

Project Evaluation Report

Report title:	Baseline Evaluation
Evaluator:	Consilient
GEC Project:	Adolescent Girls' Education in Somalia (AGES)
Country	Somalia
GEC window	Leave No Girl Behind
Evaluation point:	Baseline
Report date:	June 2020

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.

Leave No Girl Behind

Adolescent Girls' Education in Somalia (AGES)

Baseline Evaluation

Prepared on Behalf of:



**Girls'
Education
Challenge**



Zuzana Machova
Juuso Miettunen, Ph.D.
Brenton D. Peterson
Consilient Research

June 11, 2020
Version 4

ABBREVIATIONS

ABE	Accelerated Basic Education
AGES	Adolescent Girls Education in Somalia
BDR	Banadir
CEC	Community Education Committee
EGRA	Early Grade Reading Assessment
EGMA	Early Grade Math assessment
FGS	Federal Government of Somalia
FMS	Federal Member State
FGD	Focus Groups Discussion/s
FE	Formal Education
FM	Fund Manager
GEC	Girls Education Challenge
GEC-T	Girls Education Challenge - Transition
GEF	Girls Empowerment Forum
GwD	Girls with Disabilities
JSS	Jubaland
LNGB	Leave No Girl Behind
MOE	Ministry of Education
NGOs	Non-Governmental Organization
NFE	Non-formal Education (Life skills track)
PSUs	Primary Sampling Units
PPS	Probability Proportionate to Size
SOMGEP	Somali Girls' Education Promotion Project
SOMGEP-T	Somali Girls' Education Promotion Project -Transition
ToC	Theory of Change
DFID	UK's Department for International Development
SLG	Savings and Loans Groups
SWS	South West State
WHO	World Health Organization
YLI	Youth Leadership Index

TABLE OF CONTENTS

ABBREVIATIONS.....	2
TABLE OF CONTENTS	3
LIST OF TABLES.....	5
LIST OF FIGURES	8
EXECUTIVE SUMMARY.....	9
1. INTRODUCTION	14
1.1 PROJECT CONTEXT	14
1.2 TARGET BENEFICIARIES.....	18
1.3 THEORY OF CHANGE	24
2. METHODOLOGY.....	26
2.1 EVALUATION QUESTIONS.....	26
2.2 EVALUATION DESIGN	31
2.3 QUANTITATIVE METHODOLOGY	36
2.4 QUALITATIVE METHODOLOGY	41
2.5 FIELDWORK, QUALITY ASSURANCE AND ANALYSIS	43
2.6 CHALLENGES & LIMITATIONS.....	44
3. DESCRIPTION OF BASELINE SAMPLE	51
3.1 COHORT DEMOGRAPHICS	51
3.2 REPRESENTATIVENESS OF THE SAMPLE	58
3.3 EDUCATIONAL MARGINALISATION.....	63
3.4 INTERSECTION BETWEEN KEY CHARACTERISTICS SUBGROUPS AND BARRIERS.....	67
3.5 PROFILES OF COHORT GIRLS.....	71
3.6 APPROPRIATENESS OF PROJECT ACTIVITIES.....	73
4. LEARNING OUTCOME	80
4.1 ASSESSMENT DESIGN AND VALIDITY.....	80
4.2 AGGREGATE LEARNING OUTCOMES	92
4.3 SUB-GROUP LEARNING OUTCOMES	105
5. TRANSITION OUTCOME	120
AGGREGATE TRANSITION OUTCOMES	122
6. SUSTAINABILITY OUTCOME	131
6.1 INDICATOR 3: ECONOMIC STATUS OF PARENTS.....	131
6.2 INDICATOR 4: GIRLS WITH DISABILITIES.....	132
7. INTERMEDIATE OUTCOMES	135
7.1 ATTENDANCE	135
7.2 TEACHING QUALITY AND PRACTICES	156
7.3 YOUTH LEADERSHIP INDEX.....	168
7.4 SCHOOL MANAGEMENT AND GOVERNANCE	180
7.5 COMMUNITY ATTITUDES	188
8. AUXILIARY ANALYSIS	203
8.1 MENSTRUAL HEALTH AND HYGIENE.....	203
8.2 INTERMEDIATE OUTCOMES AS PREDICTORS OF LEARNING	206
9. MEASUREMENT AND INDICATOR DEVELOPMENT	216
10. CONCLUSIONS AND RECOMMENDATIONS.....	220
10.1 CONCLUSIONS	220
10.2 EVALUATION RECOMMENDATIONS	223

10.3 PROGRAMMING RECOMMENDATIONS.....	225
ANNEXES	229
ANNEX 1 – PROJECT LOGFRAME	229
ANNEX 2 – COHORT EVALUATION APPROACH	229
ANNEX 3 – BENEFICIARIES TABLE FROM SAMPLE DATA	229
ANNEX 4 – BENEFICIARIES TABLE FROM PROJECT MAPPING DATA	232
ANNEX 5 – MEL FRAMEWORK.....	237
ANNEX 6 – EVALUATION INCEPTION REPORT.....	237
ANNEX 7 – BASELINE DATA COLLECTION TOOLS	237
ANNEX 8 – DATASETS, CODEBOOKS, AND PROGRAMS	237
ANNEX 9 – LEARNING TEST PILOT AND CALIBRATION.....	240
ANNEX 10 – SAMPLING FRAMEWORK.....	243
ANNEX 13 – SAMPLING REVISIONS AND STATISTICAL POWER.....	245
ANNEX 14 – CEILING EFFECTS SIMULATION	247
ANNEX 15 – MEASUREMENT OF DISABILITY STATUS.....	250
ANNEX 16 – EVALUATION METHODOLOGY.....	252
ANNEX 17 – WEIGHTING AND SAMPLE CHARACTERISTICS.....	263

LIST OF TABLES

Table 1: Summary of direct beneficiaries from project enrolment data	20
Table 2 - Number of GwDs in cohort 1, disaggregated by type of disability	21
Table 3: Proposed intervention pathways.....	22
Table 4: indirect beneficiary groups	24
Table 5: Evaluation Questions and data sources	27
Table 6: Benchmarking for learning.....	34
Table 7: Transition benchmark age distribution	34
Table 8: Target populations and second-stage sample selection	38
Table 9: Target and achieved samples, by quantitative tool	40
Table 10: Geographic distribution of qualitative interviews	42
Table 11: Demographic composition of cohort sample.....	52
Table 12: Disability prevalence among cohort girls, across alternative coding schemes	56
Table 13: Disability rates reported by girls versus their caregivers.....	57
Table 14: Disability prevalence among formal school girls in cohort sample and sample frame	60
Table 15: Geographic composition of the sample versus the sample frame.....	62
Table 16: Characteristic subgroups of marginalized girls	64
Table 17: Main barriers faced by marginalized girls.....	65
Table 18: Key barriers to education by characteristic subgroups of marginalized girls	69
Table 19: Profiles of cohort girls	72
Table 20: Distribution of assessment scores by program track.....	81
Table 21: Distribution of numeracy and reading scores by age	82
Table 22: Item discrimination analysis.....	88
Table 23: Distribution of numeracy and literacy scores by sub-task	88
Table 24: Number of institution types in the sample, by state	92
Table 25: Aggregate learning outcomes by learning track	92
Table 26: Alternative scoring of numeracy, by learning track	93
Table 27: Foundational gaps in numeracy, FE girls	98
Table 28: Foundational gaps in literacy, FE girls	99
Table 29: Foundational gaps in numeracy, ABE girls	100
Table 30: Foundational gaps in literacy, ABE girls	100
Table 31: Foundational gaps in numeracy, NFE girls	101
Table 32: Foundational gaps in literacy, NFE girls	102
Table 33: Share of girls achieving lower primary competency, by respondent type	104
Table 34: Analysis of learning scores by disability type.....	107
Table 35: Distribution of learning scores by household and girl characteristics.....	109
Table 36: Distribution of learning scores by household educational characteristics	110
Table 37: Distribution of learning scores by household educational characteristics	111
Table 38: Distribution of learning scores by marginalized groups.....	114
Table 39: Distribution of learning scores by school resources, infrastructure and safety sub-groups ..	115

Table 40: Distribution of learning scores by formal school characteristics.....	117
Table 41: Transition outcomes.....	122
Table 42: Transition outcomes by subgroup	124
Table 43: Sustainability indicator 4.....	133
Table 44: Sustainability indicators	135
Table 45: distribution of the attendance rates by cohort group	141
Table 46: Attendance regression analysis per cohort group	142
Table 47: Attendance source by cohort group	144
Table 48: Status of attendance records when you arrive	146
Table 49: Completeness of the overall attendance records	146
Table 50: Neatness and organization of the overall attendance records	146
Table 51: Overall attendance and sub-group analysis.....	150
Table 52: formal school-specific sub-groups and the attendance scores	152
Table 53: ABE and NFE specific subgroups and attendance scores	153
Table 54: Multivariate regression analysis	155
Table 55: Teacher professionalism by educational track	157
Table 56: Classroom gender equality	160
Table 57: Gender equity by program track	161
Table 58: Disciplinary practices	162
Table 59: Disciplinary practices by program track	163
Table 60: List of question used for the analysis of pedagogical practices	164
Table 61: Use of formative assessments	164
Table 62: Pedagogical practices.....	165
Table 63: Pedagogical practices by program track	166
Table 64: list of YLI questions.	168
Table 65: YLI distribution, by cohort group	171
Table 66: sub-group analysis of the girls' characteristics	172
Table 67: sub-group analysis of the household's characteristics.....	173
Table 68: Sub-group analysis of parental attitudes	174
Table 69: sub-group analysis of the learning environment.....	175
Table 70: sub-group analysis of the disability.....	176
Table 71: Multivariate regression analysis of sub-groups and YLI	177
Table 72: Reasons for not enrolling girls to schools.....	184
Table 73: Caregiver believes girls' education worthwhile, by cohort group.....	188
Table 74: Caregiver believes girls' education worthwhile, by subgroup	190
Table 76: Caregiver attitudes toward cost of education, by cohort group	191
Table 76: Caregiver attitudes toward cost of education, by sub-group.....	193
Table 78: Caregiver aspirations for girl, by cohort group	195
Table 78: Caregiver aspirations for girl, by subgroup	196
Table 79: Caregiver attitudes toward domestic work, by cohort group	197
Table 80: Caregiver attitudes toward domestic work, by subgroup	199

Table 82: Sources of menstruation related information	206
Table 83: Analysis of learning scores and community attitudes	207
Table 83: Analysis of learning outcomes and teaching quality – formal school girls	210
Table 84: Analysis of learning outcomes and teaching quality – ABE and NFE girls	211
Table 85: Analysis of learning scores and school governance.....	212
Table 86: Characteristic subgroups and barriers of sample for portfolio level aggregation and analysis	230
Table 87: Direct beneficiaries, by age	232
Table 88: Target groups by out-of-school status.....	232
Table 89: Direct beneficiaries by dropout grade.....	233
Table 90: Other selection criteria	233
Table 91: Other beneficiaries	234
Table 92: Key variables for learning and transition analysis, household survey	237
Table 93: Key variables in other evaluation datasets	239
Table 94: Statistical power comparisons between target and achieved sample sizes	246
Table 95: Number of replacements by case	253
Table 96: Qualitative tool development	255

LIST OF FIGURES

Figure 1: Project Theory of Change.....	24
Figure 2: Sample distribution by age.....	54
Figure 4: Sample and sample frame characteristics.....	59
Figure 5: Distribution of learning scores by age.....	83
Figure 6: Scatterplot of reading and numeracy scores.....	85
Figure 7: Distribution of learning scores.....	86
Figure 8: Scores distribution by enumerator.....	91
Figure 9: Distribution of learning scores.....	94
Figure 10: Numeracy sub-task scores.....	96
Figure 11: Literacy sub-task scores.....	97
Figure 12: Mean learning scores by program track and region.....	105
Figure 13: Subgroup effect on mean aggregate learning outcomes.....	119
Figure 13: Distribution of enrolment numbers per cohort group.....	138
Figure 14: Distribution of the overall attendance rates.....	140
Figure 15: Distribution of the attendance rates by cohort group.....	141
Figure 16: Comparison of attendance rates.....	143
Figure 17: Comparison of attendance rates by cohort group.....	144
Figure 18: Comparison of yesterday's and today's attendance rates.....	145
Figure 19: YLI distribution – kernel density plot, by cohort group.....	170
Figure 22: Menstrual hygiene practices.....	204
Figure 21: YLI scores and learning outcomes.....	214
Figure 22: YLI scores and learning outcomes - lowess.....	215
Figure 23: Simulated learning effect sizes, with ceiling effects.....	248
Figure 26: Weighting and sample characteristics.....	264

EXECUTIVE SUMMARY

Project Background

The Adolescent Girls Education in Somalia (AGES) Program is being implemented in southern and south-central Somalia. Implementation began in September 2018 with the mobilisation of out-of-school girls, and will continue through August 31, 2022. The program's geographic extent covers the portions of Somalia that have been most impacted by the three decades of conflict that began in the late 1980s: areas in which pitched battles occurred, which underwent multiple rounds of displacement and "clan cleansing", which were flooded by IDPs fleeing violence elsewhere, and/or which have not been under government control for significant periods. Portions of the implementation area are also home to Somalia's most marginalised clan groups, and home to many speakers of af-Maay, the less common dialect of Somali, which puts them at a disadvantage when learning in the official language of af-Mahatiri.

The project targets 42,000 mostly out-of-school girls who face multiple, overlapping barriers to enrolment and learning, supporting enrolment into one of three tracks: Alternative Basic Education (ABE), Non-Formal Education (NFE) and formal education (FE) in primary schools. The program covers ages 10-19, with each track associated with its own age group. The first cohort enrolled 6,623 girls into FE, 7,241 girls into ABE, and 6,604 into NFE, respectively, reaching a total of 20,468 girls; the remainder up to 42,000 will be reached in a second enrolment period.

The program's objectives are to improve traditional learning outcomes and critical life skills around finances and menstrual hygiene management; prompt girls to stay in school, re-enrol in school, enrol in alternative education, or move into meaningful and dignified employment opportunities. Given the context and the barriers girls in the targeted areas face, the program takes a multidimensional approach while focusing efforts on these key barriers and the most marginalized girls. At an individual and household level, the program will sponsor Village Savings and Loans (VSL) groups, provide bursaries to girls, provide them with life skills and other training through Girls' Empowerment Forums (GEFs), mentor boys through parallel Boys' Empowerment Forums, and sensitize mothers to pro-education norms, especially around disability. At a community level, religious leaders will be mobilized for education and trained on Islamic principles that emphasize gender equality and girls' education, deliver pro-education messaging through radio and social media,

The program will train teachers and facilitators in all three program tracks, improve school infrastructure – to make it, among other things, more accessible for children with disabilities – and provide access to a mobile learning platform that will allow children who are often absent or tardy to practice key skills outside of instructional hours. Finally, at the system level, the program will develop simplified quality assurance processes around inclusive education, for use by MOE officials, work with learning institutions to develop curricular content and better school governance practices, build the capacity of MOE and other officials in terms of inclusive and special needs education, and support the review and development of government policies around ABE institutions, and inclusive and special needs education.

The three intervention tracks differ in terms of the ages targeted, the desired transition pathway and learning outcome, and the specific interventions. Girls enrolled in FE are the youngest cohort and are expected to continue through formal schooling, while ABE girls are older and the goal is to achieve Grade 4 numeracy before a portion transition into formal schools. In contrast, NFE girls are the oldest girls and the goal is to achieve functional literacy and numeracy, without a specific expectation that they will transition into further education, given their age.

Many of the program interventions are similar across tracks, including teacher training, community sensitization efforts, and the provision of bursary support and sponsorship of VSLs. Where differences arise is in the curriculum of the various learning institutions – with compressed and finance- and business-

oriented content emphasized in ABE centres – the targeting of life skills training to older girls, and the extensive support the program will provide to develop an ABE policy framework.

As the set of interventions makes clear, and consistent with most GEC and GEC-T programming, AGES views shifts in educational outcomes as most likely to occur from the confluence of multiple factors. For instance, by improving attendance, girls’ empowerment, and teaching quality, literacy and basic numeracy skills will improve. To achieve these intermediate outcomes, the program includes interventions that should improve the uptake of new skills (e.g., teacher training) and interventions that target specific barriers to attendance, such as poor sanitation facilities (improvement of girl-friendly toilets) and household workload (via community sensitization). Similarly, transition is targeted in multiple, overlapping ways: first and foremost, by providing multiple avenues for enrolment, well-suited to girls at different life stages and in varied circumstances (e.g., married, mothers); second, by inculcating pro-education and gender equality norms in key decisionmakers, such as mothers, and community influencers, such as religious leaders; and by addressing economic barriers to enrolment via VSL sponsorship and provision of bursaries. These do not represent all of the interventions, nor the full logic, of the program; however, the discussion here illustrates the extent to which the program is coupling broad interventions with highly targeted efforts to ease specific barriers known to exist in this context.

Evaluation Approach

The evaluation employs a pre-post longitudinal design, in which a cohort of girls will be tracked over time. This baseline evaluation took place beginning in November 2019, and follow-up rounds will occur in spring 2021 (midline) and 2022 (endline). Learning and transition outcomes will each be tracked for their respective cohorts of girls, and comparisons will be made to benchmarks for each outcome, in lieu of a comparison group of communities that do not receive the intervention. Data collection is mixed-methods: quantitative tools include classroom headcounts, classroom observations, household surveys, learning assessments, surveys with teachers, and surveys with head teachers; qualitative tools include FGDs with Community Education Committee (CEC) members, teachers, mothers, as well as participatory FGDs – including risk mapping – with girls. Each of the program’s intermediate outcomes – attendance, teaching quality, school management and governance, community attitudes, and life skills – are captured in the quantitative data and supplemented or triangulated using qualitative interviews. As the recommendations of this report note, there is room for improvement in the operationalization of intermediate outcomes and sustainability indicators.

Educational Marginalisation, GESI and Barriers to Education

AGES beneficiaries are among the most marginalised girls in Somalia. In the Somali context, clan membership and language spoken are significant markers of marginalisation, and 52.2 percent of AGES beneficiaries belong to either a marginalised agricultural clan (Digil-Mirifle), a minority group (“Bantu”), or occupational minority. Over one-quarter of girls speak af-Maay, putting them at a disadvantage in a school system built around af-Mahatiri instructional materials. A number of girls face personal sources of marginalisation – 7.0 percent have been married, and 5.1 percent already have a child. And, while disability prevalence rates are not extremely high overall, a considerable number of girls – 11.4 percent – experience anxiety or depression daily or weekly, and over one-quarter of sampled girls experience anxious feelings or extreme sadness monthly.

The most compelling aspect of beneficiaries’ marginalisation is possibly the low value placed on it by their families and the limited experience their families have of formal education. The vast majority (78.6 percent) of girls live in households where neither their household head, nor their caregiver, has never attended a formal school of any kind. In terms of prioritisation of schooling, 40.1 percent of caregivers believe household chores are an acceptable reason to keep their daughter out of school, and a majority of girls in the sample dedicate a few hours or more per day to chores. Combined with household financial limitations, the prevalence of school fees, the views of caregivers – most of whom say they do not have

enough money to pay for schooling, and that this is a legitimate reason to keep their daughter out of school – the low value placed on girls’ education is a significant problem.

Importantly, economic and household aspects of marginalisation overlap with clan-based marginalisation. For instance, marginalised clans or groups are more likely to experience hunger, and girls in minority groups spend more time on chores than their counterparts in less-marginalised clans. Girls from marginalised clans are also less likely to have parents or adults who have attended school.

