL	ML	Diff.	BL	ML	Diff.
<b>Explain use of subjects (<i>often</i>)</b>	71.4%	85.2%	13.9*	77.1%	83.3%	6.3	75.5%	83.0%	7.5

<b>Ideas to learn outside class (often)</b>	51.8%	60.0%	8.2	58.2%	66.7%	8.5	57.1%	61.0%	3.8
<b>Ideas to study at home (yes)</b>	93.1%	91.0%	-2.1	96.5%	88.5%	-7.9*	93.1%	89.0%	-4.1
<b>Lessons at good speed (just right)</b>	39.5%	60.5%	21.1*	40.4%	54.4%	13.9	34.9%	55.8%	20.9*
<b>Different ways of explaining (often)</b>	52.4%	60.2%	7.8	51.9%	70.8%	18.9*	49.5%	72.5%	23.1*
<b>Encourage participation (often)</b>	57.2%	69.0%	11.7	68.0%	74.0%	5.9	61.0%	74.5%	13.5*

Finally, in the below table, we analyse results reported by girls disaggregated by region. We note striking improvements at midline in Banadir, where the use of every positive teaching practice increased significantly, and Jubaland, where the use of every practice except teaching lessons at the right speed increased significantly. We note, however, that the use of positive teaching practices in Jubaland continues to lag that in Banadir, despite the significant improvement; there is thus a need to continue strengthening teaching practices in this region.

Furthermore, we find that the use of several positive teaching practices decreased significantly in South West State at midline, including teachers explaining the use of lessons for students' lives, providing students with ideas for learning outside of class, and encouraging participation. The use of positive teaching practices in SWS now falls behind both Banadir and Jubaland.<sup>89</sup> The reason for this decline in positive teaching practices is unclear, but concerning, and demonstrates a need for continued reinforcement with teachers in this region.

**TABLE 43: USE OF POSITIVE TEACHING PRACTICES REPORTED BY GIRL STUDENTS, BY REGION**

Outcome	Banadir			Jubaland			South West State		
	BL	ML	Diff.	BL	ML	Diff.	BL	ML	Diff.
<b>Explain use of subjects (often)</b>	77.6%	94.3%	.001*	58.8%	76.1%	.000*	86.1%	72.5%	0.02*

<sup>89</sup> There are few satisfying explanations for the deterioration of teaching practices in SWS over time. The region has been badly affected by the drought, but this is also true of portions of Jubaland (e.g., Gedo) where AGES works; likewise, political turmoil in SWS has caused displacement and localised inter-clan tensions, but this is true in other areas as well. The most likely explanation is a combination of drought, political conflict, and continued displacement of students into regional centres – such as Baidoa – the latter of which may have placed additional strain on schools and teachers.

<b>Ideas to learn outside class (often)</b>	54.0%	72.6%	.000*	47.7%	60.9%	0.04*	67.8%	46.0%	.001*
<b>Ideas to study at home (yes)</b>	93.3%	87.1%	0.02*	99.3%	93.3%	.003*	90.8%	89.7%	0.80
<b>Lessons at good speed (just right)</b>	33.5%	68.9%	.000*	37.2%	43.2%	0.14	48.0%	48.7%	0.90
<b>Different ways of explaining (often)</b>	52.2%	81.2%	.000*	42.5%	65.6%	.001*	58.6%	46.5%	0.13
<b>Encourage participation (often)</b>	67.7%	83.7%	.000*	45.6%	69.0%	.001*	69.6%	55.7%	0.02*

### 7.3 Leadership and Life Skills

This section will analyse progress on leadership skills, self-confidence, and life skills over the course of the programme. This is measured by self-perception and the main measure used throughout will be the Youth Leadership Index (YLI). As well as reporting headline scores, the change from baseline to midline will be disaggregated by various factors to gauge what the main influences on self-confidence and leadership are, and to understand in more detail where the programme has had successes and where it has room to make further progress. As well as analysing change over time a brief baseline analysis will be carried out for a new group of NFE girls recruited in 2022. The Life Skills Index (LSI) will also be included in the analysis for this section.

In the theory of change, the main outputs and activities that are expected to influence the acquisition of life skills are improved quality learning opportunities that are tailored to the needs of ultra-marginalized girls, and social norm change towards broader life opportunities. Improved life skills in turn are expected to improve literacy, numeracy, and financial literacy and allow beneficiaries to transition to further education and employment or self-employment.

The conceptual understanding of the importance of the girls' self-leadership skills for the improvement of the learning and transitional outcomes is based on the broader findings from the evaluations and lessons learned from the SOMGEP-T project, as well as from Phase 1 of the implementation of the broader GEC initiative. The GEC thematic review of self-esteem stipulated that interventions focused on building girls' self-esteem lead to positive changes in the girls' attitudes - more concretely, in "motivation to attend and new aspirations or a sense of school belonging"<sup>90</sup> - which in turn have a positive impact on attendance, class participation and lead to overall improvement in learning outcomes.

The Youth Leadership Index is a composite indicator based on a set of 21 questions, developed by CARE International and successfully piloted and used across several countries. The indicator measures a

<sup>90</sup> GEC & UK Aid, *Thematic Review – Girls' Self-Esteem* (March 2018), link: [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/730864/TR-Girls-Self-Esteem.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/730864/TR-Girls-Self-Esteem.pdf)

respondent's self-confidence, decision-making, voice, vision and organizational skills (including the ability to motivate others and work with them to address common issues)<sup>91</sup>. The table below contains the list of the questions used for the construction of the YLI.

**TABLE 44: LIST OF YOUTH LEADERSHIP INDEX (YLI) QUESTIONS**

Questions
I like to try new activities that I may not know how to do.
My friends ask me for advice.
I recognize when people have different skills to contribute to a task.
I am comfortable when my teacher calls on me to answer a question.
I contribute ideas to discussions at home even if they are different from others' ideas.
I ask questions at school when I don't understand something.
I can describe my thoughts to others.
The things I do set a good example for my peers.
I consider the possible outcomes of my decisions before making them.
I accept responsibility for the outcomes of my decisions.
I recognize when the choices I make today can affect my life in the future.
I can show what is important to me with my actions.
If someone does not understand me, I try to find a different way of saying what is on my mind.
I encourage others to join together to help my community.
I cooperate with others to get things done at home.
If someone treats me unfairly at school, I am comfortable telling an adult.
I am willing to work hard to achieve my dreams.
I am better able to finish a task when I plan ahead.
When I have the opportunity, I can organise my peers to do an activity.
I am interested in being a leader at my school.
I try to understand the cause of a problem before trying to solve it.

The YLI is calculated based on 21 self-reported questions on a 4-point Likert scale. All the cohort groups of girls were asked to indicate how often (rarely, sometimes, most of the time and almost always) they acted in a certain way, depending on the question asked. Lower values indicate more negative outcomes and higher values indicate more frequent instantiations of the behaviour and, by extension, more positive outcomes. The score ranges between 21 and 84 points and for the purposes of the analysis the score was standardized on the scale of 0 to 100. When a girl scored the lowest possible number of points (21) by responding 'rarely' to all questions, the standardized YLI score will take the value of 0%.

The questions for the Life Skills Index can be seen in the table below. This index is a complement to the YLI and emphasises self-esteem, and autonomy, with questions included on confidence in specific situations and control over their own time. It is calculated in a similar fashion to the YLI – the questions are self-reported and ranked on a scale with lower values indicating negative outcomes and higher values indicating positive outcomes. The scores are added and then standardised on a scale of 0 to 100. The LSI questions were not asked at baseline and in this analysis the index is only calculated for the New NFE cohort of girls recruited in 2022.

<sup>91</sup> CARE Education, *CARE's Youth Leadership Index Toolkit* (2014), link: <https://www.care.org/cares-youth-leadership-index>

TABLE 45: LIST OF LIFE SKILLS INDEX (LSI) QUESTIONS

Questions
I am able to do things as well as my friends.
I get nervous when I have to speak in front of an adult.
I get nervous when I have to speak in front of a group of people my age.
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 feel confident answering questions when I'm in a group of people.
I would like to continue learning by going back to school, learning a vocation or trade.
I recognize when choices I make today can affect my life in the future.
I can describe my thoughts to others when I speak.
I can work well in a group with other people.
When I have the opportunity, I can organize my peers or friends to do an activity.
I often feel lonely.
I ask an adult if I don't understand something.
Who decides: Whether or not you can go back to school or vocational training.
Who decides: Whether or not you will continue in school past this grade.
Who decides: If you will work after you finish your studies.
Who decides: How often you spend time with your friends.
Who decides: When/ at what age you will get married.
Who decides: What type of work you will do.
Who decides: How you spend your free time.

## Aggregate Changes in YLI Scores

This section will examine the change in YLI scores from baseline to midline. As well as looking at total figures, here we will disaggregate by the main demographic variables and characteristics, namely the main cohort groups (Formal Education, Non-Formal Education, and Accelerated Basic Education), region, and by district within region where necessary. The primary variable that will be examined is the YLI scores themselves rather than the proportion of girls meeting the target of a YLI score of 70. However, increasing this proportion is still an important aim and a key programme target and therefore will be reported as well. Throughout the section, an important caveat for all findings is that there is no comparison group. This is particularly problematic for this analysis as the main variable of interest – YLI – is expected to increase with age in any case. To demonstrate this, using the baseline data YLI scores were regressed on age while controlling for cohort, and a statistically significant positive effect was found<sup>92</sup>. This means that in the event of an increase in scores, it will be very difficult to separate the positive effects of the programme and higher scores simply due an increase in age.

The total mean score at baseline was 49.8. This rose to 53.5 at midline, which was a statistically significant increase of 3.7. This is a significant jump of around 7%, indicating some success in the programme in increasing leadership skills from a relatively low base. The proportion meeting the target of 70 on the YLI also increased over the span of the programme. Using the panel sample, at baseline it was 12.6%, which

<sup>92</sup> This regression was replicated using the cross-sectional data, i.e. the new group of NFE girls that were recruited in 2022. Again a statistically significant positive effect was found, meaning that in the same cohort YLI scores and therefore leadership skills tended to be higher for older girls



is a very low base to begin from and far from the endline target of 80% of girls reaching this target. At midline the figure stood at 17.4%, an increase of 4.8 percentage points that was statistically significant. This is a reasonably large improvement and corresponds to a percentage increase of nearly 40%. Again this points to some success in the programme of increasing youth leadership skills, which is a positive finding for the intervention. However, the new score at midline still falls far short of 80% of girls, and this ambitious target seems very unlikely to be met by the endline.

**TABLE 46: CHANGE IN YLI SCORES FROM BL TO ML BY SCHOOL TYPE**

Outcome	BL	ML	Difference	p-value
FE Centres	46.9	49.2	2.3	0.31
ABE Centres	49.3	53.6	4.3	0.05**
NFE Centres (Cohort 1)	53.0	53.3	3.3	0.07*
<b>Total</b>	49.8	53.5	3.7	0.005***

\*\*\* significant at 99% level, \*\* significant at 95% level, \* significant at 90% level

The table above disaggregates the rises in YLI scores by the main cohort groups. We can see relatively even increases for each of the cohorts, ranging from 2.3 to 4.3 from baseline to midline. The increases were statistically significant for the ABE and NFE cohorts but not statistically significant for the FE cohort. Formal Education Schools were also the lowest scoring at baseline. This is driven in part by the age effect, as they are the youngest cohort, but given that they also saw the lowest increase over the course of the programme so far they are at risk of falling behind and deserve further attention as the interventions progress. ABE centres scored lower at baseline than NFE centres and now at midline score slightly higher, but the differences between them are quite small at both rounds and they can be said to perform as well as each other.

To attempt to separate the improvement from the programme and the improvement from age, the observed mean increases are compared against the expected increase over two years from the age effect as measured from the baseline. For FE centres the regression coefficient was 2.4, meaning we expect to see a 4.8 increase over two years. For girls in ABE centres we expect a 4.6 increase, while for NFE centres the coefficient was slightly negative and not statistically significant and as such will not be used in the analysis. Overall, then, the increases are either somewhat smaller than or similar to what we would expect based on the difference in YLI scores by age at baseline.

Similar trends are also seen for the proportion of girls reaching the target of 70 on the YLI. The proportion increased for all groups and now stands at 14.4%, 17.0%, and 20.5% for the FE, ABE, and NFE cohorts respectively. Again, this shows that FE girls score the lowest and that none of the cohorts seem likely to reach 80% reaching the target by endline. This should not necessarily be seen as a failing of the programme, however, as the initial target may simply have been too ambitious for what turned out to be a very low base level of YLI scores.

The total increase in leadership skills and confidence in girls seen in the quantitative data is clearly reflected in the qualitative data as well. In focus group discussions, girls gave positive personal testimony on their experience in school and the effect it had on them. When asked what changes in life were experienced after attending school, one girl in Banadir said “When I was at home I didn't know anything but now my life has changed, and I would like to reach the university level”<sup>93</sup>, while another commented

<sup>93</sup> Vignettes FGD with Girls, Banadir, Int. 613



“I used to be unable to write my own name, but now I am completely fluent. I made a positive change”<sup>94</sup>. Answers such as these were overwhelmingly the response after being asked this question, giving a personal perspective to the positive scores seen above.

In addition to the positive answers provided when asked directly, girls also indicated that schooling would prompt positive changes in other, hypothetical girls they were asked about. They were told a story of a girl who had been unable to go to school for some years but had now enrolled in an NFE centre, and were asked what would happen to this girl. One commented that “When Hamdi registers for school, her life will change and she will be happy and she will continue her education”<sup>95</sup>, while another said “Yes, she will continue her education because later on, she will contribute to both herself and her family”<sup>96</sup>. These quotes clearly show the positive attitudes towards school and the clear understanding among the girls that it leads to beneficial changes in life paths. These positive attitudes are in turn likely to lead to increases in self-esteem.

**TABLE 47: CHANGE IN YLI SCORES FROM BL TO ML BY AREA**

Outcome	BL	ML	Difference	p-value
<b>Banadir</b>	51.4	61.0	9.6	0.00***
<b>Jubaland</b>	44.5	46.2	1.7	0.42
<b>South West State</b>	52.5	47.4	-5.1	0.04**

\*\*\* significant at 99% level, \*\* significant at 95% level, \* significant at 90% level

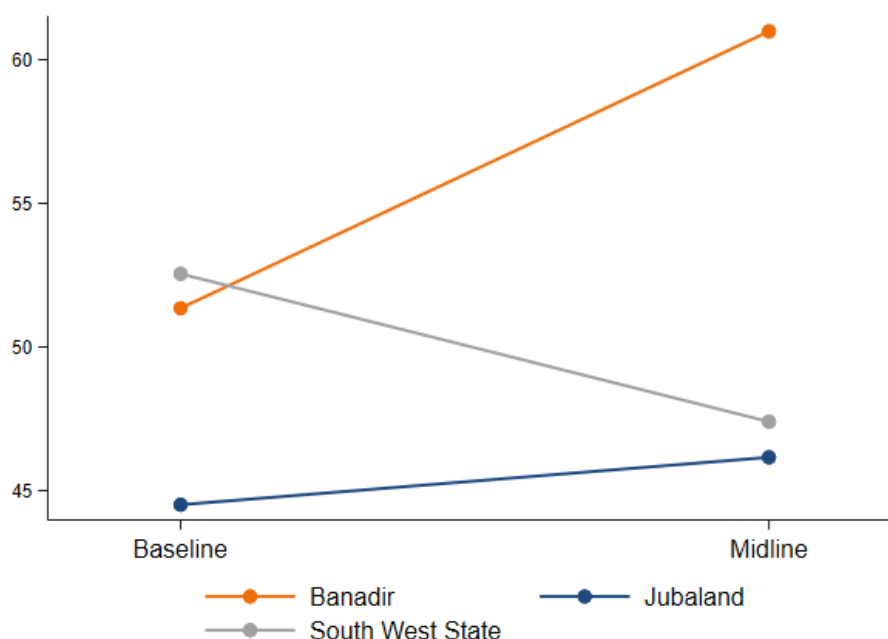
Above we see the change in scores disaggregated by zone – Banadir, Jubaland, and South West State. We can clearly see a varied impact of the programme across each of the zones. There was a strong increase in Banadir from 51.4 to 61.0. This is a statistically significant increase of 9.6, which corresponds to a percentage increase of 19%. Jubaland, however, only saw a small increase of 1.7 which was not statistically significant. Most strikingly, there was a marked decrease in scores in South West State, which is particularly surprising given that we expect this variable to increase with age. This indicates that there is probably some specific driving factor behind the trend. The results are even stronger when looking at changes in the proportion reaching the target of 70 or more. Banadir showed a strong increase in proportion from 16% to 28%, and Jubaland from 3% to 11.5% - this increase in Jubaland is promising given that the average score saw little change. But the proportion reaching the target fell sharply in South West State, from a comparatively high 16% to 3%. This is a worrying trend and will be investigated more closely below.

<sup>94</sup> Vignettes FGD with Girls, Banadir, Int. 613

<sup>95</sup> Vignettes FGD with Girls, Afgoye, Int. 610

<sup>96</sup> Vignettes FGD with Girls, Lower Shabelle, Int. 608

FIGURE 15: CHANGE IN YLI SCORES FROM BL TO ML BY ZONE



Firstly, we examine the scores by the two districts in South West State. In Afgoye there was a rise of 4.5 in mean scores from baseline to midline, while in Baidoa there was a sharp decrease of 9.8. Therefore we can see that the overall decline in South West State is driven by declining scores in Baidoa. This district also has a higher sample size of 183 compared to 90 in Afgoye, and as such will have an outsized impact on total scores. In fact, Baidoa is the district with the largest sample size across all zones in the analysis, with the next highest being Kismayo at 163. This means that the poor results in this district will affect scores reported across other demographic variables and sub-groupings.

It is difficult to account for the poor performance in this district after looking at the main demographic variables and subgroupings. YLI scores fell across every cohort in Baidoa, meaning the decrease was not driven by particularly poor scores in any of them – the decline was from 54.7 to 41.3 for formal education girls, 49.8 to 44.3 for ABE girls, and 61.3 to 50.2 for Cohort 1 NFE girls. There were also decreases in average YLI scores across minority and non-minority status. The trend is even seen at the school level – when looking at the trend for each specific school in Baidoa, falls in score were observed for 8 out of 11 of them, indicating that the trend is not driven by large changes in just one or two schools. Finally, the changes in individual girls were observed to check if the drop was driven by a small group of girls each declining by a large amount, but instead the decline was found to be more broadly shared. This means that there is some factor, specific to the district of Baidoa, that lies behind falls across all main demographic groups and sub-groupings that we cannot easily account for in the quantitative data.

There are two possible explanations for the observed trend. One is that Baidoa is a Maay-speaking area, while the official school dialect, in Maay-speaking areas, is af-Mahatiri. It is possible that this put the girls at a disadvantage and that this accounts in part for the fall in scores – for example, their self-esteem might have been harmed by struggling in a school that uses a different dialect. However, it is likely that this does not fully account for the result, as there are other Maay-speaking areas (including Afgoye) that saw rising scores over the programme.

Another possible explanation comes from the qualitative data. It is important to caveat here that there was only one focus group in Baidoa and therefore we must be careful before extrapolating trends more broadly. However, there was one clear answer when respondents were asked what the main barriers to

good teaching were in the school. Though in general they were positive about the school, many mentioned that there were not enough teachers and that better teacher training was needed.

A particular problem underlying this was that teachers had stopped being paid in the past five months and that many were voluntary and had stopped showing up. One respondent said “The school should have enough teachers, which means that if there are 700 students and only 10 teachers, it will be insufficient. . . Teachers are rare at this school because their pay has been halted. The salary they used to pay teachers, like ADRA, was stopped five months ago”<sup>97</sup>. A relatively common theme across districts was low pay for teachers and inconsistent pay schedules, but in no others was there a sudden cessation in pay, and the resulting absence of teachers and disruption to education may account for the fall in self-reported leadership skills and self-esteem.

In summary for this section, there were increases in scores for the Youth Leadership Index over the period from baseline to midline and in the proportion meeting the target of at least 70. The target set at the baseline – 80% of girls reaching the target - was likely over-ambitious and will not be met over the course of the programme, but there was an encouraging total increase from 12.6% to 17.4%. This rise occurred across the different cohorts – FE, ABE, and NFE schools – but fell short of statistical significance in FE centres, who were already starting from the lowest base. The rise in scores was not evenly shared across regions, however, and was mostly driven by a large increase in Banadir. The increase in scores in Jubaland was small and not statistically significant while in South West State there was a sharp decrease, driven by a decline in Baidoa district. This drop could not be accounted for by analysing any of the sub-groups within the quantitative data. In the focus group discussions it was mentioned that pay had been stopped in the school which greatly impacted teaching. It is difficult to extrapolate from this with further information but this is a possible factor behind the decrease. It is therefore recommended that particular attention is paid to the underperforming regions for the duration of the programme and that teacher pay arrangements are paid more attention, as this is key factor hindering sustainability of the programme and the aim of boosting institutional capacity.

### Subgroup Changes in YLI Scores

This section will further analyse the change in the Youth Leadership Index (YLI) from baseline to midline. Here we will go more in-depth using a range of variables which might be correlated with YLI scores. This will allow us to get a clearer picture into who exactly has benefitted from the programme, as well as reveal factors that might help or hinder success in developing leadership skills. The variables examined will be broken down into sections – girls’ characteristics, household characteristics, caregiver attitudes, learning environment, and participation in the programme.

TABLE 48: CHANGE IN YLI SCORES FROM BL TO ML BY GIRLS’ CHARACTERISTIC

Outcome	N	BL	ML	Difference	P-Value
Non Maay-speaker	778	49.0	56.2	7.2	0.00***
Maay speaker	284	52.0	46.2	-5.8	0.01***
Non-Linguistic minority	995	50.0	53.9	3.9	0.004***
Linguistic minority	67	47.7	47.7	0	1

\*\*\* significant at 99% level, \*\* significant at 95% level, \* significant at 90% level

<sup>97</sup> FDG with CEC Members, Baidoa, Int. 109.

In the table above we see the effect of girls' characteristics on YLI, specifically their minority status and their linguistic characteristics. One striking finding is that Maay speakers' scores declined over the course of the programme by 5.8 points while non-Maay speakers improved by 7.2. This is also reflected in the percentage reaching the target of 70 on the YLI scale – this rose from 11.6% to 22% for non-Maay speakers and fell from 15.5% to 4.9% for Maay speakers. As stated above, any decrease is worrying as we expect an upward trend in YLI due to aging effects. Further analysis showed that the fall for Maay speakers occurred across all cohorts but did not occur across all zones. In Banadir there was a rise for Maay speakers and in Jubaland there was little change, while in South West State there was a fall from 53.4 to 45.9. This suggests that the negative results for Maay speakers is driven mostly by the dynamics with Baidoa state that were discussed earlier – aside from this region Maay speakers did not seem to be greatly disadvantaged from baseline to midline.

In the qualitative data, however, it did come across that most expected Maay speakers to be disadvantaged. When told a story of a girl who speaks one language but has enrolled in a school where the teacher speaks a different dialect, the general consensus among the girls was that this would pose a strong barrier to education and enrolment. One girl said “What will happen is that Faduma and the teacher will not understand each other”<sup>98</sup>, while another said “Faduma may drop out of school if she can't understand the teacher”<sup>99</sup>. However, across some other questions there were some members with a more positive attitude who thought that the teachers would do their best to teach them the language. In response to a different but quite similar story one girl said that “her teacher will try to make her understand everything that is difficult for her, and also she will try not to speak her own dialect in the class”<sup>100</sup>. This testimony may reflect girls' experience in school of students in a linguistic minority and there is clear potential for negative effects on leadership skills, life skills, and self-esteem if a student is consistently unable to fully participate in the classroom.

Other characteristics of the girls were tested but not reported in the table above as there were no interesting trends, usually meaning that all groups increased at a similar rate. There were similar increases in YLI whether the girl had ever been married or was a mother. There were also relatively similar rises whether the girl was an IDP or not. This is also the case for the most marginalized minorities, who saw an average YLI increase of just below 2 points. Despite the negative finding above with respect to Maay speakers, it is promising that across other potentially marginalized groups that no such trend was seen.

In the table below we disaggregate mean YLI scores across the rounds for selected variables relating to household characteristics. Most of the variables included are proxies for disadvantage or poverty - for example having a poor roof or going hungry most days. In addition to these, whether the head of household was female is included to examine the effect on leadership skills in girls.

**TABLE 49: CHANGE IN YLI SCORES FROM BL TO ML BY HOUSEHOLD CHARACTERISTICS**

Outcome	N	BL	ML	Difference	P-Value
<b>Had food most days</b>	986	49.2	52.9	3.7	0.007***
<b>Went hungry most days</b>	76	57.6	62.0	4.4	0.13
<b>Did not have a poor roof</b>	853	50.8	53.8	3	0.03**

<sup>98</sup> Vignettes FGD with Girls, Banadir, Int. 605

<sup>99</sup> Vignettes FGD with Girls, Banadir, Int. 605

<sup>100</sup> Vignettes FGD with Girls, Banadir, Int. 602

<b>Had a poor roof</b>	209	45.9	52.4	6.5	0.002***
<b>Adult in house had education</b>	844	48.8	52.2	3.4	0.01***
<b>Adult in house had no education</b>	218	53.7	58.6	4.9	0.07*
<b>No female head of household</b>	671	49.9	53.8	3.9	0.007***
<b>Female head of household</b>	391	49.7	53.0	3.3	0.05**
<b>Head of household had a wage</b>	664	49.1	51.0	1.9	0.21
<b>Head of household had no wage</b>	398	51.1	57.7	6.6	0.000***
<b>Household owns land</b>	356	49.4	50.5	1.1	0.59
<b>Household owns no land</b>	706	50.0	55.0	5	0.000***

\*\*\* significant at 99% level, \*\* significant at 95% level, \* significant at 90% level

There was a higher rise for those with a poor roof, where there was an increase of nearly 7 compared to only 3 for those without a poor roof. A similar phenomenon was also seen in households without a wage. There was quite a small rise for those with a waged head of household (only 1.9 and not statistically significant), and a much larger rise for those without (a statistically significant rise of 6.6). Finally, this was also seen in land ownership. There was a high YLI rise in those who did not own land, with the rise not being significant in land-owning households – the difference is 5 compared to 1.1. The evidence is not entirely clear but there is a pattern emerging that those with some disadvantage at baseline improved more in YLI over the course of the programme so far, also often beginning with a higher base. This is a positive finding from the programme, suggesting that marginalized groups were successfully targeted and aided to a greater extent than non-marginalized groups.

Interestingly, there were similar rises in YLI scores in households that had a female head of household and those that did not have a female head of household. This is a somewhat counterintuitive finding, as it might be expected that girls in households with a female leader would benefit more from the programme in terms of picking up leadership skills. Again there were a few relevant variables that were not reported in the table above, both due to small sample sizes and no interesting pattern in the data, i.e. a similar trend was observed whether the girl was in the sub-group or not. These include having gone without water most days, being a single or a double orphan, having had no cash most days, and having savings.

Below the change in YLI scores is disaggregated by caregiver attitudes or indicators – such as household chore burden – of the value caregivers and other adults place on girls' education. This allows us to examine how dynamics within the household affected the development of leadership skills during the programme.

**TABLE 50: CHANGE IN YLI SCORES FROM BL TO ML BY CAREGIVER ATTITUDE**

<b>Outcome</b>	<b>N</b>	<b>BL</b>	<b>ML</b>	<b>Difference</b>	<b>P-Value</b>
<b>Light chore burden</b>	389	48.6	56.8	8.2	0.000***
<b>Heavy chore burden</b>	673	50.5	51.6	1.1	0.45

<b>Caregiver aspires to university</b>	933	50.1	53.5	3.4	0.02**
<b>Caregiver does not aspire to university</b>	129	47.6	53.6	6	0.01***
<b>Caregiver believes education worthwhile</b>	934	49.9	53.9	4	0.004***
<b>Caregiver does not believe education worthwhile</b>	128	49.5	51.0	1.5	0.56
<b>Caregiver prioritises work over school</b>	436	47.8	52.4	4.6	0.06*
<b>Caregiver does not prioritise work over school</b>	626	51.2	54.3	3.1	0.005***

\*\*\* significant at 99% level, \*\* significant at 95% level, \* significant at 90% level

For those with a light chore burden (possibly indicating a more positive attitude to education and more time allowed on studying) there was a large rise of 8.5, while those with a heavy chore burden saw no statistically significant rise over the course of the programme. The same trend of the girls in households with more positive attitudes increasing more was seen when the caregiver believes that education is worthwhile. Interestingly, there was less divergence based on whether the caregiver prioritised work over school. This is a less positive attitude to education but scores saw a similar rise across both, and was in fact slightly larger for those who did prioritise work over school. Whether the caregiver aspires for the girl to go to university also showed the opposite trend – though there were statistically significant rises for both groups, the rise was higher when the caregiver did not aspire for the girl to go to university.

There were also concerning findings from the focus group discussions. Members were told a few stories about a girl with responsibilities at home and a caregiver with negative attitudes to education, and expressed clearly that these were strong barriers to education. One girl said “if her mother believed that her poor daughter doesn’t need to go to school, I do not think it would be easy for her to get educated”<sup>101</sup>, while another said “Maryan can study because her mother permits it, but Nimco cannot study because her mother disallows it”<sup>102</sup>. A heavy chore burden was also seen as a barrier, with one respondent saying “helping her mother too much can make her drop out of school”<sup>103</sup> and another commenting “the chores of the house can stop her from attending the school”<sup>104</sup>. Again, however, there were some dissenting views, for example when one girl said that “Kaltuun can study or learn when she finishes the household work”<sup>105</sup> or when another commented “yes, she is capable of managing both her personal life and her education”<sup>106</sup>.

<sup>101</sup> Vignettes FGD with Girls, Banadir, Int. 613

<sup>102</sup> Vignettes FGD with Girls, Banadir, Int. 613

<sup>103</sup> Vignettes FGD with Girls, Banadir, Int. 612

<sup>104</sup> Vignettes FGD with Girls, Middle Shabelle, Int. 609

<sup>105</sup> Vignettes FGD with Girls, Middle Shabelle, Int. 609

<sup>106</sup> Vignettes FGD with Girls, Middle Shabelle, Int. 609

The trend is somewhat mixed, but overall it seems to be the case that a less positive home environment is associated with lower rises in YLI scores, as there were two instances (heavy chores and the caregiver not believing education was worthwhile) with no statistically significant rise and the girls' general consensus when asked was that these were barriers. This is worrying from the point of view of the programme success in boosting youth leadership skills and can be interpreted as a block on the programme's success. There are multiple potential reasons why these factors may influence self-esteem – a heavy chore burden and less commitment to education may mean lower attendance and therefore less exposure to the programme and to interventions specifically targeting self-esteem. It could also lead to poorer performance in school compared to those without these hindering factors, which is likely to have negative impacts on self-esteem. As the programme progresses this should be further examined and prioritised as an area for action.

In the table below mean YLI scores are disaggregated by the learning environment of the girls. In a similar way to the household-based analysis, this breakdown allows us to examine how positive or negative attitudes within the school influenced the development of leadership skills and success in the programme. The variables below can be broadly categorized into two groups – material factors such as availability of books and computers, and teachers' attitudes.

**TABLE 51: CHANGE IN YLI SCORES FROM BL TO ML BY LEARNING ENVIRONMENT**

Outcome	N	BL	ML	Difference	P-Value
<b>Computers available at school</b>	44	56.7	57.2	0.5	0.9
<b>No computers available</b>	1,018	49.5	53.4	3.9	0.004***
<b>Books/learning materials available</b>	788	51.5	55.0	3.5	0.02**
<b>No materials available</b>	274	44.9	49.2	4.3	0.08*
<b>Teachers not often absent</b>	769	51.9	52.5	0.6	0.66
<b>Teachers often absent</b>	293	44.3	56.1	11.8	0.000***
<b>Teacher encourages participation</b>	950	50.8	53.3	2.5	0.07*
<b>Teacher rarely encourages participation</b>	112	41.4	55.5	14.1	0.000***
<b>Teacher treats girls and boys similarly</b>	678	51.8	51.4	-0.4	0.77
<b>Teacher treats girls and boys differently</b>	384	46.3	57.3	11	0.000***
<b>Female teachers in school</b>	290	48.1	49.0	0.9	0.72
<b>No female teachers in school</b>	42	38.2	51.8	13.6	0.04**

\*\*\* significant at 99% level, \*\* significant at 95% level, \* significant at 90% level

The same trend was seen in schools whether or not the students had sufficient materials to learn – there was a small rise in YLI scores, of 3.5 if materials were available and of 4.3 if materials were not available.



An interesting divergence was seen with respect to whether the school had computers for use. For those with computers at baseline, no significant rise was seen from baseline to midline, while for those with no computers at baseline a statistically significant rise of 3.9 was observed.

This same trend as seen with computers – the schools that performed worse at baseline increased more from baseline to midline – was seen much more clearly for teachers’ attitudes. Those without frequently absent teachers at baseline did not see any statistically significant rise over the course of the programme, while those with absent teachers at baseline saw an average increase of 11.8 in scores. This was also clearly reflected in the numbers reaching the target of a score of 70 on the YLI. There was a small change in those without absent teachers from 15.1% to 17%, while the proportion reaching the target almost tripled from 6.1% to 18.4% for those with absent teachers at baseline. The same trend is very clearly seen across girls in schools with teachers who rarely encouraged participation with a rise of 14.1, for girls in school who treated students differently by gender with a rise of 11, and schools with no female teachers seeing a rise of 13.6 from baseline to midline. The same rise for those not experiencing these behaviours in school at baseline was either quite small or statistically insignificant.

Many of the counter-intuitive results presented in the table are actually consistent with the idea that poor teaching practices and school-based barriers strongly shape YLI scores and that – when barriers are removed and practices improve – YLI scores improve as well. As the table shows, girls with school-based factors aligned against them improved much more from baseline to midline than those without. The most straightforward explanation is that girls affected by school- and teaching-related barriers began from a lower starting point at baseline, providing more room for improvement between rounds.

A stronger explanation, though, is that the programme was successful in achieving its aims within schools to address existing issues and encourage a more positive learning environment for girls. It seems that these girls were lacking positive influences which had the potential to be addressed by the programme, and the programme was indeed successful in filling this gap and addressing the problems. Those who were already in more supportive environments were not as strongly in a position to benefit from aspects of the programme such as social norm change and leadership training.

Additional support for this explanation comes from the fact that positive teaching practices *at midline* are associated with greater gains in YLI scores. In contrast to the impact of baseline teaching practices, a teacher who encourages participation at the midline generates much greater gains in YLI scores than a teacher who does not encourage participation. Similarly, students whose teachers attend school regularly *at midline* experienced an increase of 6.0 points, while those with more-frequently absent teachers experienced a decrease of 1.5 points. In other words: teacher absenteeism *at baseline* is associated with higher YLI scores at midline because improving teacher attendance over time positively impacts student self-esteem and personal development. This is why we see opposite effects depending on how we define teacher absenteeism: when defined from the baseline, it captures students who are disadvantaged at baseline but may benefit from improvements to their learning environment; when defined from the midline, it captures students who *remain* disadvantaged at midline. This evidence more strongly supports the role of learning environment in shaping YLI scores than a cross-sectional pattern, because it links changes in learning environment over time to changes in YLI scores.

Below a table of results by participation in various aspects of the programme can be seen. This question was not asked at baseline, and as such change is not presented but simply the scores at midline. To keep the data consistent with the rest of the section the panel sample will be used, i.e. only girls who answered at baseline and were successfully recontacted at midline are included.

**TABLE 52: CHANGE IN YLI SCORES FROM BL TO ML BY PROGRAMME PARTICIPATION**

Outcome	No	Yes	Difference	P-Value
---------	----	-----	------------	---------

<b>Participated in a Girls' Empowerment Forum?</b>	50.8	59.6	8.8	0.05**
<b>Still in touch with GEF members?</b>	59.8	59.4	-0.4	0.85
<b>Community education committee supports girls education?</b>	47.5	56.3	8.8	0.01***
<b>Participated in youth groups or networks?</b>	56.4	60.2	3.8	0.12

\*\*\* significant at 99% level, \*\* significant at 95% level, \* significant at 90% level

This section must be interpreted with caution. There is no change-over-time component in the table above, meaning both groups may have had different scores at the baseline before any interaction with the programme. In addition, while we are comparing between two groups for each of the questions above, the groups have not been randomly selected, and if any differences between groups are found we will be unable to assign causality to the intervention. There are in fact likely to be confounding factors – for example, those who participated in a GEF self-selected to do so, and may have been more motivated people already and therefore more likely to have stronger self-esteem before participation.

With those caveats included, we can see that the YLI scores are higher for those who participated in a GEF. Again, we cannot confidently assign causality – it might easily be the case that more confident girls were more likely to participate in the forum – but this can be seen as weak evidence of the effectiveness of the programme. We can also check the baseline scores of those who answered at midline that they had participated in a GEF and therefore observe the change over time. The mean baseline score was 54.1, giving a mean score increase of 5.5 up to 59.6. This is a strong increase and again points to success in the programme. We can also see positive effects when the respondent said that community education committee supports girls' education and for participation in youth group or networks, although the latter finding was not significant.

To summarise this section, it is clear that the characteristics of girls, their learning environments, and their home environments influenced their youth leadership skills and self-confidence, and their progress in improving them. IDP status and minority status surprisingly had little impact but there were language effects. YLI scores for Maay speakers declined from baseline to midline and this trend was mostly driven by a decline in the Baidoa region, which is majority Maay speaking. There is some evidence that girls in households that were poorer at baseline improved more over programme so far, and clear evidence that girls in worse learning environments at baseline improved more. This is a positive finding for the programme and points to success in improving the leadership skills and self-confidence of the more marginalized girls. In contrast, there is some evidence that less positive attitudes at home towards learning hindered progress in developing these skills. Going forward the programme should focus further on home attitudes as it may constitute one of the main blocks in boosting self-esteem and life skills.

## 7.4 School Management and Governance

The next outcome assessed in this evaluation is the quality of school management and governance. Competent and effective school management offers multiple benefits for student learning, enrolment and retention, as well as student motivation and transition outcomes. Well-managed schools promote better learning outcomes by employing qualified teachers, providing them with opportunities to improve their professional skills, monitoring their teaching practices and attendance, and providing them with professional incentives, including ensuring timely pay. Well-managed schools also boost community

education levels by encouraging children and parents to enrol in school, monitoring attendance, and promoting support for education among the community, as well as finding solutions to address barriers to enrollment and retention for some of the most marginalised students. Additionally, well-managed schools demonstrate efficiency in their use of financial resources to improve infrastructure, design and implement school development plans, uphold levels of conduct that protect children and encourage inclusivity, and manage response to crises, such as COVID-19.

To achieve improvement in school governance, AGES programming targets the activities of Community Education Committees (CECs) and Ministry of Education representatives that work with program schools. The CECs are supported in building capacity to reach and assist marginalised girls, especially girls with disabilities, to enrol and remain in school. The project works with the CECs to increase their community mobilisation capabilities and adopt inclusive practices. AGES also supports relevant government officials to increase their knowledge of the needs and rights of marginalised girls and design and implement policy addressing inclusive and special needs education, including the policy framework around ABE programmes. Further, AGES assists the officials to improve their quality assurance and school monitoring procedures.

AGES tracks the changes along indicators that capture the progress the CECs and MOE have achieved to promote inclusivity and improve quality assurance procedures. The two logframe indicators are:

- Percentage of community education committees addressing barriers to the enrolment and learning of ultra-marginalised girls
- Number of MoEs' departments implementing inclusive quality assurance procedures

While these indicators are captured directly through internal M&E processes, the current evaluation collected quantitative data that, although does not directly measure the two indicators, may further illuminate CECs' efforts to promote enrolment and inclusion, as well as qualitative data that covers perception of both the CECs and MOE actions.

## Community Education Committees

Since the beginning of the program, AGES assisted CECs with recruitment of members, supported outreach to out-of-school or absent children, delivered training on school management, disseminated awareness campaign messages, and offered training on psychosocial support. Broadly, the responsibilities of the CECs include overseeing teaching and administrative practices and supporting student enrollment, with the focus on ensuring inclusion of marginalised girls. CECs are expected to contribute to their school's teaching and administrative practices by working with school administrations to create school development plans, solving disputes between administration, teachers and parents, ensuring timely teacher pay, monitoring teacher attendance and basic teaching quality, and promoting child protection. Further, CECs are responsible for monitoring student attendance, reaching out to absent and out-of-school students, raising awareness of the importance of education, and finding ad-hoc solutions for overcoming barriers individual students may face for attending school, including via fundraising. Lastly, CECs are expected to promote inclusion by conducting targeted outreach to disadvantaged families, addressing challenges specific to marginalised groups, and delivering anti-discrimination messages to students and teachers.

Inclusion of marginalised girls in the context where AGES is implemented entails navigating multiple overlapping axes of marginalisation, each one associated with unique and localised challenges. For instance, access to education for girls from pastoralist families is disrupted due to seasonal or more permanent migration, responsibilities to attend to animals, and distance to more centrally located schools from the settlements located near pastures in the town periphery. Simultaneously, marginalisation afflicts girls from minority groups in agricultural families, girls in IDP families displaced by conflict, drought, or

floods, and girls with disabilities, among others. In this environment, the CECs must consider multiple priority groups to achieve the focus on inclusion of all ultra-marginalised girls.