From a Gender Equality and Social Inclusion (GESI) perspective, the program is fairly clearly transformative. The program’s targeting emphasizes the most marginalized girls, as discussed elsewhere in this summary, and includes a strong emphasis both on including girls with disabilities (GWDs) among its beneficiaries at an outsized rate and on changing community attitudes around disability. With regard to disability, our only recommendation is that the program generate ideas for mitigating non-attitudinal barriers to GWDs’ enrolment. With regard to gender norms, the program is not explicitly – at least at this stage – making efforts to change harmful social practices, such as early marriage. However, it is working with religious leaders to promote gender equality through arguments rooted in Islamic theology and law and actively promoting norm change in lessons delivered to girls *and* boys via the GEFs and BEFs, both of which have the potential to alter gender norms and attitudes broadly

The centrality of social inclusion is clear in the program staffing and monitoring design as well. The logframe indicators include an emphasis on GWD enrolment levels, and disaggregation by disability status and other markers of social exclusion, like language. The program intends to report on GESI standards and results quarterly, and employs a Protection, Gender, and Accountability Advisor to ensure staff capacity around these issues.

Baseline Status of Program Outcomes

The program’s three primary outcomes are learning, transition, and sustainability. Girls at baseline achieved relatively high scores – considering most had never been enrolled in school previously – but this reflects the relatively low difficulty of the assessment. For instance, the numeracy assessment did not include more difficult tasks, such as multiplication or division, which would have dragged scores down significantly. Scores are correlated with age, which results in a correlation between respondent type and performance, with older girls (NFE) scoring higher than younger girls who have been recruited into formal education. For Somali literacy, average scores ranged from 33.8 percent (FE), to 41.0 percent (ABE), up to 57.9 percent among the oldest cohort (NFE). A similar pattern emerged for numeracy: 58.8 percent among FE girls, 67.0 percent for ABE girls, up to 76.6 percent for NFE girls. The analysis also revealed clear “cutpoints” in learning around specific subtasks and the underlying skillset they require. These cutpoints can be used to target learning opportunities, but identification of this kind should also be incorporated actively into classrooms, as part of a culture of formative assessment use.¹

The program’s Theory of Change links outputs or activities to learning and transition rates via five intermediate outcomes: attendance, teaching quality, girls’ life skills, community attitudes, and school governance. At baseline, attendance rates were measured using classroom headcounts, with 86 percent of enrolled girls attending, across all three institution types. In line with the expectations laid out by the

¹ In contrast to learning, transition rates were benchmarked at baseline, but measurement was not completed among the cohort, because the cohort consists of girls who have been brought from non-enrolment into various education streams, who, therefore, have 100 percent transition rates.

program's Theory of Change, girls who are responsible for more chores at home have lower attendance rates, and girls with higher self-confidence or empowerment are more likely to attend school regularly. Importantly, girls experiencing occasional or frequent bouts of depression or anxiety attend school less often; given the number of such girls in the set of beneficiaries, and the positive relationship with self-confidence, there may be particular benefits for attendance from empowerment programming delivered through Girls' Empowerment Forums. At a broader level, targeting specific schools for additional attendance-oriented interventions could be fruitful, as *all* the classrooms with attendance below 75 percent were among less than half the set of schools visited.

Discussions with CEC members seemed to confirm their intent to support marginalized girls economically, which is a key factor in generating truly inclusive education. In practice, however, it is not clear how extensive this support actually has been. Over half of CECs promote re-enrolment of OOS children, but these activities do not strictly targeted the most marginalized girls, and explicit financial assistance is almost certainly limited, given the generalized poverty in program communities. Most communities have an active CEC, but their activity levels tend to be low – one of the most active areas of CEC engagement is re-enrolment of OOS children, but engagement is lower in most other areas where their management is needed.

Beyond these school-level intermediate outcomes, the program also seeks to improve girls' leadership skills (measured with CARE's Youth Leadership Index, YLI) and pro-education attitudes within communities. At baseline, the average girl achieved a score of 49.3 out of 100 on the leadership index, with just 12 percent of girls achieving the program's threshold for success (70 points out of 100), compared to a target of 80 percent achieving this threshold. As with learning scores, YLI scores are correlated – though less dramatically so – with age, such that NFE girls achieved the highest scores.

Finally, community attitudes were broadly in favour of girls' education, with 87.8 percent of caregivers reporting that girls' education was a worthwhile investment, even when funds are limited. However, they also hinted at attitudinal barriers to girls' education, as nearly two-thirds of caregivers indicated the cost of education was an acceptable reason to keep their daughters out of school. These findings hint at the role tangible household financial limitations are likely to play in reducing enrolment and attendance rates.

The program has also defined a number of outcomes that should promote sustainability after the conclusion of activities. At the time of the baseline, little progress had been made in most of these areas, an unsurprising outcome given that programming has just begun. The tasks selected for ensuring gains are sustained after the end of the program are generally well-suited to that task, especially those surrounding the development of government policy, coordination with “umbrella” educational associations, and institutionalization of GEFs, BEFs, and VSL groups. Where sustainability is likely to present the largest hurdles is in enrolment rates for marginalized girls after the end of bursary support; however, the program is aware of this issue, and has designed economic interventions (VSLs) that tend to be sustainable to complement the shorter-run provision of bursaries.

Conclusions

In general, it is clear that the program's design is suited for its stated objectives. Where the link between activities and outcomes is definitely weakest is in the program's emphasis on policy development around ABE and inclusive education. Our concern is that policy development may have little practical benefit on-the-ground, due to a lack of government resources, overlapping jurisdiction, and conflicting priorities between district, state, and federal actors. The proposed capacity-building of MOE official around quality assurance is appropriate, and its effect is less likely to be watered down due to resource constraints or policy disputes.

At the same time, the program is on very strong footing in terms of its other activities. The program has accurately identified some of the most marginalized young girls in Somalia, in geographic, socio-cultural, and economic terms. The design has explicitly taken into account contextual nuances regarding clan and linguistic marginalization. Moreover, the particular barriers to girls' education emphasized by the program – domestic work, poor sanitation infrastructure, lack of support from religious leaders, and household financial shortfalls – appear to be accurately identified, a view that is supported by theory, by previous educational initiatives in Somalia, and by the baseline data. The program is addressing each of these barriers, though it would benefit from seeking to address the burden of domestic work more directly. If girls' domestic workload can be reduced, alongside improvements in facilities, community attitudes, and easing of the financial burden of school attendance, the program will have made substantial progress to improving girls' education.

1. INTRODUCTION

1.1 PROJECT CONTEXT

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 with militant groups and between various militias; 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; an education system composed primarily of community and privately owned schools; and political instability.

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 localized 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. Fighting in these areas has been especially intense for at least three reasons: control over the capital implies control over any nascent central government and the potential for rents that accompanies such control; the fertile areas along the Juba and Shabelle rivers are the agricultural heartland of the country, and therefore a source of wealth for those who control it; and the states of the south have been subjected to foreign influence in the war, with the Kenyan, Ethiopian, and other governments occasionally picking favoured proxies or intervening directly with armed forces.

Arguably the most important development in the conflict in the last 15 years has been the rise of an armed Islamist insurgency in the wake of Ethiopia's invasion in late 2006 and the overthrow of the Islamic Courts Union. The multi-dimensional nature of conflict in the region has a range of consequences for project implementation and the potential for the project to potentially improve girls' education. In Mogadishu, conflict and intermittent attacks have the potential to reduce school enrolment and attendance by either forcing the closure of schools or by preventing access to schools from particular neighbourhoods, especially as schools have occasionally been targeted explicitly by militants. Beyond the insurgency, inter- and intra-clan conflict is always a threat to local stability. Finally, political divergences have implications for the institutional context of education, which we discuss in greater detail below.

The second overarching contextual factor is environmental: ongoing drought and intermittent flooding, which have devastated riverine agricultural areas. According to recent reports, a moderate to severe drought is currently impacting portions of Bay, Gedo, and Lower Juba districts, while all three districts were also heavily impacted by flooding in October and November of 2019.² 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.³

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

² Food Security and Nutrition Analysis Unit, Somalia. 2019. "Quarterly Brief – Focus on 2019 Post-Deyr Season Early Warning." Available at: <http://www.fsnau.org/downloads/FSNAU-Quarterly-Brief-December-2019.pdf>.

³ Ibid.

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, 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 10.1 percent of households – are less locally numerous marginalized groups, such as the Bantu, an all-encompassing term for agricultural peoples of broad Bantu lineage who occupy farmland in riverine areas.⁴ Both groups have, to different extents, been marginalized in recent Somali history: the Digil-Mirifle are arguably the weakest of the four major clan-families and have been subject to land expropriation in times of conflict; the Bantu and other "minorities" exist more clearly on the margins of Somali society, as they fall – more or less – outside the traditional clan governance structure. In the regions evaluated here, the Digil-Mirifle are estimated to make up a majority or large minority of the population, and minority groups make up a non-trivial minority as well.⁵ The latter groups, in particular, face systematic discrimination and are poorly represented in local, state, and federal government, with downstream impacts on their access to government services.

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, beyond even the potential discrimination they face as a function of their lineage, 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

⁴ South West State is a 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.

⁵ Among the targeted locations, the Digil-Mirifle are especially numerous in the central portion of South West State, running from approximately Bardhere in the west to Dinsoor and onward to Baidoa.

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, and Mogadishu 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 organizations 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, educating far more students annually than the central government. The umbrella associations are a force for quality education, but their existence can undermine central government directives when it comes to curriculum, school monitoring, and educational policy more broadly.

In this context, the FGS has recently unveiled an official curriculum for primary education nationwide, which is the first revision of Somalia's curriculum since 1996. At the time of data collection for this baseline, instructional materials were being distributed to schools in line with the new curriculum.

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 marginalized groups, who may face a different set of barriers to educational attainment than those from Somalia's largest clan families. Fourth, while the federal government has increasingly asserted control over education administration and has developed a unified curriculum, the institutional environment in which education – and, more broadly, government in general – is situated makes administration more difficult, given the important, long-running role of private umbrella associations in the Somali educational sector. 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.

Increasing access to primary education is a major priority for the Ministry of Education, Culture and Higher Education (MOECHE) as the Gross Enrolment Rate (GER) for South Central Somalia was estimated to be as low as 22% in 2015-16⁶. In this context, access to education is marked by equity issues, with girls (20%), IDPs (17%), pastoralists (3%) and rural populations (4%) being disadvantaged⁷. Barriers to enrolment include poverty and the inability to pay school fees, as 93% of the students are enrolled in non-public schools; the opportunity cost of education for parents, given children's contribution to family income and household chores/ pastoral work; traditional gender norms restricting girls' education and prioritizing chores over school attendance; perceptions about secular education and threats from militant groups against it⁸; and the historic marginalization of subgroups of the population, including children with disabilities and those from marginalised clans and ethnic groups⁹; limited offer of education services for the most marginalised (including flexible schedule classes, accelerated education and special needs education); and limited mobility, particularly in heavily insecure areas and for girls. Poor school infrastructure further limits access, particularly for girls: 56% of the schools in South Central Somalia are

⁶ Federal Government of Somalia, Ministry of Education, Culture and Higher Education (2017) Education Sector Analysis 2012-2016, p.114

⁷ Ibid, p. 61-63; p.287

⁸ CARE (2014) *Waxbar Carurtaada* Project – Out of School Children Assessment Report, p.34

⁹ UNICEF (2016) Situation Analysis of Children in Somalia, p.16

considered to be in poor condition, 63% lack toilets and 39% have no access to water¹⁰. Inequities in access are likely to have increased due to recent shocks such as major population displacement and loss of livelihoods as a result of drought and the escalation of violence (estimated to have displaced 320,000 individuals in 2018¹¹). Additionally, Somali parents often prioritize Quranic education over formal education. This pattern results in most Somali children not starting formal primary education until completing Quranic school and therefore in late enrolment at age 8-9¹².

Official examination data was considered as unreliable until 2016¹³, although examinations were conducted in 2019. Funding for examinations remains a challenge. The first national Early Grade Reading Assessment is planned to take place in 2020. Therefore, data on learning outcomes is limited to project findings at the moment, and therefore not representative of the population in South Central Somalia. Data from a longitudinal study conducted in Puntland, Galmudug and Somaliland indicate that female students in Grade 3 could read, in average, 26 words per minute¹⁴, and reached an average reading comprehension score of 41%. Sixteen percent of the Grade 3 girls and 10% of the boys could not read a single word. Grade 3 female students reached average scores of 34% and 26% in two-digit addition and subtraction. Among students in Grades 3-6, 37% of the boys and 48% of the girls could not perform one-digit multiplication; 62% of the boys and 69% of the girls could not perform one-digit divisions¹⁵. Girls, children with disabilities and pastoralist children had lower scores in EGRA and EGMA tasks¹⁶. A regression analysis indicated that poor teaching skills, lack of learning materials, limited time for reading, lack of caregiver engagement in education and low attendance rates are predictors of poor learning outcomes in Somalia¹⁷. While no data is available, it is likely that language differences may also influence learning outcomes, particularly for Af-Maay speakers educated in Af-Mahatiri (standard Somali).

The combination of late enrolment, dropout and limited learning outcomes means that a large proportion of the students are likely to leave school before acquiring basic literacy and numeracy skills. In Jubaland and Southwest, about half of the students drop out before reaching Grade 5; survival rates are lower for

¹⁰ Federal Government of Somalia, Ministry of Education, Culture and Higher Education (2017) Education Sector Analysis 2012-2016, p.134-138

¹¹ USAID, Somalia Complex Emergency Fact Sheet #1 Fiscal Year 2019, p.2

¹² CARE (2014) Somali Girls' Education Promotion Project – Baseline Study Report, p.34; UNICEF (2016) Situation Analysis of Children in Somalia, p.64

¹³ Federal Government of Somalia, Ministry of Education, Culture and Higher Education (2017) Education Sector Analysis 2012-2016, p.131

¹⁴ Forney, J. and Ha, S. (2018) Baseline Report of the Somali Girls' Education Promotion Project – Transition, p. 105

¹⁵ Ibid, p.100; p.119

¹⁶ Ibid, p.129-130

¹⁷ Ibid, p.135

girls, compared to boys¹⁸. Research data shows that dropout is driven by a combination of school-level factors (poor school infrastructure, lack of water at school, and feeling unwelcome by teachers at school) and household factors (high chore burden, displacement and pastoralism)¹⁹. Recurrent school closure due to violence is also likely to play a role on dropout.

The results clearly indicate a pattern of gendered practices affecting education outcomes for girls. The combination of traditional gender norms; prioritisation of household chores; poverty; and safety and security concerns have a disproportional effect on girls' ability to attend school and on learning outcomes. In parallel, it is important to acknowledge that gender norms are shifting quickly in Somali society. Migration and conflict resulted in a high prevalence of female-headed households in Somalia (40-75%, depending on the region²⁰); this situation, in addition to the collapse of pastoralist and agricultural livelihoods due to conflict and climate change, resulted in women becoming the main breadwinners at home (60% of the owners of small businesses are women²¹). As a result of this situation, gender norms are rapidly shifting for adult women, who are taking on new roles and responsibilities, often without the basic literacy and numeracy skills to succeed in entrepreneurship. Qualitative data suggests that this situation has contributed to increase mothers' support to girls' education, and to create spaces for girls to engage in non-traditional roles in society, including in school.

1.2 TARGET BENEFICIARIES

As described in the previous section, AGES is being implemented in areas of South-Central Somalia severely affected by conflict, with large numbers of internally displaced people (IDPs), limited infrastructure and provision of services, and where large swaths of land remain inaccessible to the central government due to ongoing occupation by armed groups. While the education system is slowly being rebuilt by the Federal and State governments, access to education services remains extremely low, with an estimated gross enrolment rate of 22% (19.7% for girls, 24.5% for boys) across the whole of south central Somalia²². Banaadir and Jubaland's gross enrolment rates are considerably higher at 31% and 30%, respectively, compared to 12% in South West State; on the other hand, Banaadir has the largest gender gap in enrolment (35% boys compared to 26% girls)²³. A detailed analysis shows large disparities, with much lower rates among pastoralists (3%²⁴).

Enrolment rates are systematically lower for girls across all areas, reflecting social norms about girls' education as well as a complex array of barriers related to security, availability of time and gendered allocation of household resources. While gender norms are shifting across Somalia as a result of conflict and displacement, with large numbers of female heads of household taking on new gender roles, engaging in self-employment and becoming the primary provider for their families, expectations for boys' education attainment remain higher than for girls, with the latter still being disproportionately affected by early marriage and household chores, thus limiting access to education and retention in school²⁵.

¹⁸ Federal Government of Somalia, Ministry of Education, Culture and Higher Education (2017) Education Sector Analysis 2012-2016, p.116-119

¹⁹ Ibid, p.156-159

²⁰ Considering data from SOMGEP's evaluations in 2014-2019.

²¹ UNDP (2014) The Role of Somali Women in the Private Sector, p.3

²² Ministry of Education, Culture and Higher Education of Somalia (2017) Education Sector Strategic Plan, p.58

²³ Ministry of Education, Culture and Higher Education of Somalia (2017) Education Sector Strategic Plan, p.57-58

²⁴ Ministry of Education, Culture and Higher Education of Somalia (2016) Education Sector Analysis, p.xi

²⁵ Ha, S. & Forney, J. (2018) Somali Girls' Education Promotion Project – Transition – Baseline Study, p.19

Recurrent attacks and the risk of gender-based violence – disproportionately higher for girls - further limit access and attendance²⁶. The limited availability of adequate, gender-segregated toilets and water in schools also poses a barrier for the enrolment of older girls (noting that late enrolment is a phenomenon in the Somali education landscape, due to family preference for children to complete Quranic school before enrolling in formal education²⁷). A recent study indicated that 13% of the schools lack gender-segregated toilets in south central Somalia, with large regional disparities (up to 64% in Lower Shabelle)²⁸.

While all areas targeted by AGES are urban or peri-urban, socio-economic characteristics vary considerably between them. Clans historically linked to agricultural livelihoods or agro-pastoralism (Digil-Mirifle), primarily Af-Maay speakers, form the majority of the inhabitants in riverine areas and in South West State, while clans historically linked to pastoralism, speaking Af-Mahatiri, are a majority in Banaadir and Jubaland. While all targeted areas have experienced high levels of violence, some are particularly affected, such as Wanlaweyn, Bardheere and Kaxda. The proportion of IDP enrollees is much higher in Kismayo, Baidoa, Hodan and Howlwadaag, potentially reflecting (i) recent waves of displacement following the military offensive in Middle Juba and (ii) large scale displacement during the 2016-2017 drought.

AGES was designed to boost access to quality education services for ultra-marginalised girls living in urban and peri-urban areas in Banaadir, Jubaland and South West State. Specifically, AGES seeks to support out-of-school adolescent girls age 10-19 facing multiple barriers to enrolment, attendance and learning. AGES provides tailored options for enrolment, responsive to the needs of different subgroups: formal education, accelerated basic education, and a life skills course to develop basic competencies in literacy and numeracy. The mapping of beneficiaries conducted before enrolment indicated that among the girls enrolled in the first cohort of the project, 31% are IDPs; 16% belong to marginalised minority groups; 40% are Af-Maay speakers (including 18% who are Af-Maay speakers in areas where this is a minority language); 13% are orphans; 10% do not live with parents; 4% are married and 4% are divorced. The enrolment of girls with disabilities was prioritised by the project, with extensive sensitisation of stakeholders and multiple rounds of identification being conducted in order to maximise the likelihood of enrolment.