In general, the CECs' goal is to identify and resolve any students' challenges to attendance and learning, especially for marginalised girls. School fees constitute a major barrier to enrolment for the most economically disadvantaged girls, suggesting that provision of primary education free of charge could be instrumental for increasing enrolment. The number of formal schools that did not charge school fees has decreased since baseline by 17 percentage points, amounting to 40 percent of schools, according to head teachers. On average, the schools that charged a monthly fee at the time of the midline assessment, saw a small increase of the fee amount by 0.3 USD reaching 7.5 USD, although this change was not statistically significant. Together this may suggest that the traditional sources of financing for schools – including the MoE, private and agency-based support – responded negatively to COVID-19 and other economic shocks, or at least were unable to offset the increased material needs associated with distance education, causing schools to require additional fees. This, however, did not seem to have a major negative effect on the girls' ability to enrol as shown by a 19 percentage point reduction in the share of caregivers citing education costs as the reason their daughter was not in school. This reduction may be a result of the combination of the advocacy activities and targeted material support interventions discussed in detail below.

Still, during qualitative interviews, members of CECs in schools that required fees often identified the charge as a major factor decreasing enrolment, either in general or specifically for IDP girls. As expected in the context of widespread poverty, the CECs did not report universal provision of financial assistance to all students who would benefit from it. The strategies employed instead included requesting the school administration to exempt or discount the charge for one of the siblings from poor families with multiple school aged children, isolated cases of fundraising to pay school fees for individual marginalised students, and delivery of awareness campaigns to parents promoting the importance of education and calling to prioritise school fees over other expenses.

While free education would benefit all students, several CECs reported more targeted actions they have taken to support marginalised girls specifically. When discussing challenges to enrollment, pastoral families tended to be seen as having particularly unfavourable attitudes about formal education for girls due to the type of skills used in their work and labour requirement often outsourced to children. Pastoral parents were often identified as the main targets for awareness campaigns; CEC members reported visiting pastoral households to engage in discussions about girls' education and advocate for the enrolment of girls. However, other CECs also described visits to out-of-school and dropped out students' households irrespective of their demographic status as their primary responsibility, making it difficult to estimate the extent to which this activity was targeted at girls from marginalised backgrounds.

With IDP girls being seen as some of the most economically disadvantaged, CEC members highlighted their work to provide study materials and occasional fundraising efforts to support this group. A CEC in Benadir also reported providing health services at school to girls from IDP families who could not access them elsewhere, which served as an additional incentive for parents to send their daughters to school. That being said, it is difficult to estimate whether and in which direction the targeted support of marginalised girls from the CECs has changed since the baseline using only the qualitative data.

During baseline, we identified troubling patterns of bias and mistreatment of marginalised girls in schools which posed significant barriers to attendance for these students. We concluded that this discrimination was most likely to come in the form of bullying from other kids but mistreatment from teachers and institutional discrimination were also at play. The qualitative accounts of harassment and name calling recorded during baseline FGDs with CEC members were supported by the quantitative data that demonstrated higher rates at which caregivers of marginalised girls cited perceived harassment and mistreatment as reasons to keep their daughters out of school as compared to other parents. Although

serious enough to prevent girls from attending classes, these less tangible barriers, unlike the financial and accessibility issues, tend to be more malleable as they are rooted primarily in communal attitudes, rather than stable institutional structures that respond primarily to substantial material investments.

The midline FGDs with CEC members delivered a contrasting portrayal of the treatment of marginalised girls. Two of the CECs claimed that their past awareness campaigns reduced discrimination and name calling targeted at girls from minority groups, leaving no other unique challenges associated with these groups at the time of interview. However, this account may underestimate the pervasiveness of the bias faced by minority groups. Overall, the FGDs showed CEC members were more reluctant to recognise the unique barriers faced by girls from minority groups as compared to other marginalised groups such as IDPs or girls with disabilities, which may be a reflection of a cultural tendency to rhetorically downplay divisions to avoid further conflict. This may affect the CECs' ability to identify and willingness to address the challenges faced by minority girls further disenfranchising them from educational opportunities. Although some degree of discrimination was acknowledged, specifically towards IDP girls and girls from poor families who could not maintain proper hygiene, CEC members tended to downplay its impact on educational opportunities as compared to the material or accessibility barriers and continued to maintain that their awareness campaigns were effective at tackling all forms of bias.

To ground the qualitative accounts of the reduction of bias and harassment against marginalised students, we consider the changes in caregivers' perceptions of the barriers that kept their daughters out of school for several sub-groups that were found to face a degree of discrimination at the baseline. The indicators we consider capture various aspects of mistreatment children may face in school, including bullying from classmates or teachers, general perception of the lack of safety, and being refused entry to school. For marginalised groups, these barriers may be driven by discriminatory attitudes related to their identities. Initially, we suspected that the descriptions of improved student cohesion obtained during the FGDs may overstate the lack of discrimination for an average project school for at least two reasons: the social desirability of showing tolerance and acceptance of all groups during the FGDs and, since the FGDs were conducted in schools with the most active CECs, more effective than average anti-discrimination activities in these, presumably, better managed schools. However, the quantitative data supported the claim that bias and harassment of marginalised girls, or at least their perceptions by the caregivers, have largely decreased since baseline.

We can observe sharp drops in the shares of caregivers who cited bullying or discrimination as reasons they kept their daughters out of school, as well as a 10% decrease in the share of girls who have never attended school. Note, however, that the rate at which the caregivers' perceptions of mistreatment of their daughters in school declined tended to be somewhat lower for marginalised girls than other children, with the exception of af-Maay speakers, who greatly outpaced the general population. Since the baseline, girls who speak Af-Maay at home became more comfortable with the Af-Mahatiri instructions through practice, which is likely to positively impact their experience in school, which in turn influenced their perceptions of their caregivers. The perceptions of caregivers of girls from minority groups constituted somewhat of a negative exception as the perceived mistreatment from teachers has increased for this group, further suggesting that the CECs were unwilling or unable to recognise and target discriminatory attitudes directed at girls from minority groups.

Thus, there remains doubt as to what drove the decrease in perception of harassment and mistreatment, especially since the school survey did not collect quantitative data on the prevalence of anti-discrimination campaigns among the activities implemented by the CECs which means that we do not know how commonly these campaigns were implemented by the CECs that did not participate in the FGDs.

While it is possible that, despite targeting ultra-marginalised girls, the CEC awareness campaigns also benefited the general student population to an even greater extent, other factors, such as localised improvements in the security situation, increases of teachers salaries, implementation of stricter

punishments for bullying, or increases of the school capacity to accept more students may have positively impacted each of the indicators presented in the table below. Unfortunately, the level of detail in the data that would be needed to determine the relative contribution of each factor to the identified improvement across the below indicators was not available. Nevertheless, the convergence of the qualitative and quantitative data that was available, as well as the pattern of exclusion of the girls from minority groups, can be interpreted as sufficient evidence that the AGES intervention is at least partially responsible for improving the learning environment for students overall, as well as for marginalised students in particular.

**TABLE 53: REASONS CITED BY CAREGIVERS FOR A GIRL TO BE OUT-OF-SCHOOL, BY ROUND**

Subgroup	Baseline	Midline	Change from BL to ML
<b>Unsafe for girl to be at school</b>			
Overall	6.3%	1.6%	-4.7*
Has disability	8.2%	0%	-8.2*
Pastoralist	3%	0%	-3
IDP	4.2%	1.4%	-3.2
Minority group	5.8%	0%	-5.8
Speaks Af-Maay at home	16.4%	1.8%	-14.6****
<b>Teachers mistreat girl at school</b>			
Overall	5.4%	2.1%	-3.3*
Has disability	2.3%	2.3%	0.3
Pastoralist	2.1%	0%	-2.1
IDP	4%	4.2%	0.2
Minority group	2.3%	7.7%	5.4
Speaks Af-Maay at home	17.2%	1.8%	-15.4****
<b>Refused entry at school</b>			
Overall	4.6%	1%	-3.6**
Has disability	2%	0%	-2
Pastoralist	1.3%	0%	-1.3
IDP	2.3%	1.4%	-0.9
Minority group	2.3%	2.6%	0.3
Speaks Af-Maay at home	16.4%	0%	-16.4****
<b>Other students bully/mistreat girl at school</b>			
Overall	3.4%	1%	-2.4*
Has disability	2%	2.4%	0.4



<b>Pastoralist</b>	0.9%	0%	-0.9
<b>IDP</b>	2.3%	0%	-2.3
<b>Minority group</b>	0%	0%	0
<b>Speaks Af-Maay at home</b>	13.3%	0%	-13.3***

It remains clear that some of the barriers to enrolment associated with marginalised girls were not addressed by the CECs, with the members often describing them as beyond their capacity due to limited financial resources. Most commonly, CECs reported not being able to solve accessibility issues that prevent girls with disabilities from attending school. The lack of funds to finance transportation to and from school disenfranchised girls with mobility issues and visual impairments and, to a lesser degree, girls who lived far from school, including girls from pastoral families. Further, seasonal migration of pastoral families took girls away from school for prolonged periods of time. With more resources, CECs would have been able to hire teachers to move with the pastoralists allowing children to continue studying all around the year, which, as reported in one FDG, was not possible with the current budget.

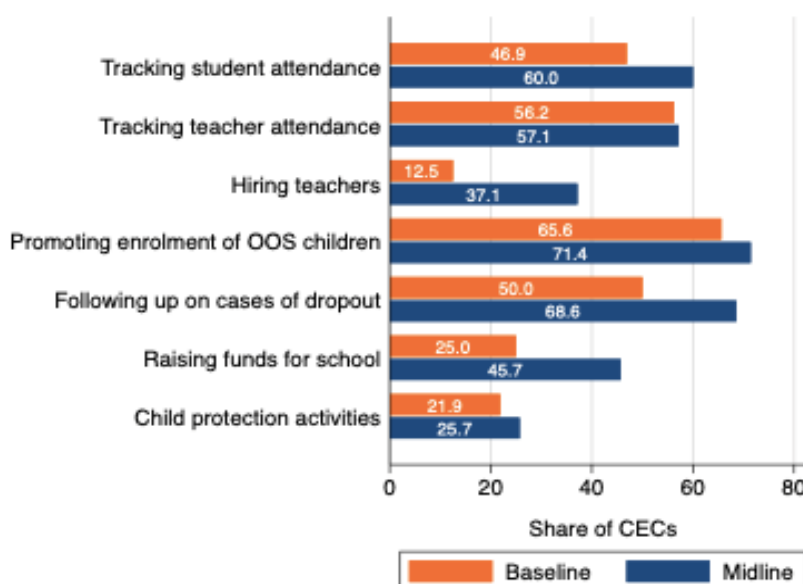
More broadly, CEC members believed that providing additional material incentives, such as meals and sanitary products, as well as improving school conditions, would be instrumental for attracting girls from poor and marginalised backgrounds by providing a rare opportunity to fulfil basic needs like food and hygiene. This shows that the needs of some of the most marginalised girls require the most resource-intensive and logistically complex solutions that have not been adopted by the CECs due to a limited financial support to schools from agencies and government, and a narrow fundraising base. Meanwhile, focusing on the more attainable and less impactful inclusion strategies may create an impression of progress despite failing to reach the most vulnerable girls.

In addition to targeting ultra-marginalised girls, the CECs engaged in activities that could benefit the entire student population. The quantitative data collected from head teachers demonstrated that all schools participating in the assessment had a CEC, meaning that the single school that did not have a CEC at the baseline managed to establish one. As the mere presence of a CEC tells us little about the extent of its impact on girls' inclusion, we consider the disaggregated data on CEC activities collected from head teachers via the school survey.

The share of schools reporting engagement of the CEC increased across all activities (see figure X), with the largest growth associated with hiring teachers (25%), fundraising (22%), and following up on dropouts (22%). While the monitoring and awareness-based activities were found to be the most common, their prevalence during baseline suggested that CEC members tended to be quite successful in implementing them even without the AGES support. This can be interpreted to suggest that the higher rate of uptake of more technically demanding tasks such as hiring and fundraising may be attributed to the program. In line with the CEC members highlighting a gap in child protection training during an FGD, increasing the implementation of these activities emerges as a priority for the next stage of the program.



FIGURE 16: CHANGE IN CEC ENGAGEMENT, BY ROUND



The share of schools visited by the CEC in the past year has increased by 14 percentage points, reaching 80 percent. While this may seem like a high number, it measures the lowest benchmark of engagement; still, 20 percent of the CEC, mostly in Banadir, failed to conduct a single visit to their school. Nevertheless, we notice improvement since the last assessment. Student attendance, which used to be overlooked by 35 percent of visiting CECs at baseline, is now the most tracked aspect of schooling by the CEC with 93 percent of committees accounting for it during their visits. The absolute number of CECs monitoring all other aspects of schooling have also increased since baseline suggesting that CEC visits became more thorough in the last two years. This improvement is likely to have positive downstream effects on school management, as well as learning and attendance outcomes.

Before we continue to discuss the qualitative insights obtained through the FGDs with CEC members, it should be once again noted that the FGDs were conducted in schools with some of the most active CECs in the sample. This suggests that the extent of the reported activities, as well as the expected impact, are unlikely to be representative of a typical project school. While the FGDs may capture positive examples of what an active CEC can achieve in the adverse environment where AGES is implemented, it is likely that the committees are substantially less engaged and impactful on average.

During the FGDs, members of most committees reported meeting once a week and visiting schools between two and six times a week. Echoing the school survey data, a common task reported by the members was monitoring of student attendance and following up on dropouts and absent girls, irrespective of their demographic background. Home visits often fed into another set of activities implemented to increase enrollment that included awareness campaigns to promote the importance of education and more informal advocacy with girls' parents or husbands to allow them to attend school. CEC members reported mixed results admitting that some of the girls whose families they talked to remained out of schools due to household responsibilities or cultural norms, however, a precise success rate of these awareness raising activities is hard to estimate using only the qualitative data.

To address the teaching quality, most CEC members reported monitoring student-teacher interactions, advising against the use of harmful practices such as physical punishment, and moderating disputes between teachers and administration or parents. Self-reported results included decrease of teacher absence and lateness. Several FGDs with CEC members included mentions of fundraising to boost teacher salaries or, in the case of one school in Lower Juba, to raise money to donate up to 50% of their agreed-upon salary to teachers, who have not been paid for months, to prevent them from leaving the school.

Even in light of these fundraising efforts, the extent of which varied substantially from school to school, CEC members tended to agree that limited resources allocated to teacher salaries continued to present a major issue for teaching quality both in terms of undermining the motivation of the currently employed teachers, as well as addressing the staffing gap that was especially glaring for female teachers who were seen as better suited to engage with girls. The school survey data demonstrated that, despite the positive trend in the reported fundraising efforts, 63% of schools did not receive contributions for teacher salaries from the CECs, only one school fewer than at baseline. The increase in an average amount contributed to teacher salaries is driven by five CECs that managed to raise between 100 and 6000 USD, while teachers in most other schools enjoyed only a small donation which, coupled with the low base salary, contributed to the issues discussed in the teaching quality section (see section on teaching quality).

This disparity further supports the picture in which the impact of a few active CECs by far exceeds the average contribution across project schools. While in some areas, such as increase in school practice oversight, student attendance monitoring, and parent engagement, the gap can be closed with the investment in CEC member technical capacity or assistance in streamlining of operational procedures of the committees, others may be far more difficult to improve. In the context of widespread poverty, the narrow fundraising base and limited personal resources of the CEC members continue to limit the contribution to school financing that most CECs are able to deliver. As established earlier, the material assistance is also key to overcoming the barriers faced by some of the most marginalised girls, suggesting that the financing aspect, despite being the most challenging, should take priority in the upcoming stages of the project. Multiple avenues, including the delivery of a major value shifting campaign promoting the importance of fundraising for girls' education targeting not only parents but also traditional and religious leaders who have greater access to resources, as well as the search for alternative financial sources, including diaspora funders, should be explored by the CECs with the assistance of AGES.

Lastly, school equipment was also monitored during school visits but only two CECs appeared to have sufficient fundraising capabilities to deliver major infrastructural investments, such as building part of the school and buying land. While other CECs also identified infrastructural shortcomings, such as hot overcrowded classrooms or no classrooms at all, as posing major barriers for girls' learning, the funds they were able to raise could only cover minor material needs, such as chairs or supplies, or were directed towards one-off emergency expenses, such as fixing broken windows.

It should also be noted that the school survey data demonstrated improvements across several indicators relevant to school management practices, including in the availability of key infrastructure and the use of supplies. Specifically, this includes the increase in the number of schools with reliable electricity, clean water, and separate toilets for girls, as well as higher rates of provision of sanitary pads, increased use of textbooks, and delivery of different types of training for teachers. While many of these targets could fall within the responsibility domain of CECs, we do not have the evidence to directly attribute these improvements to the committees' actions. Other factors are likely to have contributed, at least in part, as supported by the limitations that CECs faced outlined in the qualitative analysis above, as well as the relatively low prominence of CECs as a primary authority over budget, policies, and purchase of supplies identified in the school survey.

### MoE use of inclusive quality assurance procedures

The second indicator of inclusive school governance focuses on the use of quality assurance procedures by government officials. Broadly, the government supports schools by developing a national curriculum and helping to ensure security at schools and surrounding communities. In relation to providing quality assurance, the MoE is expected to conduct school visits to monitor teaching practices, check on administrative records, and oversee the inclusion of marginalised students, including by engaging with traditionally underserved sections of communities, such as IDPs in the area, to ensure they are included.

To assist the MoE with the tasks above, AGES planned a series of interventions, some of which could not be carried out or whose implementation was delayed due to Covid-19. AGES delivered several training sessions to MoE officials regarding gender, safeguarding, inclusive education approaches, and remote modes of monitoring adapted for the pandemic. During future stages of implementation, the project plans to review the MoE quality assurance procedures, assist with developing new policies, and continue delivering training.

The quantitative data we collected did not directly cover MoE engagement with the schools as questions about the work of ministry officials, in general or in relation to quality assurance, were not included in the school survey. In the original midline qualitative design, MoE officials also were not identified as targets for qualitative interviews. However, after identifying this gap in the collected data, the evaluation team opted to pursue remote interviews with MoE officials and our in-house senior researcher contacted 6 MoE officials to conduct short semi-structured KIIs over the phone. The interviews were transcribed and translated.

At baseline, we found that the engagement of state MoE and MoECHE officials with the schools was uniformly limited, with a potential exception in Banadir. The only avenue of engagement of government officials with schools that was tentatively supported in the qualitative data was the provision of security. While security provision plays an important role in supporting attendance and enrollment, it was clear from the context and the baseline data that this security provision was generally not specific to schools and referenced more diffuse provision of security to localities. Although the baseline findings included some accounts of monitoring visits and teacher training in certain regions, we concluded that government involvement in quality assurance procedures was quite limited overall.

During the midline KIIs, MoE officials unanimously insisted that the overall State MoE support to schools has increased significantly in the last two years. A major avenue of engagement mentioned in several KIIs focused on training for the teachers and CEC members, as well as the District Education Officers (DEOs), the MoE officials who conduct school visits. Another key improvement that MoE respondents took credit for was the increase in enrolment, including of marginalised girls. Further, the respondents maintained that the DEOs visited the schools under their respective jurisdictions between once every three months and 7 times a month. The visits typically started by touching base with the head teacher to discuss the current agenda, receive updates, and check the administrative documents, including the enrolment list and teacher and student attendance records. Then, the respondents reported walking through the school premises to record any issues with equipment, facilities, and cleanliness. Lastly, the DEOs typically meet with teachers, conduct direct observations of lessons, and talk to CEC members and mothers in the community. These activities allowed the DEOs to provide feedback on teaching practices, monitor enrolment levels, and keep track of infrastructural needs. A DEO reported:

*When we conduct school visits, we first meet with the principal to check to receive updates and check administrative records, then we move on to other activities, such as monitoring the classroom and checking in with teachers. We also monitor the cleanliness of classrooms and the conditions of the equipment. We spend between 1 and 4 hours at each school.*

-KII with MOE Official, Hirshabelle

The information gathered through monitoring visits is reportedly used to inform the assistance the MoE delivers to the schools. For example, the tracking of dropout cases directed the follow-up household visits by DEOs. The monitoring of equipment and supplies informed the material assistance provided to schools, which spanned expenses from books to desks to facility repairs. According to several respondents, many of the classrooms were furnished by the MoE. The DEOs recorded the challenges they

encountered and the issues raised during their visits, from teacher salary delays to overcrowded classrooms, in the reports that were submitted to the State MoE leadership and CARE to identify solutions.

To increase inclusion of marginalised students, DEOs described the implementation of mobilisation campaigns that spread awareness about free educational opportunities prioritising marginalised households, the hiring of teachers from minority clans, and the opening of several schools for students with disabilities. Some of these initiatives appeared to overlap with the responsibilities of CECs outlined in the previous subsection, potentially highlighting the need to define the roles of the two institutions more clearly in order to optimise the use of resources.

There was also variation in the perceptions of DEOs regarding inclusivity and the progress made. For instance, a DEO admitted that enrolling girls from disadvantaged groups remained a significant challenge, despite the MoE awareness campaigns. Another DEO, on the other hand, reported visiting minority households to encourage parents to enrol their daughters and providing financial assistance to buy books and uniforms. Several respondents noted that inclusion of children with disabilities remained an issue both in terms of how their parents view the value and suitability of education for their children, as well as the availability of essential equipment, such as wheelchairs and glasses.

While the FGDs portray a high level of engagement and oversight of the schools by the state-level MoEs, unfortunately, we do not have other sources of data to triangulate these claims or draw conclusions about the representativeness of the above accounts for the average project district or school. The ad hoc nature of the KIIs we conducted may have led to selection of more active DEOs for interviews, because DEOs were not randomly selected and sampling was likely biased toward the more active DEOs, whose contacts we obtained from head teachers. Coupled with insights from CEC FGDs that highlighted the lack of external financial support to tackle many of the same material issues – especially around infrastructure and financial aid – this suggests that the DEOs interviewed for this study may constitute a positive exception to the general state of governmental involvement in education.

## Conclusion

Overall, the CECs appear to have improved their engagement across tasks that are in line with the project goals, although the extent to which their activities specifically prioritise marginalised girls remains unclear. Following up on dropouts and promoting enrollment by engaging with parents of out-of-school students are two most direct ways the CECs can encourage better transition outcomes, while monitoring school management and teaching practices is an indirect way the CECs can support attendance and learning. With some improvement noted at this stage, the overall levels of CEC engagement in pro-active school management need to continue to increase. Most importantly, key activities of the CECs, which were found to not be sufficiently targeted at this stage, need to prioritise the most marginalised girls as they otherwise risk reproducing the existing patterns of inequality.

Compounded by the social desirability bias and in the absence of sample-wide quantitative data for government engagement with schools, the presented evidence of the MoE engagement should be interpreted with a great deal of caution. While it is clear that some gains in terms of MoE engagement and quality assurance processes have been made in at least some districts, we cannot quantify the change in the prevalence, effectiveness, and inclusivity of these procedures since the baseline, nor can we definitively conclude that the changes are positive. The assessment of MoE involvement in school quality assurance will need to be prioritised further in future rounds of the evaluation.

## 7.5 Community Attitudes

This section will assess the changes in community attitudes around girls' education that can play an instrumental role in helping marginalised girls overcome the barriers preventing them from attending and

completing school. The perceived benefits of education, economic and otherwise, influences the caregivers' assessment of the relative importance of sending their daughter's schooling as compared to housework or starting a family, as well as informs the decisions of the wider community when girls from marginalised families require additional support from the community to stay in school. Sending a girl to school may mean giving up labour resources and foregoing income-generating opportunities, as the girl could have engaged in low-skilled economic activities or taken on domestic chores to free up more time for adult family members to work. Even in the absence of economic pressures, traditional ideas around gender roles may designate young girls to the domestic realm, stigmatising public activities especially if they involve socialising with boys, or push them to marry and have children early. The environment where AGES is implemented is characterised by highly constrained resources, meaning that the investment in school fees, supplies and uniforms may often come at the cost of basic necessities for the girls or other family members, including food, medical expenses, and clothes. Thus, the decision about keeping a girl in school is not dependent simply on the caregivers' assessment of the importance of education in the abstract but on how it compares to the opportunity costs.

The attitudes towards education of other community members can affect a girl's ability to attend school when her family is unable to pay the fee or buy the supplies, in which case community members outside her family may contribute the necessary funds. The CEC fundraising is also affected by the community attitudes, with more positive perceptions of the benefits of education leading to a broader fundraising base willing to provide more generous financial contributions to support the education of the most marginalised girls. Further, the narratives around the importance of education for girls popularised by various authority figures can also influence girls' internal motivation to study as they mimic the values of their communities, which in turn can have effects on the quality of their learning and motivation.

Engagement with parents and other community members is one of the major pillars of the intervention conducted by AGES, mostly fulfilled through awareness and mobilisation activities. Through CECs, families of girls who were missing school or were not enrolled at all received messages about the importance of girls' education. Men, boys, and religious leaders constituted additional categories specifically targeted by the project. To track the project's progress, AGES assesses changes along the following indicators:

- Percentage of parents investing profits from income-generating activities in expenses related to girls' education
- Changes in mothers' attitudes and practices towards new roles for girls (reflected in changes in attendance, mobility, and early marriage)

While these indicators are tracked directly through the internal M&E processes, the current evaluation analyses quantitative data that may reveal in-depth insights related to these two aspects of community attitudes. First, we discuss the changes of family willingness to financially support girls' education, despite the household budget constraints. Then, we analyse the caregivers' aspirations for the level of education their daughter should achieve, as well as the relative importance placed on education as opposed to household chores.

## Financial Burden of Girls' Education

With many families in the AGES locations operating under tight budget constraints, we want to understand how much the project interventions were able to impact the families' willingness to invest in girls' education *despite* the burden of other household expenses. For the majority of the beneficiary households, especially those of marginalised backgrounds, sending daughters to school was likely to be associated with a material sacrifice suggesting that positive educational outcomes for the girls would only be achieved if the families were willing to forego other needs. During baseline, we asked the caregivers **whether it was worth investing in girls' education, even if financial resources were limited**

with nearly all respondents agreeing or strongly agreeing with this statement. Such a high rate of positive responses at baseline created a ceiling effect leaving limited room for improvement, which made it difficult to capture progress in caregivers' attitudes.

Overall there appeared to be a small (2 percentage points) and not statistically significant decline<sup>107</sup> in the average caregiver assessment of the value of investing in education. The share of the caregivers selecting "strongly agree" has also decreased by almost 3 percentage points amounting to 85 percent. There are reasons to suspect that the actual support of girls' education might be lower than shown in the data due to the social desirability bias which was created through two main factors: participants of the survey were aware that they were interviewed by researcher working for an education project, while a broader cultural context, including the framing of the pursuit of education as a virtue under Islam, stigmatised open opposition to girls' education. Given the incentives to avoid showing negative views of girls' education, we consider respondents who selected anything other than "strongly agree" different from those who expressed unequivocal support and analyse the changes in the share of "strongly agree" responses below. While we do not have a way to verify exactly how much lower the actual support for girls' education may be, these inferences appear realistic given the amount of out-of-school girls in the project locations.

Consider the caregivers whose daughter has never attended formal school. When they were asked to identify reasons why they did not enrol their girl in school, 36 percent cited cost of education as a major contributing factor. Among those caregivers who cited cost as one of the reasons their daughter had never been enrolled in school, 93 percent strongly agreed and 7 percent agreed that investment in girls' education was worthwhile despite limited funds. This highlights that the opinions about the value of girls' education reported in the survey do not necessarily reflect the choices that the respondents make in real life. In addition to the social desirability bias that created pressure on respondents to demonstrate their support for girls' education, factors related to the interpretation of the question may have further inflated the responses. It may have been difficult for the respondents to seriously engage with the question due its somewhat abstract nature - "limited funds" was not specific enough to make the caregivers consider whether they would keep their daughters in school at the expense of their spouse's medical bills or ability to buy more nutritious food for their other children. For some respondents, the interpretation of "limited funds" may have been divorced from their personal level of poverty, which they may see as sufficiently worse to justify not enrolling their daughter in school.

The social desirability bias affected both rounds of evaluation, which means that, despite the inflated levels of outcomes during both rounds, the change in the outcome from baseline to midline is still meaningful. The change may have been driven by a decrease in the social desirability bias itself, meaning that there would be less stigma surrounding denunciation of the benefits of girls' education. That seems unlikely given the prevalence of awareness campaigns promoting the value of education and the absence of actors spreading the opposite message. This suggests that the reduction of the "strongly agree" responses in the last two years is likely driven by an actual change in community attitude. This change could likely be attributed to the increase in financial hardships due the pandemic and droughts. A combination of supply and demand shocks, including the disruption of imports of key consumer goods, price fluctuations, decreased demand for exports, the fall of purchasing power in the domestic market, and declining remittances, negatively impacted economic opportunities for many families and made it harder to justify the education costs. It appears that the program advocacy could not offset the effects of the economic pressures of the pandemic but it is unlikely that the program itself had a negative impact.

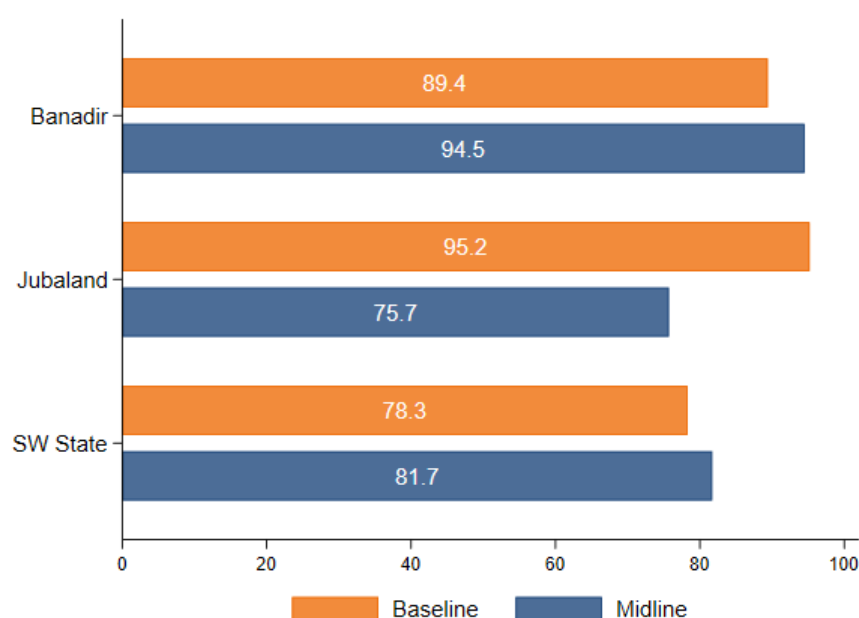
---

<sup>107</sup> Of the average score on likert scale where "1" is strongly agree and "5" is strongly disagree



The ceiling effect somewhat limited our ability to disaggregate our analysis of the changes in responses by demographic groups, however, several interesting dynamics could still be identified. First, the rate and direction of the change in support of girls' education across the three states varied significantly. In Jubaland, the share of caregivers strongly agreeing with the importance of investing in girls' education despite limited resources decreased by 20 percentage points. In Banadir and South West State they increased by 5 and 3 percentage points respectively. Jubaland is characterised by high levels of insecurity, vulnerability to droughts and floods, and the resulting major concentrations of IDPs<sup>108</sup>. It is likely that the decreasing perceptions of the value of girls' education in this state is related to the higher level of vulnerability of a large portion of its residents. In the Jubaland context, where the consequences of the pandemic and droughts may have been felt more than in the other states, a larger share of caregivers was led to doubt how additional years of schooling for their daughters can help them fulfil their basic needs like food and shelter that had to be addressed in the short term.

FIGURE 17: SHARE OF CAREGIVERS WHO SUPPORT INVESTMENT IN EDUCATION, BY STATE AND ROUND



Further, the gender of the caregiver appeared to have an impact on both the level and rate of change in support of girls' education, with female caregiver being somewhat less likely to strongly agree with the statement at midline (84 percent) and having decreased their support since the baseline by 3 percentage points, whereas men have increased their support by almost 3 percentage points. Although there were considerably fewer men in the sample, the difference in the midline levels of outcome for the male and female caregivers was statistically significant.

The difference in the midline outcome levels can potentially be explained by a greater material benefit female caregivers receive when girls are available to help around the house as domestic labour falls primarily on women. As a result of the economic challenges related to the pandemic and droughts, the burden on mothers who have to manage the household on a tighter budget while, in some cases, also

<sup>108</sup> 64% of respondents in Jubaland were IDPs, whereas only 35% and 19% were IDPs in Banadir and South West State respectively



seeking paid work outside of home, is likely to have increased in the last two years, which may be driving up their preference for keeping daughters out of school to help around the house. On the other hand, the AGES advocacy that specifically identified men in the communities as primary targets for awareness campaigns may be responsible for increasing the male caregivers' support for girls' education. Together, these two factors may explain the difference in the midline outcome change between male and female caregivers. This finding has programmatic implications to focus advocacy efforts on mothers as they tend to be the primary decision makers about their daughters' education.

When analysing the differences in the rates of change of support for girls' education across key demographic groups, we found that IDP caregivers demonstrated a sharp depreciation in the perceived value of girls' education as compared to the non-IDP households. While at baseline IDP caregivers were 7 percentage points more likely to strongly agree with the statement about investment in girls' education, at midline, they were 7 percentage points *less* likely to select the "strongly agree" response. This shift is likely linked to the disproportionate economic impact on vulnerable IDP households during the pandemic<sup>109</sup> which caused the caregivers to re-evaluate their daughters' chances of entering university given the exacerbated economic constraints and shift their preferences towards the girls engaging in income-generating activities instead of the pursuit of higher education.

During the FGDs<sup>110</sup>, the IDP and pastoral households were identified as experiencing the most prohibitive financial barriers to school enrollment and requiring most material support from CECs. Struggling to afford the basics, the opportunity cost to girls' education for these families was already relatively high, suggesting that additional pressure from economic shocks may further de-emphasise the value placed on girls' education. This implies that the AGES advocacy around the importance of schooling for girls could not offset the priority shift resulting from the changing material conditions for this group.

In contrast, caregivers of girls from several other marginalised groups including girls with disabilities and Af-Maay speakers demonstrated larger-than-average gains in support for girls' education. Thus, families of Af-Maay speaking girls increased their support for girls' education at a rate that was 17 percentage points higher than the rest of the respondents, nearly closing the gap with Af-Mahatiri speakers and reaching 85 percent of households selecting the "strongly agree" response. As the girls continued their studies from the baseline, their level of comfort with Af-Mahatiri instructions must have improved through practice, leading to a better experience in school. In turn, seeing this improvement pushed the caregiver support of schooling upwards explaining these results.

Households of girls with disabilities remained constant in their assessment of the value of girls' education, creating a 4 percentage point differential in the rate of change in the responses since baseline between the caregivers of disabled and able-bodied girls, driven by the decrease of the perceived value of education among the rest of the caregivers. Both of these groups were not necessarily economically marginalised in the way that pastoral and IDP families may have been but were still identified as key targets of the project interventions. Thus, in the context of the relatively stable perceived opportunity costs of sending girls

---

<sup>109</sup> We found statistically significant differences in self-reported effects of COVID between IDP and non-IDP households. For example, IDP caregivers were significantly more likely to report reduced food expenditures, pulling children out of school, and going hungry to bed as a result of the pandemic than their counterparts.

<sup>110</sup> FGD with CEC, Banadir

with disabilities to school, the benefits provided by the project offset the negative factors for the caregivers of girls with disabilities that affected the rest of the caregivers.

The second relevant question in the household survey that explored the implications of the financial burden of girls' education inquired **whether the cost of education was an acceptable reason to not enrol a girl in school**. The table below summarises the average midline responses and the rate of change since the baseline for major categories of respondents. Overall, the share of caregivers who believed that cost was an acceptable reason to keep their daughter out of school has declined by 17 percentage points down to 45 percent, signifying major progress in community attitudes.

Although the two questions we are considering in this subsection target related attitudes, we can observe drastic differences in the shares of respondents who affirm the importance of girls' education despite the financial burden. This may be attributed to the acquiescence bias, the tendency of survey respondents to agree with the statement or select a positive response option when in doubt. While in reality the share of caregivers committed to keeping their girls in school despite the costs is likely to be somewhere in between average responses for the two inversely phrased questions (85 percent and 45 percent) and would depend on many factors including the nature alternative expenses that constitute the opportunity cost, this divergence highlights the shortcomings of attempting to capture a real-life decision associated with material consequences with a survey question where giving either answer costs nothing.

More importantly, factors other than cost that impact the broader attitudes around girls' education are not captured by the second indicator. Thus, divergences in the outcomes of the two indicators can be related in the changes in the perceived importance of other factors that determine whether schooling is a worthwhile investment for girls, including the availability of opportunities for girls to use their education, the cultural appropriateness of school, perceived dangers that can be encountered in and on the way to school, and the benefits of the tasks the girl could engage in instead of education, such as chores or work.

Further, the results in the first question may be inflated as its formulation is somewhat more abstract in that what one thinks of as a "worthwhile" investment may vary greatly. The first indicator is directly asking about the caregivers' personal opinion, whereas the second question refers to a community-wide attitude, which may feel less targeted and make it easier to be honest if these attitudes are socially stigmatised as is not supporting girls' education. Despite maintaining doubts about the share of caregivers that would be willing to sacrifice other needs to send their daughters to school, the lower shares of positive responses in this case allow us to analyse the changes in since the baseline without the ceiling effect.

We observe that the share of caregivers who believed that the cost of education constitutes an acceptable reason for a child to not attend school had declined since baseline across all major categories of respondents. This can be interpreted as positive evidence of effectiveness of AGES advocacy as the groups that showed the most dramatic improvement, such as male caregivers, caregivers of girls with disabilities, and pastoralists were specifically identified among the key target demographics. Nevertheless, other key groups lag behind in their rate of reducing the share of respondents who believe that cost is an acceptable reason to forgo school for girls. These include the more economically disadvantaged groups such as the minority groups, IDPs, and households with non-educated caregivers for whom, it could be assumed, the wide-reaching effects of their material situations could not be offset by campaign messaging.

Also note that, unlike for the previous indicator, Jubaland experienced the highest gains, whereas South West State improved at the slowest rate. The result for the first indicator that asked about girls' education being "worthwhile", an assessment which may also be driven by several factors other than cost. Thus, the Jubaland respondents increased their likelihood of agreeing that cultural or religious inappropriateness of schooling was an acceptable reason to keep their child out of school by 26 percentage points, whereas in other states the share of caregivers agreeing with this statement has decreased since the baseline. This may indicate the propagation of more fundamentalist views among the Jubaland population, suggesting that

the decline of the perceived value of girls' education in the state captured by the first indicator is associated with cultural reasons rather than cost.

**TABLE 54: SHARE OF CAREGIVERS WHO FEEL COST IS AN ACCEPTABLE REASON FOR NOT ENROLLING GIRL**

Subgroup	Baseline	Midline	Change from BL to ML
<b>Overall</b>	<b>62%</b>	<b>45%</b>	<b>-17***</b>
<b>Banadir</b>	62%	48%	-14***
<b>Jubaland</b>	60%	37%	-24***
<b>South West</b>	64%	54%	-10*
<b>IDP</b>	64%	49%	-15***
<b>Non-IDP</b>	61%	43%	-18***
<b>Af-Maay speaker</b>	63%	50%	-13**
<b>Af-Mahatiri speaker</b>	62%	44%	-18***
<b>Disability</b>	82%	50%	-32***
<b>No disability</b>	60%	44%	-16***
<b>Minority group</b>	61%	49%	-11*
<b>Caregiver with no education or Madrassa only</b>	59%	47%	-12***
<b>Caregiver with some primary education or above</b>	75%	42%	-33***
<b>Female caregiver</b>	63%	47%	-16***
<b>Male caregiver</b>	52%	21%	-31***

## Caregiver attitudes and practices

This subsection will discuss the change in the caregivers' expectations for the girls' education and their attitudes around gender norms that may influence their decisions about their daughters' schooling. First, we consider the **caregivers' aspirations for the level of education they want their daughters to attain**. While this may not reflect the extent to which the caregivers are willing to support the girls to complete their schooling up until the desired level, including financially, the aspiration itself may affect the girls' learning motivation, help them set goals, and signify the caregivers' approval, or at least acceptance, if the girl decides to pursue education instead of choosing a more traditional domestic path.

Overall, we found no significant change in the average levels of caregivers' aspirations for their daughters' education. Partially, this may be due to the initially high results at the baseline: in 2019 we found that 88 percent of parents wanted their daughters to complete university, which increased by 1.5 percentage

points, with substantial variability, at midline. When considering a change in the aspired level of education with choice options from none to university, the average level has slightly decreased, also with significant variability. The high baseline results create a ceiling effect, limiting our ability to identify any improvements of the attitudes of the respondents who already selected the highest option at baseline. In our case, if a caregiver who reported wanting their daughter to go to university at baseline has increased their aspirations as a result of AGES advocacy<sup>111</sup> we do not have a way of tracking this change.

With only a few respondents changing their answer since midline, it is likely that, in the context of social desirability of demonstrating support for girls' education, the small share of caregivers who reported wanting their daughters to achieve less than the university level of education did so due to major material barriers that they knew would prevent their girl from attending university. The limited change is likely linked to the fact that these types of barriers require major investment and cannot be overcome through an awareness campaign delivered by the program.