Previously to identification and enrolment, AGES worked with local authorities and traditional leaders – REOs, DEOs, clan elders and religious leaders – to sensitise them on girls’ education and marginalisation characteristics. The initial identification process resulted in 22,156 girls being selected. Following enrolment in the different learning tracks, the project conducted multiple rounds of spot-checks to verify attendance against the original enrolment list. The spot-checks indicated that while a large number of the originally identified girls had not enrolled/ started to attend courses, classes were often oversubscribed by the enrolment of other out-of-school girls, with a final overall enrolment of 20,468 girls. This situation is not unexpected, particularly given the interval between identification and enrolment (leading to girls moving out of the area, families/ girls losing interest, and in some cases, families enrolling girls in other formal schools following sensitisation). In subsequent cohorts, the project will seek to identify girls immediately before enrolment, thus mitigating the risk of dropout. Additionally, the project will further

²⁶ Ministry of Education, Culture and Higher Education of Somalia (2017) Education Sector Strategic Plan, p.30

²⁷ CARE (2014) Somali Girls’ Education Promotion Project – Baseline Study Report, p.35

²⁸ Consilient (2020) Education Sector Program Implementation Grant (ESPIG) Baseline Evaluation, p.34

emphasise the sensitisation of local leaders for the enrolment of girls from occupational minorities, given the extraordinarily high levels of marginalisation observed among these groups.

TABLE 1: SUMMARY OF DIRECT BENEFICIARIES FROM PROJECT ENROLMENT DATA

Direct beneficiary numbers	Total figures
Total number of girls reached in cohort 1	20,468
Total number of girls expected to reach by end of project	42,000
Number of GwDs identified in cohort 1	6,140 ²⁹
Education level	Share of total direct beneficiaries
Never been to school	96.1%
Been to school but dropped out.	3.9%
Age groups	Share of total direct beneficiaries
10 to 15	62.7%
16 to 19	37.3%

²⁹ Calculation extrapolates the overall proportion of girls with mental health issues identified in the sample.

TABLE 2 - NUMBER OF GWDs IN COHORT 1, DISAGGREGATED BY TYPE OF DISABILITY

TYPE OF DISABILITY	NUMBER OF GWDs IN COHORT 1³⁰	NUMBER OF GWDs WITH SEVERE DISABILITY³¹
VISUAL	76	16
HEARING	49	10
ARMS/ HANDS	47	5
MOBILITY	47	12
SELF-CARE	28	8
REMEMBERING/ CONCENTRATING	64	9
COMMUNICATION	37	3
MENTAL HEALTH	6,140³²	2,333
MULTIPLE DISABILITIES, NOT INCLUDING MENTAL HEALTH ISSUES	68	
ANY DISABILITY, NOT INCLUDING MENTAL HEALTH ISSUES	253	

TABLE 3: PROPOSED INTERVENTION PATHWAYS

Intervention pathway	Which girls follow this pathway?	How many girls follow this pathway for cohort 1?	How long will the intervention last?	How many cohorts are there?	What literacy and numeracy levels are the girls starting at?	What does success look like for learning?	What does success look like for Transition?
Formal school	Girls aged 10-12	6623	4 years	2	Grade 0-1 for literacy-numeracy	Girls achieve grade 4 level for numeracy	Girls enrolled and retained in formal school
Accelerated basic education (ABE)	Girls age 13-16	7241	2 years	2	Grade 0-1 for literacy-numeracy (original estimate; present results shows otherwise)	Girls achieve grade 4 level for numeracy	Girls enrolled and retained in ABE; a proportion of the girls transition into formal school upon completion
Life skills	Girls age 17-19	6604	11 months	2	Grade 0-1 for literacy-numeracy (original estimate; present results shows otherwise)	Girls are able to read with comprehension and achieve mastery of basic operations (addition and	Girls enrolled and completing life skills training

³⁰ Except for data on the number of girls affected by mental health issues, all other data on the number of GwDs in the cohort were obtained from the beneficiary assessment conducted previously to enrolment.

³¹ “Severe disability” corresponds to the following responses under the WG response options: “Cannot do at all” and “A lot of difficulty”.

³² Calculation is based on the prevalence in the sample. Due to concerns about the quality of the data, questions about mental health (anxiety and depression) were not included in the beneficiary assessment conducted at the time of identification.

						subtraction with problem-solving)	
--	--	--	--	--	--	--------------------------------------	--

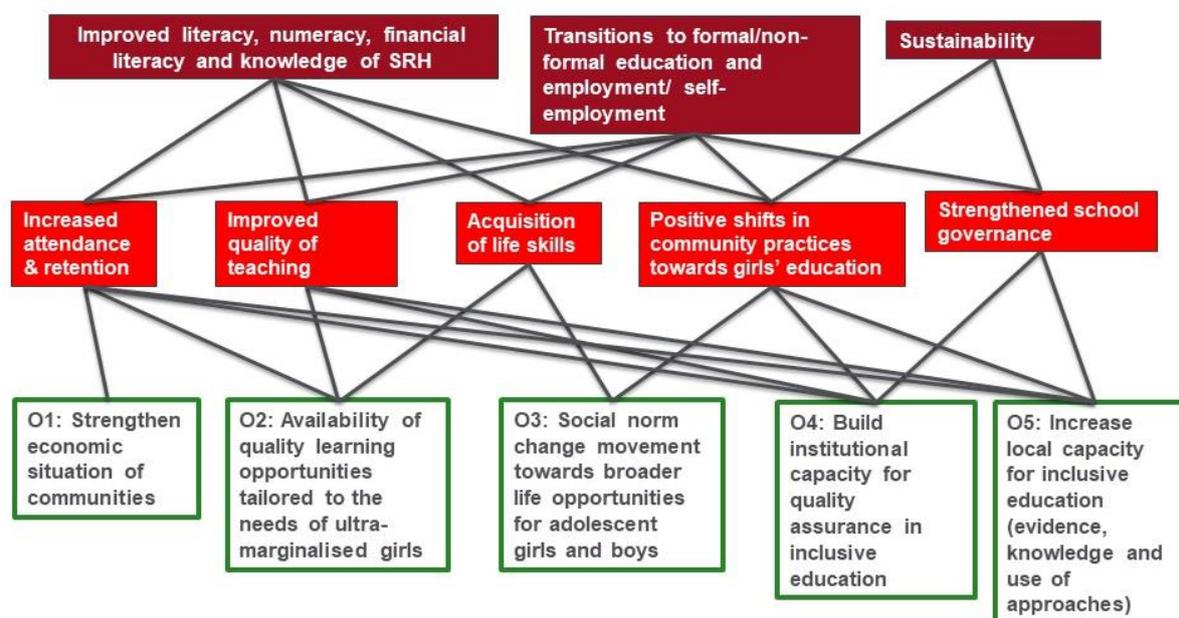
TABLE 4: INDIRECT BENEFICIARY GROUPS

Group	Interventions received	Total number reached for cohort 1
Boys	Benefitting from support to formal school	14,076
ABE facilitators	Recurrent modular trainings followed by on-site coaching	168
NFE facilitators	Recurrent modular trainings followed by on-site coaching	168
Community members	Participation in savings groups; sensitisation re: education, gender and child protection. A subset of community members also benefits from CEC trainings.	14,000 (7,000 in cohort 1)

1.3 THEORY OF CHANGE

The projects Theory of Change is outlined below in graphical form, with a more detailed explanation of project outputs, the intermediate outcomes they expect those outputs to influence, and the eventual impacts that will occur in terms of learning, transition, and sustainability.

FIGURE 1: PROJECT THEORY OF CHANGE



Impact: Extremely marginalised girls will engage in dignified employment/self-employment, delay marriage, participate in decision-making spaces in society, and build positive relationships with others. They will be able to support themselves and their families and build resilience to shocks, changing perceptions about girls' roles in society.

The following **outcomes** will contribute to the expected impact:

- Improved learning outcomes
- Increase in positive transitions
- Sustainability of interventions.

Improved learning outcomes (defined as improved literacy and numeracy skills; improved financial literacy; and basic knowledge of menstrual hygiene) will emerge from the synergy between increased attendance, improved teaching quality, development of girls' agency and positive community practices.

The increase in positive transitions (from out of school into ABE, formal education and life skills courses; from ABE to formal education; into dignified self-employment) will result from the combination of increased attendance; improved quality of teaching; positive community practices (additional investment in girls, shift in gender and social norms); acquisition of life skills and strengthened school governance (increased capacity for inclusive education and quality assurance).

The sustainability of interventions will emerge from shifts in social norms at community and girl level, associated to improved institutional capacity for quality assurance and increased local capacity for inclusive education.

The following **intermediate outcomes** will contribute to the expected outcomes:

- Increased attendance and retention of girls will emerge from the synergy of improved household finance; increased community support for education; availability of flexible education opportunities; increased capacity for quality assurance; and increased understanding and support for inclusive education.
- Improved teaching quality will result from the combination of teacher training addressing gaps in literacy and numeracy teaching; improved capacity for quality assurance; and increased knowledge of inclusive education.
- Improved life skills (leadership and negotiation skills) will result from girls' participation in platforms for social norm change and access to relevant education opportunities.
- Positive shifts in community practices towards girls' education will emerge from the participation of different community actors in platforms for social change on gender and support for girls with disabilities.
- Strengthened school governance will result from the combination of improved institutional capacity or quality assurance for inclusive education; increased capacity of CECs and community mobilisation to support education; increased knowledge about the needs and rights of GwDs leading to increased adoption of inclusive practices.

The following **outputs and activities** will contribute to the achievement of the intermediate outcomes:

- *Strengthen economic situation of communities:* VSL and scholarships will enhance family capacity to support education costs, reduce the likelihood of child labour and increase access for girls with disabilities.
- *Availability of flexible, quality learning opportunities tailored to the needs of ultra-marginalised girls,* through access to three learning tracks and an e-learning platform:
 - Accelerated basic education (ABE) including compressed primary content, financial literacy and business management. ABE facilitators will be trained by the project on curricular content; child protection; and inclusive, gender-equitable classroom management practices;
 - Increased access to quality formal education through agreements with education providers and teacher training (reading, numeracy, child protection and inclusive practices);

- Life skills course for older girls (age 16+), teaching literacy, numeracy, financial literacy, savings and basic business planning. Life skills facilitators will be trained by the project on the delivery of the life skills content in an inclusive, gender-equitable manner and on child protection;
- Access to a mobile learning platform to practice literacy and numeracy skills.
- *Social norm change movement* towards broader life opportunities for adolescent girls:
 - Religious leaders trained on Islamic principles to support gender equity and education for girls and inclusive education;
 - Mothers and mothers-in-law mobilised through VSL meetings and enrolment drives;
 - Girls trained on leadership skills and menstrual hygiene management at GEFs;
 - Messaging about education and rights of girls, in particular girls with disabilities;
- *Build institutional capacity*: Development of simplified quality assurance processes and capacity building of officials for inclusive education; support to the review of the policy framework for ABE; increased evidence and knowledge of the needs and rights of girls with disabilities.

Key assumptions include: Government, community and religious leaders are supportive of proposed interventions, in particular education opportunities for older girls and girls with disabilities; project areas remain accessible at least 50% of the time; conflict / violence does not escalate into open war, resulting in massive displacement, mass casualties, closure and potential destruction of schools; natural disasters (drought and floods) do not escalate, leading to loss of livelihoods, massive displacement, poor attendance and dropout.

Key barriers (extensively discussed above) include: Poverty; high vulnerability to natural and man-made disasters; recurrent displacement as a result of natural disasters, loss of livelihoods, conflict and violence; social marginalisation of subgroups of the population (pastoralists, minority clans, ethnic minorities), reducing social protection mechanisms for adolescent girls; limited availability of education opportunities for older girls who have never attended school; social and gender norms preventing girls from attending school and contributing to gender-based violence, including early marriage; limited awareness of the rights and needs of girls with disabilities; limited understanding of disability, particularly of non-apparent types of disability such as mental health issues; limited teacher capacity to deliver the curriculum in a quality, inclusive and child-centred manner; an emerging education system with limited capacity for quality assurance, support to inclusive approaches and child protection, in particular at local levels; intermittent access to schools due to violence/ conflict.

2. METHODOLOGY

In this section of the baseline report, we introduce the research design used for AGES. The overall pre-post evaluation design is detailed before giving an overview of the tools used in the evaluation as well as the sampling approach employed at baseline. In addition, our review is forward-looking, with an overall discussion of the evaluation approach employed by AGES from baseline through endline; similarly, much of our discussion surrounding methodology and the methodological limitations in the design are written with future evaluation rounds in mind.

2.1 EVALUATION QUESTIONS

The AGES baseline evaluation is intended to provide a rigorous starting point against which project impact can be measured going forward, in addition to providing insights that can influence and refine project delivery. Specifically, the baseline is intended to play the following roles in the context of AGES itself:

- To establish baseline values for project indicators and as benchmark to compare midterm and endline findings;
- To identify factors affecting expected outcomes and if/how those map out against the intermediate outcomes and outputs selected by the project, thus testing its Theory of Change (ToC);
- To gain up to date understanding of the project context to inform ongoing project strategy;
- To inform improvements in the delivery of the project during its lifetime;

Outside the scope of project impact and design, the evaluation will also be used to demonstrate accountability to donors, such as the UK's Department for International Development (DFID), to contribute to broader GEC programme-level reporting and evaluation, and to inform the policy debate around girls' education, girls' marginalization, and education in the Somali context.

To these ends, the AGES baseline evaluation focuses on the evaluation questions, agreed to by both the FM and CARE, outlined in the table below. For each question, we list the specific data source(s) that will be used to address it. Note that the evaluation questions in the table are organized, broadly, by thematic area (e.g., strictly outcome-oriented indicators versus analysis of subgroups and relative marginalization).

TABLE 5: EVALUATION QUESTIONS AND DATA SOURCES

Evaluation Question	Qualitative Data and Analysis	Quantitative Data and Analysis
Impact Indicators		
What impact did AGES have on the transition of highly marginalised girls into education/learning/training or work opportunities? To what extent has participation in the project affected transition pathways for girls enrolled in formal education, ABE and life skills programming?	Qualitative interviews with teachers, CEC members, and parents to assess changes in transition rates over time; participatory vignette-completion exercises to understand changing norms around transition in targeted communities	Analysis of transition rates from baseline to midline (analysis conducted at midline), compared to benchmarked transition outcomes from baseline household survey; comparative analysis of transition rates between cohorts (disaggregated by learning tracks and subgroups within the AGES population).
What impact did AGES have on learning outcomes in the targeted groups? Are students enrolled in alternative basic education acquiring literacy and numeracy skills at a similar pace as their peers enrolled in formal education?	N/A	Analysis of aggregate learning scores and learning scores across students enrolled in different types of learning institutions
Are there specific literacy / numeracy skill areas that are not improving? Why?	N/A	Analysis of subtask-specific learning scores, comparing baseline to midline
To what extent are community education committees addressing factors related to girls' learning and transition?	Qualitative interviews with teachers, CEC members, and parents to identify specific actions taken by CECs	Head teacher (School) survey data on CEC actions and monitoring of schools; teacher survey data on CEC monitoring;

		household survey (at midline and forward) on CEC actions
Is the participation of parents/ adolescent girls in economic empowerment activities contributing to increase sustainability of education outcomes, including for other members of the household?	Qualitative interviews with parents	Household survey responses regarding the perceived impact of bursaries and other economic support; at midline and final evaluation, comparison of education outcomes among girls whose parents are involved in savings groups vis a vis those whose parents are not involved in savings; subgroup analysis of changes in transition rates associated with relative economic marginalization
To what extent are girls' roles shifting at the household/ community level as they participate in the project?	FGDs with teachers and parents; participatory vignette-completion exercises with girls	Household survey data from girls on their empowerment within the household and community; data from teachers (midline forward) on the relative empowerment of girls in households and community at large
Marginalization, Targeting, and Subgroup Findings		
Are there subgroups of the targeted population for whom the interventions have been particularly effective in terms of learning and/or transition? If yes, who are they and why have they benefitted the most?	FGDs with teachers and CEC members to provide context on relative marginalization and perceived changes in learning and transition outcomes	Disaggregated analysis of learning and transition outcomes over time, focused on marginalized subgroups identified as part of the baseline
Are there subgroups of the targeted population who have not benefitted at all from the intervention, or who have benefitted to a much lesser extent, in terms of changes in learning and/or transition? If yes, who are they and why have they not benefitted / benefitted less from the interventions?	FGDs with teachers and CEC members to provide context on relative marginalization and perceived changes in learning and transition outcomes	Disaggregated analysis of learning and transition outcomes over time, focused on marginalized subgroups identified as part of the baseline
Are there any differences in learning outcomes between Af-Maay and Af-Mahatiri speakers?	FGDs with teachers to understand the barriers to learning faced by Af-Maay speakers	Comparison of learning outcomes over time between Af-Maay and Af-Mahatiri speakers, including in different states or schools in which they constitute a small minority versus plurality versus majority of students

<p>To what extent are project interventions effective in addressing barriers to the participation, learning and retention of students with disabilities? Are the interventions particularly effective in addressing barriers for some types of disability, but not others?</p>	<p>FGDs with CEC members and teachers; participatory vignette-completion exercises with girls (especially vignettes discussing girls with disabilities)</p>	<p>Teacher survey and classroom observations to understand the nature of classrooms and how they are adapted for girls with disabilities; household survey data from caregivers and girls, disaggregating learning, transition and attendance outcomes by disability status; survey with GWDs to understand specific barriers they face and how barriers have changed over time</p>
<p>To what extent are project activities effective in addressing underlying causes of exclusion, including inter-clan / ethnic dynamics and gendered practices limiting girls' voice and time to study/attend class?</p>	<p>FGDs with teachers, participatory vignette-completion exercises with girls</p>	<p>Analysis of learning and transition outcomes, disaggregated by clan membership; analysis of the extent to which girls' self-reported voice and participation in class predicts learning outcomes; classroom observations and teacher survey to understand teacher attitudes and practices of inclusion</p>
<p>Are boys' learning outcomes changing? How?</p>	<p>FGDs with teachers, parents and CEC members to understand the barriers faced by boys, and perceptions of how relative learning outcomes are changing over time</p>	<p>N/A (boys learning assessments were removed from the evaluation design to reduce costs, as discussed with the GEC Fund Manager)</p>
<p>Have state-specific differences been observed in learning and transition patterns?</p>	<p>N/A</p>	<p>Disaggregated analysis of learning and transition outcomes, by state and, if appropriate, region</p>
<p>Theory of Change</p>		
<p>Is the project's ToC adequately reflecting the factors driving learning/ transition/ sustainability?</p>	<p>FGDs with teachers, mothers, and CEC members, and participatory vignette-completion exercises, to understand the nature of barriers to learning and transition</p>	<p>Analysis testing the project's Theory of Change using quantitative data, testing the association between intermediate outcomes and key targeted subgroups/barriers, on one hand, and learning/transition, on the other</p>
<p>Which factors at household/ school level are affecting the acquisition of literacy, numeracy and financial literacy skills? How</p>	<p>FGDs with teachers and CEC members to understand the relationship between broader</p>	<p>Analysis of learning and transition outcomes across regions/states; analysis of key marginalization indicators</p>

are these factors being affected by socio-economic and political changes in South Central Somalia?	socio-political dynamics and education in target schools Contextual analysis of the social, economic, and political environment	(household economic status, migration/displacement) across states, to identify areas of increasing risk or marginalization
What works to facilitate transition of highly marginalised girls into education / training / employment and to increase learning?	FGDs with teachers and CEC members to understand the value of different project interventions	Disaggregated analysis of transition outcomes over time, focused on marginalized subgroups identified as part of the baseline, including analysis of exposure to specific project interventions
How successfully did AGES reduce barriers to full participation in education or vocational education for highly marginalised girls?	FGDs with teachers and CEC members to understand the value of different project interventions	Disaggregated analysis of transition outcomes over time, focused on marginalized subgroups identified as part of the baseline, including analysis of exposure to specific project interventions
Sustainability		
How sustainable were the activities funded by the GEC and was the programme successful in leveraging additional interest, investment, and policy change?	FGDs with CEC members and teachers to assess sustainability at the school level. KIIs with policymakers and MOE bureaucrats (in future rounds) to assess policy changes Extent to which AGES' findings will inform the new Education Sector Analysis and Education Sector Strategic Plan	Head teacher (school) survey, teacher survey, and other quantitative data sources to address specific indicators from the project's sustainability plan
Is the participation of parents/ adolescent girls in economic empowerment activities contributing to increase sustainability of education outcomes, including for other members of the household?	FGDs with parents	Household survey responses focused on economic empowerment; household survey responses regarding enrolment/transition rates for other household members (in future rounds)
Contextual Analysis		
Are there 'plateaus' in the acquisition of literacy and numeracy skills? To what extent	N/A	Subtask-specific analysis of learning scores at baseline to

are project interventions effective in addressing those?		identify learning plateaus in different respondent groups
In a context where education provision is mostly community or privately owned, what works in engaging private sector providers in increasing equality and quality?	(project learning only; not evaluation data) Qualitative analysis of the relative success of different strategies employed by the project to work with private / community-owned schools to ensure the enrolment of marginalised girls	(monitoring data only) Analysis of the profile of community/ privately owned schools with high/ low adherence to improved teaching practices
What are the opportunities and barriers emerging for girls as they take on new roles through the application of academic and leadership skills? To what extent these new roles may be harmful for girls from extremely marginalised groups, such as minority clans and ethnic groups?	FGDs with parents and participatory vignette-completion exercises with girls	Analysis of GEF participation and YLI scores from the household survey with girls; analysis of participation by girls in school committees and leadership functions, derived from the Head Teacher (school) survey.