This explanation is supported by the results of the disaggregation by demographics that showed that several major economically marginalised groups demonstrated a decline in the ambition for the girls' education at a higher than average rate, driven specifically by an increase of the share of caregivers who selected "none" as the level of education they would like their daughter to achieve. It is likely that recent economic shocks made some families realise that supporting their daughter's education was out of their reach financially, an explanation that seems plausible given the fact that parents whose daughters were not attending school for cost-related reasons reported significantly lower aspirations for their daughters' education. The caregivers of girls with disabilities and IDP families were the major groups driving this trend and also two of the groups identified by the CECs as facing the most demanding challenges that were often beyond their capacity to solve.

Banadir had the highest caregiver aspirations for their daughters' educational attainment that increased since the baseline. The average aspired level of education in Jubaland and South West State demonstrated a statistically significant decrease in caregiver aspirations. The factors defining this pattern are connected to the stable economic and political characteristics of the states that remained the same since the baseline. Being a major urban centre, Mogadishu is characterised by higher proximity to cultural capital and federal economic resources which underlies more progressive views on gender roles. On the other hand, Jubaland is a far more precarious place to live, with large portions of the state controlled by Al-Shabaab. With by far the highest share of IDPs in the sample, Jubaland respondents had fewer opportunities to challenge their ideas about gender roles and were faced with economic hardships that led them to place higher value on domestic work. Many of the same factors are present in the South West state.

Thus, while the AGES advocacy messages promoting the importance of education may be responsible for the increases in the response levels for some caregivers, especially those who were relatively better-off and less affected by COVID and droughts, some of the most marginalised families decreased their expectations for their daughters' educational attainment suggesting that economic stresses have offset the positive impact of the program on average.

Caregivers were asked a series of questions to understand what reasons were seen as acceptable to not allow girls to pursue education. One of these reasons referenced the traditional gender roles, such as that the girl needed to help with chores. AGES specifically targeted the gendered stereotypes in the project

---

<sup>111</sup> For example, by becoming more sure of this aspiration, wanting the girl to not just go to a university but also get excellent grades, wanting her to go to one of the more prestigious universities or pursue a graduate degree

awareness campaigns, which is why in this subsection we consider how the caregivers' attitudes around gendered responsibilities of their daughters have changed since baseline.

The share of caregivers who believed **it was acceptable to keep their daughter out of school so that she could help with chores at home** has decreased by 24 percentage points since baseline, now amounting to only 12 percent. In general, community attitudes tend to be one of the most malleable project targets that respond to interventions better than resource-demanding institutional or infrastructural objectives. Thus, given that the AGES messaging directly targeted the issue of domestic responsibilities for girls, this change is likely to be attributed to the program. Further, the significant improvement was possible in the context of the initially high share of caregivers accepting chores as legitimate reason to keep girls out of school and the associated lack of ceiling effect, in contrast to the previously discussed indicators.

Overall, the differences in the rates of change in adopting the more progressive beliefs about girls' education among the demographic groups were at least partially related to the group differences at baseline. We suppose that, once the information about the benefits of girls' education is widely available, an equilibrium share of caregivers who agree that education should take priority over domestic work should be achieved. Thus, the rate at which different groups update their beliefs about gendered expectations for girls' education is influenced, in part, by their distance from the equilibrium share at baseline. For example, caregivers of girls that speak af-Maay at home and girls from minority groups have changed their opinion on keeping girls at home to help with chores at a different rate than their counterparts but arrived at approximately the same midline result as the rest of the respondents, suggesting that these demographic characteristics did not appear to influence the indicator in question once more information about gender roles and girls' education became available.

At the same time, the equilibrium share of positive responses for different groups may vary according to their economic conditions and unique challenges. Two marginalised groups in particular lagged behind in caregiver attitudes towards girls' education and domestic responsibilities, namely IDPs and girls with disabilities. Although IDP caregivers have increased their support for choosing schooling over chores, at midline they were over 8 percent more likely to say that it was acceptable to keep their daughter out of school to do domestic work than their counterparts. With IDP families having less access to various facilities, they tend to experience a greater burden of domestic responsibilities, which is why they may place a higher relative importance on the help with chores as opposed to school.

The attitudes of caregivers of girls with disabilities have remained nearly constant since baseline, creating a differential in the rate of change in the response to the current question as compared to caregivers of girls without disability of over 26 percent<sup>112</sup>. As a result, the share of caregivers of girls with disabilities who believed that doing chores was an acceptable reason to keep girls out of school at midline was 10 percent higher than their counterparts. This suggests that shifting the attitudes of caregivers of girls with disabilities likely requires more targeted messaging that specifically centres the importance of education for girls like their daughters.

The rates of change in responses were similar across the three states suggesting that geography-specific factors were not significant for determining beneficiary responsiveness to the project messaging or that the project was similarly well-tailored to the audiences in the three states. This resulted in the preservation of the baseline pattern of discrepancy in the levels of responses between the states, with Banadir having the least amount of caregivers believing it was acceptable to keep girls out of school to

---

<sup>112</sup> Statistically significant at 1%

help with chores, only 4 percent, and Jubaland lagging behind at 20 percent. The factors contributing to this pattern are likely the same as described for the previous indicator.

Overall, we demonstrated the significant importance of the demographic and geographic factors for caregiver perceptions of the benefits of girls' education. We found that cultural factors and the perceived potential for the use of education for girls influence the caregiver support for their daughters' education as much or even more than the considerations of cost. These factors resulted in the decrease of support for girls' education in Jubaland and the increase in Banadir and South West State. Economically marginalised groups, and especially the IDPs, were found to respond negatively to the recent financial challenges, deprioritising their daughters' education. AGES advocacy appeared to be better received by male and educated caregivers, who did not face the same opportunity costs or were prepared to receive the campaign messages, suggesting the need to better tailor the advocacy activities to mothers and uneducated caregivers.

## 7.6 Increased Self-Efficacy

The sections below look at variables related to positive youth development and access to protection services. The reader should note that the findings presented below apply specifically for girls from the subsample of cohort girls already enrolled in non-formal education schools.

### Positive Youth Development

At the midline, data was collected for indicators extracted from the Chinese Positive Youth Development Scale (CPYDS), a globally recognized multidimensional measure of positive youth development (i.e., competence, character, confidence, and connection). The seven indicators discussed below are taken from the CPYDS subscale on self-efficacy.

#### Control of own life

The first indicator measures the degree to which girls perceive to have control of their own lives. Most girls in the old NFE cohort (58.5 percent) disagree with the statement that they have little control of things that happen in their own lives, with 33.6 percent strongly disagreeing. The perception of control of their own lives is weakest for girls in Banadir region, with only 48.5 percent disagreeing with the statement. In contrast, girls in the four other regions have a high perceived control of their own lives, with 69.8 percent of girls from these regions disagreeing with the statement. In particular, girls from Lower Shabelle have a very strong sense of control about the things that happen in their life, with 86.1 percent disagreeing with the statement.

Girls in households where the head of household has no formal education of any kind are almost equally likely to perceive having control over their own lives (57.7 percent disagreeing with the statement on whether they have little control of things that happen in their lives) compared to girls in households where the head of household has some formal education (59.0 percent). Interestingly, when we integrate heads of households who may have received Quranic education at madrasas into the category of heads of households with 'no formal education', we find girls' responses to be substantially different. In this case, girls from households where the head of household received informal education<sup>113</sup> are more likely to report having perceived control over their own lives (60.5 percent) than girls from households where the head of household has received some form of formal education (51.4 percent). The positive effect of Quranic education on girls' perceived control of their own lives is similarly reflected with caregivers. A

<sup>113</sup> Respondents who did not receive any education at all, received some form of informal education or attended madrasa were all included in the subgroup, as responded who did not receive formal education.



larger share of girls whose caregivers have received informal education perceive to have control of their own lives (60.1 percent) compared to girls whose caregivers have received some formal education (46.4 percent). Furthermore, having a female role model at home can also have a positive effect on girls' perceived control of their lives. When the head of household is female, 61.3 percent of girls in the survey disagree with the statement of having little control on the things that happen in their lives.

The share of girls who perceive to have control over their lives also varies depending on the language spoken at home. In households where af-Maay is spoken, 67.0 percent of girls disagree with the statement.

When girls report going to sleep at night feeling hungry on many or most days, they are more likely to feel that they do not have control of things that happen in their lives. 34.3 percent of girls who reported going hungry to sleep on many days (more than 10) disagreed with the statement, whilst only 29.2 percent of girls who reported going hungry to sleep on most days or always disagreed with the statement.

Girls who are married (53.1 percent) are less likely to perceive having control over their lives than girls who are not married (59.3 percent). The opinions of girls who are mothers do not vary much from girls who are not, however, young mothers are more likely to either strongly disagree (42.3 percent) or strongly agree (11.5 percent) with the statement about having little control over things that happen in their lives.

There was little variation in the results when we disaggregated by the presence of any type of disability, which suggests that overall, disabilities do not affect girls' perceptions of control over their lives. However, we do find that perceptions do vary slightly depending on the type of disability, with physical<sup>114</sup> disabilities generally having a more negative effect on girls' perceptions of control of their own lives compared to mental disabilities<sup>115</sup>.

## Solutions to problems

The second indicator measures the extent to which girls believe they can solve their own problems. Most girls in the survey disagreed with the statement that they did not have any solutions for some of the problems they are facing (59.1 percent). In Banadir region, girls are less likely than other regions to feel that they can solve their own problems (48.5 percent), whilst in Gedo and Lower Shabelle, girls are far more likely to feel that they can solve their own problems (85.2 percent and 80.1 percent, respectively).

As with the first indicator, Quranic education for the household head and to a lesser extent, the caregiver, may be related to girls' more positive perceptions regarding their problem-solving abilities. Girls whose household head has received informal education (60.1 percent) are more likely to believe they can solve their own problems than girls whose household head has received some type of formal education (54.3 percent). Girls whose caregiver has received informal education (59.7 percent) are also more likely to believe they are able to solve their own problems than girls whose caregiver has received formal education (55.4 percent).

The exception is for minorities, where 62.0 percent of girls believe that they can solve their own problem. The language spoken at home did make a difference, as 69.15 percent of girls from Maay-speaking households disagreed with the statement that they do not have solutions to their problems. Girls from

<sup>114</sup> For physical disabilities, the variable for mobility disabilities is used as a proxy. The other variables are not discussed in the text because there were very few observations (n<10).

<sup>115</sup> For mental disabilities, the variable for anxiety/depression is used as a proxy. Other variables for mental disabilities were excluded from the discussion because there were few observations (n<5).



households that own land (67.0 percent) are also more likely to believe that they can solve their own problems than girls from households that do not own land (55.7 percent).

A lack of access to basic needs can severely affect girls' perceptions of their problem-solving abilities. Girls who go to sleep without food on many days or most days are far less likely to believe they are able to solve their problems (37.2 percent and 37.5 percent, respectively). The same is true for girls who go to sleep without clean water on most days (36.0 percent). Parental support, marriage, and children have a negligible effect.

Girls with some sort of disability are less likely to feel that they are able to solve their problems (52.5 percent) than girls without any form of disability (61.43 percent). Physical disabilities appear to be less of an inhibiting factor than mental disabilities for this indicator, with 63.6 percent of girls with a physical disability believing they can solve their problems but only 50.0 percent of girls with a mental disability having the same belief.

### Ability to change

The third indicator refers to the ability to change things in life. A large majority of girls strongly disagree with the statement that they are unable to do much to change things in their own life (70.1 percent), and a large share disagree strongly with this statement (31.5 percent). This sentiment was felt most strongly in Gedo and Lower Shabelle, where 96.5 and 91.7 percent of girls respectively believe they have the ability to change things in their lives.

There appears to be little difference in opinion regarding the ability to change things in life between girls whose head of household or caregiver has received informal education compared to girls whose head of household or caregiver has received formal education, except that girls whose household heads and caregivers have informal education do not disagree as strongly with the statement that they are unable to do much to change things in their life. Female household heads may have some influence on girls' opinions, with only 67.7 percent of girls whose head of household is female disagreeing with the statement. Maay speakers also appear to have more confidence in themselves, with 77.7 percent of girls from households where Maay is spoken believing they are able to change their lives.

Girls who go for most days without basic needs such as food and water appear to have a weaker resolve, as they are far less likely to believe they can change things in their lives (48.0 percent and 50.0 percent, respectively). This trend is not observed for girls who go for many days without food and water. Marriage and children may also have a negative effect on girls' perceived ability to change their lives, as only 65.6 percent of girls who are married and 65.4 percent of girls who are mothers believe they can change. In contrast, having no parental support appears to strengthen girls' resolve. Girls with no living parents or who do not live with either parent in their households are far more likely to have the feeling that they can change things in their life (80.0 percent and 76.9 percent, respectively).

Girls with mental disabilities are far less likely to believe they can change things in their lives (66.3 percent) than girls with physical disabilities (90.9 percent). Overall, this effect is balanced out, as girls with any disability are only slightly less likely to have the belief that they can change things in their lives (69.3 percent) as girls with no disability (70.4 percent).

### Helplessness

The fourth indicator measures whether girls feel helpless when they face life difficulties. A minority of girls agree with this statement (39.9 percent). Only a small share of girls in Lower Shabelle (16.7 percent) and Bay (28.3 percent) are inclined to feel helpless in difficult situations, whereas a majority of girls in Banadir are reported to feel helplessness (50.5 percent).

Household heads' and caregivers' lack of any sort of formal or informal education is linked to stronger feelings of helplessness from girls in those households (50.0 percent and 54.6 percent, respectively). Girls from Maay-speaking households, on the other hand, are less likely to feel helpless (27.7 percent).

Feelings of helplessness are strongest in girls who most days go to sleep without food (62.5 percent) or clean water (76.0 percent). Girls with only one living parent are more likely to feel helpless (47.7 percent) than girls without parents (40.4 percent) or girls not living with their parents (40.4 percent). Girls who have been married<sup>116</sup> or who have children (44.23 percent) are more likely to feel helpless (45.3 percent) than girls who are currently married (37.5 percent). Girls with disabilities are also far more likely to feel helpless (52.5 percent), although this effect is likely driven by girls with mental disabilities (55.8 percent), who represent a far larger group than girls with physical disabilities that feel helpless (36.4 percent).

### **Fate not in hands**

The fifth CPYDS indicator for self-efficacy asks if girls feel that their lives are determined by others and fate. A majority of girls in the sample agree with this statement (55.6 percent). There is large regional variation, with an overwhelming majority of girls from Gedo feeling that their lives are not in their hands (85.2 percent), but only a minority of girls from Lower Juba (35.7 percent) and Bay (40.0 percent) sharing this sentiment.

Household heads' gender and the lack of formal education for household heads and caregivers appear to have little effect on girls' responses for this indicator. Maay language is another important factor to consider, with girls who speak the language at home being substantially less likely to feel that their lives are not in their hands (45.7 percent).

A shortage of food and clean water are linked to girls' feeling of their lives being out of their hands. The more frequently they go without these basic needs, the more likely they will report feeling that their lives are determined by others and fate. A lack of parental support has a similar effect, with 49.2 percent of girls with just one parent, 60.0 percent of girls with no parents, and 65.4 percent of girls not living with their parents feeling that their lives are not in their hands. Marriage and children have a contrasting effect, with girls who are married or have children being less likely to feel that their lives are determined by others or fate (43.8 percent and 44.2 percent, respectively).

### **Determine own life**

The sixth indicator represents girls' perceptions on how things happening in their lives are mostly determined by themselves. Most girls are in favour of this statement (65.9 percent), particularly girls in the region of Gedo (85.2 percent). Girls from Lower Juba are somewhat less confident about their ability to determine the things happening in their lives (53.6 percent).

Girls from households where the household head or caregiver has no education of any kind are very likely to believe that they determine the things that happen in their lives (74.4 percent and 74.2 percent, respectively). Girls from households where the household head is female are also more positive about their ability to determine things in their own lives (67.1 percent). Among girls from minority groups, only 56.3 percent affirm the same.

There seems to be no clear relationship between girls' perceptions on their ability to determine their own lives and a lack of access to food or clean water. The lack of parental support does appear to be linked, with girls with only one living parent more positive than the average about their ability to determine their own lives (69.2 percent), but girls with no living parents being less positive than the average (65.3 percent) and girls not living with either parent being very pessimistic (57.7 percent). Marriage does not

<sup>116</sup> Have ever been married but may be divorced now.

change girls' perceptions compared to the average, but having children does reduce the likelihood that they perceive they are able to determine the things happening in their lives. Finally, for girls with disabilities, only the perceptions of those with physical disabilities deviate notably from the mean (54.5 percent).

### Task completion

The seventh and final indicator for self-efficacy asks girls whether they believe they can finish almost everything that they are determined to do. A vast majority of girls agree with this statement (83.5 percent), with girls in Gedo being particularly confident about their ability to complete tasks (92.6 percent) and girls in Lower Juba being the least confident (67.9 percent).

Girls with household heads and caregivers that have received formal education are most likely to believe they can complete their tasks (90.0 percent and 91.1 percent, respectively). Interestingly, Maay speakers are less confident in their ability to complete tasks (80.9 percent) than non-Maay speakers (84.2 percent).

Hunger and thirst can play a role on girls' perceived ability to complete tasks, but only when girls go most days without (91.7 percent and 88.0 percent, respectively). Children and physical disabilities also reduce girls' ability to complete tasks, with only 76.9 percent of young mothers and 72.7 percent of physically disabled girls reporting being able to complete almost everything they are determined to do.

### Synthesis

When disaggregating the results geographically, certain trends emerge. In general, Banadir region appears to have girls with lower levels of positive youth development, whereas girls from regions such as Gedo and Lower Shabelle frequently score well on the CPYDS indicators. This might have to do with the fact that there are more established centres of formal education and alternative basic education in regions such as Gedo, whereas in regions such as Banadir the focus of learning centres is on non-formal education.

Household characteristics such as a formal education for the household head or caregiver can have a strong impact on girls' perceptions of what they can do. Household heads and caregivers receiving a non-formal Quranic education can also play a significant and positive role in elevating girls' perceived self-efficacy. The status of the clan girls belong to can sometimes help and sometimes hinder girls' self-efficacy, depending on the type of indicator in question. The presence of role models when household heads are female may have a slight but empowering effect on girls' self-belief. A large share of girls who speak af-Maay at home but are not from minority groups (36.2 percent), may benefit from the privileges of being part of the elite.

Individual characteristics such as a chronic shortage of food, water, and parental support have a clear influence on practically all indicators. Not being able to meet their basic needs severely reduces girls' perceptions of self-efficacy, and the lack of emotional and financial support that orphaned girls experience will reduce their levels of confidence, although in some cases it might increase their levels of independence. Marriage appears to be provide a safe space for some girls to increase their perceived self-efficacy, although divorce can entirely negate this effect. Finally, disabilities have diverse effects on girls' perceptions of self-efficacy. Girls with physical disabilities are clearly less able to do certain activities, but their disability appears to strengthen their resolve. On the other hand, mental disabilities do not affect girls' practical abilities but severely weaken their cognitive development.

### Access to protection services

The following section focuses on girls' access to protection services, both within their schools and their communities. The section will review girls' ability to access these services, as well as the types of services that are available.

## School

An overwhelming majority of girls reported that schools them with some a channel to report some form of harassment, abuse or exploitation at their school (93.9 percent). The share of schools that provide protection services is highest in Lower Shabelle (97.1 percent) and lowest in Bay (90.0 percent), but overall schools perform well in this regard. There are a few subgroups of girls who report notably less access to these services, namely girls from minorities (91.4 percent), girls who are currently married (90.6 percent) and girls who go without clean water on most days (84.0 percent).

Girls will typically report an incident to their teacher and/or head teacher, and on rare occasions to the Community Education Committee (CEC) or the Girls' Empowerment Forum (GEF). A few girls also mentioned that they can report incidents to the police as one of the menu of options available. Girls from Banadir and Lower Shabelle most frequently resort to both teacher and head teacher to report incidents, whereas girls in Gedo prefer to go straight to the head teacher and girls in Lower Juba and Bay prefer to go to either the teacher or the head teacher.

## Community

A large majority of girls also reported having access to someone to report incidents of harassment, abuse or exploitation in their communities, although this number was not as high as in schools (85.7 percent). Gedo (96.3 percent), Lower Juba (91.1 percent), and Lower Shabelle (97.1 percent) regions performed well in the provision of protection services on the community level, but Banadir was well below average in this regard (79.9 percent). Girls currently married (90.6 percent), ever married (92.2 percent), with children (94.2 percent), with disabilities (90.1 percent), and without clean water for many or most days (91.7 percent and 96 percent, respectively) are quite positive about protection services rendered to them. Girls whose head of household and caregiver has no education of any kind are somewhat less positive about the protection services available (80.8 percent and 79.4 percent, respectively).

Teachers and head teachers remain girls' first choice to report incidents of harassment, abuse, and exploitation at the community level, particularly in Banadir and Lower Shabelle. The police is the second most frequent choice to report incidents, (a popular option in Lower Juba), followed by community leaders in third (often mentioned by girls in Lower Shabelle and Bay). Community Education Committees, Girls' Empowerment Forums, and religious leaders are more frequently mentioned by girls as alternative protection services at the community level. However, the most common response that girls gave was that they didn't know who to report incidents to, a response most notable in Gedo and Bay.

## Synthesis

Overall, schools and communities score well with girls on protection services. Schools generally fare better than communities in this regard, in part because teachers and head teachers are the most preferred people to report incidents of harassment, abuse, and exploitation. The CEC and GEF are severely underutilized platforms of reporting incidents, as people from traditional societies tend to rely on protection services from local authorities such as police or community leaders. Particular attention should be paid on to protection services for subgroups that are traditionally or structurally marginalized in society.

## 7.7 Strengthened Economic Circumstances

The sections below look at variables related to economic empowerment. The reader should note that the findings presented below apply specifically for girls from the subsample of 381 cohort girls already enrolled in non-formal education schools.

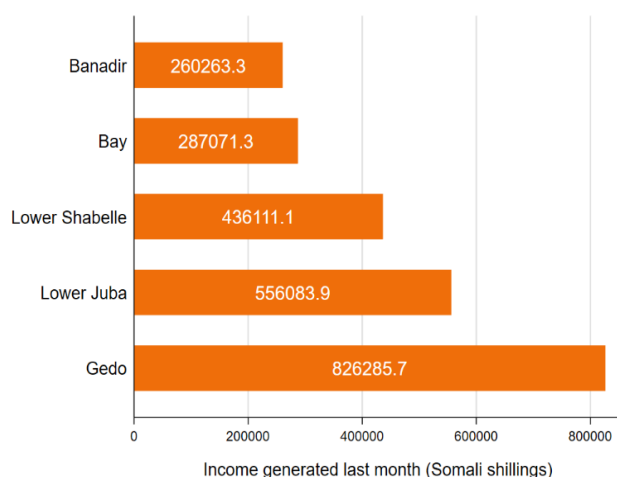
### Income generation

At the midline, quantitative data was collected for numerous indicators of income generation, including average income and type of income generation activities. In the following section, the findings for these indicators will be presented, and disaggregated for relevant variables such as region and minority and IDP status. The data will be complemented with qualitative findings from the interviews and focus group discussions that were conducted with mothers, Community Education Committee (CEC) members, and girls.

## Average income

The indicator for income measured the amount of money (in Somali shillings) that girls made in the last month.<sup>117</sup> On average, girls made about 368,128 SOS a month, with 46.0 percent reporting not having an income. As shown in the table below, there are large regional variations in income, with girls in Gedo being on average by far the highest earners (826,286 SOS), followed by girls from Lower Juba (556,084 SOS). Girls from Lower Shabelle earned just over half than girls from Gedo (436,111 SOS), whilst girls from Banadir and Bay were the lowest earners of all (260,263 SOS and 287,071 SOS, respectively).

FIGURE 18: INCOME BY REGION



Minority groups earned 355,217 SOS, in average. Girls who speak Maay at home earned on average slightly less than girls who did not speak the language (339,098 SOS vs. 376,174 SOS). When we use a more nuanced measure for linguistic divisions, the minority group outperforms the majority group. Girls belonging to a linguistic minority group earn somewhat more than girls belonging to a linguistic majority group (365,869 SOS vs. 403,786 SOS). Girls whose family is internally displaced also earn slightly more than girls whose family has not been displaced (405,284 SOS vs. 348,214 SOS). Finally, we find that a number of other factors, such as a lack of food and water, being divorced or orphaned, not living with their parents, having children and having disabilities, can negatively affect girls' income.

## Type of income generation activities

Girls were also asked about their occupation and business, as well as their contribution to agricultural work, the family business or work outside the home. Most girls reported not having an occupation (59.2 percent) or doing domestic chores inside the home (23.6 percent). A small share reported being an unskilled salesperson or service worker (7.3 percent), students (3.8 percent), and skilled salesperson or service worker (1.9 percent). Other occupation categories were rarely mentioned (< 1 percent). There

<sup>117</sup> 7 outlying values (above 3,000,000 SOS) representing less than 2% of the data were excluded from the analysis. In addition, 'don't know' and 'refused to respond' values were also dropped.

were some differences in how the main categories were distributed across regions. In Gedo, more than twice as many girls reported domestic chores as their occupation (51.9 percent) and a relatively large share reported being a skilled salesperson or service worker (7.4 percent). In Banadir (7.1 percent), Lower Juba (11.1 percent), and Bay (8.3 percent), a substantial number of girls reported working as an unskilled salesperson or service worker. Notably, 10 percent of girls in Bay were students and 10 percent of girls in Lower Shabelle worked in sustenance farming or fishing.

Minorities had quite a large share of unskilled salespeople and servicepeople (15.9 percent). Maay-speaking girls were more likely to have an occupation, be students or be unskilled salespersons and service workers in comparison to girls who do not speak Maay. These relationships also apply to girls from linguistic minority households, although they are less likely to be students than girls from majority households. Girls from families that are internally displaced are more likely to do domestic chores inside their household than those whose families have not been displaced. There was little variation in the results when disaggregating by other characteristics. Factors such as poverty do not necessarily limit girls' ability to succeed. One parent observed: "She is my neighbour - her family had very little money, they sold a sheep and with the money they got, they registered their daughters to school and bought some food"<sup>118</sup>.

Almost none of the girls surveyed have their own business (8.7 percent). This finding is reflected in the qualitative data, with only one parent mentioning that their daughter earns a living for herself<sup>119</sup>. Parents reported a number of barriers that inhibit them from starting a business, including lacking financial support<sup>120</sup> or parental guidance<sup>121</sup>, that girls are lacking knowledge<sup>122</sup>, or that they are already working for their parents' business<sup>123</sup>. Some parents consider that girls who are in school are unable to start a business<sup>124</sup>. However, parents generally acknowledge that the education that girls receive provides them with useful knowledge for starting a business and supporting their families' income. Girls who learn how to read and count from NFE courses will know how to set up a business, save money, and keep better track of their business<sup>125</sup>. Businesses suggestions include shops, beauty salons, selling food, clothing fairs, grocery stores, sewing clothes, and accounting<sup>126</sup>. One parent suggested the new skills might even lead to skilled employment; "Since she knows how to read and count, she can find job in banks or she can be

<sup>118</sup> Vignette Exercise with Girls, Bay, Int. 601.

<sup>119</sup> FGD with Mothers, Gedo, Int. 208.

<sup>120</sup> FGD with Mothers, Banadir, Int. 205; FGD with Mothers, Banadir, Int. 209; FGD with Mothers, Gedo, Int. 208; FGD with Mothers, Banadir, Int. 210.

<sup>121</sup> FGD with Mothers, Bay, Int. 204; FGD with Mothers, Gedo, Int. 207.

<sup>122</sup> FGD with Mothers, Gedo, Int. 208.

<sup>123</sup> FGD with Mothers, Bay, Int. 203.

<sup>124</sup> FGD with CEC Members, LowerJuba, Int. 105; FGD with Mothers, Bay, Int. 204.

<sup>125</sup> Vignette Exercise with Girls, Banadir, Int. 603; Vignette Exercise with Girls, Banadir, Int. 612.

<sup>126</sup> Vignette Exercise with Girls, Banadir, Int. 603; Vignette Exercise with Girls, Banadir, Int. 606.



a math teacher”.<sup>127</sup> The money that girls make from their businesses could be used to change their families’ lives and pay for their siblings’ school fees<sup>128</sup>.

In Gedo and Bay regions, girls appeared to be more entrepreneurial, with about one in five girls having a business (22.2 percent and 20.0 percent, respectively). In Banadir, the share of girls who reported having their own business was extremely low (2.0 percent). Girls from Maay-speaking and linguistic minority households were more likely to have a business than their counterparts. IDP and clan status did not have a substantial effect. From the very small subsample of 32 girls who reported having a business, most buy and sell food (56.3 percent), some sell prepared food such as setting up a tea stand or a small restaurant (21.9 percent) and the rest have a variety of businesses, including kiosks and beauty salons (21.9 percent).

Most of the girls are not asked to help with the household’s agricultural work (89.7 percent) or to support the family business or work outside their homes (75.5 percent). There were a few exceptions, however. In Lower Juba a large share of girls was working in agriculture (29.6 percent), and in Gedo the majority of girls were doing non-agricultural work for the household (66.7 percent). Finally, girls from minority groups, as well as Maay-speaking, linguistic minority, and internally displaced households are more likely to be involved in agricultural work.

## Synthesis

Girls’ reported income is quite high, as it is above Somalia’s GDP per capita of \$438<sup>129</sup>, despite most of the girls surveyed reporting not having an occupation or a business. The question is likely to be interpreted as the amount of money that the girls’ households made. The regional variations in income reflect to some degree the regional differences in girls’ self-efficacy, with girls in Gedo and Banadir representing both ends of the spectrum.

Girls belonging to Maay-speaking, linguistic minority, or internally displaced households had arguably higher incomes than their counterparts, because they were more likely to have an occupation of some kind. Perspectives on girl education and employment may be less entrenched and conservative in households which are not subjected to the dominant norms in society, even if they are marginalized by others.

Although most parents agree that girls who take NFE courses and participate in Rotating Savings and Credit Association (ROSCA)/ VSLAs will help girls set up a business and improve their families’ income, very few girls report actually achieving this ambition, with the exception of girls in Gedo and Bay. Many parents blame this fact on a lack of external investment in girls’ businesses, but many others simply do not consider girls to be old enough to set up their own businesses and others yet seem unable or unwilling to provide the necessary guidance or finances to support their daughters with their business ideas, and some are expected to support the family business. Even for girls who do receive the support to start their own business, it is mainly with selling food on the streets, as very few were able to convert their educational skills into business ideas or formal employment. ROSCA/ VSLA participation did not seem to benefit them either, with girls reporting that these savings groups were ineffective and were discontinued<sup>130</sup>.

<sup>127</sup> Vignette Exercise with Girls, Banadir, Int. 612.

<sup>128</sup> Vignette Exercise with Girls, Bay, Int. 601; Vignette Exercise with Girls, MiddleShabelle, Int. 609.

<sup>129</sup> World Bank DataBank, 2020.

<sup>130</sup> Vignette Exercise with Girls, Banadir, Int. 606.





# Baseline Results

---

New Non-Formal Education (Cohort 4)  
Girls

## 8. Learning

The present ML evaluation also serves as a BL for a new cohort of non-formal education girls, hereafter referred to as C4 NFE girls. As such, this section aims to establish benchmarks for their learning outcome levels for the numeracy and literacy assessments. Because C4 NFE girls were introduced to the study at BL, they were presented with the 11 sub-task numeracy assessment alongside other ML girls. For the first sub-section, we present the results of C4 NFE girls in terms of both the 8 sub-task and 11 sub-task numeracy assessments, in order to be able to draw initial comparisons between C4 and C1 NFE girls at their respective baselines while setting benchmarks for the AGES endline study, where we will compare changes in numeracy learning scores over time using the 11 sub-task grading scheme. The remaining sub-sections will analyse C4 NFE girls' learning outcomes based only on the 11 sub-task grading scheme.

### 8.1 Aggregate Learning Outcomes

On aggregate, C4 NFE girls tend to register very low scores on both the numeracy and literacy assessments, with the literacy scores being the lowest out of the three reported scores, with an average score of 16.90 across all three areas. C4 NFE scores are, as expected somewhat higher for the 8 sub-task grading scheme than the 11 sub-task, with average scores of 34.55 and 27.77, respectively. Notably, C4 NFE girls also register substantially lower scores than their C1 NFE counterparts' BL scores (for the 8 sub-task numeracy assessment and the literacy assessment), with a yawning gap of over 40 points on both assessments.<sup>131</sup>

TABLE 55: SUMMARY OF C4 NFE GIRLS SCORES AT BL

	Numeracy Scores, 8 Sub-Task			Numeracy Scores, 11 Sub-Task		Literacy Scores			# of C4 Girls
	C1 NFE, BL	C4 NFE, BL	Standard Deviation, C4	C4 NFE, BL	Standard Deviation, C4	C1 NFE, BL	C4 NFE, BL	Standard Deviation, C4	
<b>Banadir</b>	76.63	31.38	28.40	24.39	23.19	61.71	13.60	25.97	440
<b>South West State</b>	74.12	43.76	31.75	36.54	28.56	51.86	24.22	29.48	337
<b>Hirshabelle</b>		22.25	24.91	17.22	20.46		9.60	21.34	139
<b>Aggregate</b>	75.03	34.55	30.16	27.77	25.91	57.53	16.90	27.28	916

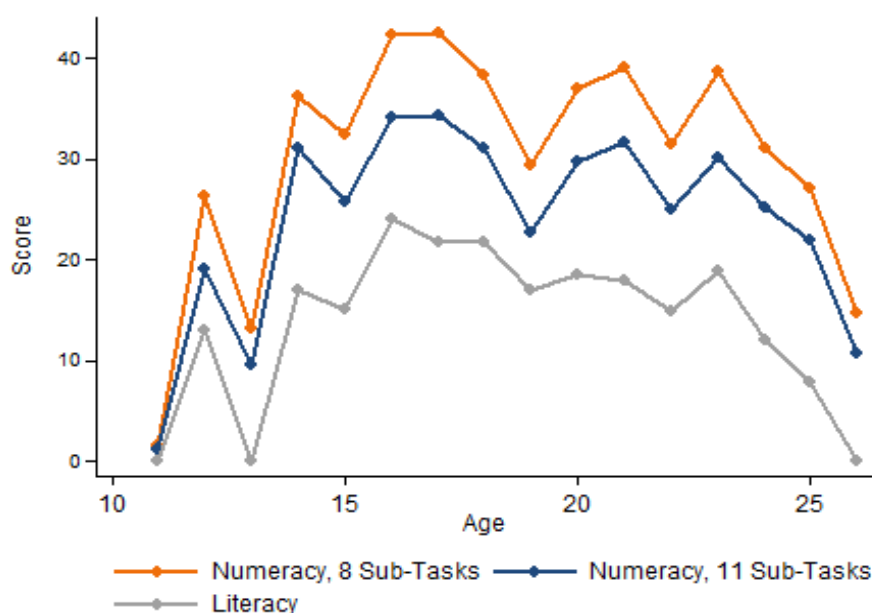
One major distinguishing factor between C1 and C4 NFE girls is the age range: while both cohorts have a median age of 18, C1 NFE girls are limited to the 17 – 20 age range, whereas C4 NFE girls range from a self-reported age of 11 years old at the youngest to 28 years at the oldest.<sup>132</sup> As previously seen in the present report and in the baseline report, age is correlated with higher assessment scores at any point in

<sup>131</sup> See previous section for a discussion of the demographic characteristics of the C4 NFE girls and how they differ from the C1 NFE girls, and more broadly, girls who were recruited into the study in 2019.

<sup>132</sup> C4 NFE girls were recruited within the age range of 15 to 25 years; however, poor administrative records and dislocation caused by conflict and drought often result in children and adolescents (and adults) being uncertain regarding their true age. Therefore, self-reported ages should be taken with a degree of caution.

time, meaning that the scores for C4 NFE girls are likely dragged down by the presence of younger, under-17 girls who tend to score lower than their older peers. Interestingly, based on the age-test score data, C4 NFE girls who are older also begin to experience decreases in their test scores: after age 16, scores on all assessments being to noticeably decline, with the drop in test scores becoming more substantial after age 20, as seen in the plot below. The different assessment score patterns with age are also similar, with literacy and numeracy patterns tracking each other closely as age increases. These trends are consistent with our expectations given our previous findings, as older girls, especially those over 20, may be expected to take on other tasks that reduce attention and time they have for school, such as supporting household economic activities, and may even face stigma when attempting to attend school.

FIGURE 19: C4 NFE GIRLS NUMERACY AND LITERACY MEAN SCORES, BY AGE<sup>133</sup>



We also observe geographic differences in scores: Hirshabelle scores the lowest in all assessments, while South West State scores the highest. In fact, geography and age are correlated: 78% of the C4 NFE girls from Hirshabelle state are over 18, compared to 68% and 66% for South West State and Banadir, respectively. When running a regression model with both age and state as regressors, we find that age is a negative, but not significant, predictor of numeracy and literacy scores, while geographic differences still have a statistically significant effect on scores.

In further examining these trends, we find that Hirshabelle C4 NFE girls were more likely than girls from other states to claim that they assist with household economic activities, including in agriculture and other activities, such as running a small business. We also find that Hirshabelle girls were less likely than South West State girls (the highest performers) to claim that the COVID-19 had had an economic impact on their household, such as loss of customers or forced closure due to lockdowns. In all, this suggests that in recent years, girls in Hirshabelle were expected to take on a greater share of household economic activities, which in turn might mean that they have less time to learn/study in a formal school setting. That the lockdown and various public health measures appear to not have impacted Hirshabelle's girls' household activities by further suggests that C4 NFE girls would be expected to help with household economic activities, while girls from other states whose household economic activities decreased due to COVID-19 restrictions might have had more available time to study. We also find that these measures,

<sup>133</sup> We omit ages 27 and 28 due to the small number of girls in those ages (1 person at each age).

too, are correlated with age, with older girls more likely to claim that they support household economic activities.

In sum, this short section has provided a quick overview of C4 NFE girls' learning outcomes at their baseline. The general image emerging from this brief analysis is that C4 NFE girls, compared to C1 NFE girls, are poorer performers in numeracy and literacy assessments. The C4 NFE cohort encompasses a greater range of ages, at the extremities of which learning outcome scores tend to be lower. In addition, age, state, and participation in household economic activities appear to all be correlated, with older girls, from Hirshabelle, more often claiming to contribute to household tasks that might detract from their learning. The above analysis thus informs our approach to subsequent sections. In the next sub-section, which further investigates differences in assessment scores among other sub-groups statuses in the C4 NFE cohort, will use state and age as control variables in order to be able to detect the independent effects of other sub-group categories.

## 8.2 Subgroup Learning Outcomes

In this subsection, we investigate whether C4 NFE girls who belong to specific sub-groups score higher or lower compared to cohort girls who do not belong to these sub-groups. Specifically, we investigate several categories of sub-groups revolving around girl's disability status, household and girl's social characteristics, household education, and household economic status. Our quantitative analysis proceeded in two steps. First, we regressed numeracy and literacy scores for C4 NFE girls on the sub-group variable of interest through a linear regression model without controls. For those sub-group characteristics that were significant at at least the 90% confidence interval, we re-ran the regression model with the addition of control variables: the girl's age and her state of residence, both of which we have found to have an effect on C4 NFE girls' learning outcomes. As previously discussed, here we use the stricter 11 sub-task grading scheme to assess girls' numeracy learning.

Due to the lengthy list of possible sub-groups to measure, this sub-section presents only the characteristics whose effects on assessment scores remained significant even after the introduction of the two control variables, though the numbers reported are from the simpler, no controls, model. A full list of characteristics considered and regressed on using the model without controls is attached in Annex 3 for the reader's consideration.

TABLE 56: SUB-GROUP IMPACTS ON BL NUMERACY AND LITERACY SCORES OF C4 NFE GIRLS

	Numeracy			Literacy			#
	BL Score	Coefficient	P-Value	BL Score	Coefficient	P-Value	
<b>ALL C4 NFE GIRLS</b>	<b>27.77</b>			<b>16.90</b>			
<b>Disability Status</b>							
<b>Has cognitive disability</b>	17.71	-10.38	0.03	4.81	-12.47	0.00	28
<b>Has cognitive disability, alternative criteria</b>	19.84	-9.20	0.00	9.24	-8.88	0.01	126
<b>Has hearing disability, alternative criteria</b>	17.56	-10.50	0.04	7.42	-9.75	0.01	25
<b>Has difficulty in caring for self, alternative criteria</b>	18.19	-9.65	0.41	5.67	-11.31	0.00	7
<b>Has visual disability</b>	54.17	26.51	0.04	28.34	11.49	0.49	4
<b>Household and Girl Social Characteristics</b>							
<b>Is or was ever married</b>	25.52	-3.63	0.55	2.31	-16.30	0.00	10

<b>Currently married</b>	33.28	4.37	0.51	3.30	-15.10	0.00	7
<b>Is a mother</b>	31.42	2.44	0.77	0.53	-17.81	0.00	5
<b>Household and Girl's Economic Status</b>							
<b>Went without income most/all days in past 12 months</b>	19.78	-11.75	0.00	11.83	-7.88	0.03	60
<b>No clean water for most days in past 12 months</b>	17.72	-12.26	0.00	6.21	-12.82	0.00	22
<b>Household owns land</b>	39.75	14.95	0.00	27.28	12.89	0.00	79
<b>Owns smartphone</b>	40.24	12.57	0.03	30.21	13.65	0.03	30

To begin, we find that several measures of disability status led to girl earning lower scores at BL compared to their peers who are not recorded as having these specific disabilities.<sup>134</sup> Specifically, having a cognitive disability, such as forgetfulness or difficulty understanding speech and having a hearing disability are strongly related to more negative scores at BL, for both numeracy and literacy measures. This is unsurprising given that cognitive and hearing skills are necessary skills in order to be able to complete one's studies well. Hearing impairments, for instance, might lead a girl to not be able to clearly understand what a teacher is attempting to explain, simply because she cannot hear well, which in turn leads to the girl having lower test scores. To a lesser extent, having difficulties in caring for one self, such as an inability to get dressed independently, or having visual impairments, also serve to decrease girls' scores, although this finding was not significant for the literacy test but significant for numeracy test. It is, however, important to note that the sample of girls whom we assessed to be in each disability category is very small (at 7 and 4 girls, respectively). In addition to disability status, we also find that a girl's marital and parental status have a considerable impact on learning. In the table above, it can clearly be seen that married girls, girls who have children, and those who are currently married all have lower averages scores than girls not burdened with these traits, though the coefficient only is only significant for the numeracy skills measure.