2.2 EVALUATION DESIGN

Overall 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 – utilizes a range of data collection tools. Qualitative data was collected using FGDs with multiple respondent groups, and participatory exercises targeting girls.³³ Quantitative data was collected using surveys with teachers, head teachers, girls with disabilities, 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:

³³ These participatory exercises include risk mapping to assess the perceived safety of the school and the surrounding community, and story-telling exercises in which participants hear a story about a girl and are asked to describe how the story might end, the barriers the girl might face in achieving one ending or another, and otherwise describe how they view the girl’s life chances.

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. The design chosen will impact the type and the strength of the analysis and inferences in subsequent phases of the evaluation.

The evaluation uses what the Fund Manager refers to as a joint-sampling approach, in which the same cohort of girls comprises both the learning and transition samples. The joint sampling approach is in line with many of the previous GEC studies, and implies that the learning and transition outcomes are both evaluated using the exact same sample of individuals. The individual cohort girls are randomly selected at the schools (or learning centres) where they first take the learning assessments and the girls survey. Family members of the same individual are then interviewed at the household. The additional respondents associated with the cohort girl include her head of household, caregiver and any female falling in the age range of interest for transition benchmark, if one is available. As all cohort girls for AGES are newly enrolled, no girl-specific attendance data is being collected at baseline for the individuals in the cohort, but this will be done in next waves of data collection.

The cohort girls will be tracked from baseline through endline to assess changes in learning outcomes over time. In the absence of a comparison group, longitudinal comparisons will be 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 will be compared to benchmarked differences in learning scores between grade levels at baseline.³⁴ 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. Girls' life skills will also be compared against benchmarks established during this baseline, in line with the grade level approach applied to learning scores.³⁵

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 become the expected change in learning, absent intervention, for girls who go from 11 to 12 years old

³⁴ These grade level benchmarks also explicitly take into account alternative learning institutions – girls in formal school will not be compared to benchmarks that include girls in ABE or NFE programs. Benchmarks are established for specific institution types and "grade" levels.

³⁵ Life skills here are defined functionally: the ability to complete reading comprehension tasks and achieving a score of 50 percent or higher on subtasks assessing addition and multiplication.

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.

Consider a situation in which flooding impacts a number of target schools and disrupts the school year between baseline and midline. A midline evaluation following this event will judge girls' progress from one year to the next against a benchmark of girls, none of whom were affected by the flooding (because they completed the learning assessment during the baseline, prior to the flooding). The cohort's performance will be judged against a benchmark that is not a fair standard, leading us to underestimate the project's impact. Alternative scenarios that yield overestimates are equally likely. A more general formulation of this limitation is that a benchmarked comparison cannot account for temporal shocks that might affect the outcome in question.³⁶As with learning, the same cohort of girls will be tracked over time for the purposes of assessing transition rates.

A key aspect of the evaluation of learning and transition outcomes is the panel nature of the analysis, in which a cohort of girls is tracked over time. Several of the intermediate outcomes are evaluated in a similar fashion, largely as a function of the fact that data collection among the learning/transition cohort overlaps with data collected for assessing community attitudes and self-esteem. For instance, girls' self-esteem is assessed using the Youth Leadership Index, which is administered to all girls in the cohort; as a result, the evaluation of changes in self-esteem over time is also derived from a panel analysis.

Two exceptions to these general rules are attendance and teaching quality, both of which utilize a cross-sectional approach. Our primary measures of teaching quality are extracted from direct observation of classrooms. This approach will be replicated at midline, ideally in the same set of schools; however, the set of teachers evaluated will, in practice, change across time. Therefore, our analysis of teaching quality and attendance will also involve pre-post comparisons, but without the advantage of a panel design.³⁷

Beyond studying changes in project outcomes over time, a core goal of the baseline evaluation is to assess the validity of the project's Theory of Change. At midline and endline, the Theory of Change can be tested more extensively by, for instance, investigating the differential impact of various project interventions, depending on girls' relative exposure to each. In this baseline evaluation, we test the Theory of Change in a cross-sectional framework, by studying the relationship between the intermediate outcomes (e.g., attendance, teaching quality) and the core outcomes, learning and transition. This analysis overlaps in many ways with the subgroup analysis performed for both learning and transition, but we investigate the

³⁶ Temporal shocks do not have to be environmental or societal-level changes. A specific scenario involves changes in the difficulty of learning assessments over time; a change of this nature impacts the performance of cohort girls at midline, but is not accounted for in calculations of the benchmarks, because no girls included in the benchmark were administered the midline learning assessment. Therefore, changes in the difficulty of the learning assessment are not accounted for by the benchmarked comparison design outlined here.

³⁷ The main drawback to a cross-sectional versus panel design is that the former involves sampling variation, which impinges on our ability to draw inferences in small samples. However, conclusions drawn from cross-sectional pre-post designs are valid under similar assumptions as those that apply to panel designs.

relationship more extensively and rigorously, employing multivariate regression models to investigate the association between intermediate and core outcomes.

Before describing the baseline evaluation design in more detail, we turn briefly to a discussion of the benchmarks against which project impact will be compared at midline and endline. In order to partially address the absence of control or comparison groups, the research design includes benchmarking (representative of treatment interventions) for learning and transition outcomes at baseline in order to provide points of comparison going for future waves of evaluation.

For learning, a total of 454 students in grades 2-5 were assessed in the intervention schools, giving between 109 and 119 girls per grade to be used as benchmark for learning. The benchmark sample was randomly drawn from girls enrolled in the sampled grades, and only took part in the learning assessments. The benchmark sample will be used to estimate the impact of the program on learning outcomes at midline. For example, for girls in Grade 3 at the midline, their scores will be compared to the scores of Grade 3 girls obtained at the baseline. As such, Grade 3 girls at the baseline serve as a kind of control group. This is because at the baseline no program activities have started, and as such any changes in learning outcomes for the girls in Grade 3 at baseline should not be influenced by the program. Consequently, any difference in learning outcomes when comparing midline Grade 3 girl cohort to baseline Grade 3 benchmark should be attributable to the impact of the AGES intervention. The following table outlines how the benchmark will be used for in learning:

TABLE 6: BENCHMARKING FOR LEARNING

Baseline grade	Benchmark at midline	Benchmark at endline
1	2	3
2	3	4
ABE ³⁸	2	4
NFE ³⁹	N/A	N/A

For transition (i.e. the rate at which students continue to the next grade level or to the next level of education), conversely, benchmark respondents were selected through the household survey. If a female community member in the age range 20-24⁴⁰ was encountered at the cohort girl’s household, she would be respond to a short set of questions to determine the transitions experienced by her. This resulted in an additional sample of 276 individuals, with the following age distribution:

TABLE 7: TRANSITION BENCHMARK AGE DISTRIBUTION

Age	N	Share of Sample
20	145	52.5
21	40	14.5
22	36	13.0

³⁸ ABE is designed to deliver a compressed curriculum corresponding to Grades 1-4 of formal education. After one year the student is expected to have acquired the skill level of those in grade 2 of formal school and after two years grade 4 level.

³⁹ According to the AGES MEL Framework: “Girls enrolled in life skills training will not be compared with a specific grade benchmark, but rather with minimum achievement levels in literacy and numeracy (able to respond reading comprehension levels at the lowest level in the adapted SEGRA; able to perform single-digit addition, subtraction, and addition and subtraction problem-solving).”

⁴⁰ Transition benchmark for younger girls is covered by the baseline cohort itself.

23	33	12.0
24	22	8.0
Total	276	100.0

Given the practical impossibility of a randomized controlled trial or a quasi-experiment difference-in-differences design, a research design combining benchmarked and non-benchmarked longitudinal comparisons is the best possible alternative. However, as we discuss in more detail in the limitations section below, a pre-post design with benchmarked outcomes faces significant limitations in the strength of inferences that can be drawn from even a panel analysis that tracks the same girls over time.

Baseline Evaluation Design

Above we have described the overarching evaluation design to be used through the lifespan of the AGES project. This section will now describe the design of the baseline phase of the evaluation, which has its own set of assessment goals.

As explained above, the evaluation utilises a mixed-methods approach. The objective of the baseline was to set the stage for the midline and the endline in establishing the baseline values for indicators and the indicator targets for the future evaluations as well as to facilitate decisions on programmatic focus. Consequently, the approach is one which allows for solid data to establish the starting point of key project outcomes such as learning and transition as well as intermediate outcomes such as attendance, teaching quality, etc. In addition, the baseline identifies key considerations and likely challenges for the project moving forward as well as considerations of key contextual factors that might play a part in these.

As outlined above, the baseline selected three main cohorts of respondents. These groups were selected randomly and independently of each other.⁴¹ The same cohorts will be tracked over time as per the longitudinal study design. As it pertains to analysis, though, these groups are often combined to look at aggregate levels of the indicators. In the subsequent phases of the evaluation, however, comparisons between these groups will become particularly interesting. But for the moment, we describe the three groups as separate to highlight the fact that they are, separately, coherent sets of respondents, and because they were sampled independently of one another.

The first cohort group is the girls attending grades 1 and 2 in formal schools that are part of the intervention. These girls, aged 10 to 12, have recently been enrolled in intervention schools and have been selected for the purposes of establishing learning and transition outcomes over time. The girls enrolled in ABE through the program, aged 13 to 16, form the second cohort group for AGES, and serve

⁴¹ The FE and ABE samples were drawn initially proportionally to the program populations in the respective regions. This sample draw, however, changed quite significantly through the fieldwork as inaccessibility and other issues detailed elsewhere in this report made some of the learning centres wash out and the evaluation team ran out of replacements sites to be used in their stead. The NFE centres, however, were initially selected to match them to either FE and ABE sites in order to ensure most overlap. However, at the end of the fieldwork essentially all NFE sites had also been used to substitute those that had to be dropped from the sample.

the same aim for the evaluation, namely establishing transition and learning outcomes. The third group, then, consists of the older girls (17-19 years of age) that have been registered onto the program's NFE track. The three groups together form the overall cohort of the program.

In addition to these groups, benchmarks for learning and transition were established as detailed above. Learning benchmark respondents were drawn from grades 2 to 5 in formal intervention schools while transition benchmark was established through interviewing respondents at the households of the cohort girls. While this is not a random sample, it is a sample of individuals who are from the same marginalised communities that produced the cohort girls themselves, and as such comparable.

The evaluation, further, specifically targeted disabled girls in the community. Disabled girls were identified primarily through household surveys, but teams were encouraged to expand their search to all interactions with community members. The primary aim of selecting an expanded sample of disabled girls was to ensure that disabled girls are adequately represented in the beneficiary selection as well as to further explore the barriers to education and challenges faced by the group.

In addition to the primary aim of establishing baseline levels of key indicators and selecting the respondent groups for the longitudinal study, the baseline evaluation is required to assess the relationships between intermediate outcomes and outcomes. In order to achieve this, the baseline first considers the project's Theory of Change from a theoretical perspective, weighing the assumptions – both explicit and implicit – therein. Further, the baseline looks at these relationships in an empirical fashion. Specifically, the data on the intermediate outcomes of teaching quality, self-esteem and empowerment, school management, and community attitudes were used to assess their effect on learning and transition. We investigate the relationship between each of these factors and the end outcomes of learning and transition formally, using regression analysis. Where possible, qualitative data is used to supplement and triangulate the findings emerging from the quantitative analysis.

The baseline made use of a variety of tools targeting a number of different populations with the aim of covering all the intermediate outcomes and outcomes and to facilitate triangulation across the board. The following section details the different tools and populations and the samples drawn from each.

2.3 QUANTITATIVE METHODOLOGY

Sample Design and Data Collection Tools

At the highest level of aggregation for the baseline evaluation of AGES are the learning centres. The three types of learning institutions were originally sampled independently of one another, with probability proportionate to size, and stratification at the regional level. To illustrate, consider NFE centres: the total population of NFE beneficiaries was calculated per region, and interviews were allocated to each region in proportion to the number of NFE beneficiaries in that region. Within regions, NFE centres were sampled with probability proportionate to size, and replacement, such that larger centres could be allocated multiple clusters of interviews. The sample design called for equal cluster sizes; given that centres were selected with probability proportionate to size, the sample of NFE girls would have been self-weighting as drawn. A similar approach was taken for ABE centres and FE schools. Because there are regional differences in the share of ABE versus NFE versus FE girls, the allocation of clusters for each institution type varied by region.

The initial sample draw included 45 sites of each kind of learning institution. For ABE and NFE centres, the sample design called for random selection of girls at the centres upon arrival, as girls enrolled into the programme at these centres had not been fully enumerated in advance. In contrast, at FE schools, the CARE programme team had compiled a list of enrolled girls who were beneficiaries, and a two-stage sample was drawn, in which schools were selected with probability proportionate to size, and girls within

schools were selected in advance using simple random sampling⁴². Learning centres of all three types were determined eligible for sample draw if a verification of their enrolment information had been completed by the CARE program team.

Prior to the start of fieldwork, the sample design underwent a significant overhaul in response, primarily, to changes in the security situation. Combined with the requirement that learning centres have a complete verification of enrolment numbers before they could be considered eligible for the evaluation, this limited the sample frame to just 55 formal schools, 44 ABE centres, and 47 NFE centres in total. As we discuss later in this report, multiple changes and challenges occurred in the fieldwork resulting in a fundamentally reduced sample for the baseline.

Immediately prior to the start of fieldwork, the sample was re-drawn. FE schools and ABE centres were drawn using the same *design* as originally specified – proportional stratification by region, and schools selected as the primary sampling unit (PSU) with probability proportionate to their population of beneficiaries. However, the sample *frame* was significantly altered, which necessitated a new sample draw. For both FE schools and ABE centres, this constituted the first-stage sample draw, beyond which a second stage of selection took place for girls drawn in the cohort, teachers selected for classroom observations, and so forth.

In contrast to FE schools and ABE centres, the sample design for NFE centres was changed substantially in response to the new situation on-the-ground. Given the small number of available sampling points in the frame, and the logistical burden of visiting many distinct sites (as ABE, NFE and FE sites do not universally overlap), CARE's Monitoring & Evaluation team and the evaluation team from Consilient agreed to visit NFE centres where they overlapped with the existing sample of FE schools and ABE centres.

At first glance, this approach seems problematic, insofar as it mimics a convenience sample of NFE centres. However, with just 47 eligible NFE centres in the sample frame, the sample very nearly constituted a census of NFE centres, and random selection was still employed to select girls within centres.⁴³ Moreover, it is important to remember that the FE and ABE PSUs were selected randomly from a list with frequent overlap with NFE centres; therefore, the set of NFE centres targeted was not a true convenience sample, but was not drawn with equal probability or probability proportionate to size (PPS). Our viewpoint regarding the NFE sample is to treat conclusions drawn from it as internally valid – just as with any panel analysis from an unknown baseline sampling method – but to be mildly cautious when extending those conclusions to the full population of NFE girls. In practice, even FE schools and ABE centres drawn with PPS are only representative of the sample frames from which they were drawn; given that the sample frames in both cases included only a portion of targeted schools and centres,

⁴² This approach was required due to the presence of both beneficiaries and non-beneficiaries in formal education classes; in order to ensure that data would be collected only from AGES beneficiaries, the project provided the list of marginalised girls enrolled through its activities in all formal schools for sampling by the external evaluator.

⁴³ Ultimately, the number of NFE centres in the sample was smaller than the number selected as overlapping with FE schools and ABE centres, because several of each institution type were dropped from the sample during fieldwork and during the data cleaning stage. At the time, however, the sample of NFE centres approached total overlap with the sample frame.

projecting results from the sample to the broader beneficiary population (as opposed to the sample frame exclusively) is unfounded even in these cases.

At each formal school, team leaders conducted an interview with the head teacher or principal (Head Teacher Survey). For all other samples, the selection of the learning centre constitutes the first stage in a multi-stage sample design whereby the learning centres were first selected, using the varied methods outlined above, preceding fieldwork. Within each learning centre, cohort girls were selected randomly. In FE schools, random selection occurred in advance of the data collection team's visit, from CARE's list of FE beneficiaries. The cluster size for FE schools was set at 12 girls per school, with four girls selected proactively from the beneficiary list, and placed in random order to act as replacements in the event that any of the original 12 girls could not be located. In practice, additional replacements were often required, and these were selected using an identical procedure, with the evaluation team employing simple random sampling at the school level from the cohort list.

In ABE and NFE centres, cohort girls were not selected in advance. An identical cluster size of 12 girls per centre was chosen, but girls were selected randomly upon arrival at the centres by the team leader for each data collection team. In each case, the team leader fully listed the eligible girls in the centre – restricting the set of eligible girls by age, in line with the criteria described in the table below – and uses a random number generator to select 12 girls with simple random sampling.

Beyond the cohort girls, samples for the remaining tools were also selected randomly, including teachers for classroom observations. All eligible teachers were recruited for the self-administered teacher survey, which was filled on paper and then entered into the database by the team leader. The second-stage samples, including the population targeted, the targeted sample size, and the data collection tools employed, are described in the table below.