Finally, we also found a positive correlation between several measures of household income, and as expected, found outstanding challenges facing C1 NFE girls when the household economy is in need of helping hands. Girls who come from household members who own land may receive more of the financial and social support that children need to grow learn, including appropriate food, tuition fees, and so on. Thus, it is not surprising to see their higher scores. Girls who owned a smartphone, as opposed to a regular phone, also registered higher assessment scores for both measures, in part because smartphones hold almost limitless capabilities and could be conveniently used by learners to improve their reading and writing skills.

Conversely, several negative traits that we hypothesized would lead to lower scores do indeed take on the same sign as expected. All else equal, girls whose households went without income or clean water for most days during the past 12 months would likely struggle with some of the basic prerequisites for

<sup>134</sup> For this sub-section, because the sample size of girls was small for disability categories, we used three separate coding schemes and tested each. The first one is based on a girl claiming to have experienced certain things that are indicative of a disability, while the second measure coded a girl as having a disability if either herself or her caregiver mentioned so. Finally, the third alternative measure relied on a lower threshold for considering a girl to have a disability.

succeeding in the classroom setting, such as maintaining focus, or even basic attendance. As such, it is recommended that the programme continue to monitor the progress of C4 NFE girls who come from more challenging household economic situations, and prepare any potential mitigation plans in case these household conditions become excessively burdensome and lead to a girl not attending school or not learning.

### 8.3 Foundational Skill Gaps

Shifting from our discussion of learning outcomes for different types of girls, we now turn to a closer dissection of how girls performed on particular aspects of the learning assessments. Our goal in this section is to understand more clearly the nature and extent of literacy and numeracy among the new cohort of NFE girls recruited by the programme in 2022. To do so, we analyse performance on individual subtasks, categorising girls according to their level of proficiency on each subtask. The categories are as follows:

- Non-Learner – was unable to complete any of the subtask (scored 0%).
- Emergent Learner – was able to answer a portion of the subtask correctly, but generally unable to perform the tasks demanded (scored 1-40% on the subtask).
- Established Learner – was able to answer around half or more of the individual test items correctly, showing facility with the subtask but not mastery (scored 41-80% on the subtask).
- Proficient Learner – was able to answer most of the individual test items correctly, showing a degree of mastery of the subtask (scored 81-100% on the subtask).

While it is also possible to analyse mean scores on individual subtasks – analysis which we briefly report in Annex 3 – using a single measure of performance, especially one measuring the average or “central tendency” of the sample, masks important patterns. The most important potential pattern is divergence between girls with regard to particular skills, in which girls are sorted into opposite ends of the spectrum, with only non-learners and proficient learners, and few girls situated in the middle range of performance. Divergence of this kind is obscured when subtask-specific scores are analysed according to their mean.

In the table below, we report the share of girls who fall into each of the four categories (listed down the first column) for each subtask (listed across the header of the table) in the numeracy assessment. Included along the header is a brief description of the skills each subtask tests. For instance, subtask 1 assesses whether a child can recognise 1- and 2-digit numbers, ranging from 1 to 99, while subtask 2 asks children to identify the larger of two numbers, with numbers ranging from 1- to 3-digits.

The first notable finding is that a significant minority of girls are unable to identify any numbers when asked to do so. Out of 916 girls, 20.3 percent were unable to identify a single number of 25 presented to them. It is important to note that some girls may be able to identify one or more numbers but wished not to participate for any number of reasons. Based on this and our past work with similar learning assessments, it seems there are occasionally girls who feel they are unable to read or do maths at higher levels and therefore do not engage with even the easier subtasks that we use to start the assessment. This viewpoint is entirely speculative, but seems consistent with the evidence, especially when the easiest subtask (subtask 1) and a noticeably more difficult subtask (e.g., subtask 2) do not stymie vastly different numbers of girls. In other words, we might expect a large increase in the number of girls who are entirely



TABLE 57: GAPS IN FOUNDATIONAL NUMERACY SKILLS AMONG NEW (COHORT 4) NFE GIRLS

Subtask	1	2	3	4	5	6	7	8	9	10	11
	Number Ident.	Number Discrimination	Missing Numbers	Addition (1 digit)	Addition (2 digits)	Subtract. (1 digit)	Subtract. (2 digits)	Word Problems (add. & subtract.)	Multiplic. (1 digit)	Division (1 digit)	Word Problems (mult & div)
Non-learner 0%	20.3	27.1	52.7	58.7	76.4	68.4	76.2	56.4	91.5	87.2	72.2
Emergent learner 1%-40%	10.3	8.4	34.6	2.6	6.1	2.3	9.8	0.8	4	2.1	12.8
Established learner 41%-80%	8.7	19.3	10.6	4.8	7.1	4.1	6.9	10.6	2.4	3.6	10.7
Proficient learner 81%-100%	60.7	45.2	2.1	33.8	10.4	25.1	7.1	32.2	2.1	7.1	4.4
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

unable to perform a subtask when we move from simple number identification to number discrimination, but we do not find this to be the case. The pattern we see below – in which the first subtask discriminates sharply between girls who are and are not able to perform the task, and the second subtask does not discriminate much more dramatically – is empirically consistent with a situation in which a subset of girls who may be able to identify *some* numbers choose not to do so.<sup>135</sup>

The second finding concerns the identification of patterns in numbers and identification of missing numbers within those patterns. Girls were presented with patterns that each required one of three skills, in addition to pattern recognition: counting, adding, and subtracting. Even the most straightforward addition-based patterns (e.g., 10, 20, \_\_, 40) were completed successfully by fewer than one-fifth (18.8 percent) of girls. A test item that required counting, albeit using two-digit numbers (78, \_\_, 80, 81) was similarly difficult, with just 16.5 percent of girls completing it successfully. A similar share of girls were able to complete most two-digit addition problems, including some that required carrying digits effectively (e.g.,  $14 + 22$ ). Girls' relatively poor performance identifying missing numbers in a pattern has one of two possible explanations: first, that the nature of pattern identification and the purpose of the subtask was confusing or unclear to girls; second, that the application of arithmetic operations to this unfamiliar use – as opposed to simply calculating sums or differences in the normal format of a mathematical problem – made the completion of what would otherwise be fairly simple addition much more difficult.

A related pattern concerns girls' performance on word problems, especially those involving addition and subtraction. Word problems are often considered more difficult than the specifics of the underlying arithmetic would suggest, because students need to conceptualise the problem, taking a story and forming an arithmetic problem from it. The process of identifying the underlying problem can reduce success rates. In this case, however, girls seem to perform comparatively well on word problems. A comparison of subtask 8 and subtasks 3 (1-digit addition) and 5 (1-digit subtraction) shows that girls performed better on word problems, despite the equivalent mathematical difficulty. For instance, a word problem whose underlying calculation required addition ( $5 + 4$ ) was answered correctly by 40.3 percent of girls, while an equivalent addition problem in subtask 4 ( $6 + 3$ ) was answered correctly by 36.5 percent of girls.

Of course, it is possible that these small differences are simply a function of random variation. If that is not the case, the most likely explanation is that many NFE girls have a rudimentary grasp of basic arithmetic but are better at using these skills in precisely the contexts in which they likely learned it – everyday issues of addition and subtraction that would be relevant in a marketplace, a shop, or elsewhere. This helps explain why girls would perform best on word problems, slightly worse on equivalent arithmetic couched in formal notation without a practical application), and somewhat worse still on equivalent arithmetic placed in an unfamiliar context whose purpose may not be very clear (number pattern identification). This explanation is especially compelling, given that the vast majority of NFE girls have not completed any schooling and likely learned the arithmetic they know from functional use.

The final pattern in mathematical performance is one that we highlighted previously among girls assessed in the AGES baseline in 2019, which is the sharp deterioration of performance as arithmetic problems begin to require application of more complex rules or procedures. The clearest demonstration of this trend can be seen in the decline in performance from subtask 4 to subtask 5. The typical NFE girl achieved a score of 37.6 percent on the former subtask, whose most difficult problems required only simple “carrying” in problems whose answers are above ten. On the latter subtask (subtask 5), girls were required to perform more complex carrying (e.g.,  $14 + 25$ ) and the typical girl's score fell to 16.0 percent. A

<sup>135</sup> Again, we emphasise that this is an explanation based on conjecture; other arguments can also explain this pattern of subtask-specific results, including the possibility that number identification is simply more difficult than we imagine.

similar decline can be seen in the context of subtraction. Our argument in previous studies has been that girls without formal schooling will often have facility with simple addition but tend not to understand the procedures that would allow them to solve more difficult problems or apply their intuitive understanding of the procedures – such as “carrying” a digit for an answer between 10 and 20 – to slightly more difficult problems. This same explanation seems to fit the patterns of numeracy skills for this new cohort of NFE girls.

By comparison to numeracy, the patterns of girls’ achievement in literacy are less complicated or nuanced. As the table below shows, performance is relatively poor across all subtasks, with fewer than 20 percent of girls achieving scores of above 40 percent on any but the first subtask. The most remarkable patterns with regard to literacy are built around this relatively poor performance, which is perhaps understated by the mean score among NFE girls of 16.9 percent.

The first finding concerns the initial subtask, in which girls were asked to identify individual letters. As the table shows, 44.5 percent of girls were unable to name a single letter. As noted above, the high rate of zeroes on this subtask may be explained, in part, by girls who feel or know they cannot read and who refuse to engage with the task of letter recognition as a result. Even so, an additional 18.2 percent of girls could identify between one and 20 (out of 50 total) letters, which indicates an extremely low level of aptitude, even when ignoring the large set (44.5 percent) of girls who could not name a single letter.

**TABLE 58: GAPS IN FOUNDATIONAL LITERACY SKILLS AMONG NEW (COHORT 4) NFE GIRLS**

Subtask	1	2	3	4	5	6
	Letter recognition	Common words	Reading fluency	Reading comp. 1	Reading comp. 3	Reading comp. 4
Non-learner 0%	44.5	66.8	76.5	78.9	79.6	81.2
Emergent learner 1%-40%	18.1	15.7	9.5	3.7	1.6	6.6
Established learner 41%-80%	16.5	9.6	8.1	10.8	12.2	9.4
Proficient learner 81%-100%	20.9	7.9	5.9	6.6	6.6	2.8
Total	100%	100%	100%	100%	100%	100%

Reading fluency levels are also potentially misleading when reported as mean scores. For NFE girls in this cohort, the average score on subtask 3 was 12.1 percent. This score corresponds to just 9.2 words read out of 76 possible in one minute. Low scores on reading fluency could indicate very low reading *speeds*, even if reading comprehension was adequate. This is unlikely – as reading speed and understanding are correlated – but it is possible, particularly in early readers. However, the EGRA results do not suggest this is the case: even when girls were given additional time to read the passage, ability to answer relatively basic questions about the passage (i.e. subtask 4) was extremely limited.

On a slightly broader level, the discussion above also makes clear that girls in this cohort have acquired rudimentary numeracy skills through necessary daily use, while the same is not true of literacy. While a majority of girls could identify most numbers and a significant minority could reliably perform basic addition, very few girls were able to identify letters and almost none were able to identify common Somali words. The much better numeracy performance among this cohort indicates that numeracy skills are learned through use and necessity, while literacy learning does not benefit from girls “learning in the background” through applied use.

## 8.4 Grade-Level Competency

In this section, we analyse the share of girls who achieve lower-primary (Standard or Grade 4) competency in Somali literacy and numeracy. Unfortunately, it is not possible to distinguish competency at a more fine-grained level – i.e. individual grade levels – due to the sparse information provided by the current FGS educational curriculum. As noted at the baseline, the current curriculum specifies learning targets only for lower-primary and upper-primary levels. As such, it is not possible to assess specific grade level competencies achieved; only competency at the conclusion of multiple grade levels. In addition, the design of the EGRA and EGMA learning assessments used at baseline test skills that only partially overlap with the set of skills selected as targets in the FGS curriculum. A number of skills specified as targets for lower primary school in the curriculum are not directly tested by the learning assessments, including multiplication and division.

With these caveats in mind, we mapped subtasks in both numeracy and Somali literacy onto the FGS curriculum for lower primary, to the extent possible. Higher-level competencies (e.g., upper primary and secondary school) are also specified by the FGS curriculum but – as with lower primary – only in terms of overall course completion, such as completing upper primary or secondary school. Moreover, the skills tested through the baseline's learning assessments are, by design, too easy to allow us to judge whether a girl has achieved competency sufficient for completing upper primary school.

To achieve lower primary and upper primary competency in literacy, per the FGS curriculum, children must meet these standards:

### Lower Primary

- Read simple two paragraph texts (for comprehension)
- Write short paragraphs about themselves and their environments

### Upper Primary

- Read short stories and factual texts
- Write their own stories and factual accounts
- Follow more complex instructions and informational sources such as textbooks

In the case of numeracy, the following standards are applied:

### Lower Primary

- Use of whole numbers
- Addition, subtraction, multiplication and division
- Simple fractions and decimal places
- Shapes, basic geometry such as properties of angles and parallel/perpendicular lines, use of measuring instruments
- Interpreting tables and graphs

### Upper Primary

- Whole numbers up to millions, and place values
- Exponents and square/cube roots of up to 3-digit numbers
- Logarithms
- Sets, unions, intersections, and subsets
- Ability to recognize number patterns
- Area, perimeter, and volume of circles, quadrilaterals, and non-planar shapes
- Ability to solve geometric problems in triangles and quadrilaterals using Pythagoras' Theorem

- Basic probability and calculations of mean, median, mode
- Solve everyday problems involving money, percentages, discounts
- Solve linear equations with two variables, simultaneous equations, and quadratic equations

The evaluation team mapped these specific skills onto the EGRA and EGMA assessments, to the extent possible. Girls are defined as achieving lower primary competence in reading if they score 80 percent or higher on all Somali literacy subtasks. All six subtasks in the assessment fall within the bounds of lower primary competency, according to the FGS curriculum, as the curriculum requires students to be able to read two paragraph texts effectively for comprehension. Note that we classify girls who meet this standard as partially competent, as the learning assessment did not include writing tasks which would be necessary to fully achieve lower primary competency. The same logic is applied to numeracy – girls must achieve 80 percent scores on all subtasks; doing so is viewed as partial competency, because the assessment only tests some of the skills targeted by the FGS curriculum.

By this standard, very few of the New (Cohort 4) NFE girls analysed here achieved lower-primary competency. This is not at all surprising, given that NFE girls generally have received no schooling in the past and certainly have not completed multiple years of formal primary school. Of the 916 NFE girls assessed, 2.4 percent achieved a literacy level equivalent to a girl who had completed lower primary school and 0.8 percent achieved an equivalent level of numeracy.

## 9. Intermediate Outcomes

### 9.1 Attendance

The first intermediate outcome for the new cohort of NFE girls (C4 NFE girls) is attendance. It is important to emphasise – throughout our discussion of the C4 NFE cohort – that enrolment of girls into the C4 NFE programmes was completed immediately prior to the start of data collection for this evaluation, which has implications for how the results should be interpreted. In terms of attendance, we might expect attendance to be unusually high at the start of a learning programme, tapering off as time goes on, resulting in a potential downward trend in attendance between now and the endline. On the other hand, we might expect training of teachers to be in its infancy, resulting in relatively poor teaching practices at this baseline and rapidly improving teaching practices as time goes on.

Attendance for the C4 NFE cohort was captured exclusively through headcounts conducted by our field researchers during their visits to each centre. At the centre, a headcount was conducted in every functioning NFE classroom, without sampling. This provides the largest possible sample size for assessing changes over time in attendance rates. Note that, in locations with both C1 and C4 NFE girls, we consider the completed headcounts to be part of the C4 NFE sample, as the C1 cohort's learning programme ended in September 2020. Our total sample includes 226 headcounts in centres participating in the C4 NFE round, spread across Banadir, Hirshabelle, and South West State.

The table below documents attendance rates by state. Headcount-derived attendance rates are defined as the number of girls present in class at the time of the visit, over the number of girls included on the enrolment register held by the head teacher or administrative staff. The value in parentheses provides the number of observations over which the attendance rate is calculated.

TABLE 59: ATTENDANCE HEADCOUNTS IN C4 NFE CENTRES, BY STATE

State	Headcount Attendance	Teacher-Recorded Attendance
<b>Banadir</b>	80.6% (102)	82.3% (83)
<b>Hirshabelle</b>	73.0% (27)	71.1% (27)
<b>South West State</b>	70.0% (89)	71.9% (78)
<b>Total</b>	75.3% (218)	76.4% (188)

We also collected information on the attendance taken by the teacher prior to our visit, as many teachers document attendance on a daily basis. This value often differs from our own headcount, for a variety of reasons, including the possibility that students left or arrived between the time the teacher took attendance and the time our team arrived in the classroom. Alternatively, teachers may take attendance poorly, introducing errors or biases into their own count. Nonetheless, the teacher-recorded attendance provides a secondary check on attendance rates, where they are recorded.

The results in the table show a sharp difference in attendance rates between the centres located in Banadir and those in Hirshabelle and South West State, respectively. It is important to emphasise that attendance rates from headcounts are heavily dependent on the records kept regarding enrolment. Incorrect enrolment numbers – either inflated or deflated – can drastically change recorded attendance rates in a manner that is opaque to outside researchers. It is possible that this is driving differences between states in this baseline, though it is also possible that attendance rates are simply higher in the sampled Banadir locations.

## 9.2 Teaching Quality

We now discuss the four dimensions of teaching quality and practices—overall professionalism, gender equity, the use of physical punishment, and pedagogical practices—measured for NFE Cohort 4 girls. These measures provide a baseline for further analysis of changes in teaching quality and practices in NFE programmes at endline.

Among NFE Cohort 4 girls, we find high average rates of girls reporting that the teacher makes them feel welcome. However, we note that rates are significantly lower in Hirshabelle than in Banadir or South West State, suggesting that it may be important to intervene with teachers in Hirshabelle to ensure a welcoming learning environment. For teacher absenteeism, we find slightly lower overall reported rates among NFE Cohort 4 girls compared to FE, ABE, and NFE schools discussed in the main section on teaching quality, with around 18 percent of girls, on average, agreeing strongly or somewhat that teachers are often absent. Results are slightly lower in Banadir, though fairly consistent across regions. Overall, although rates of reported teacher absenteeism are relatively low, these findings still suggest a need to strengthen teacher attendance, as absenteeism may have a severe impact on students' learning outcomes.

TABLE 60: TEACHER PROFESSIONALISM, C4 NFE

Outcome	Banadir	South West State	Hirshabelle	Total
<b>Feels welcome</b>	99.1%	97.0%	83.0%	95.9%
<b>Often absent</b>	14.2%	20.9%	20.7%	17.7%

Regarding gender equity, within C4 NFE programmes, there were no boy students. As such, we were unable to observe differential treatment of girls and boys by teachers. Instead, the table below reports observations of several positive and negative teaching practices used on girl students. We find that on average per class block, teachers called on girl students around 10 times; teachers called on students most frequently in Hirshabelle and least frequently in Banadir. These findings suggest a need to strengthen C4 NFE girl student participation in Banadir in particular, an issue discussed further below.

For the provision of positive and negative feedback to girls, we find that around 70 percent of teachers were observed to provide positive feedback to girls in at least two out of three class blocks. Rates were highest in Hirshabelle, at over 90 percent. However, teachers also used harsh language against girls most frequently in Hirshabelle, at an average of 1.1 times per class block compared to an overall average of only 0.2 times per class.

**TABLE 61: TEACHERS' INTERACTIONS WITH GIRL STUDENTS, C4 NFE**

Outcome	Banadir	South West State	Hirshabelle	Total
<b>Called on girl student</b>	7.7	10.2	14.9	9.7
<b>Provided positive feedback</b>	63.6%	66.7%	92.9%	69.4%
<b>Used harsh language</b>	0.02	0.2	1.1	0.2

Looking now at the use of disciplinary practices, the table below shows that less than 29 percent of C4 NFE girls, on average, stated that teachers use discipline for wrong answers and less than 9 percent stated that teachers use corporal punishment. Findings for corporal punishment are validated through results from the classroom observations, reported in the last row of the table, in which girls were observed to be disciplined physically at similar rates as reported in the survey with girls. Reported use of negative disciplinary practices was highest in Hirshabelle, particularly regarding the use of discipline for wrong answers, which was reported by over half of girls. Overall, however, these results are noticeably lower than those for FE, NFE cohort 1, and ABE girls.

We note that given these relatively low “baseline” values for NFE Cohort 4—particularly for the use of corporal punishment—it may be difficult to measure meaningful change over time, as there is relatively little room for improvement. However, it is also important to note that, particularly for corporal punishment, any value above zero percent represents children who are physically punished in classrooms, creating a hostile learning environment that may drive students away from school, reduce their confidence or interest in learning, or even permanently cause physical damage. As such, despite relatively low baseline values, it is important to continue to push teachers to eradicate this practice entirely.

**TABLE 62: USE OF DISCIPLINARY PRACTICES, C4 NFE**

Outcome	Banadir	South West State	Hirshabelle	Total
<b>Discipline for wrong answer, reported by girls</b>	16.6%	32.1%	55.4%	28.2%
<b>Use of corporal punishment, reported by girls</b>	7.8%	7.5%	13.4%	8.6%



<b>Girls disciplined physically (observed)</b>	4.6%	7.4%	14.3%	7.1%
------------------------------------------------	------	------	-------	------

Lastly, we now analyse pedagogical practices used by teachers. The below table shows the use of formative assessments and participatory teaching practices disaggregated by region. For formative assessments, we find that on average, only around one-quarter of teachers in C4 NFE programmes reported using formative assessments and even fewer—less than 12 percent—were able to produce records of these assessments, suggesting that formative assessments may not actually be used in some cases, may be used very infrequently, or that appropriate records may not be kept. It is also possible, given the proximity of the evaluation to the start of NFE programmes – that is, the learning programmes in question had just begun operations at the time of the baseline – that formative assessments were not being used yet, as teaching had only just begun. We also find large regional variations in the use of formative assessments, with nearly non-existent use by teachers in Banadir in contrast to use by around half of teachers in SWS and Hirshabelle. Interestingly, these results contrast to those found for FE, C1 NFE, and ABE teachers in Banadir, where the use of formative assessments was relatively high. Overall, however, these results show a need to strengthen teachers’ use of formative assessments as well as, potentially, their record keeping in all regions.

Examining participatory teaching practices, we find that on average, teachers most frequently involved students who were not participating in class and used open-ended questions, while the use of group work and student-centred activities or games was very infrequent. However, it is important to note that no participatory teaching practice was used by more than 60 percent of teachers observed in the classroom.<sup>136</sup> The prevalence of these teaching practices may also be somewhat overestimated in the data due to observer effects, whereby teachers may be more likely to model good practices while they know they are being observed. These results thus suggest a need for substantial improvement in the use of positive and participatory teaching practices.

The below table also reveals several gaps by region. Use of most participatory teaching practices was observed least frequently in Banadir, with the exception of the practice of allowing students to instruct each other. In contrast, participatory practices were observed most frequently in Hirshabelle (with the exception of activities or games); in particular, the use of open-ended questions and involving students who were not participating in at least two out of three class blocks was observed in over 90 percent of schools. These results suggest that while improvement in pedagogical practices is needed in all regions, the need is particularly acute in Banadir C4 NFE programmes.

**TABLE 63: USE OF PARTICIPATORY TEACHING PRACTICES IN CLASSES, C4 NFE**

Practice	Banadir	SWS	Hirshabelle	Total
<b>Formative assessments (stated use)</b>	4.6%	44.4%	57.1%	25.9%
<b>Records of formative assessments</b>	0.0%	18.5%	35.7%	11.8%
<b>Student-centred activities or games</b>	15.9%	25.9%	0.0%	16.9%
<b>Students instruct each other</b>	59.1%	18.5%	35.7%	42.4%

<sup>136</sup> Importantly, this is not an indictment of the programme, as most of the centres in question have just begun operation and training of NFE teachers had not been implemented widely at the time of data collection.

<b>Open-ended questions</b>	40.9%	46.2%	92.9%	51.2%
<b>Questions solicit student opinions</b>	40.9%	48.2%	57.1%	45.9%
<b>Involve students who are not participating</b>	45.5%	66.7%	92.9%	60.0%
<b>Group work</b>	4.6%	11.1%	14.3%	8.2%

The below table additionally reports results from the survey with girls, and shows that NFE Cohort 4 girls most frequently reported that their teachers often explain the use of lessons for their lives and provide students with ideas to study at home. In contrast, girls were least likely to state that lessons move at a speed that is “just right”, with only around half of girls agreeing with this statement while 30 percent stated lessons moved too fast and 18 percent that lessons moved too slow. Given the substantial share of girls who stated lessons moved too slow, some of the challenges related to lesson speed are likely due to the wide range of girls’ educational backgrounds and abilities within classrooms, which can make it challenging for teachers to teach at the right pace.<sup>137</sup>

The regional differences shown in this table contrast somewhat to findings in the table above. Girls in Banadir were more likely to report the use of three positive teaching practices—explaining the use of subjects to their lives, lessons moving at the right speed, and teachers encouraging participation—than girls in SWS or Hirshabelle. The latter finding particularly contrasts to findings from the classroom observations, in which teachers in Banadir were observed to encourage student participation at lower rates than teachers in SWS or Hirshabelle. These findings suggest that participatory teaching practices in Banadir may be more common than suggested by the classroom observations. However, as above, we note that the use of many positive teaching practices—especially providing students with ideas to learn outside of class, teaching at the right speed, explaining concepts in different ways, and encouraging participation—remains low, with only around 60 percent or less of girls reporting that their teachers use these practices. There is thus a need to continue improving the quality of teaching and use of positive teaching practices in all regions.

**TABLE 64: USE OF POSITIVE TEACHING PRACTICES REPORTED BY GIRL STUDENTS, C4 NFE**

<b>Practice</b>	<b>Banadir</b>	<b>SWS</b>	<b>Hirshabelle</b>	<b>Total</b>
<b>Explain use of subjects (<i>often</i>)</b>	85.7%	77.7%	76.3%	81.3%
<b>Ideas to learn outside class (<i>often</i>)</b>	57.7%	56.4%	63.7%	58.1%
<b>Ideas to study at home (<i>yes</i>)</b>	67.3%	89.6%	96.4%	79.9%
<b>Lessons at good speed (<i>just right</i>)</b>	65.2%	38.6%	38.1%	51.3%
<b>Different ways of explaining (<i>often</i>)</b>	59.4%	60.5%	69.6%	61.4%

<sup>137</sup> As noted above, it is important not to draw too firm of conclusions from this finding, given that C4 NFE classes had just begun at the time of data collection and teachers had received orientation but minimal or no training at that time. Indeed, given that this is a baseline assessment for the C4 NFE cohort, the lack of training – and accompanying issues with pedagogical practice – are expected.

Encourage participation ( <i>often</i> )	67.4%	60.5%	53.3%	62.7%
------------------------------------------	-------	-------	-------	-------

### 9.3 Leadership and Life Skills

In this section we will not assess the change in scores from baseline to midline, but will carry out a brief baseline analysis and set baseline scores for the new cohort of NFE girls that were recruited in 2022. As well as the Youth Leadership Index, scores for the Life Skills index will be analysed. This index is calculated by asking respondents a series of questions on life skills, self-esteem, and autonomy, assigning low values to negative answers and high values to positive answers, and then standardising the scores from 0 – 100. Further details including the full list of questions can be found in the introduction. Factors that may predict better or worse YLI and LSI scores will also be analysed in this section.

FIGURE 20: DISTRIBUTION OF LSI SCORES, BY STATE

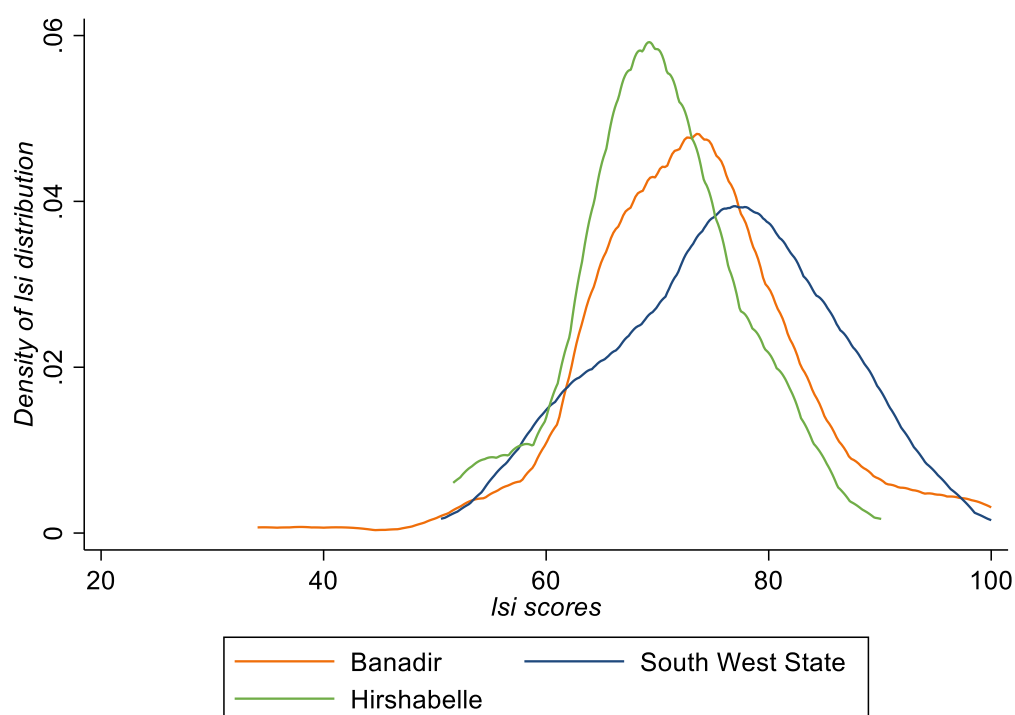


TABLE 65: MEAN LSI SCORES, BY STATE

Outcome	N	Mean LSI score
Banadir	440	73.2
Hirshabelle	139	70.0
South West State	337	75.9

In the figure above we can see the distribution of scores of the Life Skills Index, while in the table above we can see the means. The scores are disaggregated by the three zones that the New NFE cohort is from – Banadir, South West State, and Hirshabelle. The distributions are broadly similar, and between regions girls do similarly on LSI scores. Across all regions, most girls tend to do relatively well. The distributions

are centred between approximately 65-80 out of 100 and are overall quite narrow, meaning most fall within this region of scoring.

There are some differences, however, between regions in the scores. The distribution of Hirshabelle is centred slightly below that of Banadir, which is in turn centred slightly below that of South West State. This is also reflected in the means – Hirshabelle has the lowest mean of 70, Banadir’s is slightly higher at 73.2, and South West State has the highest at 75.9. The width of distributions are also different, meaning they vary less. The practical impact of this is that the wider the distribution, the higher the inequality in the Life Skills Index within that region. Hirshabelle has the narrowest distribution, Banadir has a slightly wider one and also has the lowest minimum, and South West State has the widest.

It is also interesting to note the correlation between this score and the Youth Leadership Index scores. These both capture aspects of self-esteem and life skills and some questions are quite similar, meaning we would expect a high correlation between the two. The correlation is calculated to be 0.41. This is a reasonably strong association, but it also makes clear that the two indexes are not capturing exactly the same characteristics – for example there are several questions probing autonomy in the LSI with no parallel in the YLI. This highlights the value of having both indexes to give a fuller picture of self-esteem and life skills.

FIGURE 21: DISTRIBUTION OF YLI SCORES, BY STATE

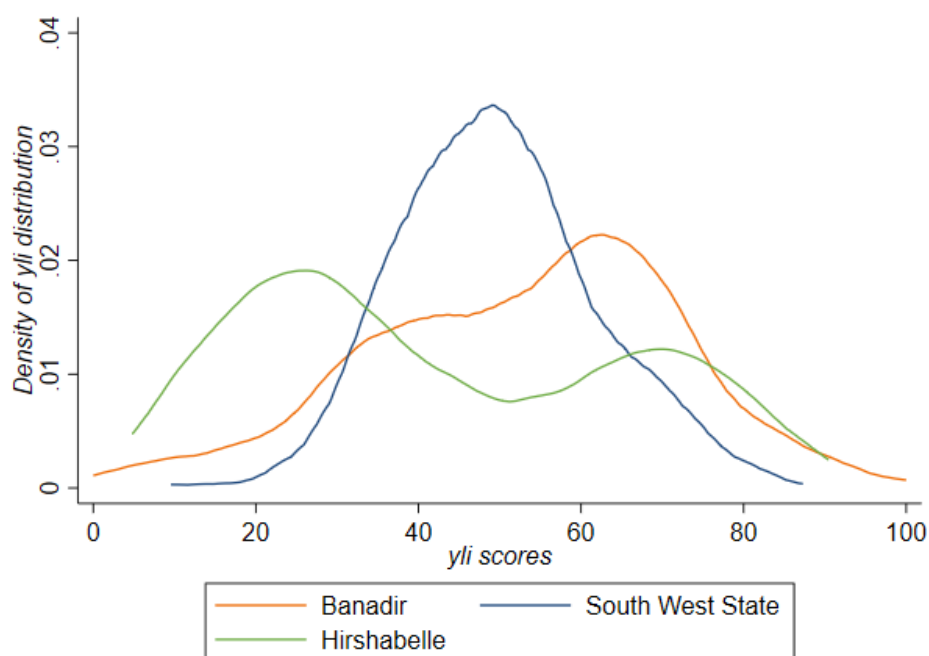


TABLE 66: MEAN YLI SCORES, BY STATE

Outcome	N	Mean YLI score	Proportion achieving target
Banadir	440	52.9	14.3
Hirshabelle	139	42.3	16.5
South West State	337	49.6	5.3

In the figure above we can see the distribution of the Youth Leadership Index, while in the table we can see the means. Again, the scores are disaggregated by zone. The distributions for each of the zones are distinct from each other and show very different patterns. South West state shows a nearly normal distribution, with scores centred around 50 and some girls doing better or worse than this. In Banadir, the distribution is skewed and wider. Some girls scored higher than in South West State, and there is a clear peak to the right of girls who scored between 60 and 70 – however, in this skewed distribution, we can see that many girls did substantially worse than this. Interestingly, there is another mode to the left, just around 40, indicating that there was a group of girls scoring worse. Overall this points to higher inequality in scores within this region than in South West State.

Hirshabelle state shows this phenomenon to an even greater extent. There is one very high-scoring group of girls scoring above 70, better than most in either of the other two states. However, there is a larger group of girls that scored much lower at between 20 and 30. This indicates that the regional inequality in life skills is higher in this region compared to the other two. The overall pattern and non-normal distributions is reflected in the mean scores and proportion achieving the target. Banadir has the highest mean score of all zones and in this zone 14.3% reach the YLI target. Hirshabelle, on the other hand, has a much lower mean score but has a higher proportion of its students reaching the target at 16.5%. As seen in the distributions, this does not reflect an equally shared high level of leadership skills but is driven by a smaller group of girls who score very highly.

In order to further investigate the multi-modal distributions in scores, the scores by the districts within Hirshabelle and Banadir state were analyzed. Disaggregating by district, in Banadir there was high variance between districts, with some scoring above 60 as seen in the distribution above and some scoring closer to 40. This suggests that the non-normal distribution with two distinct peaks can be accounted for by variance across districts. This is not the case for Hirshabelle, however. Each district is relatively close to each other in mean score with a high range, suggesting that the non-normal structure is caused by another factors that cuts across each district. The pattern is also apparent across all age groups. The means by age are more similar than would be needed to account for the two peaks, and the ranges within each age are very large. The same structure was again observed for grade. The above suggests that there is some factor behind the non-normal distribution, leading to a group of girls with high scores and a group of girls with low scores, that we cannot account for with the available quantitative data.

**TABLE 67: MEAN YLI SCORES, BY GIRL AND HOUSEHOLD CHARACTERISTICS**

Outcome	No	Yes	Difference	p-value
IDP	51.5	47.9	-3.6	0.03**
Any disability	51.0	46.5	-4.5	0.07*
Disability (excluding mental health-related)	51.1	37.3	-13.8	0.000***
Mental health-related disability	50.1	50.0	-0.01	0.963
Linguistic minority	50.5	42.5	-8	0.004***
Maay speaker	51.8	46.6	-5.2	0.006***
Teacher encourages participation	39.3	50.7	11.4	0.000***
Went hungry most days	52.3	48.6	-3.7	0.011**
Went without household water most days	51.9	48.8	-3.1	0.07*

Poor roof in house	51.1	46.3	-4.8	0.009***
--------------------	------	------	------	----------

\*\*\* significant at 99% level, \*\* significant at 95% level, \* significant at 90% level

In the table above we disaggregate YLI scores by various characteristics of the girls, their household, and their learning environment. This allows us to examine the effect of marginalization on leadership skills and self-confidence. Minority status was analysed but surprisingly had no significant effect on scores. However, there are multiple other characteristics that are associated with lower scores which the programme should attempt to address. Scores were statistically significantly lower for IDPs, girls with a disability, those who were Maay speakers and those that were in a linguistic minority (either Maay speakers in a non-Maay area or Mahatiri speakers in a Maay-speaking area). They were much lower for girls in school where the teacher rarely encouraged participation and also seemed to be lower for those from poorer households, measured by those going hungry most days, going without water most days, and those living in a house with a poor roof. Interestingly, those with a mental health disability did not differ in score from those without a mental health disability, while those with a non-mental health disability suffered much lower scores than those without. The sample sizes are small but it is still an interesting finding. It might be the case that non-mental health disabilities, given the smaller number of girls reporting one, were more severe and therefore had a stronger exclusionary impact, leading to poorer self-esteem.

The lower scores among marginalized girls is a negative finding in this baseline analysis. However, as seen in the midline analysis above, the programme has been successful in targeting this exact group of girls. At baseline the same trend was occurred across many of the same sub-groups, where marginalised girls scored poorly, but by midline the scores had equalled and in many cases were now higher for those who started off with factors against them. The programme should ensure it cements its earlier successes and brings about the same result for this new set of girls in the programme.

**TABLE 68: MEAN LSI SCORES, BY GIRL AND HOUSEHOLD CHARACTERISTICS**

Outcome	No	Yes	Difference	p-value
IDP	73.9	73.4	-0.5	0.53
Any disability	74.1	72.3		0.09*
Disability (excluding mental health-related)	74.0	70.3		0.015**
Mental health-related disability	73.8	73.3		0.7
Linguistic minority	73.9	71.0		0.01***
Teacher encourages participation	71.0	73.9		0.09*

\*\*\* significant at 99% level, \*\* significant at 95% level, \* significant at 90% level

We carry out the same analysis for the LSI scores and report the results in the table above. Again minority status had no significant impact on scores, and for LSI being an IDP did not significantly impact the scores either. The overall number of significant variables for LSI was lower than for YLI, which is likely caused by the narrower distributions seen earlier, meaning that the scores simply vary less overall. However, the general trend is still the same – for any results that were significant, it shows that marginalised girls tend to do worse. The same finding occurs for disability as for YLI. Those with any disability scored significantly worse on LSI than those without one, and this result is mostly driven by disabilities not related to mental health. Those in a linguistic minority – either a Maay-speaker in a non-Maay area or a Mahatiri speaker in a Maay-speaking area – scored significantly worse, and those in classrooms where the teacher rarely encourages participation also score worse. Again it is a negative finding that marginalised girls tend

to score lower, but the programme's track record of successfully targeting these girls leave significant potential for a levelling in scores at the next round of evaluation.

## 9.4 School Management and Governance

This section will establish the baseline levels of the indicators related to school management and governance for the New (Cohort 4) NFE girls. The rationale for assessing this aspect of schooling, as related to the positive downstream effects for learning and enrollment, is the same as outlined in the midline section on school management and governance (see Section 7.4) – simply put, that well-managed schools increase enrolment, promote transition, and improve learning outcomes. AGES interventions targeting the quality of school management focus on building the capacity of CECs to promote the enrolment and learning of ultra-marginalised girls and the capacity of MoEs to ensure the quality of teaching, school administration, and inclusivity of schools.

When approaching the analysis of the baseline levels of engagement of the CECs and MoE, we faced challenges when trying to separate the outcomes of the New NFE (Cohort 4) cohort and the midline (Old NFE) cohort, due to the structure of our data and because the CECs and the MoE do not differentiate between the two in their programming. For instance, the analysis of the quality assurance procedures of the MoE presented in the midline section was based solely on qualitative interviews with the District Education Officers (DEOs) that work with the NFE centres where both cohorts were sampled. As one DEO stated: “the work with the new NFE cohort is not different in any way”. Thus, in order to avoid repetition, we do not include the discussion of the MoE engagement in this section but the reader may refer to the midline School Management section for relevant insight on the work of the MoE.

One of the AGES logframe indicators related to school governance measures the progress achieved by the CECs:

- Percentage of community education committees addressing barriers to the enrolment and learning of ultra-marginalised girls

While we do not track the indicator directly, this section analyses data collected from the new cohort of NFE girls, their caregivers, and head teachers to better understand the baseline levels of the CEC efforts to promote enrolment and inclusion. First, we establish the level of barriers to girls' education and their variation across the marginalisation axes by analysing the caregiver perceptions. Then, describe the CECs' activities targeted at supporting marginalised girls using the quantitative data collected from girls and head teachers.

The first key component of understanding the changes along the current indicator that we will consider is related to the *perceived* barriers to enrollment, especially those that can be expected to affect marginalised girls at a disproportionately higher rate. The change in this perception will be tracked over time to determine whether the CECs were successful in creating an inclusive environment and assisting marginalised girls to overcome these barriers.

One of the major barriers that reduces enrolment across the board, and especially for the economically disadvantaged students, is the cost of education. Among the caregivers of girls in the new NFE cohort who have never attended formal school, 52 percent cited the cost of education as a major reason they could not send their daughter to formal school. The NFE program being free of charge alleviates a portion of this financial burden, making schooling more accessible; however, the costs associated with supplies, uniforms, and transport may still prevent some students from enrolling or completing an NFE program. The significance of this barrier will be measured in the next round of assessment by surveying the girls who dropped out before finishing the course. IDP girls appear to be particularly vulnerable to disenfranchisement due to financial reasons as a higher than average share of caregivers, 59 percent, cited



the lack of finance as the reason their daughter never attended formal school, generally consistent with what we know about the lack of economic opportunities accessible to the IDPs. During the FGDs, CEC members reported occasionally fundraising to help students purchase uniforms and supplies, although it is difficult to estimate the reach of this support in an average project school and extent to which this initiative targeted marginalised girls specifically.

Other barriers that may disproportionately afflict marginalised girls and are targeted by the CEC awareness campaigns are related to bias and discrimination. The indicators we present in the table below capture different aspects of mistreatment that may arise from discriminatory attitudes, including harassment by teachers or other students at school, refusal of enrollment, and the perception of the lack of safety. Regression analysis of these indicators generally shows that minority girls were somewhat more likely to be subject to these forms of mistreatment, however, with quite a lot of variation in the data and small sample sizes, most of the differences were not statistically significant.

**TABLE 69: SHARE OF CAREGIVERS CITING REASONS THEY DID NOT ENROL THEIR DAUGHTER IN FORMAL SCHOOL**

Barrier	Overall	Disability	IDP	Minority group
n	183	29	87	31
Unsafe for girl to be at school	3.8%	6.9%	6.9%	0%
Teachers mistreat girl at school	6.6%	6.9%	6.9%	9.7%
Refused entry at school	0.5%	0%	0%	0%
Other students bully girl at school	2.7%	3.5%	2.3%	3.2%

Caregivers of girls with disabilities whose daughters have never attended formal school were more likely to cite safety concerns, bullying by other students, and mistreatment from teachers as deterring factors to their daughters' enrollment than other parents. It is difficult to adjudicate whether this perception is derived from an actual discrimination pattern or the caregivers' preconceived notions about their child's ability to integrate with other children and learn effectively despite their disability. Remember that the qualitative findings from interviews with the CEC members summarised in the midline chapter identified the attitudes of parents who did not think that their daughters were suited for education as a major deterring factor for enrollment of girls with disability.

With all the girls in the new cohort currently enrolled in the program, we are not able to assess how many girls with disabilities were unable to join the cohort as a result of their family's biases or any other specific barrier. To measure the impact of caregiver attitudes on the disabled girls' enrollment, we plan to identify girls who did not complete the NFE course during the next assessment and analyse the reasons they dropped out. Still, based on the findings from the midline cohort, accessibility barriers – such as the lack of school transport and assistive equipment – are expected to be more damaging to the enrolment and attendance outcomes of the new NFE cohort than attitudinal barriers, due to the limited availability of the financial resources required to address them.

Other patterns shown in the table above were not statistically significant. Due to very small sample sizes of the interviewed caregivers whose daughters did not attend formal school because of one of the reasons above, our data does not allow us to make meaningful inferences about our results. Note that the insights from the CEC member FGDs that identified discrimination-based barriers faced by IDP girls, girls with

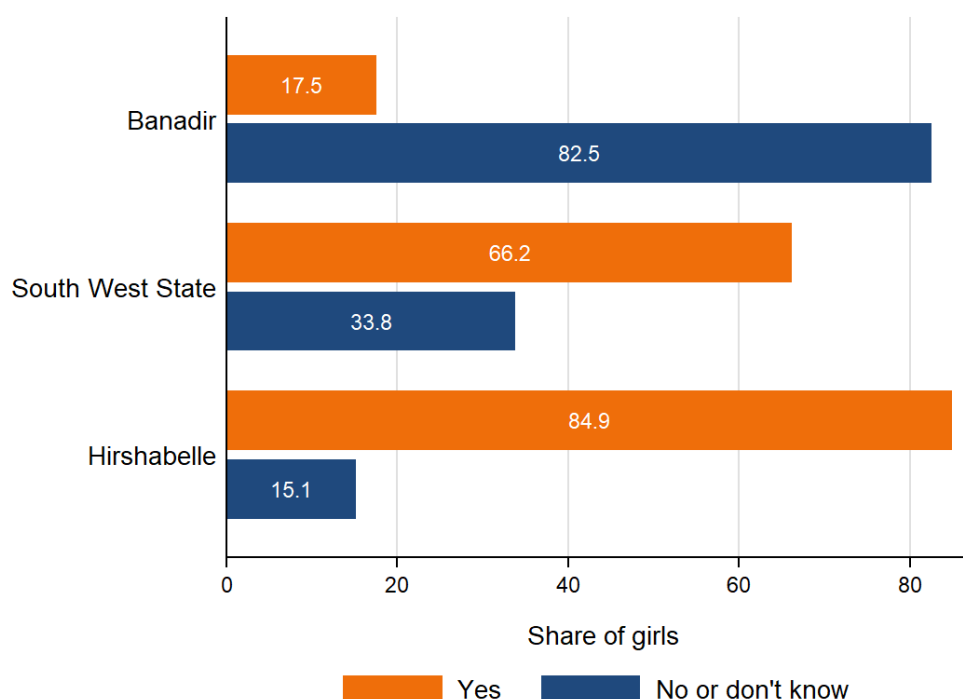
disabilities, and girls from economically disadvantaged families described in the midline chapter are likely to apply to the new NFE cohort as well.

Note further that the FGDs with CEC members identified a number of additional infrastructural or material challenges that required investments beyond their fundraising capabilities and thus were seen as largely unresolved. Several of these issues, such as the inability to accommodate the pastoral migration patterns or provide transportation or assistive equipment to girls with disabilities, were described as disproportionately affecting marginalised girls. More data needs to be collected to determine the applicability of these insights to the new NFE cohort.

To ensure that these barriers do not exclude marginalised girls from pursuing education, the CECs are tasked with supporting school administration, parents, and students through school monitoring visits, awareness campaigns, attendance tracking, and more. To establish the baseline levels of the CEC activity in the centres where the new NFE girls were sampled, we asked the girls **whether they thought that the CECs at their centres did anything to support girls' education**. The plurality, 46 percent, responded positively, 37 percent did not know, and 17 percent responded negatively. The results varied greatly across districts. The Hirshabelle CECs were perceived to be the most supportive with 85 percent of respondents thinking that the CECs support girls' education. In Banadir, 71 percent of girls selected "don't know" as their response, which can be interpreted to signify low levels of engagement of the CECs with girls.

While not all of the CECs tasks aimed at improving opportunities for girls would require interaction with students, students not being aware of the CEC activities aimed at supporting them is likely an indication of the absence of such activities and thus should be considered as the partial evidence for the CECs failing to address the barriers to education faced by girls. Thus, CECs in Banadir schools, primarily sampled in precarious peripheral districts, may require the most support from the project, although South West state CECs also need additional support.

FIGURE 22: DOES THE CEC DO ANYTHING TO SUPPORT GIRLS' EDUCATION, BY STATE



When disaggregating girls' responses by demographic group, we found that girls with disabilities and girls who spoke Af-Maay at home were statistically significantly more likely to say that the CECs support girls' education. This result can be interpreted to suggest that the CECs were better able to support these groups, partially fulfilling their goal of addressing barriers to education of marginalised girls specifically.

On the other hand, girls from minority groups were statistically significantly less likely to agree that CECs were supportive of girls' education suggesting that their particular needs were underserved. This is in line with the qualitative findings that showed that CEC members were more reluctant to acknowledge and address barriers specific to girls from minority groups as compared to other marginalised groups. The IDPs and girls from pastoral families did not differ significantly in their responses. This result still means that the CECs have not been able to specifically emphasise the needs of these groups as should have been their goal.

The most common activity the CECs engaged in, as reported by 86 percent of girls, was the enrollment of out-of-school girls, followed by supporting dropouts to return to school (59 percent) and checking on student absences (34 percent). As girls should be expected to be most aware of the CEC activities that engage them directly, it is not surprising that the most common activities are student-facing. Girls are also well positioned to provide reliable evidence on whether the CECs address harassment and abuse cases and provide hygiene products, both of which were found to be uncommon, with 12 and 9 percent of girls reporting them, respectively. The reported rates of the activities that may not have been visible to the girls directly or fully, such as monitoring teacher absenteeism (1 percent), hiring female teachers (2 percent), monitoring teaching quality (10 percent), and raising awareness in the community (25 percent) should be assumed to be somewhat undervalued, although likely still relatively low. As the girls were likely to be unaware of the full extent of the CEC engagement in the activities above, community advocacy emerged as another commonly implemented CEC activity.

Given the negative bias in the girls' perception of the levels of some of the CEC activities, one way to cross reference this data is by considering the head teacher responses to the school survey. Our sample included 11 centres where the new NFE cohort was sampled. Out of these 11 centres, 7 centres also had the old NFE girls and 3 centres had other types of midline cohort girls and were included in the midline analysis. The remaining 1 NFE centre only had the new NFE cohort and was not included in the midline analysis.

All 11 schools where new NFE girls were sampled had a CEC but only 64 percent reported having at least one monitoring visit in the past year. With annual visits being a liberal measure of CEC engagement as active committees could be expected to visit schools at least weekly, this finding indicates a relatively low baseline level of CEC involvement.

The most common activities reported by the head teachers were promoting enrolment of out-of-school children (82 percent) and following up on dropout cases (73 percent), consistent with the results of the girls' survey. Overall, all other activities were reported at higher rates than in the girls' survey, which is likely linked to the head teachers' positionality as the focal point for coordinating with the CECs which makes them more aware of the committees' work. Child protection activities and hiring teachers were each reported only by 3 out of 11 schools (27 percent), emerging as the key areas for improvement in the next project stage.

Although the analysis of the baseline caregiver perceptions of barriers to enrollment did not identify major discrepancies for marginalised girls, with a potential exception of IDPs, the more negative perceptions of the CECs' support by girls from minority clans, IDPs, and girls from pastoral families suggested the needs of these groups were not successfully targeted by the CECs. The baseline levels of the CEC engagement across most activities were relatively low, with child protection and teacher hiring emerging as priority areas. More data is needed to understand the CECs' ability to target non-discrimination based barriers

that may affect marginalised girls but, based on the FGDs with CEC members and girls' reports of the lack of hygiene kit provision, it can be assumed to be quite limited.

## 9.5 Increased Self-Efficacy

The sections below look at variables related to positive youth development and access to protection services. The reader should note that the findings presented below apply specifically for girls from the subsample of cohort girls newly enrolled in non-formal education schools.

### Positive Youth Development

At the midline, data was collected for indicators extracted from the Chinese Positive Youth Development Scale (CPYDS), a globally recognized multidimensional measure of positive youth development (i.e., competence, character, confidence, and connection). The seven indicators discussed below are taken from the CPYDS subscale on self-efficacy.

#### Control of own life

The first indicator measures the degree to which girls perceive to have control of their own lives. Most girls in the new NFE cohort disagree with the statement that they have little control of things that happen in their own lives (56.4 percent). Girls' perceived control over their own lives is weakest in Middle Shabelle region, with only 14.4 percent of girls disagreeing with the statement. In contrast, girls in Lower Shabelle have a very high perceived control of their own lives, with 88.3 percent of girls disagreeing with the statement.

Girls whose household head or caregiver has not received formal education<sup>138</sup> are more likely to perceive having control over their own lives (55.7 percent and 56.8 percent, respectively). These figures decrease substantially when we drop the subgroup of people that received Quranic education and examine only household heads and caregivers who have no education of any kind is dropped (29.8 percent and 30.2 percent, respectively), which suggests that Quranic education may drive this positive effect. Female heads of household for the new NFE cohort appear to have a strong negative effect, as only 41.2 percent of girls with female household heads perceive to have control on the things that happen in their lives.

Girls from minority groups are less likely to perceive having control of their own lives (51.2 percent). Perceptions of girls from locally dominant groups do not deviate much from the average. In addition, land ownership may also factor in. In households that own land, 65.8 percent of girls feel they have control over what happens in their lives.

Food and water are intrinsic to people's sustenance and well-being, and therefore strongly influence girls' perceived self-efficacy. Increasingly chronic shortages of food and water will increase girls' perception of having little control over what happens in their lives. Parental support is another key factor for well-being, as girls with no living parents are far less likely to perceive having control over their lives (33.33 percent) than girls with one (49.1 percent) or both parents (51.5 percent). However, girls who do not live with their parents are far *more* likely to perceive having control over their own lives (66.7 percent). In addition, girls who are currently married (28.6 percent) are less likely to perceive having control over their lives than girls who are not currently married (51.7 percent). Girls with children have even less perceived control over their lives (20.0 percent).

<sup>138</sup> Respondents who did not receive any education at all, received some form of informal education or attended madrasa were all included in the subgroup, as responded who did not receive formal education.

Girls with disabilities are less likely to perceive having control over their lives (44.39 percent). This is the case for mental disabilities<sup>139</sup> (38.0 percent) and to a lesser extent for physical disabilities<sup>140</sup> (45.5 percent).

### Solutions to problems

The second indicator measures the extent to which girls believe they can solve their own problems. A minority of girls in the survey disagreed with the statement that they did not have any solutions for some of the problems they are facing (46.7 percent). In Middle Shabelle region, girls are far less likely than other regions to feel that they can solve their own problems (10.1 percent), whilst in Lower Shabelle, girls are far more likely to feel that they can solve their own problems (86.1 percent).

Quranic education for the household head and the caregiver may be related to girls' more positive perceptions regarding their problem-solving abilities. Girls whose household head or caregiver has received informal education (53.6 percent and 53.4 percent, respectively) are more likely to believe they can solve their own problems than girls whose household head has received some type of formal education (30.2 percent and 31.8 percent, respectively). Female household heads appear to have a negative effect, with girls from households with female heads being far less likely to believe they can solve their own problems (42.7 percent) than girls from households with male heads (65.9 percent).

For the new NFE cohort, non-minority status increases girls' perceptions of being able to solve their problems. Girls from households that own land (67.1 percent) are also more likely to believe that they can solve their own problems than girls from households that do not own land (43.3 percent).

A lack of access to basic needs can severely affect girls' perceptions of their problem-solving abilities. The longer that girls go without food, the less likely they are to believe they are able to solve their problems. This is largely also true for girls who go to sleep without clean water. The lack of parental support has a notable effect, as girls with no parents being less likely to believe they can solve their own problems (33.3 percent) than girls with only one parent (45.3 percent). Girls who are currently married or have children are also less likely to believe they can solve their own problems (28.6 percent and 20.0 percent, respectively).

Finally, girls with any type of disability are less likely to feel that they are able to solve their problems (28.9 percent) than girls without any form of disability (51.3 percent). Physical disabilities appear to have a reverse effect compared to mental disabilities for this indicator, with 59.1 percent of girls with a physical disability believing they can solve their problems but only 25.4 percent of girls with a mental disability having the same belief.

### Ability to change

The third indicator refers to the ability to change things in life. A large majority of girls strongly disagrees with the statement that they are unable to do much to change things in their own life (58.8 percent). This sentiment was felt most strongly in Lower Shabelle, where 91.2 percent of girls believe they have the ability to change things in their lives. Girls from Middle Shabelle reported very low perceived ability to change their own life (28.1 percent).

<sup>139</sup> For mental disabilities, the variable for anxiety/depression is used as a proxy. Other variables for mental disabilities were excluded from the discussion because there were few observations (n<25).

<sup>140</sup> For physical disabilities, the variable for mobility disabilities is used as a proxy. The other variables are not discussed in the text because there were very few observations (n<5).

Similarly to the previous indicator, girls whose household head or caregiver has received informal education (57.4 percent and 58.5 percent, respectively) are more likely to believe they can solve their own problems than girls whose household head has received some type of formal education (46.5 percent and 40.9 percent, respectively). Female household heads may have some influence on girls' opinions, with only 50.0 percent of girls whose head of household is female believing they can change their lives. Girls from land-owning households also appear to have more confidence in themselves, with 68.4 percent of girls from these households believing they are able to change. In contrast, girls from minorities groups are less likely to believe in their ability to change.

The availability of food and water is again closely interlinked with girls' perceived self-efficacy. Girls who go without these basic needs for extended periods are increasingly less likely to believe they can change things in their lives. Marriage and children may also have a negative effect on girls' perceived ability to change their lives, as only 28.6 percent of girls who are currently married and 20.0 percent of girls who are mothers believe they can change. Having less parental support appears to weaken girls' resolve. Girls with one parent are more likely to have the feeling that they can change things in their life (56.6 percent) than girls with no parents (33.3 percent).

Girls with mental disabilities are less likely to believe they can change things in their lives (45.8 percent) than girls with physical disabilities (50.0 percent). Overall, girls with any disability are somewhat less likely to have the belief that they can change things in their lives (50.8 percent) than girls with no disability (60.9 percent).

### Helplessness

The fourth indicator measures whether girls feel helpless when they face life difficulties. A minority of girls agree with this statement (46.9 percent). Only a small share of girls in Lower Shabelle (19.0 percent) are inclined to feel helpless in difficult situations, whereas a majority of girls in Middle Shabelle are reported to feel helplessness (84.2 percent).

Household heads and caregivers' lack of any kind of education is linked to stronger feelings of helplessness from girls in those households (53.2 percent and 53.5 percent, respectively). A female household head and land ownership are factors that are likely to reduce girls' feeling of helplessness (41.7 percent and 72.2 percent, respectively). Minority status has no noticeable effect on self-efficacy.

Feelings of helplessness are strongest in girls who most frequently sleep without food or clean water. Girls with only one living parent are more likely to feel helpless (41.5 percent) than girls without parents (66.7 percent). Girls who have been married<sup>141</sup> (50.0 percent) or who are currently married (57.1 percent) are less likely to feel helpless than girls who have children (60.0 percent). Also, girls who do not live with their parents are less likely to feel helpless (30.0 percent) than girls who have one or no living parents (41.5 percent and 66.7 percent, respectively). Girls with disabilities are also far more likely to feel helpless (70.6 percent), although this effect is likely driven by girls with mental disabilities (75.4 percent), who represent a far larger group than girls with physical disabilities that feel helpless (50.0 percent).

### Fate not in hands

The fifth CPYDS indicator for self-efficacy asks if girls feel that their lives are determined by others and fate. A majority of girls in the sample agree with this statement (58.0 percent). There is large regional variation, with a large majority of girls from Middle Shabelle feeling that their lives are not in their hands (69.1 percent), but only a minority of girls from Bay (47.0 percent) sharing this sentiment.

<sup>141</sup> Have ever been married but may be divorced now.



A lack of formal education for household heads and caregivers appears to have little effect on girls' responses for this indicator. Household heads' gender is a pertinent factor, with 69.3 percent of girls in households with female heads feeling that their fate is determined by others, compared to only 56.8 percent of girls in male-led households. Girls belonging to minorities are somewhat more likely to perceive lives are determined by others (65.6 percent), and girls who belong to a household that does not own land are very likely to believe that their fate is determined by others (71.1 percent).

A shortage of food and clean water are related to girls' feeling of their lives being out of their hands. The more frequently they go without these basic needs, the more likely they will report feeling that their lives are determined by others and fate. A lack of parental support also increases this perception, with 56.7 percent of girls not living with their parents, 60.4 percent of girls with just one parent, and 100.0 percent of girls with no parents feeling that their lives are not in their hands. Marriage and having children have a similar effect, with girls who are currently married or have children being less likely to feel that their lives are determined by others or fate (71.43 percent and 80.0 percent, respectively).

### **Determine own life**

The sixth indicator represents girls' perceptions on how things happening in their lives are mostly determined by themselves. Most girls are in favour of this statement (55.6 percent), particularly girls in the region of Middle Shabelle (72.7 percent). Girls from Banadir are somewhat less confident about their ability to determine the things happening in their lives (45.9 percent).

Girls from households where the household head or caregiver has no education of any kind are very likely to believe that they determine the things that happen in their lives (61.7 percent and 62.8 percent, respectively). On the other hand, girls from households where the household head is female are less positive about their ability to determine things in their own lives (47.4 percent). With respect to minority status, no clear trend can be identified between power dynamics and girls' ability to determine things in their own lives. Land ownership also appears to have no effect on the indicator.

There seems to be no clear relationship between a lack of access to food or clean water and girls' perceptions on their ability to determine their own lives. The lack of parental support does appear to be linked, with girls not living with their parents feeling less likely to be able to determine their own lives (40.0 percent) than girls living with one parent (41.5 percent) or girls with no parents (100.0 percent). Marriage and children increase girls' belief that they are able to determine the things happening in their lives. 71.4 percent of girls who are currently married, 50.0 percent of girls who were ever married, and 80 percent of girls who have children believe they can determine things in their own lives. Finally, disabilities strengthen girls' beliefs that they can determine things for themselves (68.5 percent), and this applies for both mental and physical disabilities.

### **Task completion**

The seventh and final indicator for self-efficacy asks girls whether they believe they can finish almost everything that they are determined to do. A vast majority of girls agree with this statement (74.9 percent), with girls in Middle Shabelle being particularly confident about their ability to complete tasks (80.6 percent) and girls in Lower Shabelle being the least confident (65.0 percent).

For the new NFE cohort, girls with household heads and caregivers that have received no education of any kind are most likely to believe they can complete their tasks (78.7 percent and 83.7 percent, respectively). The presence of female household heads have a negligible effect. Interestingly, girls whose household owns land are less confident in their ability to complete tasks (64.6 percent) than non-land owners (70.2 percent).

There is no clearly identifiable trend between the lack of food or water and girls' perceived ability to complete tasks. Girls with one parent are less likely to believe they can complete their tasks (64.2 percent) than girls who do not live with their parents (73.3 percent) or girls with no parents (100.0 percent).



Children likely reduce girls' ability to complete tasks, with only 60.0 percent of girls with children deeming themselves able to complete their tasks.

## Synthesis

When disaggregating the results geographically, certain trends emerge. In general, Middle Shabelle region appears to have girls with substantially lower levels of positive youth development compared to the rest of the sample, whereas girls from regions such as Lower Shabelle frequently score well on the CPYDS indicators. Middle Shabelle is a region included at midline with girls from a new NFE cohort, so these girls have not been exposed to two years of programming like many of the others in the sample, and it is therefore not surprising that they should score so low on the indicators.

Household characteristics such as a formal education for the household head or caregiver can have a strong impact on girls' perceptions of what they can do. Household heads and caregivers receiving a non-formal Quranic education can also play a significant and positive role in elevating girls' perceived self-efficacy, as it teaches basic tenets of fairness and justice, and the Quran also includes statutes that provide a degree of protection to women. Land ownership is associated to wealth and power, so families and households who are land owners are more likely to be able to provide a platform for empowerment to their girls. Female heads of household have a mixed effect for this cohort, and this may be due to the lack of programming and community sensitization campaigns in new programme areas.

Individual characteristics such as a chronic shortage of food, water, and parental support have a clear influence on many of the indicators. Marriage can sometimes create allow girls to become more empowered in certain aspects of their lives, whilst more inhibited in others. A chronic shortage of food and water affects wellbeing and the ability for people to function properly, and needs to be addressed before cognitive youth development can continue. Furthermore, the lack of emotional and financial support that orphaned girls experience can severely affect their growth and development as a person, although for some it can become a source of resilience. Finally, disabilities have diverse effects on girls' perceptions of self-efficacy. Girls with physical disabilities are clearly less able to do certain physical activities, but their disability appears to strengthen their mental resolve. On the other hand, mental disabilities do not affect girls' practical abilities but severely weaken their cognitive development.

## Access to protection services

The following section focuses on girls' access to protection services, both within their schools and their communities. The section will review new NFE cohort girls' ability to access these services, as well as the types of services that are available.

### School

An overwhelming majority of girls reported that schools them with some a channel to report some form of harassment, abuse or exploitation at their school (93.0 percent). The share of schools that provide protection services is highest in Banadir (95.2 percent) and lowest in Middle Shabelle (89.21 percent), but overall schools perform well in this regard. There are a few subgroups of girls who report relatively less access to these services, namely girls from land-owning households (89.6 percent), girls who go without food on just one day (88.0 percent) or most days (90.2 percent), and girls with physical disabilities (86.4 percent).

Girls will typically report an incident to their teacher and/or head teacher, and on rare occasions to the Community Education Committee (CEC) or the Girls' Empowerment Forum (GEF). A few girls also mentioned that they can report incidents to the police as one of the options available. Girls from Banadir, Lower Shabelle, and Middle Shabelle most frequently resort to both teacher and head teacher to report incidents, girls in Bay prefer to go to either the teacher or the head teacher.

## Community

A large majority of girls also reported having access to someone to report incidents of harassment, abuse or exploitation in their communities, although this number was not as high as in schools (81.8 percent). Lower Shabelle (93.9 percent) and Bay (90.9 percent) regions performed well in the provision of protection services on the community level, but Middle Shabelle was well below average in this regard (67.6 percent).

Girls in dominant groups (79.8 percent), girls with only one parent (78.9 percent), girls who never go to sleep without food (73.4 percent) report having relatively little access to protection services compared to their counterparts. On the other hand, girls who are currently married (100.0 percent), ever married (100.0 percent), mothers of children (100.0 percent), girls belonging to households with heads who have received informal education (90.7 percent), or girls belonging to households with caregivers who have received formal education (93.2 percent) have almost all access to protection services.

Teachers and head teachers remain girls' first choice to report incidents of harassment, abuse, and exploitation at the community level, particularly in Lower Shabelle. The police are the second most frequent choice to report incidents, (a popular option in Bay), followed by community leaders in third. Community Education Committees, Girls' Empowerment Forums, and religious leaders are sporadically mentioned by girls as alternative protection services at the community level. However, the most common response that girls gave was that they didn't know who to report incidents to, a response most notable in Banadir, Bay, and Middle Shabelle.

## Synthesis

Overall, communities score well with girls on protection services, and schools generally fare better. This is likely because teachers and head teachers are the most preferred people to report incidents of harassment, abuse, and exploitation. The CEC and GEF are severely underutilized platforms of reporting incidents, as people from traditional societies tend to rely on protection services from local authorities such as police or community leaders. Particular attention should be paid on to protection services for subgroups that are traditionally or structurally marginalized in society, such as minorities and internally displaced persons, as well as girls with disabilities. Furthermore, more should be done about communicating information about the protection services available to girls, since they appear to be chronically uninformed.

## 9.6 Strengthened Economic Circumstances

The sections below look at variables related to economic empowerment. The reader should note that the findings presented below apply specifically for girls from the subsample of 916 cohort girls newly enrolled in non-formal education schools.

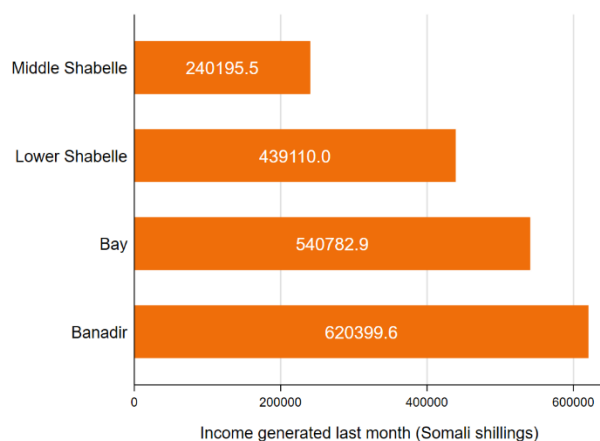
### Income generation

At the midline, quantitative data was collected for numerous indicators of income generation, including average income and type of income generation activities. In the following section, the findings for these indicators will be presented, and disaggregated for relevant variables such as region, minority, and IDP status. The data will be complemented with qualitative findings from the interviews and focus group discussions that were conducted with mothers, Community Education Committee (CEC) members, and girls.

## Average income

The indicator for income measured the amount of money (in Somali shillings) that girls made in the last month.<sup>142</sup> On average, girls made about 509,741 SOS a month, with 51.4 percent of all girls reporting no income. The table below shows some regional variation in income, with girls in Banadir being the highest earners (620,400 SOS), followed by girls from Bay (540,783 SOS) and girls from Lower Shabelle (439,110 SOS). Girls from Middle Shabelle were the lowest earners of all (240,196 SOS).

FIGURE 23: INCOME BY REGION



Girls from minority groups earned a lower average income (259,242 SOS). Girls who speak Maay at home (536,807 SOS) and girls from linguistic minority households (554,278 SOS) earned on average slightly more than girls who did not speak the language (497,090 SOS and 507,835 SOS, respectively). However, girls whose family is internally displaced earn less than girls whose family has not been displaced (447,527 SOS vs. 549,869 SOS). Finally, we find that a number of other factors, such as a lack of food and water, ever being married or orphaned, not living with their parents, having children and having disabilities, can negatively affect girls' income.

## Type of income generation activities

Girls were also asked about their occupation and business, as well as their contribution to agricultural work, the family business or work outside the home. Most girls reported not having an occupation (52.4 percent) or doing domestic chores inside the home (17.3 percent). A substantial proportion of girls were students (15.0 percent). A small share reported being an unskilled salesperson or service worker (8.8 percent), pastoralists (1.1 percent), or mentioned other occupations (1.3 percent). Other pre-categorised occupations were rarely mentioned (< 1 percent). There were some differences in how the main categories were distributed across regions. In Banadir and Lower Shabelle, girls were more likely to report domestic chores as their occupation (22.5 and 29.8 percent, respectively), in contrast to Bay (8.0 percent) and Middle Shabelle (4.6 percent). In Bay, almost a third of girls were students (29.7 percent), whereas in Middle Shabelle, there were no students and a large majority did not have an occupation (76.2 percent). Notably, 4.6 percent of girls in Middle Shabelle were pastoralists and 4.3 percent of girls in Lower Shabelle worked in sustenance farming or fishing.

Minority girls had the lowest relative share of students (6.8 percent). Maay-speaking girls were more likely to have an occupation or be students in comparison to girls who do not speak Maay. Girls from linguistic minority households are also more likely to have an occupation than girls from linguistic majority households, but they are less likely to be students and more likely to do domestic chores. Girls

<sup>142</sup> The highest value was 3,000,000 SOS, so only 'don't know' or 'refused to respond' values were excluded from the analysis.

from families that are internally displaced are more likely to have an occupation, do domestic chores inside their household, and work as unskilled salespeople or service workers than those whose families have not been displaced.

Almost none of the girls surveyed have their own business (8.2 percent). In Middle Shabelle, girls were more likely to set up a business (15.6 percent), particularly in comparison with Lower Shabelle (5.3 percent) and Banadir (6.1 percent). Girls from Maay-speaking and linguistic minority households were less likely to have a business than their counterparts. IDP did not have a substantial effect. From the 52 girls who reported having a business, a large share of them buys and sells food (34.6 percent), some have a kiosk or sell items at home (19.2 percent), others sell prepared food such as setting up a tea stand or a small restaurant (15.4 percent), and a large share have a variety of other businesses (30.8 percent).

The qualitative data concurs with this result, as only one parent mentioned that their daughter is working<sup>143</sup>. Parents reported a number of barriers that inhibit them from starting a business, including lacking financial support<sup>144</sup> or parental guidance<sup>145</sup>, that girls are lacking knowledge<sup>146</sup>, or that they are already working for their parents' business<sup>147</sup>. Some parents consider that girls who are in school are unable to start a business<sup>148</sup>. However, parents frequently recognise that the schooling that girls receive provides them with useful knowledge for starting a business and supporting their families' income. Girls who learn mathematics and reading from NFE courses will know how to save money, and set up and keep track of their business<sup>149</sup>. Parents propose that girls start businesses such as shops, beauty salons, selling food, clothing fairs, grocery stores, and sewing clothes<sup>150</sup>, and find employment in accounting, teaching or banking<sup>151</sup>. The money that girls make from their businesses could be used to change their families' lives and pay for their siblings' school fees<sup>152</sup>. One parent in Banadir provided an example in which the idealized scenario occurred: "Yes, there are young girls who have started businesses in this area and are giving money to their parents, there is also an inspection when they open shops to look at what has

<sup>143</sup> FGD with Mothers, Gedo, Int. 208.

<sup>144</sup> FGD with Mothers, Banadir, Int. 205; FGD with Mothers, Gedo, Int. 208; FGD with Mothers, Banadir, Int. 209; FGD with Mothers, Banadir, Int. 210.

<sup>145</sup> FGD with Mothers, Bay, Int. 204; FGD with Mothers, Gedo, Int. 207.

<sup>146</sup> FGD with Mothers, Gedo, Int. 208.

<sup>147</sup> FGD with Mothers, Bay, Int. 203.

<sup>148</sup> FGD with CEC Members, LowerJuba, Int. 105; FGD with Mothers, Bay, Int. 204.

<sup>149</sup> Vignette Exercise with Girls, Banadir, Int. 603; Vignette Exercise with Girls, Banadir, Int. 612.

<sup>150</sup> Vignette Exercise with Girls, Banadir, Int. 603; Vignette Exercise with Girls, Banadir, Int. 606.

<sup>151</sup> Vignette Exercise with Girls, Banadir, Int. 603; Vignette Exercise with Girls, Banadir, Int. 612; Vignette Exercise with Girls, Banadir, Int. 613.

<sup>152</sup> Vignette Exercise with Girls, Bay, Int. 601; Vignette Exercise with Girls, MiddleShabelle, Int. 609.

increased and decreased their business, so if they get supported and put in the effort, they can also achieve their goal”.<sup>153</sup>

Most of the girls are not asked to help with the household’s agricultural work (82.8 percent) or to support the family business or work outside their homes (71.2 percent). In Middle Shabelle and Bay, the share of girls supporting the household with both agricultural and non-agricultural work was higher whilst in Banadir, the share was very low.

## Synthesis

Reported (household)<sup>154</sup> income for girls in the NFE Cohort 4 is quite high, particularly compared to Cohort 1 NFE girls. The largest differences are observed in Banadir and Bay, where girls earn 2 to 3 times as much as their Cohort 1 NFE counterparts. These differences hold even when we include the outliers in the Cohort 1 NFE sample. The surveys for both cohorts were taken in different schools, and often in different districts, so they may reflect the socioeconomic differences between the geographical locations targeted. As with the findings on self-efficacy, Middle Shabelle has low levels of girl economic empowerment, possibly because girls from this region have only recently been enrolled in GEF courses.

Girls who belong to linguistic minorities, Maay-speaking, or internally displaced households often earn more and are more likely to have an occupation than girls from majorities or those not displaced. This result mirrors the findings for the previous cohort, and reinforces the premise that perspectives on girl education and employment may be less entrenched and conservative in households which are not subjected to the dominant norms in society, even if they are marginalized by others. The relationship between minority status and income or occupation does vary notably to the relationship for Cohort 1 NFE girls. The differences are likely the result of data collection occurring in different clusters, in which respondents represent different compositions of subgroups. For example, girls from linguistic minorities represent 52.1 percent of girls from minorities in the Cohort 1 NFE, but only 20.8 percent of girls from minorities in the NFE Cohort 4.

With respect to owning a business, we can observe the same emerging trend in the data, where girls from Maay-speaking, linguistic minority, and IDP households generally fare better than their counterparts. Particularly for the Cohort 4 NFE subsample, where these groups are highly correlated and represent similar sets of girls, girls from these groups appear to have different norms and values in comparison to girls from ‘host’ or majority groups. There are less expectations by their families on them to stay and work within their households or start their own businesses, and there seems to be more opportunities for them to study and find an occupation.

## 9.7 Enhanced social support for female youth

This section discusses the opportunities that ultra-marginalised girls in the new NFE cohort can gain access to with support from the AGES project. As these girls have only just joined the program, we are looking to establish their baseline level of engagement with various development opportunities that will be tracked in the future evaluations. These opportunities may include involvement with youth groups, engagement in community and government forums, and receiving humanitarian aid.

<sup>153</sup> Vignette Exercise with Girls, Banadir, Int. 606.

<sup>154</sup> The question is likely to be interpreted as the amount of money that the girls’ households made, since most girls report having no business or occupation.

Our analysis will complement the indicator AGES uses to track this outcome, which is the percentage of youth connected to other development opportunities via AGES support. While we do not evaluate this indicator directly, the quantitative data we collected captures the baseline share of girls engaged in each of the four opportunities that will be tracked in future assessments with consideration of factors that may drive the identified changes, including the AGES intervention.

We begin by describing the types of opportunities analysed in this section, highlighting how they can contribute to the girls' life outcomes. Then, we report the average and disaggregated outcomes for each of the four indicators, explaining important patterns in our findings.

First, we consider the share of girls who reported being part of a youth group or network. These types of opportunities can benefit the girls by strengthening their social connections, increasing their self-esteem and leadership skills, and – depending on the nature of the group – develop other useful skills. The downstream effects of youth network membership can include improvement in transitional outcomes as better integration in the community, as well as enhanced confidence and skills, can enhance the girls' access to income generating opportunities and motivation and ability to study.

Next, we asked the girls if they were able to participate in any discussions to improve service delivery in their communities, which include community meetings and consultations with local leaders or NGOs implementing projects locally, among others. Engagement in such forums could benefit girls by allowing them to voice their priorities and advocate for changes in service delivery that would directly address their needs. More broadly, the ability to influence community decisions can contribute to girls' confidence and leadership skills, pushing them to set ambitious goals in their education and professional life.

Further, we considered girls' involvement in discussions and activities related to local governance. This could be an avenue to address various barriers young women in the region face in different aspects of life, including in education, as well as to inform policy that represents the needs of girls and their communities. Involvement in youth politics may also be empowering for girls, boosting their confidence and building transferable skills that could be applied in academic and professional contexts.

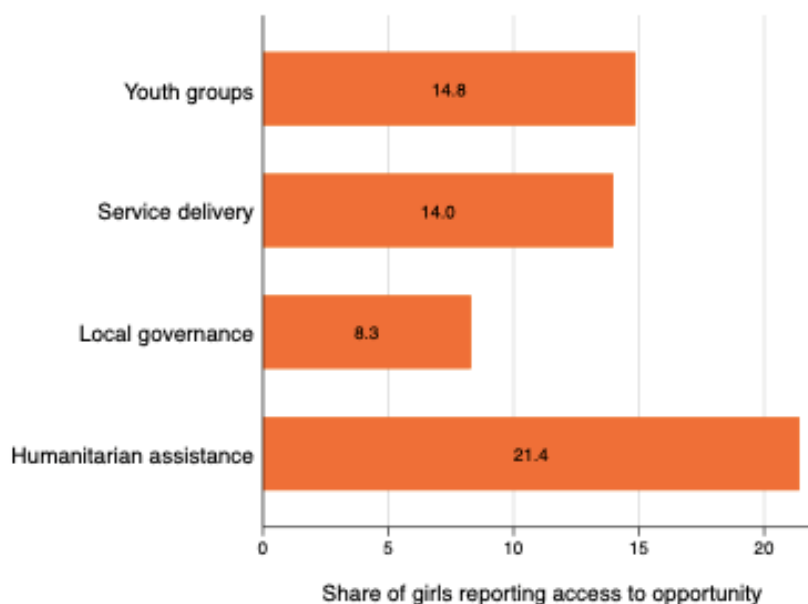
Lastly, we asked the girls whether they have received any form of humanitarian assistance during the past year. This can be a vital source of support for marginalised families that can make a difference for girls' ability to continue their education. Whether the humanitarian assistance targeted school-related costs in particular or helped families fulfil other basic needs – thereby freeing up funds that could be used for schooling – it is likely to have an impact on girls' learning and transition outcomes.

The average outcomes for the four indicators are illustrated in the figure below. At 21 percent, receiving humanitarian assistance was the most common of the opportunities studied. The opportunities associated with this indicator were delivered through external actors and thus were limited by donor constraints around budgets and delivery. While COVID-19 has disrupted some of the avenues of aid delivery – as donor countries began to prioritise domestic recovery or faced logistic barriers created by public health restrictions – the post-COVID restoration of humanitarian assistance would be key to ensuring that a greater share of girls can receive help and gain access to associated opportunities in the future.