TABLE 8: TARGET POPULATIONS AND SECOND-STAGE SAMPLE SELECTION

Respondents	Sample Population	Quantitative Tools Completed
Cohort Girls – Formal schools	<ul style="list-style-type: none"> • Grades 1-2 • Ages 10-12 • Random sample drawn from enrolment information before fieldwork 	<ul style="list-style-type: none"> • Girls School Survey • Learning Assessments
Cohort Girls – ABE	<ul style="list-style-type: none"> • Girls aged 13-16 in ABE • Random sample selected from enrolment list at facility 	<ul style="list-style-type: none"> • Girls School Survey • Learning Assessments
Cohort Girls – NFE	<ul style="list-style-type: none"> • Girls aged 17-19 in NFE • Random sample selected from enrolment list at facility 	<ul style="list-style-type: none"> • Girls School Survey • Learning Assessments
Households	<ul style="list-style-type: none"> • Joint sample: households of the cohort girls in FE, ABE and NFE • Includes transition benchmark with women aged 20-24 	<ul style="list-style-type: none"> • Adult module (any adult) • Caregiver module • Transition module for

		females aged 20-24
Learning Benchmark Girls	<ul style="list-style-type: none"> • In-school girls, program schools • Grades 2-5 • Minimum 100 girls sampled per grade 	<ul style="list-style-type: none"> • Learning Assessments
Girls with Disabilities (GWDs)	<ul style="list-style-type: none"> • Sampled through household surveys. Researchers asked respondents of any disabled girls they knew in the community and proceeded to contact the caregiver of the girl. • All identified individuals were sampled 	<ul style="list-style-type: none"> • Disability module
Head Teachers	<ul style="list-style-type: none"> • Head teachers at all sampled formal schools 	<ul style="list-style-type: none"> • School (Head Teacher) Survey
Teachers & Classrooms	<ul style="list-style-type: none"> • Classroom observations: <ul style="list-style-type: none"> ○ Randomly selected 2 teachers in formal schools teaching grade 1-2 Somali or Math • Teacher survey: <ul style="list-style-type: none"> ○ All eligible teachers in formal school (grade 1-2) and facilitators in all learning centres interviewed through a self-administered survey • Headcount tool: <ul style="list-style-type: none"> ○ Attendance spotcheck and comparison against enrolment register in all grades of formal school and streams of ABE/NFE 	<ul style="list-style-type: none"> • Classroom Observations • Teacher Survey • Headcount Tool

As we discuss in more detail in the next section, various learning centres needed to be replaced during fieldwork due to inaccessibility, irregularities in enrolment data and other factors. Learning centres or sampling units were replaced with another learning centre of the same type in the same region. For example, a formal school in Banadir would be replaced using another formal school also in Mogadishu. However, the evaluation team exhausted the number of facilities available for replacement as the CARE program team had not completed the requisite verification of enrolment registers for all the intervention sites. As such, there is no systemic bias coming from a discrepancy between the sampling frame and the sample, as the two are essentially the same. However, it is possible that there are sources of bias coming from the order in which the program team had conducted their verification, i.e. in the construction of the sampling frame itself. This is something that should be kept in mind in future evaluation points.

Achieved Sample

The sampling methodology outlined above was the result of a number of adjustments made during the inception phase. In the early stages of planning, it was assumed that formal schools, ABE centres and NFE centres operating within the AGES programme would be sampled together – a sample of 45 locations, each of which would include all three types of learning institutions. Upon further review between CARE, its implementing partners, and Consilient, it became clear that a sample of this kind would not work in practice, because some of the centres were more distant from each other than expected; in other cases, one type of institution in a sample location was fully prepared for baseline data collection, while other

types were not, making it impossible to sample all three types of institutions in each sample location. These practicalities necessitated a move toward sampling formal schools and ABE centres separately, and visiting NFE centres where they overlapped geographically with the formal schools and ABE centres sampled.

The shift in sampling methodology produced a less cohesive overall sample, but the overall sample size was relatively unchanged. Prior to the start of fieldwork, it was agreed that Consilient would visit the following learning institutions:

- 45 formal schools
- 43 ABE centres
- 44 NFE centres

The institutions – hereafter, "sites" or PSUs – were spread over 61 unique sample locations. At each of the 132 sites (45 formal schools, 43 ABE centres and 44 NFE centres), field teams were assigned to sample 12 girls, producing a maximum potential sample of 1,584 girls. This represents an oversample relative to the targets developed during the inception stage, as it is common for some sites to have too few girls to comprise a full cluster.⁴⁴

Note that, throughout our discussion of sampling, we refer to "household surveys", which are precisely coterminous with the sample of cohort girls. Every cohort girl selected into the sample completed a learning assessment and survey, while her caregiver completed a household survey.⁴⁵ We refer to household surveys, in general, for the sake of brevity, but may occasionally refer to learning assessments where the focus of the discussion is on learning outcomes. The sample being analysed is the same – unless otherwise noted, such as in the case of benchmark learning girls – regardless of the terminology used.⁴⁶

The only sample for which the target size was fully achieved was the benchmark learning sample. The table below describes the sample achieved for each tool and/or sampled group, versus the targets set at inception.

TABLE 9: TARGET AND ACHIEVED SAMPLES, BY QUANTITATIVE TOOL

Tool	Inception Target	Achieved
Head Teacher (School) Survey	45	38
Classroom Observations	90	70
Attendance Headcount	N/A	373
Teacher Survey	N/A	310
Household Surveys	1,530	1,420

⁴⁴ Schools report enrolment numbers, but this generally does not represent the true number of girls available for sampling, especially on a randomly-selected day when field teams visit the school. In previous education projects in Somalia, including those as part of the Girls' Education Challenge (GEC and GEC-T) programme, we have encountered this problem repeatedly, typically for a small number of schools in each sample.

⁴⁵ The survey completed by each cohort girl included the YLI questionnaire, but also covered other topics, such as teaching quality.

⁴⁶ It is important to emphasize the precise overlap between household surveys and learning assessments among the cohort girls. In past GEC projects, the learning and transition cohorts have occasionally differed substantially, such that analysis of transition and learning outcomes employ entirely or largely different sets of cohort girls. This is not the case for AGES, where the learning and transition cohorts are part of a joint sample, with complete overlap; the only other samples of girls are one-time samples at baseline (i.e. these girls do not constitute a "cohort" in the traditional sense) for benchmarking learning and transition outcomes.

GwD Survey	N/A	125
Transition Benchmark Surveys	314	293
Learning Benchmark Assessments	400	454

The failure to achieve the targeted samples, illustrated in the table above, stems from two sources. The first are changes in the sample of sites that occurred during fieldwork. As noted above, the target sample covered 45 formal schools, 43 ABE centres and 44 NFE centres. However, during fieldwork, adjustments were made to the sample, especially when data collection teams found sites with insufficient numbers of eligible girls available to constitute a full cluster. The full set of schools that were dropped from the sample – and the reasons for their removal – is provided in Section 2.5 below. For the purposes of this discussion, the critical point is that some schools were not ultimately replaced, either because no replacement school was available, or because the replacement school also lacked a sufficient number of eligible cohort girls.

Unfortunately, similar issues applied to ABE and NFE centres. Overall, the sample of formal schools was reduced from 45 to 38; the sample of ABE centres was reduced from 43 to 34; and the sample of NFE centres was reduced from 44 to 36. In response, ABE and NFE cluster sizes were increased – typically from 12 to 17 girls – in an effort to preserve statistical power necessary for making inferences regarding project impact.

The second reason for failing to achieve the targeted sample was poor administration of the learning assessments on the part of two enumerators. Concerning patterns in the learning assessments were flagged early in fieldwork by both CARE's Monitoring & Evaluation team and Consilient's technical evaluation staff. As a result of poor administration of the learning assessments, the evaluation team removed 91 cohort girls from the sample in total.

The changes in sample composition described above have consequences for the statistical power of the evaluation and for how we conduct the analysis. For instance, different cluster sizes may necessitate the use of survey weights. In the annexes, we provide supplemental analysis and additional discussion meant to illuminate these issues. In Annex 12, we formally investigate the impact of these sample size reductions on the evaluation's statistical power. In practice, we find a small but meaningful reduction in statistical power as a result of lost sampling sites or clusters, and a smaller reduction as a result of lost interviews due to poor administration.

In Annex 16, we discuss the use of survey weights and clustered standard errors. Survey weights can be deployed in a number of ways and with a number of goals, and we describe the alternatives and how we selected from among them. In brief, we use survey weights to adjust for differential cluster sizes only, and do not use weights to make other sample adjustments. Similarly, we justify our decisions around the definitions of clusters.

2.4 QUALITATIVE METHODOLOGY

Qualitative Tools and Sampling

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. The tools were reviewed and commented upon by the GEC Fund Manager prior to fieldwork. 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 let 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.

Overall, we have conducted 40 qualitative interviews across 10 locations. Each of the teams was assigned a specific number of the qualitative interviews based on the number and variety of geographic locations covered during the fieldwork. Team leaders were instructed to select a school with an active and available CEC for the qualitative data collection – which was used as the main criteria for sampling purposes.

The table below summarizes the numbers of qualitative interviews per state and type of tool.

TABLE 10: GEOGRAPHIC DISTRIBUTION OF QUALITATIVE INTERVIEWS

Type of the interview	Banadir	Jubaland	South West State
FGD CEC	4	3	3
FGD teachers	4	3	3
FGD mothers	4	3	3
Risk mapping/vignette participatory exercises	4	3	3
Total	16	12	12

Annex 16 provides additional information regarding the qualitative methodology. With regard to the tools, we provide a brief mapping of each tool to the specific outcomes they were intended to capture, and a statement of who developed the tool and whether it was piloted. Annex 16 also describes the

procedures by which qualitative interview participants were recruited and a description of methodological issues encountered specific to the qualitative data.

2.5 FIELDWORK, QUALITY ASSURANCE AND ANALYSIS

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. Note that many details of enumerator selection, training, quality assurance, and data cleaning are provided in Annex 16.

Fieldwork and Replacement Selection

During the fieldwork, several changes were made to the sample, mostly related to our teams' inability to locate 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.⁴⁷ In this section, we provide more details on the specific circumstances under which the school-level changes were made to the sample and how replacement schools were selected. Over the course of fieldwork, 5 formal schools were dropped from the sample without any replacements, 4 in Banadir and 1 in Jubaland, as the teams could not locate any of the 12 pre-selected cohort girls. Due to the lack of available replacement schools in the original sample frame, only one school was successfully replaced – in the second replacement school selected, as the team could not locate any cohort girls in the first replacement school. A second school in Banadir was replaced by a school not initially included in the sample, where all the cohort girls were re-located, according to CARE and CARE's implementing partner. However, as no pre-selected cohort girls could be located in this new replacement school, the formal school was excluded from the sample.

Moreover, due to the additional problems related to the sampling and assessment scores of the girls, an additional 3 formal schools, 9 NFE, and 9 ABE centers were removed from the sample, as requested by CARE Somalia.

To compensate for the reduced sample size stemming from the loss of sampling points, 5 additional girls were recruited into the cohort sample in each of 21 ABE and NFE centers. As the changes to the sample were made toward the end of fieldwork, locations (cities or districts) where all data collection was already completed were excluded from the list of sample points where additional data would be collected. Also, locations with both ABE and NFE centres in the same area were prioritised, to facilitate timely collection of the replacement data. Beyond these restrictions – around geography and prioritisation of overlapping sampling points – the 21 ABE and NFE centres to be re-visited were randomly drawn from the list of available sample points.

⁴⁷ We discuss individual-level replacement selection (and the reasons why this was occasionally necessary) in Annex 16.

Data Analysis

Quantitative data collected during the baseline was analysed in a traditional survey sampling framework. In other words, 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, as discussed elsewhere, 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 modeling choices.

Qualitative data analysis was completed using the qualitative data analysis package RQDA. Following transcription and translation of the qualitative interviews, the evaluation team developed a coding scheme based on the evaluation questions, contextual knowledge of marginalized groups in Somalia, and an understanding of the educational context in which AGES is being implemented. The evaluation team read a portion of the interviews to refine the coding scheme, and then proceeded to code each qualitative interview, facilitating systematic analysis of the qualitative data.

In contrast to the FM's recommended approach, qualitative and quantitative data collection occurred at the same time, due to logistical and budgetary constraints. However, the evaluation team made particular effort to separate the qualitative and quantitative data analysis, to ensure that insights from the qualitative data were not colored – at least in the first round of analysis – by the quantitative results. Specifically, a first round of qualitative analysis was conducted, exploring the data for interesting themes; a second round was completed after the quantitative data was analyzed, with the goal of explaining unusual results, contextualizing quantitative findings, and triangulating insights between the two data sources. By analyzing the qualitative data in two rounds, the evaluation team was able to glean additional insight from the FGDs and participatory exercises, insights which fed into and influenced the statistical analysis.

2.6 CHALLENGES & LIMITATIONS

The evaluation discussed in this report takes a broadly rigorous approach to assessing project impact. Nonetheless, it is subject to a number of methodological limitations that should be borne in mind when reviewing the results from the baseline or future rounds of evaluation, and when considering how to adjust the evaluation going forward. We do not revisit in detail challenges that occurred during fieldwork – as replacement and loss of schools from the sample was already discussed extensively earlier in this section. However, we do discuss the methodological implications of those same issues, to the extent they

impact the rigor and validity of the results presented here and have consequences for midline and endline design.

Sample Size and Statistical Power

The AGES evaluation design took into account explicit power analysis when setting a sample size and structure, with discussions taking place between CARE and the FM. Unfortunately, the target sample size was not met at baseline, due to a loss of sampling points (both schools and centres) and improper survey administration by two of Consilient's enumerators. As a result, the statistical power of the evaluation has been reduced, as we discuss and demonstrate in more detail in Annex 12. Given that this is the first evaluation of AGES, it is difficult to estimate the extent of panel attrition that should be expected over time, but the reduction in sample size at the baseline may combine with panel attrition to significantly undermine the power of the evaluation. This challenge is partially mitigated by the potential for carrying out a second baseline following the enrolment of the second cohort.

The available sample is sufficiently large to enable high-quality analysis at baseline, given that the baseline focuses on establishing targets and assessing the Theory of Change, both of which rely on cross-sectional analysis. However, looking forward to the midline, special efforts should be made to reduce panel attrition as much as possible. A similar recommendation applies to school sampling and replacement – every effort should be made to maintain the school sample, with explicit, pre-defined criteria for school replacement in advance of midline data collection. Finally, the midline should include a stipulation that, if panel attrition rises above a pre-set threshold, additional clusters and/or children will be recruited into the midline sample, to ensure an adequate sample size for midline-to-endline comparisons.

Lack of a Comparison Group

Perhaps the most serious threat to valid inferences regarding the impact of AGES on learning, transition, and other outcomes is the lack of a comparison group against which intervention schools can be compared. In contrast to randomized experimental or quasi-experimental difference-in-differences approaches, the pre-post design employed by AGES requires stronger assumptions for attributing causation to the intervention itself. In a classic difference-in-differences design – with a comparable, but not randomized comparison group – attributing causation requires assuming that the intervention and comparison groups would have developed similarly over time.⁴⁸ Importantly, external shifts in the condition of intervention schools are not necessarily problematic, if the external shifts also affect comparison schools.

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. Consequently, 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. If community attitudes in Somalia are generally becoming more supportive of girls' school completion, the evaluation may attribute impact to the program that is actually part of a broader trend in Somali society.

⁴⁸ This is the assumption that exogenous, time-variant shocks apply equally to both groups, often referred to as the "parallel trends assumption."

It is important not to overstate the drawbacks of a pre-post design. For outcomes that change slowly and do not have a natural maturation effect (i.e. do not naturally rise over time), pre-post designs can be effective. We would consider the assessment of deep-seated community attitudes and norms, which tend to change slowly, as one case where a pre-post design assessed over one or two years to be a sufficiently strong design. However, where shifts can occur quickly – as in the case of attendance, which is highly susceptible to temporal shocks – and where outcomes have a natural growth curve – as in learning scores for girls as they age – pre-post designs are less effective. In the context of learning and transition outcomes, the analysis will apply benchmarks among older girls, established at baseline. But this approach has its own share of drawbacks, particularly in the context of changing learning assessments, as we discuss below.

In general, the evaluation should consider utilizing secondary data sources to facilitate comparisons to a broader sample of "comparison" schools. For instance, if the evaluation could access enrolment rates from a subset of non-intervention schools, possibly provided by the FGS or state governments, it would be possible to triangulate the impact of AGES on enrolment rates (a plausible, if imperfect, proxy indicator for transition). Combining targeted comparisons of this kind with pre-post analysis would strengthen the midline and endline evaluation.

Comparability of Learning Assessments across Rounds

The evaluation of AGES employs a Somali literacy and numeracy learning assessment which was designed prior to baseline data collection. At midline and endline, we assume new versions of the learning assessments will be employed, to avoid the possibility of students recalling answers from the previous assessment.⁴⁹ In past education projects in this context, we have tested whether assessments are of similar difficulty by recruiting a subsample of students to complete both the baseline and midline assessment, and testing for similarity in scores, generating an equating coefficient where applicable.

Unfortunately, changes in the learning assessments over time complicate and weaken inferences regarding learning scores. Even minor changes to the assessment between rounds – changing the response options or providing a new question with assumed equal difficulty – introduce the potential for differences in assessment difficulty. In standard difference-in-differences designs, changes in assessment difficulty are not problematic, because they apply to both the intervention and comparison groups. A difference-in-differences regression model returns unbiased estimates of the project's effect on learning in this context, because changes in assessment difficulty are exactly equivalent to a temporal shock – it applies only in the second round – that is applied to both groups equally. The difference-in-differences method is explicitly designed to accommodate such shocks.

The absence of a comparison group, on the other hand, undermines this analysis, making it impossible to distinguish between two possible interpretations of changing learning scores. One interpretation places the responsibility for a change in scores on the project itself – if scores are higher at midline, it is because the project caused this outcome. An equally plausible explanation, however, is that the test became easier, and this caused an upward shift in scores.

Unfortunately, the methods designed to prevent these problems are not particularly effective. For instance, as part of the midline and endline evaluations, the different iterations of the learning assessments will be tested for comparable difficulty among a subsample of students, as noted above. But even relatively small differences in assessment difficulty over time represent systematic bias, and testing difficulty on a small subsample of students is insufficiently precise to detect small differences in difficulty. For instance, a "difficulty equating procedure" using a subsample of 50 students will detect a difference in difficulty of two percentage points in learning just 8.1 percent of the time. Increasing the true difference in difficulty

⁴⁹ Assessments in later rounds may also be adjusted to add more difficult subtasks.

to 5 percentage points increases power to just 16.7 percent, far below conventional standards for properly-powered studies. In other words, the equating procedure was insufficiently powered to actually detect differences in difficulty that can be substantively meaningful.⁵⁰

Further, potential changes in the difficulty of the learning assessment undermine the benchmarking setup employed by this evaluation. To guard against maturation or growth effects, we performed learning assessments with a benchmark group of girls in grades 3-5. At midline, changes from baseline to midline in the cohort sample (most of whom have advanced a year in school) will be compared to the differences in scores between, for instance, grade 3 and grade 4 benchmark girls. The benchmark girls provide a basis for assessing what "normal" maturation effects look like, with cohort girls being compared against them. The FM's technical guidance suggests utilizing a modified difference-in-differences procedure, constructing a pseudo-comparison group from the benchmark sample. However, the utility of this method is reduced by the fact that assessment difficulty may change over time. Changes in the cohort sample are calculated using assessments of potentially different difficulty; but the same shift in difficulty does not apply to maturation rates calculated in the benchmark sample, because *all* benchmark girls receive the baseline learning assessment. The modified difference-in-differences approach does not, therefore, maintain the fundamental benefits of the difference-in-differences design (accounting for temporal shocks that apply to all groups equally), because learning scores among benchmark girls are all generated at the same time, at baseline. And it cannot account for changes in assessment difficulty, because the benchmark girls all received the baseline assessment.

The importance of implementing the baseline evaluation prior to the start of programming likely made it impossible to design and pilot three distinct assessments before baseline data collection. We strongly recommend that, prior to the midline, two assessments are developed and piloted. Rather than using one new assessment at midline and a third assessment at endline, we recommend employing both new assessments at midline, randomly assigning girls to receive one or the other. This will not resolve the issue of assessment difficulty for baseline-to-midline comparisons, because both new assessments could be of different difficulty compared to the baseline. However, at endline, girls can be given the assessment they did not complete at midline; the random assignment of girls at midline provides the best possible protection against changes in assessment difficulty from midline to endline, and should facilitate stronger causal inferences about project impact.⁵¹

⁵⁰ Full results of this analysis are available upon request. In short, we performed simulations in which we assumed a baseline assessment score of 60 percent, and a 2-point decrease in difficulty, with learning outcomes distributed normally with a standard deviation of 24 points. The standard deviation selected is a fairly moderate assumption, based on previous education evaluations, but is lower than that observed for Somali literacy scores at the AGES baseline. A higher standard deviation implies even lower statistical power, exacerbating the issue noted here. We employed a paired t-test, as we have done in practice for equating procedures in the past.

⁵¹ Randomization alone does not guarantee equivalent aggregate difficulty between midline and endline, but randomizing such that precisely half of the girls at midline receive each assessment will protect against changes in difficulty under normal circumstances.

Ceiling and Floor Effects

As we discuss in more detail in our presentation of the learning results, the learning assessment for Somali literacy exhibited significant ceiling and floor effects. Ceiling effects are also present in the numeracy assessment, though to a lesser extent. Both ceiling and floor effects can produce bias toward a null result. In other words, both effects can reduce the gains in learning scores we are able to calculate from the data below the true gains that might be observed if students completed a test that assessed them on a more expansive spectrum of difficulty.⁵²

It is worth noting that there is a wide variation in the results of the Somali assessment, with particular subgroups of students reaching atypically high scores. This phenomenon was more marked on some NFE classes and seems to be particularly evident in some locations. As communicated to the GEC Fund Manager in December 2019, there is a likelihood that some of the girls enrolled in such classes may actually be dropouts from upper grades, despite claims to the contrary. While the baseline was conducted five months after classes started, it would be virtually impossible for an illiterate and innumerate girl to achieve these learning levels within such a short period of time. On the other hand, the possibility of free education may have led some girls and their families to misrepresent their schooling levels in order to return to school. Therefore, there is a likelihood that there is no actual ceiling effect – but rather that a subgroup of respondents does not meet the criteria. This issue is discussed further in section “Previous Exposure to Educational Interventions” below.

At the midline, we recommend incorporating more difficult subtasks on both the numeracy and Somali literacy assessments; it may also be worthwhile to add a single easier subtask to the literacy assessment, though it is difficult to envision a subtask that would assess literacy-relevant skills more basic than letter identification. Doing so will not obviate the problem for baseline-to-midline analysis, because that analysis should only employ the set of comparable subtasks. However, adding more difficult subtasks will have benefits for midline-to-endline comparisons, by putting the midline learning assessments on firmer methodological footing.

Panel Attrition

The design of AGES relies on a panel design, in which the same girls are re-contacted from baseline through endline, assessing changes in learning scores and tracking transition outcomes over time. The advantages of this design are significant: in pre-post comparisons, sampling variation is reduced to zero within the panel of girls, increasing statistical power and allowing firmer conclusions regarding project impact, because changes from one round to the next cannot be attributed to random changes in sample composition.

Panel studies in conflict-affected settings are notoriously difficult, however, and few contexts are as difficult as South-Central Somalia, which has a long history of conflict, ongoing environmental distress, and a range of push-pull factors – including cultural traditions – encouraging migration. Re-contact rates in such settings tend to be low, which poses a serious risk to the panel design. Similar studies we have participated in elsewhere in Somalia and Somaliland have typically produced wave-on-wave attrition rates of around 20 percent. However, we would expect somewhat higher attrition rates in the case of AGES, for three reasons: first, sampled girls are newly enrolled in the project schools, and have shallower connections to the schools, which reduces the likelihood they will be successfully re-contacted in the future, both because they are more likely to drop out, and because the school may have less knowledge

⁵² Importantly, the floor and ceiling effects observed are not likely the result of poor test design. Rather, because this is the first time CARE and its implementing partners have worked with this particular group of girls, there was a lack of information regarding their pre-existing reading and mathematics skills. There is also a fundamental tension between designing an assessment that can be implemented in a budget-conscious way and without generating survey fatigue amongst respondents, and designing an assessment that tests a broader range of difficulty levels.

of their location, their families, etc.⁵³ Second, the project is being newly implemented, and CARE does not have as deep of connections in the communities as they do in the context of the Somali Girls Education Promotion Project (SOMGEP and SOMGEP-T), which has been running for approximately six years. Third, the sampled schools are primarily in urban areas; in rural areas, teachers and staff are part of the communities in which they work, and tend to know students and their parents, and where they live; in larger urban settings, girls may travel long distances to school, and these types of interpersonal connections are less common, making it more difficult for field teams to locate girls who are no longer enrolled.⁵⁴

In later rounds of the assessment, girls who cannot be re-contacted will be replaced in the panel to ensure a sufficient sample size for analysis, but these girls may be fundamentally different from those initially selected into the sample. At that time, we will perform parallel sets of analysis: a panel-only, or "true panel" analysis that employs only the set of girls who were contacted in both rounds; and a cross-sectional analysis that includes the "true panel" as well as the girls who were replaced from baseline and their replacements selected at midline. Of course, the process of replacing girls in the panel is fraught, because girls need to be selected in a way that does not produce bias; however, we defer that discussion to the midline evaluation, as replacements are not relevant to the baseline analysis.

Previous Exposure to Educational Interventions

AGES aims to work with the most marginalized girls in South-Central Somalia, focusing predominantly on girls who have not previously been enrolled in any formal education. The cohort girls sampled from formal schools, ABE centres, and NFE centres were all expected to have minimal exposure to previous formal education.⁵⁵ The reality during school visits was often different, as some girls sampled from ABE or NFE centres indicated that they were simultaneously enrolled in formal education⁵⁶. In other cases, girls indicated they had been in school previously and were now re-enrolled or enrolled in the NFE centre.

⁵³ Panel attrition rates among out-of-school girls and those who have dropped tend to be much higher than among girls who remain in school, because teachers and administrators do not systematically track the whereabouts of girls who have dropped out.

⁵⁴ During the baseline, field teams collected GPS coordinates at the households of cohort girls. However, while this is standard practice for panel studies, GPS coordinates are not the most common method by which our teams traditionally locate households of girls. In practice, it is more common for field teams to inquire at the schools in question, where they normally receive significant assistance locating the girls in question, thanks to the knowledge of school staff.

⁵⁵ This fact partially explains the decision to design a somewhat easier set of learning assessments.

⁵⁶ Overall, 16% of caregivers of NFE and ABE girls (101 cases of NFE girls and 102 cases of ABE girls) indicated that girls were simultaneously enrolled in a formal school. Also, 53% of ABE and NFE caregivers indicated they girls have been previously enrolled in a formal school, representing 325 cases for NFE girls and 327 cases for ABE girls. These numbers

As part of the standard household survey, caregivers were asked whether their girl had ever attended school in the past, and 54.5 percent of respondents indicated that they had. However, this result likely overstates the extent of previous enrolment, as caregivers may have understood the question to include *current* enrolment. Previous exposure is not limited to formal schooling, however. In response to surprisingly high learning scores midway through fieldwork, the evaluation team incorporated follow-up questions targeted at girls who achieved scores of 50 percent or higher on either assessment. First, girls were asked where they had learned to read or do maths: a significant share of the cohort indicated they had learned one or both skills during a previous project.⁵⁷ There is a likelihood that many of these girls may have been exposed to previous interventions on emergency education, which typically run for a short period of time due to funding limitations; additionally, there is a possibility that girls have been exposed to education previous to displacement, particularly in the case of those who moved during the 2016-2017 drought. While this is not a population-level study and findings should not be generalised, the findings also suggest that some assumptions regarding access to education in south central Somalia may not be correct, indicating that there may be a pattern of intermittent enrolment rather than complete exclusion from education.

In practice, previous exposure to education does not present a specific methodological threat to the evaluation. While AGES aims to enrol the most marginalized girls, it is not clear that exposure to an education project means a girl is not extremely marginalized, as the vast majority of girls had not been enrolled in formal education previously. Moreover, even if girls have been exposed to previous education, this will not influence the inferences drawn regarding project impact, except insofar as it may be a driving force behind the ceiling effects observed at baseline.

Social Desirability Bias in Survey Responses

A core problem of public opinion research is soliciting honest and forthright answers to survey questions that may have social stigma attached to them. It is important to note that social stigma relates not only to community norms or expectations, but can also be generated through the expectations – or perceived expectations – of enumerators or their sponsoring organization. The well-known "enumerator effect" can bias results when characteristics of the enumerator drive changes in response patterns – this can occur for any number of reasons, only one of which is the possibility that the respondent believes the enumerator will judge their response in some way.

Overall, social desirability bias is a relatively minor problem in this evaluation. Few of the outcomes are likely to involve socially-desirable or socially-stigmatized behaviours or attitudes. However, where social desirability may play an outsized role is in the analysis of community attitudes toward girls' education. Respondents likely know that expressing support for girls' education is the socially desirable response, and may alter their responses because they know community norms, the expectations of the sponsoring organization, or the expectations of the enumerator dictate it. For instance, we spend considerable time analysing the extent to which caregivers aspire for their daughters to attend university. Admitting that one has a preference for their daughter to *not* attend university may be considered a typical socially-stigmatized response.

In general, our approach to this issue is to triangulate responses across a number of questions, which vary in the extent to which they are hypothetical versus tangible, and the extent to which they ask respondents

remain indicative, as caregivers could be referring to other types of education or may not distinguish formal schools from ABE or NFE while responding to the question.

⁵⁷ In total, 254 girls reported learning reading skills from a project, while 352 reported gaining numeracy skills from a project. We do not report percentages, because only a subset of girls were asked these questions – those who scored well on one of the assessments, and who completed the assessment in the latter half of fieldwork. Nonetheless, the sheer quantity of girls exposed to education via previous projects is noteworthy.

about themselves versus their community more broadly (or employ generic terms, in lieu of asking about a respondent's personal opinions). With that said, greater efforts could be made to incorporate questions that specifically seek to reduce bias, by posing specific hypothetical scenarios, asking respondents about the attitudes of other people, and triangulating more across other respondents by, for instance, asking girls to assess the attitudes of community members. In preparation for the midline, a review of the tools and previous evaluations of girls' education projects should be conducted to identify promising approaches to these attitudinal questions.

Inaccuracy of School Enrolment Records

One of AGES' intermediate outcomes is attendance rates in formal schools, ABE centres, and NFE centres. Our primary indicator of attendance rates are physical headcounts performed by team leaders during visits to the schools. Calculating attendance rates from headcounts is vastly preferable to relying on school attendance records, which tend to be inaccurate and largely incomplete. However, calculating attendance rates from physical headcounts still relies on the school providing accurate enrolment numbers per class, as this is the denominator against which the physical headcount is measured. Enrolment counts tend to be the most accurate records kept by most schools, because they are not completed daily (as with attendance records), because they influence funding from communities, NGOs, or government, and because they are tied to the collection of school fees. Nonetheless, enrolment counts may not be completely accurate, in which case attendance rates based on them will also be inaccurate. Our expectation is that any inaccuracy in enrolment rates will not systematically change from baseline to midline, but this may be unfounded. We recommend, at midline, that additional measures of attendance be incorporated in the household survey (assessing recent attendance rates for each cohort girl) and, possibly, in the teacher survey.

3. DESCRIPTION OF BASELINE SAMPLE

3.1 COHORT DEMOGRAPHICS

In this section, we describe the characteristics of the cohort sample in general terms, highlighting issues that are likely to impact analysis throughout the report. In the section that follows, we empirically assess the extent to which the sample is representative of the broader population of beneficiaries, based on comparisons to the sample frame.

The table below describes the composition of the cohort sample in terms of respondent type, state, and age. For clarity, respondents are broken into three categories, depending on the type of learning institution in which they were enrolled at baseline.⁵⁸ In contrast to the plan early in the inception stage – which entailed equal allocation of interviews across the three respondent types – the sample is skewed toward NFE and ABE girls, with comparatively fewer (29.6 percent, versus 36.3 and 34.1 percent,

⁵⁸ These categorizations are important on an individual level, as transition rates at midline and endline will be calculated based on girls' "starting points" at baseline. For instance, formal school girls who transition into ABE centres would not be considered successful cases of transition, whereas ABE girls who transition into formal schools would be.

respectively) girls enrolled in formal schools. It is important to consider this sample composition when reviewing aggregate learning outcomes, as learning scores predominantly reflect girls who are not enrolled in formal schools.

TABLE 11: DEMOGRAPHIC COMPOSITION OF COHORT SAMPLE

Characteristics	Percentage share of the cohort sample (in %)
Learning Institution Type:	
Formal School Girls	29.6
NFE Girls	36.3
ABE Girls	34.1
Age:	
10	10.2
11	6.5
12	12
13	15.8
14	8
15	5.7
16	4.6
17	15.9
18	11.7
19	8.7
20	0.9

Cohort girls range in age from 10 to 20, by design. There does not appear to be any systematic relationship between age and inclusion in the sample, with the exception that very few (0.9 percent of the sample) girls aged 20 years are included (which is to be expected as one of the selection criteria was age 10-19; only a fraction of the 19-year olds would have turned 20 within a short period of time). Otherwise, the age distribution is uneven, but not in any way that suggests issues in sample selection or representativeness.

At the same time, the age distribution is more complicated than the table above suggests. During the project's inception stage, girls were registered and enrolled into different learning streams or institution types on the basis of age. The sampling criteria employed during the baseline evaluation reflected these limits: 10-13 years for formal school girls, 13-17 years for ABE girls, and 17-20 years for NFE girls. This fact has minimal impact of the sample's representativeness of the underlying beneficiary population, because the sampling criteria match the criteria for official inclusion in the program.⁵⁹ The result is that

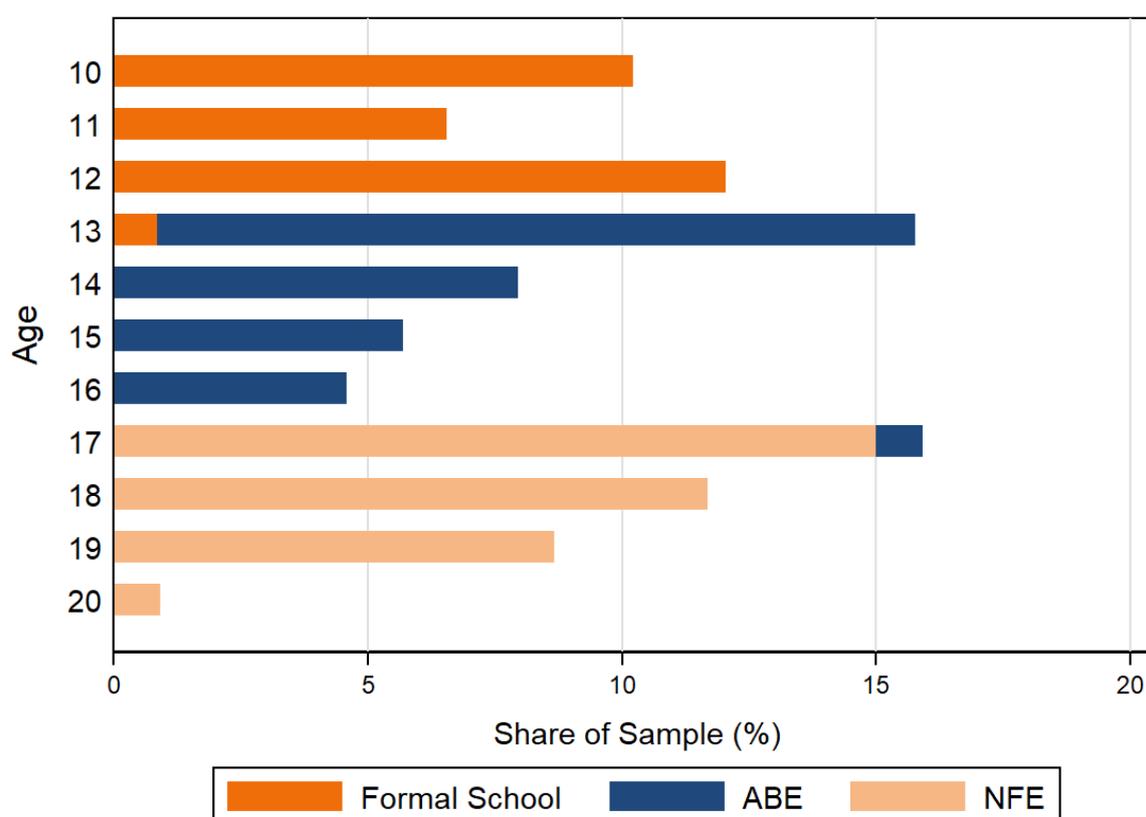
⁵⁹ The sampling criteria are only different from the program's registration criteria in that the sampling criteria add one year of age on the upper bound for each type of girl, because some girls have had a birthday since the registration period. In reality, some girls attending on the day of data collection may have fallen outside of these age ranges, because the program does not prevent girls from attending the school or centre, even if they are not registered (e.g., a girl who is not registered, attending with her sibling who is registered). In addition, some girls may have provided incorrect ages when asked by the evaluation team, as some girls do not know their precise birth day and year. The former issue does not reduce the

there is a very close correlation between respondent type and age, as shown in the figure below. As with the correlation between geographic location and respondent type, strong correlation between respondent type and age threatens to bias any analysis focused on differences in outcomes between respondent types or across age groups. For instance, age is generally a strong predictor of dropout and transition rates; because NFE girls tend to be older, we would expect their transition rates to be lower than other girls – regardless of the project's interventions, these girls face barriers to transition not confronted by younger girls. Throughout this report, our analysis will often control for age or respondent type, in an effort to avoid bias of this kind; in other places, we will perform analysis on subsamples of girls defined by their respondent type, which is a more drastic – and more effective – method to account for differences across subgroups.⁶⁰

representativeness of the sample – it makes it more likely that the sample captures registered beneficiaries. The latter issue may exclude some girls who do not know their age from the sample, but the impact of this on sample representativeness is almost certainly very small.

⁶⁰ In general, estimating statistical tests separately on different subsamples is equivalent to regression models utilizing interaction terms, but with the advantage of being more straightforward to interpret. It is also the case that some statistical tests, such as chi-square and t-tests, do not lend themselves to interaction terms like those used in multivariate regression, which means that performing the tests on distinct subsamples is the only solution available.

FIGURE 2: SAMPLE DISTRIBUTION BY AGE



Beyond geographic location and age, a number of demographic and other characteristics are of interest. However, we limit our attention in this section to disability status, as most other characteristics are discussed extensively elsewhere in this report (e.g., subgroup characteristics and barriers in the sample, later in this section). The evaluation employed the standard Washington Group questions on child functioning to measure the number of girls with disabilities in the sample. The survey module on this topic included 22 questions, addressing physical, cognitive, behavioural and mental health disabilities⁶¹. The module was completed by both girls and their caregivers. The full set of questions applied is available in the annexed version of the household survey, and a slightly broader description of the measurement methodology – useful as a supplement to the discussion here – is provided in Annex 14.

In the table below, we provide disability prevalence rates for the cohort sample, disaggregated by type of disability. Importantly, we provide three estimates of prevalence (columns 2-4), using alternative methods of coding disability from the Washington Group questions. The rates presented in column 2 are the most straightforward and adhere most closely to the coding dictated by the Washington Group. Specifically, the results in column 2 are derived by coding the indicators exactly in line with the Washington Group instructions, based on the responses of caregivers, with a single exception of disabilities indicated as "affect."⁶² We elected to use caregiver responses, rather than those of the girls

⁶¹ CARE also included a question on hand/arm disability (not part of the standard Washington Group questions) as previous field observations suggest that this is a common type of physical disability, potentially resulting from birth issues (considering the high prevalence of Type III FGM) and poorly healed fractures (in a context where access to health services is extremely limited).

⁶² The affect domain covers experiences of anxiety and depression. According to Washington Group guidance, girls should be coded as having an affect-oriented disability if they experience anxiety or depression daily. In consultation with CARE's

themselves, because – at least in the case of younger girls – we consider caregivers to be more reliable respondents – they are more likely to understand the meaning and subtext of the questions, and they are less likely to conceal disabilities due to social stigma, as the questions are not specifically about themselves.⁶³

In general, the prevalence rates reported in the table are extremely low, except for those concerning anxiety and depression (issues of "affect"). Overall rates of non-affect disability, by this standard, are just 0.6 percent. The rate of physical disability is even lower, at just 0.1 percent. In contrast, the best research on worldwide disability rates suggest a cross-national prevalence of 15.6 percent among adults 15 years and older.⁶⁴ Due to a lack of representative household survey or census data from Somalia, there is little evidence on which to base estimates of national disability rates, but there are strong reasons to believe that the World Health Organization's (WHO) estimate of 15.6 percent is likely to be approximately correct. On one hand, disability rates are higher among older populations, which suggests the expected disability rate in our sample of adolescent girls should be lower than the worldwide mean; on the other hand, disability rates are higher in conflict-affected regions, especially those that have experienced long-term instability and internecine conflict, which would suggest the disability rate in our sample should be, overall, higher.⁶⁵

Monitoring and Evaluation team, we elected to use daily *or* weekly experiences of anxiety or depression as our standard, because these rates tend to be underreported. We provide more detail in Annex 14, which describes the complete coding scheme used for all disability types, and how we group types of difficulties.