The other three indicators were associated with a different set of constraints that were internal to the community, including the availability of the relevant forums, girls' awareness and motivation to participate, and a range of bureaucratic and attitude-based barriers to participation, including familial and community support for young women's involvement in the public domain. As demonstrated by the lower rate of access, the latter set of barriers appears to be more restrictive in the absence of intervention. This is especially the case for girls' participation in activities related to local governance – only 8 percent of respondents reported to have access to. In the context of limited community engagement in governance and high concentration of political power among traditional leaders and a narrow political elite, it is not surprising to see that young women had few opportunities to contribute to discussions around local

governance, especially in light of young women's broad exclusion from power. It is likely that affecting these three indicators would require an integrated approach targeting girls, families, and local institutions to make these public fora and spaces more accessible to young women.

**FIGURE 24: SHARE OF GIRLS REPORTING ACCESS TO DEVELOPMENT OPPORTUNITIES**



There was substantial covariation between the first three outcomes, meaning that many of the same girls engaged in two or more of the activities (youth groups, discussions about service delivery, and local governance). This suggests that the same factors may contribute to access to these opportunities, resulting in benefits that are concentrated among a relatively small minority of girls. To identify the profiles of these girls, we considered factors that may influence the girls' access to these opportunities, including household economic characteristics, caregiver attitudes, girls' leadership skills, and demographic factors. The results are summarised in the table below.

**TABLE 70: PARTICIPATION IN DEVELOPMENT OPPORTUNITIES, AMONG SUBGROUPS**

Subgroup	Youth group	Service delivery	Local governance	Humanitarian assistance
<b>Overall</b>	14.9%	14%	8.3%	21.4%
<b>Demographic</b>				
<b>Has disability</b>	13.4%	13.4%	10.7%	18.7%
<b>Minority group</b>	13.6%	14.4%	11.2%	20%
<b>IDP</b>	11.2%	10.1%	6.1%	27.2%
<b>Economic</b>				
<b>Has poor roof</b>	14%	12.3%	7%	19%
<b>Gone a day without eating</b>	8.8%	8.8%	3.7%	16.2%



<b>Reduced food expenditure in the last month</b>	9.5%	8.8%	3.4%	16.3%
<b>Caregiver attitudes</b>				
<b>Acceptable to keep girl out of school to help at home</b>	13.7%	17.6%	9.8%	13.7%
<b>Acceptable to keep girl at home if school is too costly</b>	7.7%	4.2%	2.4%	12.5%
<b>Strongly believes that girls are as likely to use their education as boys</b>	8.7%	6.1%	3%	16%
<b>Girl's leadership skills</b>				
<b>Feels confident answering questions in class</b>	25.1%	24.7%	14.7%	26.3%
<b>Can organise friends to do an activity</b>	26.7%	24.4%	14.2%	33.5%
<b>Can ask adult if doesn't understand something</b>	17.8%	16.2%	9.6%	21.4%

First, we disaggregated the results by demographic groups to analyse how various axes of marginalisation influence participation in the development opportunities. Girls from IDP households were the least likely group to be involved in youth networks and political fora, with significantly lower participation levels than their counterparts for each of the three indicators. With many of the IDP girls in the sampled settlements being relatively recent arrivals, their access to social opportunities is likely to be restricted by their low levels of community integration and limited social networks that are necessary for engagement in public life. However, discrimination or the perceived threat thereof may further contribute to lower levels of involvement by IDP girls, as they may not feel comfortable or welcome to join these initiatives.

Although there were large variations in the data, girls with disabilities and from minority groups had higher average rates of participation in activities or consultations related to local governance. With inclusive policy making being promoted by donor agencies, implementing NGOs, and increasingly prioritised by Somali government officials as well, these groups' status as a special stakeholder may be used to target them for consultative activities. In contrast, girls with disabilities and minority girls were no more likely to receive humanitarian assistance, unlike IDP girls, who were significantly more likely to benefit from it. Our hypothesis is that this is linked to the programmatic priorities, mentioned above, of implementing NGOs and donors, and the often visible and intense needs of IDP households who have been displaced by recent environmental shocks.

Geographic differences in reported access to and participation in social opportunities were stark. Overall, girls in Hirshabelle were much more involved in public life, whereas Banadir lagged behind in participation in all three types of opportunities. These differences are likely linked to a combination of factors related to the states' political processes, demographic composition and cohesion, and recent

conflict dynamics. Lower engagement of Banadir girls in youth networks and political forums may be explained by the relatively high tensions around the elections in the capital region.. With an influx of internal migrants, Banadir is the most diverse region in the country, which tends to be associated with weaker community integration and may discourage formation of youth groups even outside the election period. The Banadir sample primarily included girls from peripheral districts that were characterised by a high share of IDPs and substantial insecurity, which can make political activism dangerous or inaccessible, compounding the barriers to participation for the girls.

Hirshabelle, the youngest state in Somalia, was formed in 2017 and the government formation process is still ongoing. Programs implemented by the federal government and NGOs at the state and district levels target institutional formation, public service delivery, and youth engagement, which may explain the greater youth engagement shown in our results. For instance, the sample in this state included significantly fewer IDP, suggesting that, on average, Hirshabelle respondents were better integrated and had stronger community ties, which was helpful for accessing development opportunities.

The districts sampled in South West State, and especially its capital Baidoa, are known as centres of NGO activity, which may explain the medium to high outcomes in the first two indicators. It is likely that, since the state-building process has been completed, encouraging citizen engagement in governance discussions was outside the key priorities of development actors in the region, explaining the lower score on the third indicator. The girls interviewed in South West State received significantly more humanitarian assistance than their counterparts which, while related to the high NGO presence, is likely connected to the large share of IDPs in the sample.

The table above presented three indicators that are indicative of household wealth, with the goal of understanding whether participation is a function of household socioeconomic status. Overall, girls from more economically disadvantaged households tended to have less access to youth groups and political forums. The financial position of girls' households could influence their ability to access development opportunities in one or more of the following ways: poor girls may need to spend more time on income-generating activities or engage in household chores to free up time for adult family members to work leaving less time for other activities; not being able to afford new clothes, hygiene kits, or other items can undermine girls' confidence to participate in social activities; and higher incomes may be associated with the ability to gain access to opportunities through personal networks. We also found that girls from poorer households were less likely to receive humanitarian assistance. This may be a result of donors targeting assistance towards specific demographic groups, such as the IDPs, rather than according to household income, partially because the former method is operationally easier to implement and requires less complicated eligibility procedures for the beneficiaries.

We also considered how three indicators that captured the caregivers' attitudes around gender roles influenced girls' ability to participate in youth networks and political forums as they could reflect the caregivers' support for their daughters' engagement in public life. The results show that more traditional caregiver attitudes tended to be associated with lower participation in the development opportunities for girls. This can be interpreted to suggest that girls' access to these opportunities requires family support in the form of permission or encouragement to participate.

Finally, we identified a positive relationship between certain indicators associated with girls' leadership skills and an increased likelihood of accessing development opportunities. As girls with better leadership skills are likely to be more effective in their pursuit of development opportunities, this finding tentatively supports the argument that girls' internal motivation and social skills may be a major factor impacting participation in youth groups and forums at baseline. In other words, the beneficiaries of these opportunities are not fully externally determined and girls have some ability to increase their chances of accessing youth groups and political forums by exerting effort and showing initiative. Note, however,

that there were other leadership-related indicators that were not positively correlated – or were only weakly correlated, – with access to development opportunities.

Thus, the baseline findings demonstrate that, in the absence of programmatic interventions, characteristics that indicate economic marginalisation, such as low household income and IDP status, were associated with limited access to youth networks and political forums. Further, we found that girls who had the family support and social skills to actively and effectively pursue these opportunities were more likely to benefit from them. To increase access to development opportunities, AGES would have to design an intervention that empowers girls to seek out participation, targets caregiver attitudes to increase their support, and reduces various economic and discrimination-based barriers that disenfranchise marginalised groups specifically, with tailoring to local context. Finding effective ways for humanitarian assistance to reach the poorest households also emerged as a priority area. The geographic analysis showed that programming targeting girls' political engagement should focus on Banadir and South West State, while interventions increasing access to humanitarian assistance should prioritise Banadir and Hirshabelle.

## 10. Ancillary Analysis

### 10.1 Menstrual health and hygiene

One goal of the AGES programme is to promote better sexual and reproductive health and hygiene among girls. While this outcome falls outside of the standard outcomes in the broader Girls' Education Challenge programme discussed above, menstruation can have a large impact on girls' attendance and thus learning outcomes. Menstruation is a common challenge for girls' attendance; girls often stay at home during their period due to shame, a lack of menstrual hygiene products, a lack of hygienic places at school to manage their periods, for religious reasons, or due to period-related pain.

To better understand these dynamics, the evaluation included a survey module on menstrual health and hygiene directed at girls who had begun menstruating. Girls being interviewed by a female enumerator<sup>155</sup> were asked whether they had begun to menstruate; for those who had, they were asked further questions about menstrual health practices, such as use and reuse of pads or rags; their information sources about menstruation; the impact of menstruation on school attendance; and their experience with menstruation-related illness. Four-hundred and fifteen girls at baseline and 890 girls at midline stated that they were menstruating and were asked subsequent questions. Of these girls, 296 girls within the baseline group and 390 girls within the midline group were a part of the panel sample—girls who were surveyed at both baseline and midline.<sup>156</sup> Below, we analyse results for all surveyed girls and, where changes between baseline and midline are investigated, for the panel sample of girls surveyed during both rounds.

<sup>155</sup> In some cases, the majority of the survey was completed by a male enumerator, but the menstrual health module was completed by a female enumerator or team leader who was available. Male enumerators did not administer this module.

<sup>156</sup> There are more girls who answered menstrual health-related questions within the midline panel sample than the baseline panel sample for two possible reasons: First, as girls age, more begin menstruating and thus qualify to be administered the module. Second, more female enumerators or team leaders may have been available to administer this module at midline than baseline.

At baseline, we found that menstruation was stated to present a barrier to attendance by girls less frequently than predicted: Only 9.6 percent of all 415 surveyed girls felt they could not attend school during their period. At midline, we similarly find that of all 809 surveyed girls, only 7.1 percent felt they could not attend school during their period. While any rate above zero percent represents girls who face monthly barriers to school attendance and learning, these findings suggest that menstruation may be a less significant barrier to attendance than qualitative data around attendance rates might suggest.

These findings also suggest that menstruation may have become a less salient barrier to attendance since baseline. To investigate this, we analyze the change over time in the panel sample of girls. The below table presents these results, and shows that there was no significant change in girls reporting menstruation as a barrier to attendance between baseline and midline. To check for robustness, we further control for age<sup>157</sup>; the results remain insignificant, suggesting little programmatic impact on girls' access to school during menstruation from baseline to midline.

**TABLE 71: CHANGE IN GIRLS REPORTING THAT MENSTRUATION PREVENTS ATTENDANCE**

Outcome	BL	ML	Difference	p-value
<b>n</b>	296	390		
<b>Menstruation prevents attendance</b>	11.2%	7.2%	-4.0	0.14

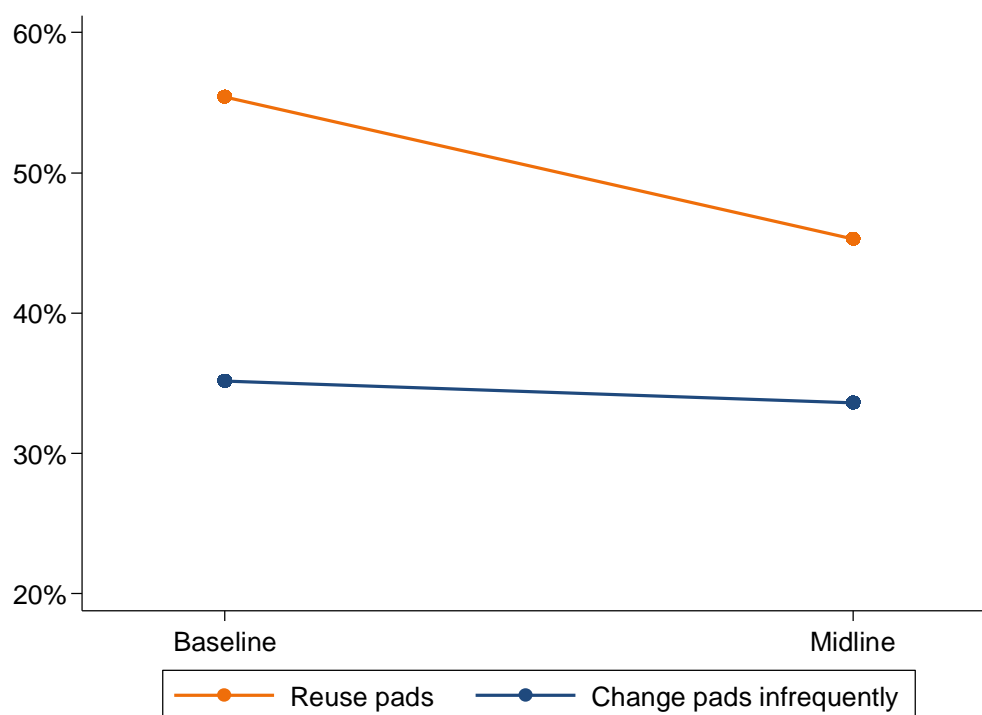
Girls who reported that menstruation prevents their attendance at school were then asked the reasons why this occurs. Given that few girls reported an impact on attendance, sample size is low: 40 girls responded to this question at baseline and 57 at midline. As such, we do not compare results across rounds. At midline, the majority of girls—over 70 percent—reported that pain or discomfort prevented their attendance at school; this is followed by girls reporting a lack of sanitary pads (19 percent). While low in absolute numbers—only 11 total girls at midline stated that a lack of sanitary pads prevented their attendance—this suggests that there may still be some need to continue strengthening girls' access to menstrual products.

This is reinforced by the finding that, at both baseline and midline, many girls reported reusing regular menstrual pads or changing pads less than three times a day, among other unhygienic practices. At midline, for example, over 41 percent of all surveyed girls stated that they had reused a regular menstrual pad twice or more, and nearly 32 percent of girls reported changing their pad less than three times a day. It is possible that the prevalence of these unhygienic practices is due to lack of regular access to sanitary pads; it may also be due to lack of access to information, discussed more below.

However, we note that although these unhygienic practices are still common, use of these practices has decreased from baseline to midline among the panel sample of girls. The figure below shows a significant decline for reuse of regular menstrual pads ( $p < 0.05$ ), and a small but insignificant decline in girls reporting to change their pad less than three times a day. This may reflect an improvement in knowledge about these unhygienic practices, an improvement in access to menstrual hygiene products, or some combination of both.

<sup>157</sup> Older girls may be more comfortable and experience managing their periods, and may thus find menstruation to be less of a barrier to attendance. However, older girls may also have heavier periods or more regular periods which present a larger barrier to school attendance.

FIGURE 25: CHANGE IN REUSE OF PADS AND FREQUENCY OF PAD REPLACEMENT



Outside of these hygienic practices, girls were also asked about their access to water to clean during menstruation, as well as to clean reusable menstrual products (including rags and reusable pads). The table below shows that among the panel sample, the majority of girls reported that they had access to water to wash during menstruation all or most of the time; there was no significant change at midline. It is worth further noting that less than 1 percent of all surveyed girls at both baseline and midline said they had no access to water to clean. While these are positive results, findings regarding cleaning rags are less positive. Among girls who reported using rags in the panel sample, at midline, almost one-quarter reported using only water—not soap and water—to clean the rags. This is a significant decrease of nearly 13 percentage points from baseline, and suggests that more work is needed to either improve girls’ knowledge about the need to clean with soap or to improve girls’ access to soap.

TABLE 72: CHANGE IN MENSTRUAL PRODUCT CLEANING PRACTICES

Outcome	BL	ML	Difference	p-value
<i>Do you always have access to water to wash regularly during menstruation?</i>				
<b>n</b>	296	381 <sup>158</sup>		
<b>All or most of the time</b>	90.5%	92.7%	2.2	0.57
<b>Sometimes or no</b>	9.5%	7.4%	-2.2	0.57
<i>If you are using rags, do you clean them with soap or just with water?</i>				
<b>n</b>	201	214 <sup>159</sup>		
<b>Soap and water</b>	89.1%	76.2%	-12.9	0.03*

<sup>158</sup> Eight girls refused to answer and one stated that she did not know.

<sup>159</sup> Excludes girls who stated they used other cleaning processes, discarded rags, did not use rags, or refused to answer.

<b>Water only</b>	11.0%	23.8%	12.9	0.03*
<b><i>Do you sometimes have to re-use a rag (or reusable sanitary pad) before it is fully dry after being washed?</i></b>				
<b>n</b>	295	380 <sup>160</sup>		
<b>Yes</b>	26.4%	20.0%	-6.4	0.13

Lastly, the table also shows that at midline, one-fifth of the panel sample of girls reported reusing a rag before it was fully dry after cleaning it. This represents a substantive but insignificant decrease from baseline, but is still fairly high in absolute numbers. This practice is problematic because reusing rags (or pads) before they dry can cause infections—as can washing with water alone. Indeed, at midline, 82 percent of girls reported that they had experienced an illness related to menstruation. While the wording of this question may have prompted girls to report experiencing an illness if they face cramps or other period-related pains, rather than solely infections, this finding is still concerning, particularly given that the percentage has increased substantively (though not significantly) from baseline, when only 74 percent of girls reported experiencing a menstruation-related illness.

As discussed above, these issues related to menstrual hygiene may be due to two primary factors: first, a lack of access to hygienic materials (such as soap, sufficient numbers of pads, or clean water), and second, a lack of information/knowledge. To investigate the second factor, girls were asked from whom they receive information regarding menstruation. The majority of all girls surveyed at midline—82 percent—stated that they received information from their mother, with another 6 percent stating they received information from a female relative. While many mothers may be able to provide accurate and effective information about menstruation to their daughters, given the relatively low education levels of most caregivers in our sample, the reliance on mothers and female relatives as a source of information may mean that some girls receive inaccurate or outdated information about menstruation.

## 10.2 COVID-19 impacts in AGES communities/schools

The largest single barrier to girls learning since the program's inception has been the COVID-19 pandemic. In mid-March 2020, Somalia had its first documented case of COVID-19, and public health authorities quickly closed down international and domestic air travel, closed schools, and put in place limits on internal movement and group gatherings. Taking students out of the classroom placed the burden of instruction on the students and their family, reduced accountability for students and teachers alike, and created an economic downturn that reduced family's ability to afford tuition. To assess the ways that the pandemic affected CARE's beneficiaries, a question series was incorporated into the household survey asking students about their ability to respond to pandemic-induced changes and a variety of ways that their lives had changed over the past two years.

When asked whether they had been able to study at home during school shutdowns, 56% girls answered affirmatively, and 66% of those also noted that they had some in-person assistance doing so – whether by family, neighbours, or other students. Family members were by far the most common source, with 53% of girls who could study at home reporting that they had received some assistance from them. However, the percentage of girls who had family members *willing* to help educate them is almost certainly much higher. In reality the 53% value indicates that approximately half of girls had a family member who was willing to educate them, capable of doing so, and also had the spare time for it. Many live in households where their potential tutors are each working all day or who may not have the basic numeric or literacy

<sup>160</sup> One girl at baseline and one girl at midline stated she did not know, and nine girls at midline refused to answer.

skills to help them even if they were available and motivated to do so. That so many still did get help despite these barriers is a reassuring consideration during this period where buy-in from inside the home is more valuable than ever in getting girls back to school.

However, access to study spaces necessitates having spare rooms in the house, assistance with studying requires the household to have an adult at home during the day, and homework help can only come from those who have completed some amount of educational attainment themselves. This means that groups that have been historically marginalized, whether through limited access to land, economic opportunities, and schooling, will have been more constrained in their ability to make up for the lack of formal education during 2020 and 2021. The data bears this out – only 52% of minority girls were able to study at home, while 64% of those from locally-dominant groups were able to. This relationship continues to hold when looking into whether those who were able to study at home received assistance doing so, where historically marginalized minorities were the least likely to have had help (61%), fully six percentage-points less than locally dominant groups.

**TABLE 73: COVID-RELATED OUTCOMES ACROSS REGIONS**

Outcomes	Banadir	Gedo	Lower Juba	Lower Shabelle	Bay
<b>Was girl able to study at home</b>	57%	84%	70%	55%	40%
<b>Did girl have assistance studying</b>	84%	60%	26%	88%	72%
<b>Did girl receive teacher support</b>	49%	64%	62%	79%	43%
<b>Did the teacher check on absentees</b>	50%	83%	53%	57%	36%
<b>Child workload (% who reported an increase)</b>	32%	50%	5%	34%	22%
<b>Child no covid effect</b>	49%	71%	83%	56%	64%
<b>Caretaker no covid effect</b>	48%	45%	82%	58%	54%
<b>Child covid econ impact</b>	42%	27%	12%	19%	30%
<b>Caretaker covid econ impact</b>	50%	55%	12%	25%	37%

There is meaningful regional variation in these outcomes as well. As can be seen in Table 1, when it came to girls' ability to study at home Gedo and Lower Juba meaningfully outperformed while Bay languished. The other survey outcomes point toward the value of teachers engaging with students outside of the classroom during school closures; fully 43% of the girls from Bay stated that a lack of support from their teacher was a primary challenge for their at-home studying, compared to only 15% for the other regions. Looking at these two outcomes more broadly, within regions there appears to be a negative correlation between whether the girl had been able to study at home and whether they had in-person assistance. Since the latter question is only asked to those who stated they had been able to study, it is plausible that those without assistance were the first to forfeit studying from home and consequently be filtered out of our



survey, leaving only those with support. This evidence points to the value of at-home buy in if girls are to continue their schooling, whether school is being held in classrooms or not.

These hurdles which reduce families' ability to educate their own children are widely recognized, and teachers and educational systems exist precisely to fill that role in their stead. The shuttering of schools in 2020 and 2021 removed not only the physical building for learning but, in many cases, the teachers as well. Since most education providers in Somalia are privately funded through the collection of tuition from students, closing them halted teachers' salaries in these communities as well. This had the dual-effect of removing one of their incentives to help students learn from home while also displacing them into other forms of labor, consequently reducing the time they could spend contributing to helping students learn.

Despite these barriers, 56% of the girls who had been able to study at home reported receiving some sort of teacher support to do so; 44% of the time this support came in the form of teacher-provided practice materials. Teacher support may have been an equalizer for those who received less assistance at home. Teachers, who are aware of their student's home conditions, may have intentionally assisted those they knew were more at risk of learning losses or permanent dropout. However, it may have also simply been a function of demand – those without the resources at home to complete their schoolwork did so through other means. Either scenario provides reason for optimism – in the former teachers recognized and actively worked to address marginalization in their communities, and in the latter the girls displayed an earnest interest in learning.

Teachers have shown to be persistent not only in making at-home-school more plausible for their students but also in facilitating and encouraging their return to the classroom. Just over half of the surveyed students reported their teacher checking in on them when they were absent from school, an accountability mechanism for attendance that may not exist in households where the returns to girls' education are undervalued or where parents leave for work before their daughter leaves for school. Inter-regional analysis suggests a correlation between this mechanism and girls having been able to study at home – Gedo received the highest marks on each measure with Bay receiving the lowest on each. However, the directionality or causality of this relationship is not entirely clear. The narrative may be one of returns on investment, where girls who stayed on top of their studies during lockdown were viewed as more capable of catching up or completing the year and therefore it was considered more worthwhile to check in on them. It could also be an institutional correlation where school systems that hire, train, and promote teachers who/to care about attendance to the point that they will beyond their classroom duties may have been more apt at supporting students to study during the school closures.

High chore burdens and declining household financial conditions are also widely asserted as drivers of school absenteeism and, eventually, dropout. Household chores because they compete with studying for scarce free time, and declining financial conditions because shrinking disposable incomes means tough spending decisions will need to be made. When asked whether their workload has increased, decreased, or stayed the same since two years ago, 27% of girls stated that it had increased and 34% that it had decreased, suggesting a median that is approximately the same as before the pandemic's onset. That, on average, the COVID-19 epidemic did not erect this additional barrier is assuring to a degree. But disaggregating the data to individual regions provides some evidence for caution, as fully 50% of girls in Gedo, the region with the highest rate of girls having been able to study at home, reported rising housework. The lessened learning losses due to at-home studying may be fragile. Having children take on greater burden around the home may have been a response to financial instability, as Gedo was also the region where the greatest share of caretakers reported a COVID-induced economic impact (55%), compared to 36% in the rest of the sample.



# 11. Recommendations

## Geographic Anomalies and Follow-Up

Many of the trends tracked since baseline have been similar across locations and even across cohorts. Even where improvements have been modest – such as in the development of girls’ leadership skills or YLI scores – this suggests relatively uniform application of the programme’s interventions. Where this has not been the case, additional attention should be given to assessing the divergence in programme outcomes. The most prominent case in this round is the performance of NFE centres and the C1 NFE cohort in Jubaland, the latter of which experienced a sharp decline in numeracy since baseline and no gain in literacy, compared to substantial gains in every other cohort and state. This is coupled with other problematic signals, such as a massive increase in absenteeism among Jubaland’s teachers that is not matched in other locations. While the evaluation cannot definitively pinpoint the reason for divergent geographic trends, it can and should be used to guide further investigation by programme staff into the context in, for instance, Jubaland’s NFE centres, to determine whether the poor outcomes observed are a function of exogenous factors – such as the worsening drought – or internal factors within the centres.

## Classroom Disciplinary Practices

In line with the more general recommendation above, we recommend focused targeting of particular areas and schools for additional training and sensitisation to reduce the use of corporal punishment and other negative disciplinary practices. The use of corporal punishment has declined significantly in formal schools and ABE centres tracked since the baseline; however, corporal punishment remains relatively common in Jubaland’s formal schools and ABE centres and – even more so – in the new NFE centres (C4) brought into the programme in Hirshabelle. This suggests the possibility of tailored interventions, with particular attention to basic behaviour change on disciplinary practices focused on Jubaland and Hirshabelle centres.

Tailoring of this kind would also benefit the schools and centres where corporal punishment has already declined dramatically. In these schools, more nuanced forms of still-negative practices persist, particularly the use of harsh language when interacting with students and punishing incorrect answers in class, which can be a major deterrent to active student participation and classroom inclusivity. By shifting attention in these schools and centres away from a focus on stamping out corporal punishment, additional training time and monitoring efforts could be spent on reducing these more nuanced – but, nonetheless, problematic – practices.

## Poor Menstrual Hygiene with Cost-Effective Avenues for Improvement

Section 10 of the report documented the relatively poor menstrual hygiene practices among girls in the sample, particularly in the form of insufficient washing of reusable pads (i.e. washing reusable pads without using soap) and infrequent changing of pads, both of which can lead to infection. Similarly, a significant minority of girls indicate that they occasionally have to reuse a pad before it is fully dry from being washed. These issues have fairly straightforward and cost-effective remedies, however, when we consider how many girls are already comfortable washing and reusing pads. By providing a small number of supplemental reusable pads to girls – as opposed to entire sanitary kits, with their greater costs – girls would be able to wait longer for pads to dry and change their pads more frequently. Combined with the provision of soap, a relatively modest investment could significantly improve girls’ menstrual health.

## Changing Perceptions of the Value of Education

Caregivers in much of the sample tracked since baseline report marginally more positive views of the *value* of girls’ education. Specifically, when asked whether girls’ education is a worthwhile investment of scarce resources, positive responses increased slightly in Banadir and South West State. In contrast, the notion that girls’ education is a worthwhile investment fell sharply in Jubaland, which may reflect both shifting

attitudes and a sense that the poor labour market or other issues reduce the utility of girls' education. To the extent possible, attitudinal change programming in Jubaland should continue work with religious leaders and community leaders while also focusing attention on the economic and non-economic benefits of girls' education. The latter could be accomplished by hosting female business owners, NGO employees, or government staff, to provide girls and their parents with a visible example of the benefits of their education.

### **Coordinate with, and Demonstrate Targeting to, Other Programmes**

AGES has done an admirable job of targeting ultra-marginalised girls in the communities where it works. For instance, in the overall midline sample, just 34.8 percent of beneficiaries hail from the locally dominant group, and a full 42.3 percent are from minority groups. Compared to many other programme evaluations we have performed in Somalia, AGES outperforms nearly all of them in terms of targeting – a majority of beneficiaries in most other programmes are drawn from dominant clans and overwhelmingly exclude the most marginalised. This argument is supported by the fact that marginalised girls in the AGES sample are less likely to have benefitted from *other* humanitarian interventions, despite being drawn from objectively poor households.

We would recommend that AGES work with other NGOs in Banadir, Hirshabelle, Jubaland, and SWS to illustrate how their own targeting could be improved and to highlight the overrepresentation of dominant groups and – more generally – the “non-marginalised” – that is often present in other programmes. By opening discussions of this kind and demonstrating how more effective targeting can be completed cost-effectively, AGES could provide important spillover benefits, expanding its impact among marginalised communities by helping other NGOs broaden their own beneficiary base.

### **Promote a Role for Girls in Local Governance**

While girls' participation in the public sphere, generally, is limited, this evaluation found that girls were especially unlikely to participate in discussions and forums related to local governance. Even among adult women in our sample – those NFE girls 18 years and above – just 11.0 percent have participated in public discussions regarding local governance. Increasing this participation will require addressing four types of barriers described by girls in qualitative data: a lack of motivation on their part, a lack of caregiver support, economic barriers to participation, and limited space in a political system dominated by elders, traditional leaders, and older men. If girls are to have a more active role in governance, it is not sufficient to promote the importance of their participation to them; rather, spaces need to be explicitly opened for their participation, with incentives provided to NGOs and government bodies to more fully incorporate female voices in meetings and public fora.

### **Promote Inclusion by Recognising Bias Openly**

Given the programme's focus on inclusivity and promoting enrolment among marginalised girls, the issues of bias, discrimination, and special needs for ensuring inclusion were a frequent theme in surveys and qualitative interviews. While teachers, CEC members, and community members generally recognise the teasing, stigma, and outright discrimination faced by girls with disabilities, there is very little recognition or openness about discrimination against groups based on their social identities. Indeed, the majority of CEC members and teachers – to provide one example – deny that ethnic minorities face any mistreatment in schools at all.

Denial of the problem, however, forecloses the possibility of improving the situation. School staff typically argue that the school treats all students equally, which prevents them from seeing that mistreatment at the hands of other students and even subtle biases among teachers can have important, deleterious impacts on marginalised girls, especially when those actions are mirrored by widely-known but seldom-acknowledged bias in society writ large. The programme should focus attitudinal change efforts on helping school staff to recognise the bias that exists against minority students. While other marginalised groups, such as girls with disabilities, face massive challenges, their challenges are often recognised by school staff

and community members and are difficult to remedy due to resource limitations. By focusing any interventions related to attitudinal change on bias against minorities, the programme would not be detracting from efforts to help GWDs; rather, they would be focusing one aspect of programming on a group (minority students) who could benefit from attitudinal change and acknowledgment of discrimination, while reserving any available financial resources for reducing practical barriers to enrolment and education for GWDs.

### **Develop Specific Protection Mechanisms within Schools and Centres**

The girls interviewed in this round overwhelmingly indicate that they could report cases of abuse, harassment, or exploitation occurring at school to a member of the teaching staff. They are similarly confident that they could report such issues occurring outside of school to teachers, community leaders, or their parents, among others. However, a wealth of research on sexual and gender-based violence and access to justice in Somalia indicates that women are widely hesitant to report instances of sexual assault and that they lack confidence a report would be handled either confidentially or competently. This suggests that AGES beneficiaries may be too sanguine about their ability to report or that the survey question did not adequately capture the potential for hesitation or concern around reporting.

In general, the idea that girls could report instances of abuse to teaching staff is problematic, not least because many schools lack any female staff members and staff members themselves may be perpetrators of abuse. In our view, schools should have an established focal point for child protection and abuse allegations, even if that focal point is a female member of the CEC, rather than a teacher. An established and advertised channel for reporting is essential to increase confidence when girls are actually faced with this dilemma.

# Annexes

## Annex 1 - Evaluation Methodology

This annex describes the research design used to conduct the midline evaluation of the AGES project. We will review the evaluation questions pursued in the assessment, explain the tool design, detail the fieldwork procedures, describe the analytical approach, and outline the limitations and challenges. Broadly, this section explains how the midline assessment fits into the overall pre-post evaluation design of the multi-year longitudinal evaluation consisting of several rounds of assessment. Additionally, this round of assessment also included a baseline evaluation for a new cohort of NFE girls who enrolled in the program this year.

Broadly, the AGES midline assessment intends to estimate the change in key project indicators since the beginning of the project. To do this, we began by establishing the midline values of the relevant indicators provided by CARE. We then compared the current levels of the indicators to the values established at baseline and identified the key factors driving the changes related to the progress along the intermediate outcomes and contextual changes. The analysis was disaggregated by major beneficiary characteristics to determine the relative impact of the project on different groups and inform project recommendations targeting better inclusion of the marginalised beneficiaries.

### Evaluation Design

The AGES evaluation uses a longitudinal, pre-post evaluation design to draw conclusions regarding project impact over time. The evaluation is mixed-methods, and – within broad qualitative and quantitative data divisions – utilises a range of data collection tools. Qualitative data was collected using FGDs with multiple respondent groups, participatory exercises targeting girls, and KIIs with religious leaders. Quantitative data was collected using surveys with head teachers, a full cohort of randomly sampled girls and their households, and direct observation of classrooms, in addition to learning assessments with the aforementioned cohort girls.

There are no control or comparison groups in the study. This is due to the practical and security related concerns of using control groups in the volatile and conflict-affected areas of Southern Somalia. This is best articulated in CARE’s MEL framework for the project:

*“A pre-post design was selected due to the extremely fragile, conflict-affected context where AGES will be implemented. In this context, the random assignment of a control group entails the risk of unequal distribution of aid to rival clans/ sub-clans, potentially triggering violent reaction against beneficiaries and education facilities. Even the use of a non-random comparison group (as in QED) may result in backlash against beneficiaries and/or project staff/partners in this volatile context.”*

As such, the design does not use the standard difference-in-differences approach, which would use a set of control schools for inferring impact in intervention schools. Instead, the same cohort of girls and schools that are part of the intervention are being tracked over time.

For the midline evaluation, we followed up on the same girls we randomly selected and assessed at baseline to deliver the same learning assessment and a series of survey questions. If the girl was under 18, her caregiver and head of household were also interviewed as part of the survey. These girls will continue to be tracked until the endline to measure the changes in the learning outcomes over time. In addition to the learning outcomes, data to evaluate changes in community attitudes, girls’ self-esteem and leadership skills, economic and demographic indicators was also collected.

In the absence of a comparison group, longitudinal comparisons are subject to bias from secular trends – for instance, the same girls might be expected to improve their learning performance as they age, due to maturation or growth effects, even in the absence of intervention. To control for maturation effects, gains in learning from baseline to midline or endline are compared to benchmarked differences in learning scores between grade levels at baseline. To illustrate, consider girls age 11 at baseline: the difference in their performance between midline and baseline is the project's naïve impact on learning scores. The difference between 11- and 12-year old girls at the baseline serves as the benchmark – the difference in performance for these girls becomes the metric against which the naïve change in scores is assessed. Therefore, if the gains from one round to the next (as a girl goes from 11 to 12 years of age) are larger than the differences between 11- and 12-year old girls observed at baseline, the benchmarked impact of the project is positive.

It is important to highlight the shortcomings of a benchmarked pre-post design of this kind. The most important is that the design cannot account for changes in learning outcomes under the counterfactual (i.e. what would have occurred in the absence of the project's interventions) from year to year. Benchmarked comparisons are not equivalent to a true difference-in-differences design, because the pseudo-counterfactual or comparison group is constructed exclusively from girls who took the learning assessment at baseline (e.g., 11- and 12-year old girls at baseline, the difference between these two groups becomes the expected change in learning, absent intervention, for girls who go from 11 to 12 years old from baseline to midline). Therefore, the benchmark does not account for broad societal- or community-level changes that might influence learning outcomes, such as shifts in enrolment patterns, exogenous shocks – such as conflict, drought, or flooding – that impact attendance rates, and myriad other factors.

Further, the changes in attendance and teaching quality are assessed using classroom observation and headcount tools that utilise direct observation of classrooms. We revisited the same schools as in baseline, with the exception of a few replacements due to changes in the security situation that will be described below. However, as some of the teachers may have changed in the last two years, the approach to analysing the teaching quality and attendance is cross-sectional rather than panel.

Beyond studying the changes in project outcomes over time, a core goal of the midline evaluation is to assess the validity of the project's Theory of Change. This is done, for instance, by investigating the differential impact of various project interventions depending on girls' exposure to each of them, as well as studying the relationship between the changes in the intermediate outcomes (attendance, teaching quality) and the learning outcomes.

## Quantitative Methodology

The quantitative data was collected across 38 formal schools, 34 ABE centres, 36 old NFE centres, and 46 new NFE centres. The original proposed sample for the household surveys was comprised of 2,340 girls, including 1,420 recontact girls for the midline assessment and 920 new cohort girls for the baseline assessment. The new cohort of girls included only new NFE students, while the recontact sample contained 421 FE girls, 484 ABE girls, and 515 old NFE girls.

As several schools had to be dropped due to security concerns, the maximum possible sample was reduced. The formal schools that were dropped were replaced and the sample of students was randomly drawn, the replacements are described in more detail below. As a result, the updated target sample for household surveys included 421 FE girls, 468 ABE girls, 506 old NFE girls (also excluding one girl that was listed twice on the recontact list), and 920 new NFE girls.



TABLE 74: SAMPLE TARGETS BEFORE AND AFTER SCHOOL-LEVEL ATTRITION

Cohort type	Original target sample	Updated target sample	Notes
FE	421	421	
ABE	484	468	Target sample reduced by 16 as a result of dropping two schools in insecure districts
Old NFE	515	506	Sample reduced by 9 as a result of dropping one school in an insecure district and one duplicate girl
New NFE	920	920	

The achieved sample is shown in the table below, including its value as a share of the updated target sample. Note that household surveys included a series of questions for the girls and their caregivers, as well as a learning assessment for the girls.

TABLE 75: ACHIEVED SAMPLE, RELATIVE TO TARGET

Tools and target groups	Achieved sample	Completion rate versus updated target
Household surveys with girls in the formal education	417	99.0%
Household surveys with ABE girls	363	77.6%
Household surveys with old NFE girls	381	75.3%
Household surveys with new NFE girls	917	99.7%
Classroom observation	219	76.6%
Headcount	463	
School survey	37	97.4%

For the midline assessment, we recontacted the same girls who were interviewed at baseline. To facilitate the recontacting process, extensive tracking materials have been developed, including tracking sheets for each individual girl and school cohort lists for each school. The recontact procedure involved multiple steps that needed to be completed before the enumerator could mark the girl as “not found”. These steps included:

- Asking the head teacher and other teachers
- Calling all available phone numbers three times
- Visiting the girl’s household twice
- Asking the community, including the girls her age

The FE girls who could not be recontacted were replaced with girls in the same school selected from the replacement list. For each formal school, 20 replacement girls were randomly sampled among the pupils not included in the original sample. Out of 65 FE girls who could not be recontacted, 63 have been replaced. The other two girls could not be replaced because the school they used to go to was permanently closed and we could not draw a replacement sample. New NFE girls from the baseline sample list who could not be found or have dropped out of the programme were also replaced with randomly selected girls in the replacement lists, with 20 replacement girls randomly selected for each NFE centre; in total, 96 new NFE girls have been replaced.

In many instances, girls were replaced due to the following reasons:

- Moved away (abroad or another district)
- Could not be located in the school/school records and could not be reached even after several attempts
- New NFE girl dropped out of the programme

ABE and old NFE girls were not replaced as the girls in these cohorts tend to have more diverse backgrounds and educational histories which makes it more difficult to select a comparable girl, compromising the strength of the longitudinal comparison.

## Qualitative Methodology

For the qualitative data collection, CARE provided qualitative guides for the focus group discussions (FGDs) with mothers, teachers, and CEC and guides for the risk mapping and vignette exercises with the girls, as well as guides for the KIIs with religious leaders. Upon the reception of the tools, the evaluation team conducted a second check to verify the accuracy of the Somali translations.

The guides for the FGDs with CEC aimed to collect information about CEC members' experiences with school and CEC management, as well as their attitudes towards girls' education. The guides for the FGDs with teachers collected information about the teachers' attitudes and perceptions toward gender differences in classes and their experiences teaching students and working with the CEC, school management, and their colleagues. The questions in the guides for the FGDs with mothers asked about decision-making processes in the households, their attitudes towards girls' education, security, and more general perceptions of the community attitudes towards girls' education and opportunities.

Compared to other tools, the risk mapping and vignette exercises were a more participatory and innovative tool to better understand girls' attitudes, perceptions, and experiences related to their educational opportunities, barriers to learning and attending school, learning environment and safety. During the risk mapping exercises, girls were first asked to draw a map of their community, mark their way to school and any other important landmarks, then we asked to mark the places where they feel safe or happy and explain why. After this, girls were asked to draw the map of their school grounds and also mark where they felt safe or happy and to explain why. Consequently, they were asked to mark the places where they felt the least safe or unhappy on the maps of the community and the school grounds and also explain why they felt that way. Also, girls were asked if they think that these places are less safe for girls than boys or equally safe for girls and boys and to explain why they thought that.

The vignette exercise aimed to better understand the girls' perceptions of the value of education, barriers to learning and attending the school and also how they could be overcome. Based on four short stories about girl characters facing different education-related challenges, the girls were asked a couple of questions (included in the guides) to complete the stories and to get their opinions about what they think would happen with characters and what would help them to face their challenges.