⁶³ The one exception concerns visual disabilities, which are defined based on responses from the girls themselves, due to an error in the scripted survey's logic/filtering. When caregivers were asked questions regarding their girl's vision, only a subset received the correct follow-up questions. This error was not present in the filter logic when the girls were presented with the same questions.

⁶⁴ World Health Organization. 2011. *World Report on Disability*.

⁶⁵ Evidence from Somalia confirms the idea that overall prevalence rates are higher than worldwide averages, with one study finding that 42 percent of surveyed households in Somaliland contain at least one member with a disability (CEVSI and Handicap International. 2012. "Children with Disabilities in Somaliland: A Knowledge, Attitudes and Practices Household Survey." Available at <https://www.humanitarianresponse.info/sites/www.humanitarianresponse.info/files/assessments/Somaliland%20Children%20with%20Disabilities%20KAP.pdf>). However, another study found that just 4 percent of Somali youth had a disability, highlighting the role of age in disability rates (UNDP. 2012.

At the same time, rates of anxiety and depression are much higher in the sample. Survey questions engaging with these topics are difficult, because respondents may not understand the terminology used and may misunderstand how severe anxiety or sadness should be to meet the threshold for reporting it. As an example, the text of one question asks how often a girl seems "very sad or depressed." Naturally, respondents will have different views of the term "very sad" and apply different standards in their answers. High rates of anxiety and depression would be in line with our expectations, however, given that AGES is being implemented in areas beset by long-running internecine conflict.

It is possible that the low general prevalence of disability among the AGES cohort girls is a result of social stigma or cultural barriers to "admitting" one has a disability, which might be seen as admitting weakness. It is also possible that the cohort sample is non-representative of Somali girls aged 10-19, because it was drawn from within schools, ABE centres, and NFE centres. Children with disabilities are less likely to be enrolled in school or even alternative educational pathways, implying that prevalence rates may be lower among the AGES cohort than among the broader population of girls aged 10-19 years.⁶⁶

TABLE 12: DISABILITY PREVALENCE AMONG COHORT GIRLS, ACROSS ALTERNATIVE CODING SCHEMES

Disability Type	Standard Approach (caregiver responses)	Standard Approach (caregiver or child responses)	Liberal Approach (including "some difficulty")
Vision	0.1% (1)	0.1% (1)	1.9% (27)
Hearing	0.1% (1)	0.1% (2)	0.4% (5)
Mobility	0.0% (0)	0.3% (4)	0.7% (10)
Arms/Hands	0.0% (0)	0.0% (0)	0.1% (2)
Self-care	0.0% (0)	0.1% (1)	0.0% (0)
Communication	0.0% (0)	0.0% (0)	0.2% (3)
Cognitive	0.1% (1)	0.4% (6)	3.7% (52)
Behavioral	0.4% (5)	0.5% (7)	2.8% (40)
Affect	11.4% (162)	15.4% (218)	30.1% (428)
Any Disability	11.9% (169)	16.3% (232)	34.6% (491)
Any Non-Affect Disability	0.6% (8)	1.5% (21)	7.8% (110)

In light of measurement complexities involved, we consider two alternative methods of defining disability. The second option defines a girl as having a disability if either her *or* her caregiver's responses meet the Washington Group definition.⁶⁷ This is a more expansive definition, because girls and their caregivers often disagree about the severity of a possible disability.

Somalia Human Development Report 2012 – Empowering Youth for Peace and Development. Available at <http://www.undp.org/content/dam/undp/library/corporate/HDR/Arab%20States/HDR-Somalia-Factsheet-2012-E.pdf>.

⁶⁶ To be clear, this is not an indictment of the sampling approach, as the sample was not intended to be representative of the full population of adolescent girls. The sample was designed to be representative of the population of AGES beneficiaries specifically.

⁶⁷ Again, we apply the Washington Group standards to all questions except those for affect, where we apply a slightly more liberal definition based on either daily or weekly feelings of anxiety or depression.

The third option is the most expansive, by far, as demonstrated by the prevalence rates reported. In the standard Washington Group definition, a girl has a disability if she has "a lot of difficulty" doing something, such as walking 500 meters, or "cannot do it at all." In our third approach to defining disability, we expand this definition to include girls who have "some difficulty" with those same tasks. In the context of affect-oriented disability, here we apply a standard of daily, weekly, or even monthly feelings of anxiety or depression. Naturally, this coding approach results in significantly higher prevalence rates, at 17.7 percent overall and 2.9 percent in the context of physical disabilities alone.

The results from our three measurement approaches make clear that the reported rate of physical disability is very low, and that the overall rate of non-affect disabilities is fairly low, depending on the definition or standards used. The largest shifts in disability prevalence in the table above come from a liberalization of the standard applied, i.e. when we move from considering only girls who experience a lot of difficulty or cannot complete a task at all, to considering those who can complete a task with *some* difficulty. By comparison, the source of the assessment – caregivers versus girls – does not make a dramatic difference in prevalence. For example, moving from a reliance exclusively on caregiver responses to a consideration of either caregivers or girls increases the non-affect disability rate from 0.6 to 1.5 percent – a change, but not a significant one.

In practice, girls and their caregivers report disability at approximately equal rates, though the specific girls who were classified as having a disability did vary significantly, depending on whether reports from girls or caregivers were used. The table below provides disability prevalence rates as reported by girls and their caregivers, separately. The columns report the following values:

- Girls' self-reports, based on a stricter definition (in line with column 2 and 3 in the table above) that adheres closely to the Washington Group
- Caregiver reports, based on the same stricter definition
- Girls' self-reports, based on a more liberal definition (in line with column 4 in the table above) that includes girls who experience "some difficulty" completing tasks
- Caregiver reports, based on the same liberal definition

The results in the table below suggest that, on average, girls are marginally more likely to respond in ways that classify them as having a disability, when compared to their own caregivers. However, this difference is genuinely small in substantive terms. Where there is any significant disagreement between girls and their caregivers, it occurs in the more liberal coding, specifically in the context of cognitive and mobility disabilities.

TABLE 13: DISABILITY RATES REPORTED BY GIRLS VERSUS THEIR CAREGIVERS

Disability Type	Girl Reports Strict Standard	Caregiver Reports Strict	Girl Reports Liberal	Caregiver Reports Liberal
Vision	0.1% (1)	N/A	1.9% (27)	N/A
Hearing	0.1% (1)	0.1% (1)	0.5% (7)	0.4% (5)
Mobility	0.3% (4)	0.0% (0)	1.5% (21)	0.7% (10)
Arms/Hands	N/A	0.0% (0)	N/A	0.1% (2)
Self-care	0.1% (1)	0.0% (0)	0.3% (4)	0.0% (0)
Communication	0.0% (0)	0.0% (0)	0.9% (12)	0.2% (3)
Cognitive	0.4% (5)	0.1% (1)	0.4% (5)	3.7% (52)
Behavioral	0.3% (4)	0.4% (5)	3.0% (43)	2.8% (40)

Affect	11.6% (164)	11.4% (162)	30.7% (436)	30.1% (428)
---------------	-------------	-------------	-------------	-------------

This analysis provides some support for one of our approaches outlined above: using responses from *either* girls or caregivers to define disability status (i.e. girls are considered to have a disability if either they or their caregiver indicates so). Given the generally similar prevalence rates between the two, girls and caregivers are approximately as likely to report disability, and taking the union of their reports as one measure of disability is a reasonable approach. On the other hand, there is often disagreement regarding the disability status of specific girls – the correlation in classification of affect disabilities, for instance, is 0.81 – which illustrates the potential for measurement problems in the Washington Group questions that go beyond the chosen definition employed here.⁶⁸

Our goal in discussing and utilizing varied definitions is not necessarily to suggest one measure is more valid than the other. In the absence of broader population-based measures or objective verification exercises among a subsample, we cannot empirically assess validity. However, we believe it is instructive to consider the varying definitions, because it suggests the number of children who may have some milder form of disability that is not detected by the Washington Group questions, but which may still impact their life outcomes. To the extent that physical and cognitive disabilities impact learning and transition outcomes, tracking their impact via more liberal definitions of disability should be useful, even if doing so does not fit precisely with standard approaches.

3.2 REPRESENTATIVENESS OF THE SAMPLE

Building on the characteristics of the sample outlined in the previous section, this section reports a deeper assessment of the extent to which the cohort sample for AGES at baseline is representative of the broader beneficiary population. Our analysis is broken into two parts: the first concerns the sample and population of formal school girls, as CARE provided a full listing of girls in this group prior to fieldwork, and the initial sample was drawn directly from this list (i.e. there is a full, traditional sample frame from which the sample was drawn). The sample frame in this case includes extensive details about each girl, including her age, demographic details, and disability status, among other data, allowing us to formally test the similarity between the sample frame and achieved sample.

The second portion of the analysis concerns ABE and NFE girls, for whom a full sample frame was not available. Girls in these cohort groups were selected randomly by Consilient field teams upon arrival at ABE and NFE centres; thus, the sample provides the most detailed information on the beneficiary population available. With that said, we are able to check the geographic distribution of the ABE and NFE samples against the geographic distribution of their populations provided by CARE.

The sample frame of formal school girls includes 6,623 beneficiaries. As mentioned previously, the sample frame provides extensive demographic details about these girls, many of which overlap with data collected from the baseline sample. Our analysis focuses on the overlapping variables, testing their similarity.⁶⁹ A statistical test of the difference between a well-designed sample and its sample frame should yield statistically significant differences in approximately 1 out of every 20 tests.⁷⁰ That is, even if no systematic differences exist between the sample and sample frame – the sample was designed in an unbiased fashion,

⁶⁸ The correlation between responses from girls and their caregivers in the context of behavioural disabilities (using the liberal definition) is lower, at 0.54. The correspondence between responses from children and their adult caregivers is a place where significant further research may be warranted, but it is beyond the scope of this report.

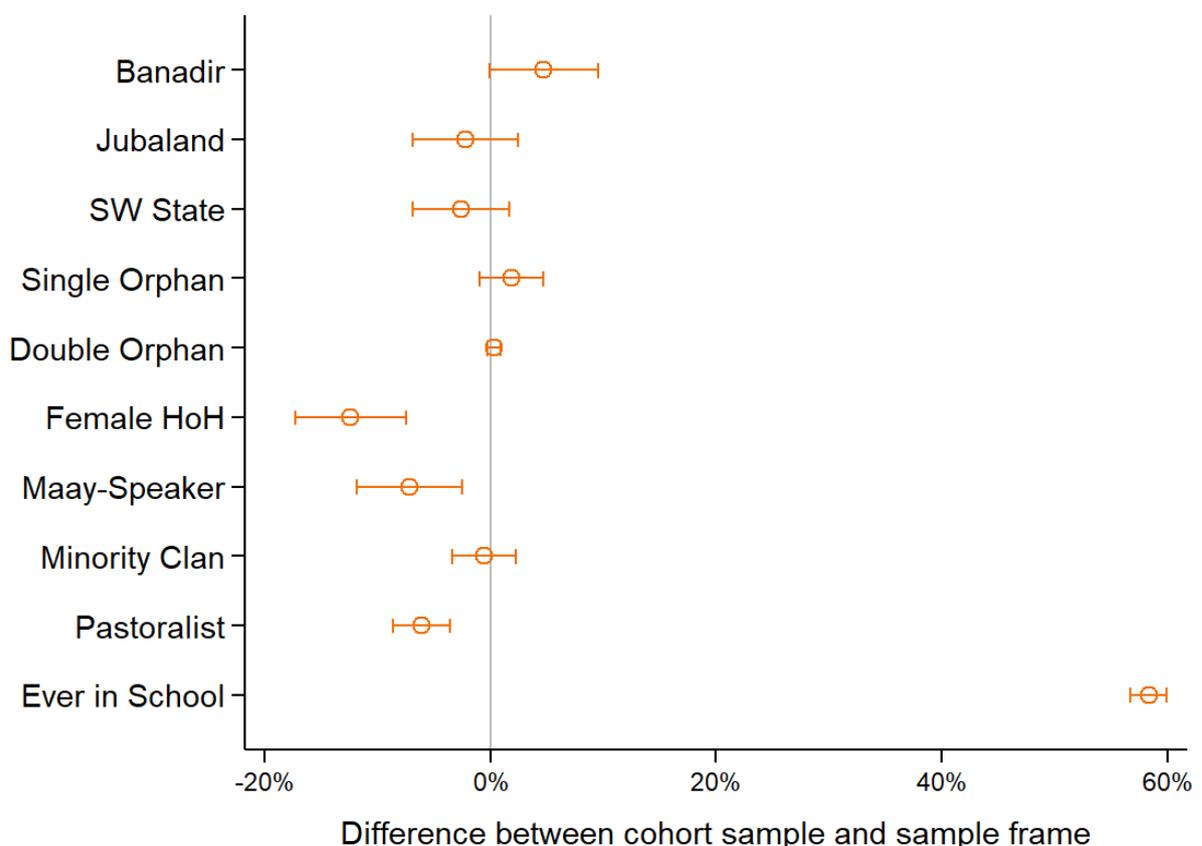
⁶⁹ For this purpose, we test similarity in a regression framework, testing the statistical significance of a variable that distinguishes members of the sample from members of the sample frame.

⁷⁰ Assuming the tests are conducted with a 5 percent (i.e. 0.05) significance level.

and data was collected in line with the design – setting the alpha level to 0.05 means that, across many tests, 5 percent will yield statistically significant differences as false positives. It is important to bear this fact in mind when considering the results below, as one false positive out of every 20 tests is consistent with an unbiased sample.

The figure below reports the difference between the sample and sample frame for a number of key demographic variables, capturing aspects of relative marginalization and/or deprivation. The hollow circles represent the regression coefficient in a bivariate regression, while the horizontal bars represent the 95 percent confidence interval around that coefficient. The vertical line represents a null result, i.e. no difference between the sample and the sample frame. Characteristics in which the 95 percent confidence interval overlaps the vertical line may have different values in the sample and the sample frame, but those differences are too small to be statistically distinguished from a null effect; put differently, the gap calculated is not statistically significant. Note that coefficients falling on the left side of the vertical "null line" are those in which the sample has a smaller share of a characteristic than the sample frame or full beneficiary population. For instance, 47.3 percent of formal school girls in the sample frame come from a female-headed household, compared to 34.9 percent of the achieved sample. The 12.4 point gap is indicated by the coefficient being placed at -12.4 percent on the graph.

FIGURE 3: SAMPLE AND SAMPLE FRAME CHARACTERISTICS



A number of the findings suggest the sample is not representative of the population of formal school beneficiaries. At the most basic level, the sample is disproportionately based in Banadir. Girls in Banadir schools represent 40.9 percent of the sample frame, but 45.6 percent of the cohort sample, a difference that is marginally significant ($p =$

.055). Beyond geographic distribution, the sample appears to underrepresent the number of girls in female-headed households and girls whose families speak Af-Maay at home. Both tend to be indicators of relative disadvantage, which suggest the girls in the sample are less marginalized – at least by these metrics – than the typical girl in the broader beneficiary population. This is reaffirmed by the fact that girls in the sample are much less likely to be pastoralists or from nomadic/pastoralist households – 7.5 percent of girls in the sample frame hailed from pastoralist households, compared to 1.4 percent of the sample.⁷¹ On the other hand, minority groups, such as the groups commonly referred to as Bantu, and coastal groups like the Bajuni, are equally represented in the sample and the sample frame from which it was drawn.

The final result in the figure – and the most compelling – concerns the share of girls who say they have ever attended school. In the sample frame, just 1.3 percent of girls are recorded as having been enrolled in school previously. In contrast, 59.6 percent of caregivers in the sample report that their girl has been enrolled in school. This finding implies that targeting of the AGES program may not have been as effective as expected, since AGES eligibility criteria generally stipulated that girls should not have been enrolled in the school in the past. An important caveat applies to this analysis, however, as the household survey simply asked caregivers whether their girls had ever been enrolled in school, without specifying "other than their current enrolment as part of AGES" – and in the case of ABE and NFE, girls had been enrolled for five months prior to the baseline. In other words, many respondents likely believed that their girl's current enrolment qualified, and answered affirmatively as a result. If prior enrolment is an issue of interest, it may require further data collection to verify the previous educational experience of AGES enrollees.

Data from the sample frame also shed light on our earlier discussion regarding disability prevalence within the sample. Rates of disability are extremely low within the sample – based on Washington Group standards, no girls have physical impairments, and relatively few have cognitive, communication, behavioral, or mental health disabilities. As the table below shows, this is also true of the sample frame – just 0.2 percent of girls in the sample frame report a vision disability, and rates of physical disability are low across the board. While it is possible that both the baseline survey and the project's internal data collection are subject to some form of bias around disability – for instance, if girls and their parents are hesitant to report disabilities – the fact that disability rates in the sample are so low in the broader sample frame suggests disability prevalence may, in fact, be very low among the beneficiary population.

TABLE 14: DISABILITY PREVALENCE AMONG FORMAL SCHOOL GIRLS IN COHORT SAMPLE AND SAMPLE FRAME

Disability Type	Share in Cohort Sample (formal school girls)	Share in Sample Frame (formal school girls)
Vision	0.0%	0.2%
Hearing	0.0%	0.1%
Arms/Hands	0.0%	0.1%
Mobility	0.0%	0.1%
Self-care	0.0%	0.1%
Cognitive	0.0%	0.1%

⁷¹ Note that this finding is based on the share of girls whose head of household engages in pastoralism, and is not a measure of how many girls come from pastoralist clans. In the baseline survey, households were categorized based partially on their membership in pastoralist clans, who tend to hold greater political power and social status in Somalia than the Digil-Mirifle and minority groups. Pastoralism is emphatically different from membership in pastoralist clans, though there is minor overlap between the two characteristics.

Finally, the rich data provided by the sample frame allowed us to test similarities in three additional characteristics of the girls: their age, marriage rates, and motherhood status. The mean age in the sample frame was 11.0 years, compared to 11.1 years in the sample. While this difference is statistically significant, this is primarily a result of the large number of observations available for analysis – the difference is not substantively meaningful, in our view. Given the low age of girls in the formal school cohort, the share who are married and the share who have children is extremely low, and any gap between the sample frame and the achieved sample is too small to be of interest.

Interpreting the results of this analysis should be done with caution. There are three reasons why the sample may not mirror the sample frame, and they are not mutually exclusive:

- Natural sampling variation
- Bias in sample selection
- Differences in responses by the same girls/households in two different rounds of data collection

Natural sampling variation could explain differences between the sample frame and sample, as all samples are merely approximations of the population from which they are drawn. Some degree of sampling variation is always expected. Bias in sampling is also possible. On one hand, we know the precise process by which the sample was selected, as the evaluation team conducted the random sampling themselves. On the other hand, some schools were dropped from the sample after data collection – typically because cluster size requirements could not be fulfilled from the number of available girls in the cohort list, as some girls could not be found – which could result in gaps between the sample and the full sample frame, if dropped schools are fundamentally different from those that remain in the sample. Finally, survey respondents can alter their responses from wave to wave for any number of reasons, including differences in how they understood the questions. Even survey questions with objectively correct answers (such as age) can be subject to such problems.⁷²

Usefully, we are able to quantify the extent to which this latter problem drives our results, because sampled girls were matched against the sample frame provided by CARE. In some cases, responses to our survey match those in the sample frame: very similar shares of girls are identified as members of minority groups (8.3 percent versus 8.6 percent) in the sample frame as in the sample. In other cases, however, there are discrepancies between the data provided by the same girls in the sample as in the sample frame: more girls (32.3 percent) are identified as Af-Maay speakers in the sample frame than in the sample (27.9 percent), and the share of households headed by a woman is substantially higher in the sample frame (47.8 percent) than in the sample (35.0 percent), among the exact same group of girls.