For most qualitative interviews, the locations were randomly selected with consideration to proportionately represent the assessed districts. However, interviews with some types of respondents, such as the CEC members, were selected based on the presence of an active CEC at the chosen school. This choice was motivated by the improved quality of data, members of active CECs would be able to give more expansive insights about the work of the committee than those who are less engaged, at the cost of sample representativeness. This bias was factored into the analysis.

While selecting the participants for the interviews, team leaders were instructed to select the girls from the same age category corresponding to the cohort groups, to make it easier for girls to interact with each other during the exercise, make girls feel more comfortable and to avoid that the age gap will be too big between individual participants.

To identify the teachers for the FGD sessions, the team leaders were asked to select the teachers teaching the classes with the girls participating in the AGES program. The desired number of participants was 6 and the list of all the eligible teachers was provided to the team leaders by the school principal. On several occasions, the selected NFE centre would only have two or three teachers employed, which is fewer than a typical number of FGD participants. In these cases, the FGDs had to be conducted with only 3 teachers as it was important to represent the NFE centres and replacement centres with more teachers were not available.

For the FGDs with mothers, team leaders would receive a list with the contact details of the mothers of the girls who were interviewed for the household survey. Team leaders would individually contact each of the mothers and invite them to participate in the FGD sessions.

Altogether, we planned to collect 68 qualitative interviews across 19 locations. The qualitative sample is specified below:

- 12 FGDs with CEC members
- 12 FGDs with teachers
- 12 FGDs with mothers
- 12 vignette FGDs with girls
- 12 risk mapping FGDs with girls
- 8 KIIs with religious leaders

We were able to collect 67 out of the 68 planned interviews. The achieved sample contains one extra vignette but is missing one iFGD with mothers and one with CEC members. No changes have been made to the geographic locations of the interviews that were determined before the start of fieldwork.

After the qualitative data was collected and the report writing began, we found that more information was needed about the work of the Ministry of Education officials who support the project schools. One of our staff collected phone numbers of Education officers who work with the schools included in the assessment and conducted 6 reduced KIIs over the phone. The MoE officials were asked about their engagement with project schools, quality assurance procedures, and efforts to include marginalised girls. The interviews were transcribed and translated and used to support the analysis.

## Fieldwork

In this section, we describe critical aspects of data collection and discuss how the data was analysed. With regard to data collection, this section includes details on the number and reasons for school-level replacements and removals (i.e. replacement or removal of entire sampling points or clusters), and how replacements were selected. In terms of data analysis, we describe our general approach to the qualitative and quantitative data, and how their analyses relate to one another. Further details on enumerator selection, training, quality assurance, and data cleaning are also provided.

## Enumerator selection

Nine teams were deployed to collect data, each with one team leader and 3 enumerators. Enumerators and team leaders were selected based on their experiences, gender, and the language requirements for the fieldwork locations. Altogether, 14 female and 22 male team leaders and enumerators were deployed to conduct the data collection. While assembling the fieldwork teams, we considered the gender balance. Two team leaders were women. While we prioritised the experiences of the team members, each team had at least one experienced female member, to administer gender-sensitive tools.

The focus group discussions (FGD) with mothers, risk mapping exercises and vignette exercises (RMV) with girls required to be administered by a female researcher, as female responders could be censoring their responses in the presence of male researchers. During the FGDs, mothers could feel uncomfortable to discuss certain topics in the presence of a male researcher, such as decision-making dynamics of their households, community perceptions of the girls education, risks girls face on their way to school/at school and barriers to education related to the sensitive topics, such as girls' menstrual health. In the same way, the participatory nature of the RMV tools and the sensitivity of certain questions, such as mapping the sites where girls feel insecure and explaining the reasons, requiring that only female researchers could conduct these sessions. The specificities of the qualitative tools would be discussed in the later sections, dedicated to the design of the qualitative tools and the achieved qualitative sample.

The same criteria for the implementation of the menstrual hygiene module in the household surveys were applied. During the tool development, additional controls were scripted based on the enumerator's gender and only female researchers could access the module questions. In Somalia's context, the topic of menstrual hygiene and health represents a social taboo, and due to high levels of stigmatisation, girls may feel especially uncomfortable or even refuse to respond to these questions, in the presence of men.

## Training

All teams participated in six days of training, including one day dedicated to a pilot test. The pilot test was followed by a feedback session involving the evaluation team and enumerators, to clarify issues encountered during the pilot, highlight remaining errors in the programmed surveys, and answer questions. The training covered the following areas:

- Administration of the learning assessments & mock learning assessments
- Review of all the tools & their administration
- Child protection & safeguarding
- Selection of respondents, fieldwork management & assignment of the team leaders' responsibilities

During the training, all the teams were acquainted with the purpose of the study, tools and ODK application, used for the data collection. Also, the team leaders were trained in fieldwork management (i.e. filling and keeping the tracking sheets). In addition, one day of the training was dedicated to the administration of the learning assessment and enumerators were given extra time to practice during the training.

A portion of the training time was set aside for a dedicated session on child protection and research ethics. The discussion was led by a child protection specialist from CARE's Somalia team, and it highlighted issues specific to the administration of a GEC evaluation, including: obtaining consent from children to participate in the study, ensuring children are kept safe during survey administration, ensuring children cannot be overheard when responding to sensitive questions (e.g., questions about their teachers or households), minimising the pressure felt by children when completing the learning assessment (e.g., making sure they are aware that the assessment does not impact their ability to stay in school), and the

importance of keeping information collected confidential, among other topics. All enumerators and team leaders signed CARE's child protection policy, as well as Consilient's internal child protection and research ethics standards. For the purposes of safeguarding data confidentiality, data with information that could be used to identify individual respondents – either children, community members, teachers, or head teachers – was provided exclusively to a single technical evaluation focal point at CARE, and was password-protected.

During the training, female team leaders and experienced female enumerators participated in a separate training session on the menstrual hygiene section of the household survey and on how to conduct the risk mapping/vignette exercises, led by our experienced female staff member. While conducting qualitative interviews was a responsibility assigned to the team leaders, when the team leaders were male, appointed female team members were trained on how to conduct the FGDs with mothers and the RMVs with girls. Only the most experienced female researchers were selected for the administration of these tools. All the selected female members had experiences in conducting participatory exercises and qualitative interviews with girls and mothers for projects of similar scope as AGES. Besides, the female team leaders and researchers were trained by one of our experienced female researchers, who implemented these tools during previous GEC-T evaluations.

## Pilot

Prior to the start of fieldwork for the baseline, a pilot exercise was conducted with all team leaders and enumerators participating in data collection for AGES. Following five days of training, a pilot was held in three primary schools. During the pilot, each enumerator completed a minimum of one learning assessment and one household survey. Team leaders, meanwhile, completed one classroom observation, one headcount and one school (head teacher) survey. The evaluation team decided to have team leaders focus on completing their own data collection exercises during the pilot – rather than observe their team members completing household surveys and learning assessments – for two reasons. First, some of the team leaders selected were new to GEC and GEC-T evaluations, as the evaluation team had not completed a GEC assessment in South Central Somalia before. Second, the data collection tools team leaders are tasked with can be fairly complicated, and require strong organisational skills. They also, at least in the case of the classroom observation, require some degree of subjective judgement. Completing these tools for the first time during live data collection can be extremely challenging; thus a full pilot of these tools for each team leader was critical. The qualitative tools were not completed during the pilot exercise in Hargeisa.

## Replacement schools

During the fieldwork, several changes were made to the sample, mostly related to our teams' inability to recontact pre-selected girls in formal schools. All the teams in nearly all the locations encountered similar problems and changes were made both at the level of the school, and the level of individual respondents. This subsection details the specific circumstances under which the school-level changes were made to the sample and how replacement schools were selected.

In total, two formal schools from the recontact sample were replaced due to insecurity using random selection methods. In our research design, ABE and old NFE girls who could not be recontacted are dropped from midline analysis, which is why we did not replace the ABE and NFE centres located in the two schools. Three schools in the new NFE sample were also replaced due to insecurity. An additional school initially selected in the new NFE sample only had 12 students, which was not enough to meet the target 20 and have a replacement list; it was replaced by another school. All replacement schools were in the same states as the original schools.

## Quality Control

For the quantitative data, several quality checks were scripted into the survey tools to reduce the data-entry related errors and ensure only eligible respondents would be interviewed, such as choice filters, age restrictions, constraints for the numeric values, and calculations for the learning assessment scores.

During the fieldwork, teams were provided with several tracking tools, such as individual tracking sheets and tracking sheets for each sample point, containing the identifier and demographic information for the cohort girls for each sample point, allowing us to verify the survey data. The tracking sheets will supplement the electronic data for tracking purposes during future evaluations.

A quality control tracking tool was specifically developed and used on the daily basis by the full-time staff members (assigned research officer and fieldwork manager) to track the number of submitted surveys, specific cohort groups, individual respondents and any changes/information related to the quantitative and qualitative data collection, as well as time it took to complete the surveys. The timestamps scripted throughout the survey allowed us to identify the enumerators who tended to rush and provide feedback to team leaders to ensure they monitor the pace at which their team members deliver the surveys.

In addition, we have conducted quality control checks of the submitted data on a daily basis, as well as regular and more extensive data cleaning. All the inconsistencies and mistakes were discussed with the teams in the field, and if necessary, corrected in the data.

To ensure the quality of the qualitative data, team leaders were assigned to conduct the FGDs, as they were the most experienced team members in the qualitative data collection. In addition, during the training female team leaders and selected female enumerators were trained by our experienced staff on how to conduct the risk mapping/vignette exercises with the girls. Teams were also provided with the rosters for the participants in the qualitative interviews, to track the age and gender of the participants.

During the fieldwork, a team of full-time staff members was assigned to regularly review the recorded audio files of the FGDs and risk mapping/vignette exercises and flag the inappropriate administration of the tools from the start. One of the team members was especially experienced in the implementation of these tools and had a good understanding of how they should be administered, as the researcher participated in the past evaluations of the CARE's educational projects. All the received qualitative interviews were systematically reviewed after their reception.

## Data management and cleaning

For the quantitative data, to ensure secure data management, the evaluation team used an online data management platform – ONA – and all teams were required to submit the surveys to the ONA servers once they were completed. The submitted data were downloaded on a daily basis for regular quality control and data cleaning.

Daily data cleaning focused on general inconsistencies and the duplicate unique ids/observations, age variables, the respondent types, school grade variables, phone numbers, spelling of string variables and learning assessment scores. While household survey and learning assessment data were reviewed daily, the review and cleaning of the data from other surveys were done bi-weekly.

On a weekly/bi-weekly basis, depending on the specific survey data, a more in-depth data cleaning was conducted by our team. All the variables were separately examined and cross-tabulated to identify any possible inconsistencies in the data. If logical inconsistencies were discovered, we contacted team leaders to double-check the answers in case they included typos or accidental mistakes.

As far as the qualitative interviews were concerned, team leaders were required to share audio recordings with our team controlling the quality of the data. Once reviewed, all the qualitative interviews were transcribed and translated by our full-time staff members and externally contracted staff, using specifically developed templates. The process of transcription and translation was supervised by our full-time staff



member and the quality of the English translation was reviewed by international full-time staff members. Subsequently, the quality of the translations was reviewed and corrected.

## Analysis

To analyse the quantitative data, survey weights were applied wherever appropriate, and clustering in the sample design was accounted for by clustering the standard errors in our analysis. Survey weights were used in the context of aggregate household- and girl-level analysis, with weights adjusting for unequal cluster sizes across sampling points within the same institutional category (i.e. formal schools, ABE centres, and NFE centres). Standard errors were clustered at the level of individual learning institutions, meaning that error adjustments were applied separately to, for instance, an NFE centre and formal school, even in cases where they share a building. The logic of this approach is that clustering of outcomes is driven largely by the similarity of girls within a given learning institution, rather than by their geographic proximity. We empirically confirmed this assumption by assessing intra-cluster correlation levels within geographic sampling points (aggregating different learning institution types) and within individual learning institutions, finding that intra-cluster correlation was dramatically higher in the latter case. Our approach to clustering is justified by this empirical finding. Note that all analysis with more than one observation in an institution employed clustered standard errors, but weights were only applied to aggregate analysis of household- and girl-level data.

Most of the quantitative analysis was conducted in a linear regression framework, allowing us to account for clustering that is not possible to incorporate into a traditional t-test. The regression framework also made it simple to dig deeper into key findings by incorporating critical control variables and exploring how the substantive results responded to such modelling choices.

Throughout the analysis, we cluster standard errors wherever appropriate. Specifically, we utilise clustered errors wherever multiple observations or data points were collected from the same learning institution. Therefore, we cluster errors when analysing headcounts, classroom observations, teacher surveys, and household surveys or learning assessments. The only exception are head teachers (school) surveys, only one of which was completed per school.

Decisions around clustering are based on our expectations regarding the relative similarity of girls within learning institutions versus between learning institutions. We view clustering as occurring along two axes: geographic, insofar as girls who live in the same area are broadly subject to the same experiences and have similar backgrounds; and institutional, insofar as girls enrolling in each distinct type of learning institution are more similar to one another than to girls in other types of institutions. The latter point is obvious from the age distribution of the sample in the different institution types, as institution type is highly correlated with age. Girls in the same type of learning institution are also likely to share other similarities, such as their trajectory through or outside the educational system up to the point of the baseline. For this reason, we opt to define formal schools, ABE centres, and NFE centres as distinct clusters, even if they are situated within the same school building.

The qualitative data, after being translated and organised in a master spreadsheet, was analysed systematically to identify the insights that supported the findings in the quantitative data. Instances where the qualitative data supported or ran counter to quantitative findings were also recorded.

## Challenges and Limitations

This subsection details the challenges encountered during fieldwork, as well as the strategies employed to mitigate them whenever possible. Limitations related to the research design are detailed in methodology section of the main report body.



During fieldwork, a number of challenges, related to security and otherwise were encountered. Security challenges, as outlined in detail in the fieldwork report, included elections-related tensions, an explosion a day before the start of fieldwork, and a transportation curfew imposed after 5 pm that created challenges for our fieldwork teams who were returning late from remote schools. More severe and ongoing security challenges also prevented the inclusion of schools altogether in two locations.

Additional fieldwork challenges included:

**Closures of ABE and NFE programmes:** 15 ABE and NFE programmes were closed at the time of assessment as the original cohort of girls has graduated and the programme was not renewed. The girls were recontacted at their homes but classroom observations and headcounts for these centres could not be completed.

**Permanent school closures:** One school stopped all operations (FE, ABE, and NFE classes) two years ago. The baseline girls were recontacted at their homes, but classroom observations and headcounts could not be completed. School closure also meant that FE girls who were not found could not be replaced.

**Ongoing CARE training:** Teachers in several new NFE centres in Mogadishu were participating in a CARE training at the time of the assessment, resulting in the research teams having to come back to the centres at a later date.

**Tense relationship between caregivers and schoolteachers:** In several schools in Banadir, the tension between parents and schools made it difficult to ensure support for locating and interviewing the girls.

**Mixed classes with NFE, ABE and FE students:** In several schools, students from different programme streams were sharing the same classroom due to the limited number of classrooms in the school, requiring additional help from teachers to conduct the headcounts.

**Limited availability of NFE teachers for FGDs:** New NFE centres that were randomly selected as the teacher FGD sites only had two to three teachers, which was less than the typical number of FGD participants. Ultimately, FGDs were conducted with three teachers, as there were no replacement options with more teachers available.

**Caregiver availability:** In one school, mothers initially refused to participate in the FGD as most of them worked from 8am to 5:30pm. Additional effort was required to encourage FGD recruitment and participation.

**Lack of cooperation from the school authorities:** In one school, the school's head teacher did not allow the team to conduct classroom observations.

**Suspicion from local authorities:** One school is located near a military compound. The military personnel requested the researchers to leave the school immediately after the classes were finished, but this was resolved after the team leader spoke to the commander.

**Building trust with families:** On several occasions, girls' husbands refused to let the girls participate in the assessment, primarily when it required the girls to come to the school. In all reported cases, team leaders managed to convince the husbands to allow the interview, but the negotiations could take up to three hours.

**New teachers:** Several schools had new teachers who did not know the girls who have dropped out since the baseline assessment and were unable to help with locating them. In one school, the head teacher was also new, resulting in his limited knowledge of some of the information relevant for the school survey, as well as the AGES project more broadly.

**Weather conditions:** In several locations, very hot weather and water shortages posed major inconveniences for the enumerators. For example, in one school, classrooms were made of iron sheets creating a very hot environment, making the implementation of headcounts and classroom observations very challenging.

## Annex 2 - Learning Tables - Midline Cohorts

In this annex, we provide additional tables documenting learning outcomes among the “midline cohorts” – the FE, ABE, and NFE girls who were first interviewed in the 2019 baseline. The tables provide more detailed learning results that may be of interest but are tangential to an assessment of programme impact. For instance, we report results on subgroup learning outcomes at midline. While these results may be of interest for understanding current patterns in how marginalisation or demographic characteristics affect learning outcomes, this is better assessed using baseline data and readers would be better-served by a review of the baseline report, where subgroup learning outcomes were the subject of lengthy discussion. Moreover, if readers are interested in subgroup-specific *changes* in learning outcomes, this question is addressed in the main body of the report.

### Subgroup Learning Outcomes

The tables below report literacy and numeracy outcomes among specific subgroups of the midline cohorts. Membership in each subgroup is binary, and we report learning outcomes among the members of these subgroups, which can be compared to the “overall” results provided in the first row of the table. We refer to the FE, ABE, and NFE groups, respectively, as cohorts. For each cohort, we report the size (sample size) of the subgroup within that cohort and the mean learning score for that subgroup. We do not report aggregations across cohorts – e.g., results for all married girls, averaged across all three cohorts – because the three cohorts are so different. Averaging results across the cohorts makes little sense, given their differences in location, age, and intervention exposure. Note that panels within each table clearly delineate the type of subgroup to be reported, such as state, family characteristics, disability status, and so forth.

The first table reports scores for Somali literacy among subgroups that are defined in a straightforward fashion utilising either responses from the girl, responses from her caregiver, or both.<sup>161</sup> The second table also reports Somali literacy scores among particular subgroups, but focuses on subgroups derived exclusively from data reported by a girl’s head of household or caregiver; because the use of data from a girl’s head of household or caregiver alters the overall sample for comparison slightly – because not all caregivers completed surveys – we have elected to separate the results into two tables for the sake of clarity. The tables for numeracy scores follow an identical pattern: two tables report numeracy results drawn from the abbreviated version of the numeracy (EGMA) assessment that was used at baseline. This assessment consists of the first eight numeracy subtasks, with the first table providing results from our main subgroups and the second reporting results for subgroups derived from data collected during the household survey from heads of household and caregivers. A third set of tables reports results from the full numeracy assessment, using all 11 subtasks.

<sup>161</sup> In a few cases, noted in the table with a single asterisk, we use baseline data on the girls to determine their membership in a particular subgroup. Note, however, that this is only done in cases in which: a) membership in the subgroup is very unlikely to change over time and b) the midline data collected was less complete, due to missing responses from household heads or caregivers.

TABLE 76: SOMALI LITERACY AMONG SUBGROUPS OF FE, ABE, AND NFE GIRLS (FULL PANEL SAMPLE)

Subgroup	FE Girls		ABE Girls		Old (Cohort 1) NFE Girls	
	Sample Size	Somali Lit.	Sample Size	Somali Lit.	Sample Size	Somali Lit.
All Girls in Midline Panel Sample	332	63.8	364	62.1	366	62.9
Banadir	142	66.4	175	64.6	187	70
Jubaland	119	70.1	83	57.1	83	52.8
South West State	71	48.2	106	62.1	96	57.7
<b>Family Characteristics and Parental Educational Attainment</b>						
Lives without parents	26	60.8	42	62.9	86	53
Female-headed household*	111	65.9	125	58.9	155	59.2
HoH has no education*	75	56.8	85	59.2	78	66.4
HoH has no formal education*	102	58.1	111	64.5	97	65.2
CG has no education*	265	64.5	306	62.8	296	60.8
CG has no formal education*	285	64.8	314	63.2	310	61.9
Neither adult has completed any education*	69	56.8	76	61.2	73	66.8
Neither adult has completed any formal education*	259	64.7	292	63.1	283	60.4
<b>Household Wealth &amp; Economic Security</b>						
Poor quality roof	41	57.2	74	52.5	41	51.6
Gone without food most/all days, last 12 months	29	66.5	31	56.6	24	49.5
Gone without clean water most/all days, last 12 months	31	53.7	29	52.6	25	52.6
Gone without medicines most/all days, last 12 months	65	62	75	56.4	76	64
Gone without cash income most/all days, last 12 months	70	55.2	94	59.2	92	66.8
Household owns land, either jointly or independently	125	61.5	100	68.1	108	66.1
<b>Disability Status</b>						

Vision disability	1	48.4	3	61.7	2	74.1
Vision disability, liberal definition	19	68.1	26	50.7	41	70.1
Hearing disability	1	88.5	2	28.5	1	89.2
Hearing disability, liberal definition	10	58.2	8	29.4	9	62.7
Mobility disability	3	69	3	38.4	11	61.6
Mobility disability, liberal definition	36	60.5	34	53.7	33	59
Arms/hands disability**	3	70.1	1	36.6	0	
Arms/hands disability, liberal definition**	9	68.1	3	67.4	0	
Self-care disability	0		2	39.2	1	0
Self-care disability, liberal definition	13	63.5	6	45	6	37.7
Communication disability	1	71.2	0		1	0
Communication disability, liberal definition	13	61.9	9	63.4	5	23.9
Cognitive disability	1	0	3	58.1	2	0
Cognitive disability, liberal definition	23	68.1	23	54.2	26	47.8
Behavioural disability	4	53.7	1	67.2	3	46.1
Behavioural disability, liberal definition	30	58.7	16	36.8	16	49.3
Mental health disability	40	67.5	70	59.9	78	65.9
Mental health disability, liberal definition	113	67.6	151	65.7	155	65.6
Physical disability, any kind (self-reported)	8	69.3	11	43	15	61
Physical disability, any kind (self- or caregiver-reported)	16	56.9	16	48.7	15	61
Cognitive, communication, or behavioural disability, any kind (self-reported)	6	47.7	4	60.4	6	23.1
Cognitive, communication, or behavioural disability, any kind (self- or caregiver-reported)	11	52.2	6	53.4	6	23.1

**Vectors of Marginalisation – Identity and Language**

Girl speaks af-Maay at home	79	48	106	64.8	82	55.1
Linguistic minority	11	46.3	27	58.7	19	35.4
IDP household	125	66.2	157	58.5	127	55.5
Somali Bantu and other ethnic minorities	39	60.9	86	53	71	51.3
<b>Quality of School Facilities</b>						
Doesn't use drinking water facilities	14	63.1	18	59	37	65.3
Doesn't use toilet at school	25	56.8	34	51.7	48	64.6
No computers at school	291	63.6	318	62.7	337	62.8
Cannot use books or other learning materials at school	9	35.5	13	50.4	20	59.2
Not enough seats for all students	18	42.7	14	48.7	32	61
<b>Teacher Characteristics and Quality</b>						
Disagrees teachers make them feel welcome	35	55	39	47.4	40	53.5
Agrees teachers often absent from class	71	72.3	68	59.3	75	52.2
Teacher does not encourage student participation	15	45.7	13	65.1	11	42.3
Teacher punishes students who get things wrong in a lesson	187	63.6	167	62.8	130	61.8
Teacher uses physical punishment****	56	72.3	37	70.7	6	85.5
<b>Miscellaneous</b>						
Girl feels route to school is unsafe	7	35.2	23	59.6	27	73.2
Girl spends half or more of her day completing chores	133	64.1	262	63	354	62.8
Girl feels she cannot attend school if menstruating***	5	34.1	13	72	10	35.1

\*Based on the full sample of panel girls; however, parental education was not captured at midline, so these results are derived from baseline survey responses for the same girls. Because parental educational attainment generally does not change year-on-year, baseline values have been employed to provide maximum sample coverage.

\*\*Disability affecting the hands/arms is coded on the basis of caregiver reports only, as girls were not directly asked about this form of impairment.

\*\*\*Based on subsample of girls interviewed by female enumerator and who reported they had begun menstruation.

\*\*\*\*Based on the subsample of girls who were enrolled in school at midline.

TABLE 77: SOMALI LITERACY AMONG SUBGROUPS OF FE, ABE, AND NFE GIRLS (SUBGROUPS DERIVED FROM HOUSEHOLD SURVEY)

Subgroup	FE Girls		ABE Girls		Old (Cohort 1) NFE Girls	
	Sample Size	Somali Lit.	Sample Size	Somali Lit.	Sample Size	Somali Lit.
<b>Subgroups Derived from Household Head Responses</b>						
All Girls in Panel Sample with Complete Household Head Module	325	64.1	214	59.7	12	48.9
Partial Orphan (one parent deceased)	49	67.3	27	55.2	1	54.9
Full Orphan (both parents deceased)	4	46.9	1	18.9	0	N/A
HoH has no wage-earning occupation	156	65	104	63.5	7	49.4
HoH engaged primarily in pastoralism	4	62.6	1	100	0	N/A
<b>Subgroups Derived from Caregiver Responses</b>						
All Girls in Panel Sample with Complete Caregiver Module	322	64.1	211	60.2	12	48.9
Girl is married	4	66.9	10	73.6	2	70.8
Girl was ever married	5	53.5	11	75.1	2	70.8
Girl is a mother	2	46.9	4	75.6	1	76.7
Household owns a phone	290	64.4	188	60.1	11	46.6
Household owns a smartphone	98	71.6	24	69.1	2	34.6
Household has savings	29	73.5	14	66.9	2	73.3



TABLE 78: NUMERACY (ABBREVIATED VERSION) AMONG SUBGROUPS OF FE, ABE, AND NFE GIRLS  
(FULL PANEL SAMPLE)

Subgroup	FE Girls		ABE Girls		Old (Cohort 1) NFE Girls	
	Sample Size	Numeracy (8 tasks)	Sample Size	Numeracy (8 tasks)	Sample Size	Numeracy (8 tasks)
All Girls in Midline Panel Sample	332	73.9	364	71.5	366	70.6
Banadir	142	76.6	175	73.6	187	75.6
Jubaland	119	73.4	83	61.2	83	53.2
South West State	71	69.4	106	76	96	75.9
<b>Family Characteristics and Parental Educational Attainment</b>						
Lives without parents	26	75.5	42	72.7	86	66.4
Female-headed household*	111	74.5	125	69.2	155	66.2
HoH has no education*	75	72.8	85	74.5	78	74.1
HoH has no formal education*	102	72.3	111	75.1	97	74.9
CG has no education*	265	74	306	72	296	69.8
CG has no formal education*	285	73.9	314	72.1	310	70.1
Neither adult has completed any education*	69	71.9	76	74.7	73	74.8
Neither adult has completed any formal education*	259	74	292	72.1	283	69.6
<b>Household Wealth &amp; Economic Security</b>						
Poor quality roof	41	75.5	74	66.7	41	62
Gone without food most/all days, last 12 months	29	73.3	31	65.4	24	57.5
Gone without clean water most/all days, last 12 months	31	64.4	29	69.3	25	65.9
Gone without medicines most/all days, last 12 months	65	71.8	75	65.1	76	69.3

Gone without cash income most/all days, last 12 months	70	66.8	94	69.6	92	75.4
Household owns land, either jointly or independently	125	72.6	100	76.4	108	75.6
<b>Disability Status</b>						
Vision disability	1	78.3	3	72.2	2	66.1
Vision disability, liberal definition	19	72.1	26	61.4	41	73.6
Hearing disability	1	66	2	72.8	1	81.6
Hearing disability, liberal definition	10	60.3	8	49.2	9	59.2
Mobility disability	3	81.1	3	78.2	11	70.3
Mobility disability, liberal definition	36	72.4	34	66.1	33	64
Arms/hands disability**	3	65	1	81	0	
Arms/hands disability, liberal definition**	9	68.9	3	61.5	0	
Self-care disability	0		2	31	1	0
Self-care disability, liberal definition	13	68.8	6	60.6	6	40.9
Communication disability	1	67.4	0		1	0
Communication disability, liberal definition	13	65.6	9	61.9	5	24.2
Cognitive disability	1	2.5	3	62.5	2	0
Cognitive disability, liberal definition	23	71.6	23	66	26	55.9
Behavioural disability	4	58.2	1	84.6	3	64.2
Behavioural disability, liberal definition	30	63.4	16	48.2	16	55.6
Mental health disability	40	77	70	69.9	78	72
Mental health disability, liberal definition	113	75.1	151	73	155	71.9
Physical disability, any kind (self-reported)	8	72.9	11	67.2	15	65.8

Physical disability, any kind (self- or caregiver-reported)	16	70.8	16	63.9	15	65.8
Cognitive, communication, or behavioural disability, any kind (self-reported)	6	50.5	4	68	6	32.1
Cognitive, communication, or behavioural disability, any kind (self- or caregiver-reported)	11	53.1	6	61.9	6	32.1
<b>Vectors of Marginalisation – Identity and Language</b>						
Girl speaks af-Maay at home	79	69.2	106	74.5	82	71
Linguistic minority	11	65.7	27	62.1	19	43.6
IDP household	125	72.3	157	65.6	127	62.1
Somali Bantu and other ethnic minorities	39	77.7	86	60.5	71	60.5
<b>Quality of School Facilities</b>						
Doesn't use drinking water facilities	14	69.1	18	69.3	37	68.3
Doesn't use toilet at school	25	71.6	34	64.4	48	70.2
No computers at school	291	74.3	318	71.5	337	70.6
Cannot use books or other learning materials at school	9	67	13	64.3	20	67.1
Not enough seats for all students	18	59.5	14	55.1	32	62.6
<b>Teacher Characteristics and Quality</b>						
Disagrees teachers make them feel welcome	35	66.4	39	63.2	40	57.7
Agrees teachers often absent from class	71	72.3	68	64.7	75	57
Teacher does not encourage student participation	15	63.6	13	75	11	71

Teacher punishes students who get things wrong in a lesson	187	74.1	167	71.1	130	69.8
Teacher uses physical punishment****	56	77.8	37	80.3	6	85.9
<b>Miscellaneous</b>						
Girl feels route to school is unsafe	7	55.2	23	64.7	27	80
Girl spends half or more of her day completing chores	133	73.2	262	73.1	354	70.7
Girl feels she cannot attend school if menstruating***	5	52.8	13	83.4	10	54.4

\*Based on the full sample of panel girls; however, parental education was not captured at midline, so these results are derived from baseline survey responses for the same girls. Because parental educational attainment generally does not change year-on-year, baseline values have been employed to provide maximum sample coverage.

\*\*Disability affecting the hands/arms is coded on the basis of caregiver reports only, as girls were not directly asked about this form of impairment.

\*\*\*Based on subsample of girls interviewed by female enumerator and who reported they had begun menstruation.

\*\*\*\*Based on the subsample of girls who were enrolled in school at midline.

**TABLE 79: NUMERACY (ABBREVIATED VERSION) AMONG SUBGROUPS OF FE, ABE, AND NFE GIRLS (SUBGROUPS DERIVED FROM HOUSEHOLD SURVEY)**

Subgroup	FE Girls		ABE Girls		Old (Cohort 1) NFE Girls	
	Sample Size	Numeracy (8 tasks)	Sample Size	Numeracy (8 tasks)	Sample Size	Numeracy (8 tasks)
<b>Subgroups Derived from Household Head Responses</b>						
All Girls in Panel Sample with Complete Household Head Module	325	73.8	214	67	12	62.7
Partial Orphan (one parent deceased)	49	74.2	27	67.6	1	52.9
Full Orphan (both parents deceased)	4	69.2	1	92.7	0	N/A
HoH has no wage-earning occupation	156	72.6	104	67.5	7	62.6

HoH engaged primarily in pastoralism	4	81.3	1	98.5	0	N/A
<b>Subgroups Derived from Caregiver Responses</b>						
All Girls in Panel Sample with Complete Caregiver Module	322	74.0	211	67.7	12	62.7
Girl is married	4	70.4	10	68.2	2	75.2
Girl was ever married	5	56.3	11	70.1	2	75.2
Girl is a mother	2	46.9	4	74.3	1	74.8
Household owns a phone	290	74.1	188	67.5	11	60.1
Household owns a smartphone	98	78.1	24	73.9	2	59.9
Household has savings	29	79.9	14	80.2	2	65.5

TABLE 80: NUMERACY (FULL VERSION) AMONG SUBGROUPS OF FE, ABE, AND NFE GIRLS (FULL PANEL SAMPLE)

Subgroup	FE Girls		ABE Girls		Old (Cohort 1) NFE Girls	
	Sample Size	Numeracy (11 tasks)	Sample Size	Numeracy (11 tasks)	Sample Size	Numeracy (11 tasks)
All Girls in Midline Panel Sample	332	66.4	364	63.9	366	62
Banadir	142	68	175	65.1	187	65.3
Jubaland	119	67.8	83	55.2	83	47.8
South West State	71	61.2	106	68.7	96	67.6
<b>Family Characteristics and Parental Educational Attainment</b>						
Lives without parents	26	66.1	42	63.9	86	57.7
Female-headed household*	111	67.5	125	61.3	155	57.9
HoH has no education*	75	65	85	65.5	78	64.9

HoH has no formal education*	102	64.5	111	66.9	97	65.2
CG has no education*	265	66.8	306	64.4	296	61
CG has no formal education*	285	66.7	314	64.7	310	61.6
Neither adult has completed any education*	69	64.1	76	66.1	73	65.3
Neither adult has completed any formal education*	259	66.8	292	64.7	283	61
<b>Household Wealth &amp; Economic Security</b>						
Poor quality roof	41	67.5	74	56.9	41	57.8
Gone without food most/all days, last 12 months	29	65.4	31	55.7	24	49.2
Gone without clean water most/all days, last 12 months	31	55.8	29	60.3	25	59.5
Gone without medicines most/all days, last 12 months	65	61.9	75	54.9	76	57.7
Gone without cash income most/all days, last 12 months	70	57	94	58.4	92	63.8
Household owns land, either jointly or independently	125	65.6	100	69	108	67.1
<b>Disability Status</b>						
Vision disability	1	57	3	59.3	2	55.1
Vision disability, liberal definition	19	62.6	26	51.3	41	62.7
Hearing disability	1	75.3	2	53	1	77.1
Hearing disability, liberal definition	10	57.9	8	39	9	48.8
Mobility disability	3	71.1	3	59.6	11	64.6
Mobility disability, liberal definition	36	66	34	56.2	33	56.5
Arms/hands disability**	3	64.4	1	65.3	0	
Arms/hands disability, liberal definition**	9	65.8	3	52.2	0	
Self-care disability	0		2	30.5	1	0

Self-care disability, liberal definition	13	65.9	6	53.6	6	40.2
Communication disability	1	68.6	0		1	0
Communication disability, liberal definition	13	62.9	9	58	5	21.1
Cognitive disability	1	1.8	3	60.3	2	0
Cognitive disability, liberal definition	23	68.3	23	58.7	26	49.7
Behavioural disability	4	51.8	1	71.1	3	61.7
Behavioural disability, liberal definition	30	56.7	16	41.8	16	51
Mental health disability	40	71.8	70	63.7	78	65.1
Mental health disability, liberal definition	113	68.7	151	66.2	155	63.8
Physical disability, any kind (self-reported)	8	67.4	11	53.5	15	59.8
Physical disability, any kind (self- or caregiver-reported)	16	64.7	16	52.6	15	59.8
Cognitive, communication, or behavioural disability, any kind (self-reported)	6	46.3	4	63	6	30.8
Cognitive, communication, or behavioural disability, any kind (self- or caregiver-reported)	11	48.1	6	56	6	30.8
<b>Vectors of Marginalisation – Identity and Language</b>						
Girl speaks af-Maay at home	79	60.9	106	67.2	82	63.5
Linguistic minority	11	57.5	27	54.8	19	38.6
IDP household	125	64.8	157	57.8	127	54.9
Somali Bantu and other ethnic minorities	39	70.3	86	52.2	71	53



Quality of School Facilities						
Doesn't use drinking water facilities	14	62.6	18	59.6	37	61
Doesn't use toilet at school	25	62.3	34	55.4	48	61.8
No computers at school	291	66.8	318	63.7	337	61.6
Cannot use books or other learning materials at school	9	57.9	13	59	20	60.5
Not enough seats for all students	18	48.4	14	47.8	32	54.9
Teacher Characteristics and Quality						
Disagrees teachers make them feel welcome	35	58.2	39	54.2	40	51
Agrees teachers often absent from class	71	67.9	68	58.6	75	53
Teacher does not encourage student participation	15	63.4	13	68.1	11	59.2
Teacher punishes students who get things wrong in a lesson	187	67.2	167	65.3	130	63.7
Teacher uses physical punishment****	56	72	37	73.2	6	84.4
Miscellaneous						
Girl feels route to school is unsafe	7	45.9	23	56.6	27	70.2
Girl spends half or more of her day completing chores	133	67.5	262	65.2	354	62.1
Girl feels she cannot attend school if menstruating***	5	40	13	76.7	10	45.1

\*Based on the full sample of panel girls; however, parental education was not captured at midline, so these results are derived from baseline survey responses for the same girls. Because parental educational attainment generally does not change year-on-year, baseline values have been employed to provide maximum sample coverage.

\*\*Disability affecting the hands/arms is coded on the basis of caregiver reports only, as girls were not directly asked about this form of impairment.

\*\*\*Based on subsample of girls interviewed by female enumerator and who reported they had begun menstruation.

\*\*\*\*Based on the subsample of girls who were enrolled in school at midline.

**TABLE 81: NUMERACY (FULL VERSION) AMONG SUBGROUPS OF FE, ABE, AND NFE GIRLS  
(SUBGROUPS DERIVED FROM HOUSEHOLD SURVEY)**

Subgroup	FE Girls		ABE Girls		Old (Cohort 1) NFE Girls	
	Sample Size	Numeracy (11 tasks)	Sample Size	Numeracy (11 tasks)	Sample Size	Numeracy (11 tasks)
<b>Subgroups Derived from Household Head Responses</b>						
All Girls in Panel Sample with Complete Household Head Module	325	66.3	214	59.8	12	51.9
Partial Orphan (one parent deceased)	49	66.7	27	58.4	1	40.8
Full Orphan (both parents deceased)	4	55.8	1	69.7	0	N/A
HoH has no wage-earning occupation	156	66.4	104	61.3	7	54
HoH engaged primarily in pastoralism	4	76.9	1	71.6	0	N/A
<b>Subgroups Derived from Caregiver Responses</b>						
All Girls in Panel Sample with Complete Caregiver Module	322	66.5	211	60.4	12	51.9
Girl is married	4	60.8	10	63	2	65.4
Girl was ever married	5	48.6	11	64.1	2	65.4
Girl is a mother	2	47.7	4	63.9	1	62.1
Household owns a phone	290	66.6	188	59.9	11	50.3
Household owns a smartphone	98	70.1	24	67.2	2	45.8
Household has savings	29	69.8	14	70.8	2	48.8

## Foundational Skill Gaps

In this section we analyse girls’ performance on individual subtasks, using the “foundational skill gaps” framework common to Girls’ Education Challenge evaluations. The idea of this analysis is to identify the skills – the narrower subskills required for, e.g., reading, such as letter identification – that girls lack, or have begun acquiring, or have acquired with proficiency, and to understand how girls are distributed across this spectrum of skill acquisition. This analysis can be useful for identifying where girls’ learning tends to be “hung up”, where there is divergence between girls with high and low literacy scores, and –

therefore – where targeted interventions by teachers may be able to have outsized impacts on student learning.

In each table below, we report the share of girls who fall into each of four categories:

- Non-learner
- Emergent learner
- Established learner
- Proficient learner

These categories are associated with specific outcomes on a subtask, such as achieving a score between 41 and 80 percent (“established learner”) on that subtask. Results for each subtask are reported in separate columns.

We provide separate tables for each cohort. The first table below captures numeracy performance of FE girls; the top panel includes all FE girls re-contacted at midline, while the bottom panel includes only those FE girls who remained in-school at midline. The next tables repeat this analysis for ABE and NFE girls, though we do not report separate results for those girls who remain enrolled, because few ABE and NFE girls from the baseline remain enrolled in their respective programmes at midline. Each of these tables are repeated for Somali literacy.

TABLE 82: FOUNDATIONAL GAPS IN NUMERACY AT MIDLINE, FE GIRLS

Subtask	1	2	3	4	5	6	7	8	9	10	11
	Number Ident.	Number Discrimination	Missing Numbers	Addition (1 digit)	Addition (2 digits)	Subtract. (1 digit)	Subtract. (2 digits)	Word Problems (add. & subtract.)	Multiplic. (1 digit)	Division (1 digit)	Word Problems (mult & div)
<b>All FE Girls</b>											
Non-learner 0%	1.5	2.7	12	7.5	17.8	13.3	21.4	10.8	45.5	38.3	23.8
Emergent learner 1%-40%	0.3	0.9	53.6	1.5	11.1	3.9	20.2	2.1	10.8	10.8	21.4
Established learner 41%-80%	1.2	9	28.3	11.7	27.7	11.7	24.4	12.3	21.7	17.2	28.6
Proficient learner 81%-100%	97	87.3	6	79.2	43.4	71.1	34	74.7	22	33.7	26.2
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
<b>FE Girls Remaining Enrolled at Midline</b>											
Non-learner 0%	0	1.1	9.5	4.6	13	8.8	16.5	6.7	40.8	33.1	18.3
Emergent learner 1%-40%	0.4	0.7	54.2	1.1	12	3.9	20.1	2.5	11.3	11.3	21.8
Established learner 41%-80%	1.1	9.2	29.2	12.3	28.5	10.9	26.4	12.3	24.6	18.3	31
Proficient learner 81%-100%	98.6	89.1	7	82	46.5	76.4	37	78.5	23.2	37.3	28.9
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

TABLE 83: FOUNDATIONAL GAPS IN NUMERACY AT MIDLINE, ABE GIRLS

Subtask	1	2	3	4	5	6	7	8	9	10	11
	Number Ident.	Number Discrimination	Missing Numbers	Addition (1 digit)	Addition (2 digits)	Subtract. (1 digit)	Subtract. (2 digits)	Word Problems (add. & subtract.)	Multiplic. (1 digit)	Division (1 digit)	Word Problems (mult & div)
Non-learner 0%	3.8	4.7	14.8	9.1	22.3	18.1	24.7	15.7	45.3	43.4	29.4
Emergent learner 1%-40%	0.8	0.5	45.6	3	10.7	3.6	16.2	2.5	15.1	7.1	16.2
Established learner 41%-80%	1.4	9.3	33	9.1	23.6	10.2	26.4	11	17.9	18.7	33.2
Proficient learner 81%-100%	94	85.4	6.6	78.8	43.4	68.1	32.7	70.9	21.7	30.8	21.2
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

TABLE 84: FOUNDATIONAL GAPS IN NUMERACY AT MIDLINE, OLD NFE (COHORT 1) GIRLS

Subtask	1	2	3	4	5	6	7	8	9	10	11
	Number Ident.	Number Discrimination	Missing Numbers	Addition (1 digit)	Addition (2 digits)	Subtract. (1 digit)	Subtract. (2 digits)	Word Problems (add. & subtract.)	Multiplic. (1 digit)	Division (1 digit)	Word Problems (mult & div)
Non-learner 0%	7.4	7.9	13.7	11.7	22.1	17.8	25.1	13.1	51.4	47	27.6
Emergent learner 1%-40%	0.3	1.1	55.7	1.1	9.8	2.7	20.8	1.1	18.6	9.8	20.8
Established learner 41%-80%	1.9	6.6	26	6.8	26.5	9.6	23.2	11.2	14.2	13.7	32.2
Proficient learner 81%-100%	90.4	84.4	4.6	80.3	41.5	69.9	30.9	74.6	15.8	29.5	19.4
Total	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%

TABLE 85: FOUNDATIONAL GAPS IN SOMALI LITERACY AT MIDLINE, FE GIRLS

Subtask	1	2	3	4	5	6
	Letter recognition	Common words	Reading fluency	Reading comprehension 1	Reading comprehension 2	Reading comprehension 3
<b>All FE Girls</b>						
Non-learner 0%	6.6	12.7	19.3	22.6	24.7	29.2
Emergent learner 1%-40%	4.5	11.4	12.3	5.4	3.9	15.4
Established learner 41%-80%	16.3	25.3	24.4	31.3	28.6	31.6
Proficient learner 81%-100%	72.6	50.6	44	40.7	42.8	23.8
Total	100%	100%	100%	100%	100%	100%
<b>FE Girls Remaining Enrolled at Midline</b>						
Non-learner 0%	3.9	8.1	14.1	16.5	18	22.9
Emergent learner 1%-40%	3.2	10.2	11.3	4.6	4.2	16.9
Established learner 41%-80%	16.2	27.1	27.1	33.1	31	34.2
Proficient learner 81%-100%	76.8	54.6	47.5	45.8	46.8	26.1
Total	100%	100%	100%	100%	100%	100%



TABLE 86: FOUNDATIONAL GAPS IN SOMALI LITERACY AT MIDLINE, ABE GIRLS

Subtask	1	2	3	4	5	6
	Letter recognition	Common words	Reading fluency	Reading comprehension 1	Reading comprehension 2	Reading comprehension 3
Non-learner 0%	8.5	16.5	22.3	22.8	24.7	28.6
Emergent learner 1%-40%	4.9	11.5	12.4	4.9	4.4	14
Established learner 41%-80%	11.8	24.5	20.6	34.6	34.3	34.9
Proficient learner 81%-100%	74.7	47.5	44.8	37.6	36.5	22.5
Total	100%	100%	100%	100%	100%	100%

TABLE 87: FOUNDATIONAL GAPS IN SOMALI LITERACY AT MIDLINE, OLD NFE (COHORT 1) GIRLS

Subtask	1	2	3	4	5	6
	Letter recognition	Common words	Reading fluency	Reading comprehension 1	Reading comprehension 2	Reading comprehension 3
Non-learner 0%	10.7	16.7	21	23	24.6	27
Emergent learner 1%-40%	4.1	9	10.9	3	4.1	13.9
Established learner 41%-80%	12	28.1	21.6	33.3	33.1	39.6
Proficient learner 81%-100%	73.2	46.2	46.4	40.7	38.3	19.4
Total	100%	100%	100%	100%	100%	100%

## Grade-Level Competency

The tables below report the share of FE, ABE, and NFE girls who have achieved lower-primary literacy and numeracy skills, where lower-primary skills are defined in line with the FGS curriculum. In Section 8.4, we reported equivalent achievement rates among the New NFE (Cohort 4) girls recruited in this evaluation round; in that discussion, we described the FGS curriculum, its limitations, and how the skills indicated map onto the EGMA and EGRA assessments used in this evaluation. Interested readers should refer to that discussion for a comprehensive description of these results.

The tables below report the share of girls who achieved the skills expected of girls who have completed lower-primary school at baseline and midline, disaggregated by cohort (FE, ABE, NFE). The tables utilise the panel sample, the set of girls who were successfully re-contacted at midline.

**TABLE 88: LITERACY COMPETENCY (PARTIAL) AT LEVEL OF LOWER PRIMARY COMPLETION, BY COHORT**

Cohort	Baseline	Midline
<b>FE Girls</b>	4.8%	24.4%
<b>ABE Girls</b>	12.4%	24.2%
<b>Old NFE (Cohort 1) Girls</b>	26.0%	20.2%

**TABLE 89: NUMERACY COMPETENCY (PARTIAL) AT LEVEL OF LOWER PRIMARY COMPLETION, BY COHORT**

Cohort	Baseline	Midline (Comparable Subtasks)	Midline (Expanded, Difficult Subtasks)
<b>FE Girls</b>	4.2%	4.5%	2.4%
<b>ABE Girls</b>	5.2%	4.7%	3.3%
<b>Old NFE (Cohort 1) Girls</b>	9.0%	3.6%	2.5%

## Annex 3 - Learning Tables - Baseline New NFE Cohort

In this annex, we provide additional tables that were not included in the primary learning analysis for the New NFE cohort in the main body of the report. The goal of these tables is to ensure completeness in our analysis and reporting of results, without distracting from key findings highlighted in the main report.

### Subgroup Learning Outcomes

The tables below document Somali literacy and numeracy outcomes among specific subgroups of the new NFE cohort, disaggregating results by demographic characteristics, disability status, and so forth. For each subgroup, we report the number of girls who fall into the subgroup and the mean literacy or numeracy score for girls in that subgroup. Comparisons can be made to the first two rows of the table, which report the overall mean score for all NFE girls in the cohort.

Note that many NFE girls in the sample were 18 years of age or older. As a result, their families did not complete household surveys and the girls themselves answered an abbreviated set of questions about their household characteristics. This approach to survey design means that – for girls 18 or over – there are subgroups for which we cannot determine a girl’s status. For instance, questions regarding educational attainment of the head of household were only asked as part of the full household survey; for girls 18 years or over, we do not know the educational attainment of their head of household and cannot identify whether they are from a household in which the head of household completed no formal education. For those subgroups that are based on this smaller sample – the set of girls who are under 18 years old – we mark the subgroup name with a single asterisk (\*). This partially accounts for the very small sample size for some subgroups in the tables below.

TABLE 90: SOMALI LITERACY AMONG SUBGROUPS OF NEW NFE GIRLS COHORT

Subgroup	Sample Size	Somali Literacy
All New (Cohort 4) NFE Girls	916	16.9
All New (Cohort 4) NFE Girls who completed a full household survey*	280	18.0
<b>State/Region</b>		
Banadir	440	13.6
South West State	337	24.2
Hirshabelle	139	9.6
<b>Family Characteristics and Parental Educational Attainment</b>		
Partial Orphan (one parent deceased)*	53	19.6
Full Orphan (both parents deceased)*	3	11.1
Lives without parents*	30	17.7
Girl is married*	7	3.3
Girl was ever married*	10	2.3
Girl is a mother*	5	0.5
Female-headed household*	192	17.4
HoH has no education*	47	12.1
HoH has no formal education*	237	17.9
CG has no education*	43	10.6

CG has no formal education*	236	18.2
Neither adult has completed any education*	40	9.4
Neither adult has completed any formal education*	226	17.8
<b>Household Wealth &amp; Economic Security</b>		
Poor quality roof	201	12.1
Gone without food most/all days, last 12 months	61	13.3
Gone without clean water most/all days, last 12 months	74	11.7
Gone without medicines most/all days, last 12 months	129	14.3
Gone without cash income most/all days, last 12 months	184	11.7
Household owns land, either jointly or independently	211	25.5
Household owns a phone*	258	18.2
Household owns a smartphone*	30	30.2
HoH has no wage-earning occupation*	128	18.6
Household has savings*	9	28.9
HoH engaged primarily in pastoralism*	3	23.6
<b>Disability Status</b>		
Vision disability	4	28.3
Vision disability, liberal definition	63	14.3
Hearing disability	1	4.7
Hearing disability, liberal definition	25	7.4
Mobility disability	22	12.7
Mobility disability, liberal definition	76	13.2
Arms/hands disability*	1	16.7
Arms/hands disability, liberal definition*	2	8.3
Self-care disability	2	8.3
Self-care disability, liberal definition	7	5.7
Communication disability	0	
Communication disability, liberal definition	19	15.6
Cognitive disability	28	4.8
Cognitive disability, liberal definition	126	9.2
Behavioural disability	15	7.9
Behavioural disability, liberal definition	61	14.7
Mental health disability	142	12.6
Mental health disability, liberal definition	284	15.9
Physical disability, any kind (self-reported)	29	14.3
Physical disability, any kind (self- or caregiver-reported)	31	13.8
Cognitive, communication, or behavioural disability, any kind (self-reported)	40	6.3
Cognitive, communication, or behavioural disability, any kind (self- or caregiver-reported)	51	10.3

<b>Vectors of Marginalisation – Identity and Language</b>		
<b>IDP household</b>	375	12.3
<b>Girl speaks af-Maay at home</b>	305	24.6
<b>Linguistic minority</b>	53	11.6
<b>Somali Bantu and other ethnic minorities</b>	159	7.3
<b>Quality of School Facilities</b>		
<b>Doesn't use drinking water facilities</b>	157	15.9
<b>Doesn't use toilet at school</b>	191	18.8
<b>No computers at school</b>	812	17.9
<b>Cannot use books or other learning materials at school</b>	59	8.2
<b>Not enough seats for all students</b>	35	10.2
<b>Teacher Characteristics and Quality</b>		
<b>Disagrees teachers make them feel welcome</b>	117	13.7
<b>Agrees teachers often absent from class</b>	159	14
<b>Teacher does not encourage student participation</b>	48	14.1
<b>Teacher punishes students who get things wrong in a lesson</b>	258	19.3
<b>Teacher uses physical punishment</b>	70	21.1
<b>Miscellaneous</b>		
<b>Girl feels route to school is unsafe</b>	15	9.3
<b>Caregiver feels route to school is unsafe*</b>	4	26.6
<b>Girl spends half or more of her day completing chores</b>	553	17.2
<b>Girl feels she cannot attend school if menstruating**</b>	29	13.5

\*Subgroup identified only among girls (n = 280) who completed full household survey (under 18 years old).

\*\*Subgroup identified only among girls (n = 396) who were interviewed by a female enumerator and reported having begun menstruating.

TABLE 91: NUMERACY AMONG SUBGROUPS OF NEW NFE GIRLS COHORT

Subgroup	Sample Size	Numeracy
All New (Cohort 1) NFE Girls	916	27.8
All New (Cohort 1) NFE Girls who completed a full household survey*	280	29.0
<b>State/Region</b>		
Banadir	440	24.4
South West State	337	36.5
Hirshabelle	139	17.2
<b>Family Characteristics and Parental Educational Attainment</b>		
Partial Orphan (one parent deceased)*	53	26
Full Orphan (both parents deceased)*	3	19.7
Lives without parents*	30	26.2
Girl is married*	7	33.3
Girl was ever married*	10	25.5
Girl is a mother*	5	31.4
Female-headed household*	192	26.9
HoH has no education*	47	28.1
HoH has no formal education*	237	29.4
CG has no education*	43	28.8
CG has no formal education*	236	29.6
Neither adult has completed any education*	40	27.5
Neither adult has completed any formal education*	226	29.2
<b>Household Wealth &amp; Economic Security</b>		
Poor quality roof	201	25.3
Gone without food most/all days, last 12 months	61	17.8
Gone without clean water most/all days, last 12 months	74	24.6
Gone without medicines most/all days, last 12 months	129	21.5
Gone without cash income most/all days, last 12 months	184	20.5
Household owns land, either jointly or independently	211	39.8
Household owns a phone*	258	29.4
Household owns a smartphone*	30	40.2
HoH has no wage-earning occupation*	128	27.5
Household has savings*	9	29.7
HoH engaged primarily in pastoralism*	3	32.4
<b>Disability Status</b>		
Vision disability	4	54.2
Vision disability, liberal definition	63	24.4
Hearing disability	1	20.3
Hearing disability, liberal definition	25	17.6



Mobility disability	22	28.6
Mobility disability, liberal definition	76	25.6
Arms/hands disability*	1	37.1
Arms/hands disability, liberal definition*	2	18.6
Self-care disability	2	18.6
Self-care disability, liberal definition	7	18.2
Communication disability	0	
Communication disability, liberal definition	19	19.6
Cognitive disability	28	17.7
Cognitive disability, liberal definition	126	19.8
Behavioural disability	15	15.9
Behavioural disability, liberal definition	61	21.6
Mental health disability	142	25.9
Mental health disability, liberal definition	284	27.4
Physical disability, any kind (self-reported)	29	31.2
Physical disability, any kind (self- or caregiver-reported)	31	30.3
Cognitive, communication, or behavioural disability, any kind (self-reported)	40	17.4
Cognitive, communication, or behavioural disability, any kind (self- or caregiver-reported)	51	18.3
<b>Vectors of Marginalisation – Identity and Language</b>		
IDP household	375	25.7
Girl speaks af-Maay at home	305	38.8
Linguistic minority	53	25.4
Somali Bantu and other ethnic minorities	159	18.9
<b>Quality of School Facilities</b>		
Doesn't use drinking water facilities	157	25.6
Doesn't use toilet at school	191	27.9
No computers at school	812	29.1
Cannot use books or other learning materials at school	59	20.3
Not enough seats for all students	35	27.9
<b>Teacher Characteristics and Quality</b>		
Disagrees teachers make them feel welcome	117	26.4
Agrees teachers often absent from class	159	23.3
Teacher does not encourage student participation	48	27.2
Teacher punishes students who get things wrong in a lesson	258	28.6
Teacher uses physical punishment	70	32.5
<b>Miscellaneous</b>		
Girl feels route to school is unsafe	15	24.7
Caregiver feels route to school is unsafe*	4	34.7

<b>Girl spends half or more of her day completing chores</b>	553	29.8
<b>Girl feels she cannot attend school if menstruating**</b>	29	26.7

\*Subgroup identified only among girls (n = 280) who completed full household survey (under 18 years old).

\*\*Subgroup identified only among girls (n = 396) who were interviewed by a female enumerator and reported having begun menstruating.

## Annex 4 - Logframe

---

The project's logframe is included as a separate annex.

## Annex 5 - Data Collection Tools

---

The data collection tools – quantitative surveys and qualitative interview guides -- are included as a separate annex.

## Annex 6 - Datasets and Replication Files

---

The replication files, which include datasets and Stata .do files (statistical syntax files) for replicating the primary results, are included as a separate annex.

## Annex 7 - Project Management Response

---

### Reflection on trends

**Literacy:** The substantial gains observed among formal education students, surpassing the benchmark, suggest that the program has been particularly successful in accelerating literacy skills development among girls who scored zero or had very low scores at the baseline. The more limited gains observed among ABE and NFE students potentially reflect the larger proportion of girls with higher scores among those subgroups. While a decrease in the proportion of non-learners was observed across all subgroups, this decrease was markedly higher among formal education students (27 percentage points compared to 18 pp for ABE and 2pp for NFE), which also had the highest proportion of zero scores at the baseline (46% non-readers compared to 40% among ABE students and 23% among NFE<sup>162</sup>).

**Numeracy:** While numeracy gains among formal education students have exceeded the benchmark, they were more limited than changes in literacy. The gains were also below the benchmark for ABE students. The results potentially reflect the challenges related to acquiring numeracy skills through remote learning, which will require further adaptations of training of teachers/ facilitators and learning materials, particularly to support non-learners. The loss of numeracy skills among NFE students is, on the other hand, a surprising result even though they had completed the course more than a year before the assessment, therefore being (mostly) out of school since then. Moreover, learning losses have occurred across all EGMA tasks, including simple addition and subtraction. NFE graduates are expected to regularly practice numeracy in markets/ income generation activities, which would have reduced potential losses. The result may, however, reflect cognitive losses due to hunger, and in particular, teacher absenteeism in Jubaland. A more refined analysis shows that the proportion of NFE non-learners in specific numeracy tasks has not changed – thus indicating the need for adaptations catering specifically to this subgroup.

**Learning and Disability:** Learning gains were significantly higher among GwDs compared to non-GwDs, indicating that the program has been successful in its approach to inclusion. Nonetheless, qualitative data shows that school staff still tend to perceive disability largely as posing challenges beyond their means to support students, potentially due to a narrow definition of “students with disabilities” limited to those with severe physical or cognitive disabilities, thus requiring access to specialised equipment or methodologies. Further sensitisation of school staff on types of disability and inclusive approaches is therefore necessary.

---

<sup>162</sup> Considering the results in the passage reading task at the baseline.

**Higher Learning Gains Among the Poorest:** Girls who lack regular access to water had significantly higher learning gains compared to their peers, suggesting that the program has been successful in supporting the poorest students and in training facilitators to respond to their needs. The result should be shared with facilitators as a successful example, triggering discussions on which other adaptations could be made to support disadvantaged girls, in particular those who are struggling with numeracy (as noted above).

**Learning Plateaus:** For both literacy and numeracy, the project has had limited impact on students who already had some skills. This is in part by design – the project was set up to target adolescents who had never attended school, although some of the C1 NFE and ABE students clearly had some previous exposure to education prior to enrolling in the courses. Nonetheless, this situation poses a challenge in the case of high performers, whose facilitators may not be adequately prepared to support them to learn new skills.

**Learning Gains by Geography:** Students in Jubaland had far higher learning losses than in other states, potentially reflecting the high teacher absenteeism rate, higher rates of child labour and low consumption of protein, as well as issues related to conflict, drought and political disputes affecting access to resources. The results suggest that adaptations to improve teacher attendance/ remote learning may be particularly necessary in this location, as well as additional oversight of schools in the state to identify and address issues with teacher absenteeism.<sup>163</sup> While the impact to the FE cohort (the only subgroup for whom implementation will continue in Jubaland) was far more limited, the major decline seen in ABE and NFE in the state indicates that similar issues may affect FE classes in the future as well. Additionally, while the use of corporal punishment has declined across all states, the reduction was far more limited in Jubaland, again indicating the need for additional oversight of schools.

**Learning/ Married Girls:** Learning gains in literacy were significantly lower among married girls and young mothers, potentially due to time poverty but also to social norms driving lower expectations of education for those subgroups. The sample sizes for both subgroups were also considerably small, thus increasing the potential bias in results. Further adaptations will be required to ensure adequate facilitator support to this subgroup, particularly among the new NFE cohort, which includes a much larger proportion of young wives and mothers.

**Remote Learning:** The ability to take learning materials home resulted in the highest gains in literacy and numeracy among students. Further coaching of teachers/ CECs/ school principals is necessary to ensure that all students will be able to borrow materials to take home, particularly girls at a higher risk of absenteeism, such as those who are married, mothers, or working.

The results confirm the gains associated with participation in remote learning observed in a study conducted in July 2020. Girls who were able to study remotely for at least half of the day had much higher gains in literacy and numeracy, compared to their peers who only studied for about an hour a day – a difference of 10 pp for literacy and nearly 9pp for numeracy. The results support the project’s approach of investing in a ‘permanent’ remote learning strategy – not just during lockdowns but as a permanent feature, responding to the high absenteeism rates. The results also pose future questions to the project on how to make remote learning more efficient for those facing time poverty, such as young mothers, and how to make the approach more effective for numeracy.

**Teaching Quality:** Some of the highest gains in literacy and numeracy were observed among girls whose teachers were applying poor classroom practices at the baseline – not encouraging participation, not

---

<sup>163</sup> While teacher absenteeism is explained in part by insecurity and illness, including COVID-related illnesses, the exceedingly high rates in Jubaland suggest the need for additional oversight.

welcoming students, and /or using physical discipline. The results suggest that the program has been particularly successful in improving the performance of the worst teachers, with a positive impact on learning outcomes. This is corroborated by the sharp decline in the use of corporal punishment, for example. The results should be disseminated to the schools, REOs and DEOs, serving as an example of what can be achieved by adopting improved classroom practices.

While substantial gains have been observed in teaching practices, the adoption of games and group work remains limited, and a large proportion of teachers does not keep records of formative assessments. The results indicate the need for further training/ coaching in those areas.

**Life Skills/ Social-Emotional Learning:** The participation in Girls' Empowerment Forums had a significant impact in increasing literacy gains (6.6 percentage points over and above non-participants) and numeracy (9.1 percentage points above non-participants). The impact on literacy outcomes increased over time, with girls who continue to participate in GEFs scoring 7.4 percentage points above their peers. GEF participants were also 12.5 percentage points more likely to remain in school, compared to their peers, thus confirming the project's assumptions about the crucial contribution of leadership skills development on marginalised girls' learning outcomes.

The acquisition of leadership skills, measured through the YLI, predicts higher learning scores and transition rates, as expected. Similar gains in leadership skills development were observed among IDPs, married girls and those experiencing food insecurity, but YLI improvements were lower among Af-Maay speakers and occupational minorities.<sup>164</sup> On the other hand, gains in leadership skills development were minimum among the girls with a heavy chore burden, potentially reflecting time poverty and limited capacity to participate in activities. The finding suggests the need for additional support to this subgroup. As expected, the participation in Girls' Empowerment Forums (GEFs) led to significantly higher gains in leadership skills development, as well as having a supportive Community Education Committee (CEC), thus confirming the project's ToC.

Attending a school with a GEF was a predictor of successful re-contact (a proxy for retention), increasing the average re-contact rate by 4.7 percentage points. Interestingly, the groups with the highest exposure to GEFs were ABE and NFE students, who were actually less likely to be recontacted, further suggesting that the GEF was the actual factor increasing re-contact rates. Despite the ongoing crisis, 50% of the girls who were previously involved in GEFs among the ABE and NFE graduates remained in contact with group members even after one year, and reported continuing to participate in activities to enrol out-of-school girls, reflecting the sustainability of the approach. Interestingly, 15% of the previous ABE/NFE students reported conducting joint businesses with other GEF members, despite the ongoing crisis, in line with the project's assumptions that the approach would strengthen peer-to-peer linkages for economic empowerment.

The proportion of girls with mental health disability – anxiety or depression – remains substantially high, a finding already observed during quarterly surveys. Nonetheless, the current prevalence rates of 11-17% for mental health disability are lower than the 30% observed at the baseline, potentially due to the contribution of support platforms, such as GEFs and mentorship.

**Transitions:** Considering the multiple crises in Somalia at the moment, including near-famine conditions in some implementation sites, transition rates are higher than expected, with 84% of the FE girls retained in school and 16.5% NFE progression into employment. Food insecurity predicts lower transition rates,

<sup>164</sup> The lower YLI scores observed among Af-Maay speakers may be a result of the language used in the administration of the assessment.

as expected, highlighting the need for humanitarian assistance to prevent the erosion of education gains as a result of the ongoing drought.

### **New NFE cohort**

**Demographics:** The new cohort is substantially more marginalised than the initial NFE cohort, reflecting the progressive refinement of the project's targeting, particularly in relation to GwDs and minority girls, as well as the deterioration of living conditions. Students are facing severe food insecurity—23.7% of the new NFE students reported going to sleep hungry many / most days, compared to 16.3% of the previous cohort. A large proportion of the students – 41% - have been displaced. Only 21% of the girls currently receive humanitarian assistance and involvement in local governance platforms and youth groups is extremely low (8% and 15%, respectively).

**Learning outcomes:** The new cohort has extremely limited literacy and numeracy skills, compatible with their lack of previous exposure to formal education. While the results are as expected, further coaching of facilitators on supporting non-learners will be necessary, particularly in numeracy, given the findings for C1 NFE.

The relatively large proportion of zero scores in the lowest EGMA tasks – number identification (20%) and quantity discrimination (27%) – is unusual, particularly considering that a large number of NFE students are older and likely to be engaging in small businesses or at least market transactions on a regular basis. It is possible that the zero scores in those tasks are, at least in part, related to the presence of cognitive disabilities, thus requiring an alternative approach to teaching mathematics.

It is also unusual that a very large proportion of girls – 44.5% - have scored zero in the letter recognition task. While this would not be unusual in other contexts, Somali girls typically have a higher literacy baseline due to their prior exposure to Qur'anic classes, having learned to decode at least some combinations of letters and short words even if they have never attended formal school. The result indicates that this subgroup may have only learned to memorise the Qur'an at Qur'anic school, and/or may face learning/cognitive disabilities, thus potentially requiring adaptations to teaching approaches.

**Teaching practices:** The NFE students reported that a substantial proportion of facilitators in Hirshabelle and South West disciplined them for providing incorrect responses (28% overall). While reports of corporal punishment are very limited, they still exist. The finding requires the project to coach facilitators on positive discipline and approaches to teaching struggling students. The use of formative assessments and games remains limited, as expected for a baseline.

### **Other factors of note**

**Safety in school and on the way to school:** While the proportion of girls reporting unsafe journeys to school remains small, it has increased from 1% to 6% among ABE respondents – a particularly important change as this group was already formed by adolescents at the baseline. The result suggests that the security conditions faced by the girls have indeed deteriorated, and not just as a function of the additional harassment faced by girls as they become older. The increased insecurity in Banaadir may have a substantial impact on future program outcomes, particularly considering the higher proportion of targeted girls living in that area. Interestingly, however, the proportion of caregivers reporting that school is unsafe declined sharply among those who have children with disabilities and those who are Af-Maay speakers, indicating that while the overall security situation has deteriorated, safety at school has improved for some disadvantaged groups, reflecting the program's investment in sensitisation of CECs and GEFs. The proportion of occupational minority caregivers reporting mistreatment of girls at school has increased since the baseline, however, in contrast with the sharp decline of Af-Maay speaking caregivers reporting the same. The finding shows that additional oversight of schools will be necessary to address teacher behaviour towards minority children. Despite this finding, however, the vast majority of the respondents

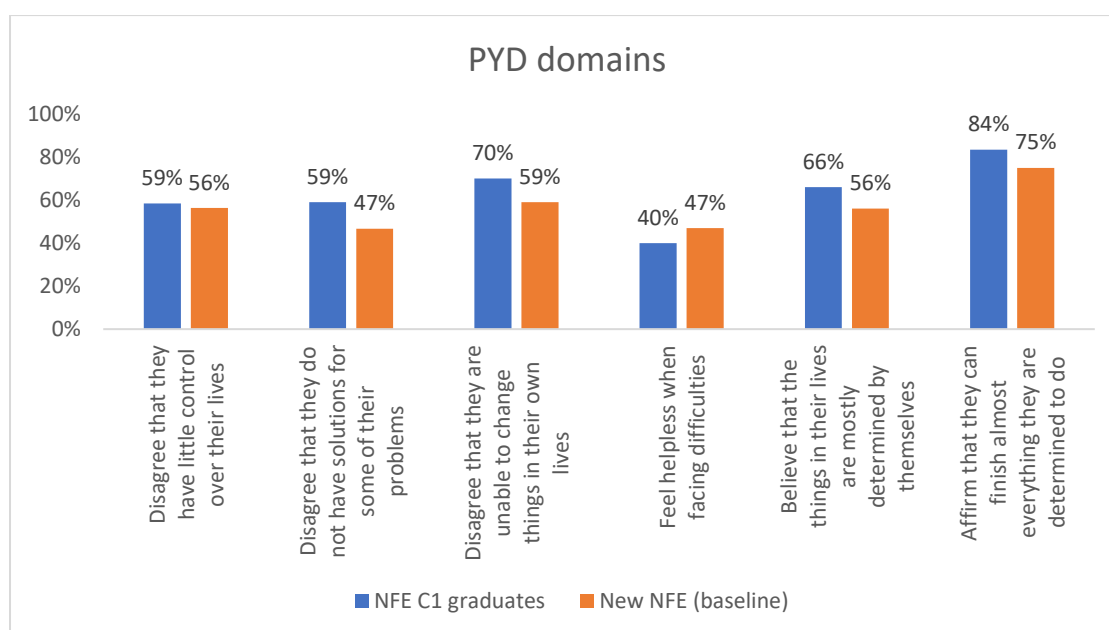
affirmed having access to protection services at school, with teachers and head teachers being primary points to report cases of abuse.

**Minority status:** While girls from an ethnic/occupational minority background had more limited gains in literacy, the difference is no longer statistically significant, suggesting that the learning gap for this group is starting to close. Minority language speakers actually had marginally higher gains in literacy, compared to the average, although the difference is again non-significant. On the other hand, there is a significantly lower gain in numeracy for minority language speakers, indicating that despite the progress made, language-related learning gaps remain in this area.

**VSLAs:** While the program has formed a relatively large number of groups to date, there is a clear need to further expand the VSLA footprint among parents. The limited proportion of the midline sample participating in VSLAs (8.8%) may indicate that the distribution of the groups is not consistent across geographies. Adult groups have mostly remained active (83%) while only 36% of the adolescent groups remain so. While adolescent groups are more vulnerable to shocks – given the limited proportion of girls who are employed/ have an income – the result indicates the need to strengthen the approach for youth VSLAs.

**Phone/ smartphone ownership:** Phone coverage and ownership of devices have increased systematically in Somalia over time, but progress has been limited among the AGES population since 2020, particularly in the case of older girls. Considering the widespread use of mobile cash in country and the social desirability of owning a smartphone among youth, this is a surprising finding, even when taking into consideration the economic crisis triggered by COVID and drought. The finding has implications for programming, highlighting the continued need for blended learning approaches as well as the risk of exclusion during phone-based surveys.

**Positive Youth Development:** While only baseline data has been collected at this point, interesting results have been observed among NFE graduates in the scale used to measure PYD. While the new NFE cohort and the C1 NFE graduates are not directly comparable groups due to age and location differences, the NFE graduates have substantially different, and more positive, results than the new NFE baseline, suggesting that the program is indeed boosting self-efficacy, as seen in the figure below.



### Logframe – Proposed Changes



The indicators and targets remain largely valid, with the exception of the increase in the proportion of girls reaching the expected YLI score (70). Given the low baseline and the ongoing crises, it is likely that the project will need to reduce the target for this indicator.

The NFE students' reported income is surprisingly high, suggesting that it may represent the *household* income and not the girls' own earnings. The results require additional validation. If the data indeed refers to the household income, it may hinder the project's ability to demonstrate impact in this area.

## **Response to Recommendations**

### *Geographic Anomalies and Follow-Up*

*Many of the trends tracked since baseline have been similar across locations and even across cohorts. Even where improvements have been modest – such as in the development of girls' leadership skills or YLI scores – this suggests relatively uniform application of the programme's interventions. Where this has not been the case, additional attention should be given to assessing the divergence in programme outcomes. The most prominent case in this round is the performance of NFE centres and the C1 NFE cohort in Jubaland, the latter of which experienced a sharp decline in numeracy since baseline and no gain in literacy, compared to substantial gains in every other cohort and state. This is coupled with other problematic signals, such as a massive increase in absenteeism among Jubaland's teachers that is not matched in other locations. While the evaluation cannot definitively pinpoint the reason for divergent geographic trends, it can and should be used to guide further investigation by programme staff into the context in, for instance, Jubaland's NFE centres, to determine whether the poor outcomes observed are a function of exogenous factors – such as the worsening drought – or internal factors within the centres.*

The project agrees with the recommendation. As noted above, the trends in Jubaland indicate the need for additional oversight of schools and strengthening teacher training/ coaching.

### *Classroom Disciplinary Practices*

*In line with the more general recommendation above, we recommend focused targeting of particular areas and schools for additional training and sensitisation to reduce the use of corporal punishment and other negative disciplinary practices. The use of corporal punishment has declined significantly in formal schools and ABE centres tracked since the baseline; however, corporal punishment remains relatively common in Jubaland's formal schools and ABE centres and – even more so – in the new NFE centres (C4) brought into the programme in Hirshabelle. This suggests the possibility of tailored interventions, with particular attention to basic behaviour change on disciplinary practices focused on Jubaland and Hirshabelle centres.*

*Tailoring of this kind would also benefit the schools and centres where corporal punishment has already declined dramatically. In these schools, more nuanced forms of still-negative practices persist, particularly the use of harsh language when interacting with students and punishing incorrect answers in class, which can be a major deterrent to active student participation and classroom inclusivity. By shifting attention in these schools and centres away from a focus on stamping out corporal punishment, additional training time and monitoring efforts could be spent on reducing these more nuanced – but, nonetheless, problematic – practices.*

The project agrees with the recommendation. While strong gains have been made in reducing corporal punishment, the practice persists among a minority of teachers and facilitators, and the use of discipline against students providing incorrect answers is an issue with the new NFE teachers. The findings indicate the need to strengthen oversight / coaching of facilitators and increase awareness of and engagement of CECs in addressing safeguarding issues.

### *Poor Menstrual Hygiene with Cost-Effective Avenues for Improvement*

*Section 10 of the report documented the relatively poor menstrual hygiene practices among girls in the sample, particularly in the form of insufficient washing of reusable pads (i.e. washing reusable pads without using soap) and infrequent changing of pads, both of which can lead to infection. Similarly, a significant minority of girls indicate*

*that they occasionally have to reuse a pad before it is fully dry from being washed. These issues have fairly straightforward and cost-effective remedies, however, when we consider how many girls are already comfortable washing and reusing pads. By providing a small number of supplemental reusable pads to girls – as opposed to entire sanitary kits, with their greater costs – girls would be able to wait longer for pads to dry and change their pads more frequently. Combined with the provision of soap, a relatively modest investment could significantly improve girls' menstrual health.*

The project agrees that menstrual hygiene management is a major issue in Somalia and that under the current crisis, the provision of reusable pads could be a necessary solution. That being said, it is also important to acknowledge that menstrual health management (MHM) is a much more complex issue in South Somalia than assumed in the recommendation. Reusable pads require water for washing, and the limited access to water may result in additional issues. The high prevalence of type III FGM also affects MHM practices. In a different project, CARE observed limited use of the sanitary pads provided and limited impact of the activity, potentially due to cultural perceptions as well as the practical issues posed by type III FGM. The provision of handouts, as suggested here, can be implemented by the project as a one-off activity, but remains a temporary, non-sustainable solution. The project is currently working with GEFs to form youth VSLAs (to boost income and the ability to purchase sanitary pads, among other essential items) and in parallel, to increase access to information on MHM through linkages with community health workers.

#### *Changing Perceptions of the Value of Education*

*Caregivers in much of the sample tracked since baseline report marginally more positive views of the value of girls' education. Specifically, when asked whether girls' education is a worthwhile investment of scarce resources, positive responses increased slightly in Banadir and South West State. In contrast, the notion that girls' education is a worthwhile investment fell sharply in Jubaland, which may reflect both shifting attitudes and a sense that the poor labour market or other issues reduce the utility of girls' education. To the extent possible, attitudinal change programming in Jubaland should continue work with religious leaders and community leaders while also focusing attention on the economic and non-economic benefits of girls' education. The latter could be accomplished by hosting female business owners, NGO employees, or government staff, to provide girls and their parents with a visible example of the benefits of their education.*

The project agrees with the recommendation.

#### *Coordinate with, and Demonstrate Targeting to, Other Programmes*

*AGES has done an admirable job of targeting ultra-marginalised girls in the communities where it works. For instance, in the overall midline sample, just 34.8 percent of beneficiaries hail from the locally dominant clan, and a full 42.3 percent are from minority groups. Compared to many other programme evaluations we have performed in Somalia, AGES outperforms nearly all of them in terms of targeting – a majority of beneficiaries in most other programmes are drawn from dominant clans and overwhelmingly exclude the most marginalised. This argument is supported by the fact that marginalised girls in the AGES sample are less likely to have benefitted from other humanitarian interventions, despite being drawn from objectively poor households.*

*We would recommend that AGES work with other NGOs in Banadir, Hirshabelle, Jubaland, and SWS to illustrate how their own targeting could be improved and to highlight the overrepresentation of dominant groups and – more generally – the “non-marginalised” – that is often present in other programmes. By opening discussions of this kind and demonstrating how more effective targeting can be completed cost-effectively, AGES could provide important spillover benefits, expanding its impact among marginalised communities by helping other NGOs broaden their own beneficiary base.*

The project agrees with the recommendation. Since its baseline, the project has disseminated its findings and approaches widely, in particular regarding the marginalisation of minorities and minority language speakers, including with the MOECHE and State MOEs as well as the Education Sector Committee, and partner NGOs.

### *Promote a Role for Girls in Local Governance*

*While girls' participation in the public sphere, generally, is limited, this evaluation found that girls were especially unlikely to participate in discussions and forums related to local governance. Even among adult women in our sample – those NFE girls 18 years and above – just 11.0 percent have participated in public discussions regarding local governance. Increasing this participation will require addressing four types of barriers described by girls in qualitative data: a lack of motivation on their part, a lack of caregiver support, economic barriers to participation, and limited space in a political system dominated by elders, traditional leaders, and older men. If girls are to have a more active role in governance, it is not sufficient to promote the importance of their participation to them; rather, spaces need to be explicitly opened for their participation, with incentives provided to NGOs and government bodies to more fully incorporate female voices in meetings and public fora.*

The project agrees with the recommendation, which is already part of its design and one of its intended intermediate outcomes.

### *Promote Inclusion by Recognising Bias Openly*

*Given the programme's focus on inclusivity and promoting enrolment among marginalised girls, the issues of bias, discrimination, and special needs for ensuring inclusion were a frequent theme in surveys and qualitative interviews. While teachers, CEC members, and community members generally recognise the teasing, stigma, and outright discrimination faced by girls with disabilities, there is very little recognition or openness about discrimination against groups based on their social identities. Indeed, the majority of CEC members and teachers – to provide one example – deny that ethnic minorities face any mistreatment in schools at all.*

*Denial of the problem, however, forecloses the possibility of improving the situation. School staff typically argue that the school treats all students equally, which prevents them from seeing that mistreatment at the hands of other students and even subtle biases among teachers can have important, deleterious impacts on marginalised girls, especially when those actions are mirrored by widely-known but seldom-acknowledged bias in society writ large. The programme should focus attitudinal change efforts on helping school staff to recognise the bias that exists against minority students. While other marginalised groups, such as girls with disabilities, face massive challenges, their challenges are often recognised by school staff and community members and are difficult to remedy due to resource limitations. By focusing any interventions related to attitudinal change on bias against ethnic minorities and marginalised clans, the programme would not be detracting from efforts to help GWDs; rather, they would be focusing one aspect of programming on a group (minority students) who could benefit from attitudinal change and acknowledgment of discrimination, while reserving any available financial resources for reducing practical barriers to enrolment and education for GWDs.*

The project agrees with the recommendation. This has in fact been one of its key areas of work, first through deconstructing narratives of non-marginalisation among CARE/ partners' own staff, and then through a substantial investment in coaching school staff and local authorities on the disparities of treatment and conditions faced by minorities. The project will continue to evolve and adapt its approaches in social norm change.

### *Develop Specific Protection Mechanisms within Schools and Centres*

*The girls interviewed in this round overwhelmingly indicate that they could report cases of abuse, harassment, or exploitation occurring at school to a member of the teaching staff. They are similarly confident that they could report such issues occurring outside of school to teachers, community leaders, or their parents, among others. However, a wealth of research on sexual and gender-based violence and access to justice in Somalia indicates that women are widely hesitant to report instances of sexual assault and that they lack confidence a report would be handled either confidentially or competently. This suggests that AGES beneficiaries may be too sanguine about their ability to report or that the survey question did not adequately capture the potential for hesitation or concern around reporting.*

*In general, the idea that girls could report instances of abuse to teaching staff is problematic, not least because many schools lack any female staff members and staff members themselves may be perpetrators of abuse. In our view, schools*

*should have an established focal point for child protection and abuse allegations, even if that focal point is a female member of the CEC, rather than a teacher. An established and advertised channel for reporting is essential to increase confidence when girls are actually faced with this dilemma.*

The project agrees with the recommendation, and this is an already existing intervention.