Tellingly, the gaps identified between the same group of girls in the sample and the sample frame mirror very closely the gaps identified between the composition of the sample and the overall composition of the sample frame. This suggests that the gaps in composition between the sample and the sample frame are not exclusively, or even primarily, the result of natural sampling variation or bias in the sample. Instead, much of the discrepancy stems from differences in how girls or their families answered questions or how

⁷² Answers can be influenced by the nature of the questioner (i.e. "enumerator effects") or could be systematically altered by data collectors for the purpose of gaining access to greater project resources.

they were classified. To the extent this is correct, it suggests there are no specific problems with how representative the sample is of the broader population, but that there may be issues with how beneficiaries were classified during initial enrolment.⁷³

Moving beyond the sample of formal school girls, we also investigate the geographic distribution of the ABE and NFE samples. ABE centres were selected through an independent sampling procedure, while NFE centres were selected where they overlapped with the existing formal school and ABE centre samples. Therefore, some degree of discrepancy between the geographic distribution of girls in the beneficiary population and the sample should be expected, especially in the NFE sample.

The table below reports the state-by-state composition of the sample and the sample frame for ABE girls (top panel) and NFE girls (bottom panel). In contrast to the formal school sample, where Banadir was marginally overrepresented, the ABE and NFE samples both show more dramatic departures from their respective populations. The difference in state-level composition between the sample and the broader sample frame are statistically significant in nearly every case; to illustrate, South West state makes up just 32.6 percent of the sample of ABE girls, but the population counts provided by CARE suggest South West state accounts for 47.7 percent of all girls enrolled into ABE centers programme-wide.

TABLE 15: GEOGRAPHIC COMPOSITION OF THE SAMPLE VERSUS THE SAMPLE FRAME

	Share of Sample	Share of Sample Frame
ABE Girls		
Banadir	45.0%	33.3%
Jubaland	22.3%	19.0%
SW State	32.6%	47.7%
NFE Girls		
Banadir	52.2%	64.6%
Jubaland	20.8%	13.5%
SW State	27.0%	22.0%

At first glance, discrepancies around ABE girls are surprising, given that the ABE sample was drawn at the school level, with probability proportionate to size. The sample should, therefore, be self-weighting and representative – within the bounds of normal sampling variation. However, the sample became distorted, at an aggregate level, during fieldwork due to issues encountered with some ABE centres. Specifically, due to concerns surrounding the enrolment/registration of girls, some ABE centres were dropped from the sample during fieldwork. To preserve the target sample size as much as possible, CARE and Consilient agreed to complete additional interviews at ABE centres that were still to be visited, and re-visit some ABE centres where data collection had already been completed, to recruit additional ABE girls into the sample. In general, 5 ABE girls were added at each site visited after this decision was made, and several sites were re-visited.

Unlike the ABE sample, the NFE sample was selected according to its overlap with sampled formal schools and ABE centres. Therefore, it was never likely to match the geographic distribution of the NFE population. Even in the original sample, Banadir was underrepresented relative to the number NFE centres it contains, while Jubaland was overrepresented, largely because Banadir has an outsized share of

⁷³ Important caveats apply to this discussion because, as mentioned previously, survey respondents often give contradictory answers over multiple survey rounds, and this fact does not inherently reflect problems of data quality.

NFE centers, and comparatively fewer ABE centres and formal schools. Thus, if the selection of NFE centres is driven by the population distribution in ABE centres and formal schools, Banadir is underrepresented in the NFE sample. When additional data collection was completed at NFE centres, it did not appreciably alter the sample composition, as it was still driven largely by the geographic distribution of ABE centres and formal schools.

The discrepancies between the samples of each respondent type, and their respective populations, may seem deeply problematic. However, from the perspective of drawing inferences regarding project impact, they are in many ways less severe than one might imagine. Given the panel structure of the design, it will still be possible to make inferences regarding the impact of the project, restricted to the respondents in the sample.⁷⁴ The lack of representativeness has consequences for our ability to draw conclusions regarding impact in the entire population, but has no effect on the *internal* validity of conclusions regarding the sample itself. Finally, given the fact that we employ weights throughout much of our analysis, it is useful to note that the use of weights does not appreciably alter the demographic composition of the sample.

3.3 EDUCATIONAL MARGINALISATION

The intended beneficiaries of AGES programming are ultra-marginalized girls across South Central Somalia, who are influenced by a context of internecine conflict, environmental distress, and an education system that is fragmented at best. The environment in Banadir, Jubaland and South West state described at the outset of this report makes clear the difficulties all girls in the project areas face, regardless of their personal characteristics or household situation. However, AGES seeks to target girls who are especially marginalized, even beyond the fragile situation in which they live. In this section, we expand on our earlier general description of the sample's characteristics, by analysing the extent and nature of marginalization of girls within the sample. Our analysis focuses on identifying demographic subgroups that are at particular disadvantage and likely barriers to educational attainment, and studying the extent to which girls in the sample fall into these subgroups and face these barriers.

The table below breaks down the sample by characteristic subgroups (loosely, demographic characteristics, or traits that are difficult or impossible to change about oneself). As shown previously, the rates of reported disability – for most types of disability – are low in the sample, with just 0.6 percent of girls subject to a disability that is *not* related to anxiety or depression. However, a fairly large share of girls are affected by anxiety and depression: 11.4 percent experience daily or weekly bouts of anxiety or depression, and 30.1 percent experience bouts monthly or more often. Other characteristics are similarly common: 13.0 percent of girls have lost one parent, and a further 1.0 percent are orphans, having lost both parents. A substantial share (7.0 percent) of girls in the sample have been or are currently married, and 5.1 percent of girls have given birth to at least one child. This is not entirely surprising, as the sample includes older girls who have been enrolled into NFE programmes administered through AGES, but – despite their age ranging up to 20 years in a small number of cases – the number of mothers in the sample

⁷⁴ Under the same assumptions necessary for any inferences drawn from a pre-post design, as discussed in the limitations section (e.g., assumptions regarding the stability of outcomes over time, in the absence of the intervention).

has consequences for project implementation and the potential for girls to continue their education, or enter the paid workforce.

TABLE 16: CHARACTERISTIC SUBGROUPS OF MARGINALIZED GIRLS

Subgroup	Proportion of sample with this characteristic
Vision disability	0.1%
Hearing disability	0.1%
Cognitive Disability - trouble remembering, understanding, etc.	0.1%
Behavioral Disability - trouble making friends, or controlling behaviour	0.4%
Affect Disability - frequent anxiety/depression	11.4%
Any type of disability	11.9%
Any non-affect disability	0.6%
Girl has only one living parent	13.0%
Girl has no living parents	1.0%
Girl does not live with either parent in her HH	5.8%
Girl is currently married	3.7%
Girl has ever been married (even if divorced now)	7.0%
Girl has a child	5.1%
Girl lives in a pastoral HH	1.8%
HoH belongs to agricultural clan	42.1%
HoH belong to marginalized group or occupational minority	10.1%

The final three characteristics described in the table require explanation for readers not intimately familiar with the Somali context. Broadly, Somalia's four largest clan groupings are divided into those that are traditionally pastoralist (Darood, Dir, Hawiye, and Isaaq), which are occasionally referred to as the "noble pastoralist" clans. The fifth major clan is the Digil-Mirifle (occasionally referred to as the Rahanweyn), which is a largely agro-pastoralist clan that is – relative to the four pastoralist clans – politically marginalized. Finally, a large number of other minority groups, comprised of non-Somali ethnic groups, exist largely at the margins of Somali society. These groups are typically referred to under the blanket term "minority groups."⁷⁵

This brief discussion makes clear that belonging to the "agricultural clan" (Digil-Mirifle) can be seen as a source of marginalization, especially depending on the particular region in which one lives. More universally, belonging to a minority group is an indicator of economic, political, and social marginalization. At the same time, households that engage in pastoralism also exist in a fragile economic position, given the range of environmental and other stresses that impact their livelihoods. It is important to distinguish between membership in a pastoralist clan – a marker of relative political inclusion and social status – and engagement in pastoralism as a livelihood – a predictor of relative poverty and insecurity. In

⁷⁵ Somalis of Bantu descent are the largest such group, but it includes coastal groups that are descended from non-Bantu peoples as well.

practice, many of the households engaged in pastoralism in our sample are actually members of the Digil-Mirifle or minority groups.

Beyond demographic characteristics, girls in our sample face a number of barriers to educational attainment, many of which are a function of their family/household structure, the economic environment in which they live, cultural traditions and norms, as well as the limitations of the schools in their respective areas. The table below describes less indelible characteristics of the girls in the sample, but characteristics that typically represent significant barriers to achieving a full education.

In terms of girls' household situation, most girls in the sample are being raised by parents who have not received any formal education. The vast majority (81.8 percent) of girls live in households where the head of household never attended formal school of any kind; in fact, 78.6 percent of girls live in households where *neither* their caregiver nor their head of household ever attended formal schooling. This is a particularly problematic finding, as parents who have completed at least some formal schooling are generally more supportive of educational attainment for their children. Children living in households without parents who have received a formal education of any kind may have less appreciation for its value and no readily available role model that can promote education as a worthwhile pursuit, a problem exacerbated in the case of girls, who face a dearth of educated female role models in Somali society in general.

The next barriers are primarily economic in nature, and make clear that many girls exist in a state of economic insecurity, even if most do not experience routine and persistent hunger, lack of water, and so forth. For instance, just 7.0 percent of households in the sample report that members went to bed hungry most or all nights during the last 12 months, but the vast majority of households did experience hunger at least once or twice during the year – only 20.3 percent of households stated that they *never* experienced hunger of this kind during the last 20 months. Even more problematic is the fact that so many households live in a state of roughly continual economic insecurity: just 1.9 percent of caregivers interviewed had any savings, and just one-third of households own land of any kind. In the regions where AGES is being implemented, insecurity of land tenure is a significant problem, as displacement due to conflict or drought can lead to dispossession.

TABLE 17: MAIN BARRIERS FACED BY MARGINALIZED GIRLS

Barrier	Proportion of sample affected by this barrier
Female-headed household	37.3%
HoH has no education of any kind (no Quranic)	21.3%
HoH has no formal education (may have Quranic)	81.8%
Caregiver has no education of any kind (no Quranic)	27.0%
Caregiver has no formal education (may have Quranic)	85.1%
HH has poor roof (mud/thatch/cardboard/plastic)	20.0%
Went to sleep hungry some/many/most nights, last 12 months	27.1%
Went without water for home use most/all days, last 12 months	5.2%
Went without medicines most/all days, last 12 months	16.8%
Went without cash income most/all days, last 12 months	14.8%
HH owns land, either solely or jointly	32.5%

Caregiver has no savings	98.1%
HH owns phone	86.5%
HH owns smartphone	18.0%
HoH does not have occupation or does not earn a wage	39.3%
Girl spends a few hours or more per day doing HH chores	64.7%
Caregiver aspires to university education for girl	87.4%
Caregiver believes girls' education worthwhile, even if funds are limited	87.7%
Caregiver believes work or HH chores are acceptable reason to not attend school	40.1%
Caregiver believes cost of education is acceptable reason to not attend school	62.3%
There isn't enough money to pay the costs of girl's schooling	56.4%
No computers available for use at school	95.7%
Girl cannot use books/learning materials at school	26.2%
Textbooks are shared between students	39.9%
Girls are able to take textbooks/materials home at night	34.7%
Not enough seats for every student in class	13.5%
School has reliable electricity	66.7%
School has water access within 1 km	54.6%
School provides at least one meal for students	28.5%
School has only cement floors (no dirt)	67.7%
School has separate toilets for girls	71.5%
School charges school fees	51.8%
My teacher's lessons move too fast for me	46.0%
Teacher used corporal punishment in last week	21.1%
No female teachers, either FT or PT	15.2%
Girl does not feel safe traveling to school	1.8%
Caregiver does not feel it is safe for girls to travel to the school	1.6%
It is unsafe for girl to be in school	5.5%
Child says teachers mistreat her at school	4.3%
Girls does not feel she can attend school if she is menstruating	9.6%

Beyond generalized economic insecurity, many girls face additional financial barriers to enrolling in school. These financial barriers tie into broader issues of household priorities and trade-offs between paying for schooling and alternative uses of limited funds. Most caregivers interviewed value educational attainment for their girl in the abstract, but simultaneously acknowledge that financial constraints may force them to keep their girl out of school. The results in the table above are stark: 87.4 percent of caregivers hope their daughter will complete university education, and 87.7 percent believe investing in

girls' education is worthwhile, even in the context of limited funds. However, in the same set of respondents, 62.3 percent of caregivers said that the financial cost of schooling is an acceptable reason to keep one's child out of school. Given that over half of respondents say their household has insufficient funds to pay for their girl's education, this is particularly problematic, as it is clearly a trade-off faced by most households. Perhaps more disconcerting is the fact that most caregivers aspire for their daughter to complete university education, but simultaneously place a heavy chore burden on their shoulders, preventing them from studying and almost certainly impacting their attendance rates.

Finally, the last set of barriers described in the table focus on issues girls may face at their schools or educational institutions. Unsurprisingly, the schools in question face a number of infrastructural and financial limitations: just two-thirds of schools have reliable electricity available, and just over half (54.6 percent) have access to water within 1 kilometer of the school – which often produces an additional burden on female students, who are frequently tasked with collecting water for the school. Within classrooms, many girls indicate that their teacher's lessons move too quickly for them, and 21.1 percent of girls reported that their teacher used physical punishment in the classroom within the last week. Teaching practices which discourage girls who have fallen behind and that inhibit a welcoming classroom environment can reduce attendance rates, impinge on learning, and even encourage girls to dropout altogether.

In light of the context of AGES programming, the share of caregivers who cite security concerns around their daughter's attendance at school is lower than we would have expected. However, a number of girls indicate that they are mistreated at school, and the overall security situation may be underreported, because girls and their caregivers may have become accustomed to the insecure environment in which they live. It is worth noting that security concerns were broadcast clearly during focus group discussions and participatory risk mapping exercises conducted at baseline. FGDs with CEC members often mentioned security provision as a key way in which the government helps schools, indicating that insecurity is a legitimate barrier to attendance or enrolment and that security is not a foregone conclusion. Even more telling were responses given by girls themselves, who cited a wide range of security concerns in a wide variety of areas around their schools and towns, including outright conflict between warring parties. Girls also emphasized the threat of violent sexual assault in some parts of their towns, in areas that seemed familiar to all of the girls present, and discussed petty crime, the risk of being hit by passing cars, and a range of other safety issues.

3.4 INTERSECTION BETWEEN KEY CHARACTERISTICS SUBGROUPS AND BARRIERS

In the previous section, we illustrated the myriad barriers that girls in South Central Somalia face with regard to enrolling in and attending school. We also discussed, previously, the extent to which girls in the sample face a number of life disadvantages – orphanhood, early marriage, and being raised in households that are economically insecure, at best. Many of the girls in the sample – and, indeed, in South Central Somalia more generally – face significant barriers to educational attainment; but these barriers are often exacerbated by the baseline marginalization they face as a result of their clan heritage, the deaths of their parents, or other broader demographic factors.

In the table below, we disaggregate barriers to educational attainment by key demographic characteristics, identifying places of overlap that present particularly pernicious obstacles to education. The figures in the table below indicate the percentage of girls with specific characteristics (subgroup along the top) who also face specific type of barrier (subgroup along the left side). For instance, girls who have lost one parent are more likely to live in households in which neither their head of household nor their caregiver has any formal education.

These results do not show a single group of girls who are uniquely disadvantaged, but they do demonstrate a general trend for certain classes of girls to face specific sets of barriers. As an illustration, girls hailing from minority groups are slightly more likely to have parents who received a formal education. But, at the same time, they have more than double the likelihood of having experienced persistent hunger over the past year (30.1 percent of minority households have experienced hunger many, most or all days in the last year, compared to 7.0 percent in the overall sample). This group generally faces greater economic insecurity: they are far less likely to own land, and are somewhat less likely to have savings. Given that they also exist outside the informal, but generally effective, clan-based system of social insurance and protection, their economic insecurity is even more pronounced than these numbers suggest.

Married girls and mothers also show interesting patterns with respect to marginalization. These girls are not particularly disadvantaged financially, relative to the broader sample from which they were drawn. However, they do face a higher chore burden than other girls, as do girls who reside in a one-parent household. Clearly, responsibility for household chores fall heavily on girls who are married, as they are now responsible for keeping a household or – in some cases – helping their mother-in-law maintain a household. In the case of girls in one-parent households, a heavier burden also falls on girls, as they become responsible for things like household upkeep and cooking, while their sole parent works outside the home.

TABLE 18: KEY BARRIERS TO EDUCATION BY CHARACTERISTIC SUBGROUPS OF MARGINALIZED GIRLS

Barrier	Girls with disabilities	Single orphans	Married girls	Mothers	Agricultural clans	Minority groups
Female-headed household	30.2	66.3	42.4	46.6	31.8	39.9
HoH has no education of any kind (no Quranic)	23.3	27.2	21.2	19.2	20.1	30.1
HoH has no formal education (may have Quranic)	74.4	85.3	82.8	82.2	86.1	78.3
Caregiver has no education of any kind (no Quranic)	23.3	29.9	22.2	19.2	26.6	32.9
Caregiver has no formal education (may have Quranic)	76.7	86.4	84.8	83.6	88.6	81.8
HH has poor roof (mud/thatch/cardboard/plastic)	34.9	16.3	24.2	23.3	24.7	18.2
Went to sleep hungry many/most/all nights, last 12 months	87.5	31.0	21.2	15.1	38.6	30.1
Went without water for home use most/all days, last 12 months	14.0	6.0	7.1	5.5	6.2	12.6
Went without medicines most/all days, last 12 months	20.9	22.3	19.2	19.2	20.7	19.6
Went without cash income most/all days, last 12 months	20.9	20.7	6.1	6.8	18.7	23.1
HH owns land, either solely or jointly	34.9	24.5	31.3	31.5	34.9	18.9
Caregiver does not have savings	100.0	97.8	100.0	100.0	97.3	99.3
Household owns a phone	90.7	91.3	83.8	84.9	88.3	82.5
Household owns a smartphone	25.6	20.1	18.2	16.4	18.4	11.2
HoH does not have occupation or does not earn a wage	41.9	37.0	37.4	37.0	40.3	26.6
Girl spends a few hours or more per day doing HH chores	58.1	77.7	92.9	93.2	64.4	76.2
Caregiver aspires to university educ for girl	86.0	86.4	79.8	79.5	83.9	90.2
Caregiver believes girls' educ worthwhile, even if funds are limited	79.1	85.3	89.9	91.8	80.4	86.7
Caregiver believes work or HH chores are acceptable reason to not attend school	39.5	22.8	28.3	26.0	40.3	30.8
Caregiver believes cost of education is acceptable reason to not attend school	69.8	63.0	62.6	58.9	66.6	61.5

There isn't enough money to pay the costs of (name)'s schooling	65.0	57.4	52.4	46.9	59.4	45.5
No computers available for use at school	100.0	97.8	96.0	97.3	95.8	96.5
Girl cannot use books/learning materials at school	23.3	23.9	29.3	28.8	20.7	31.5
Textbooks are shared between students	22.2	37.8	100.0	N/A	28.6	41.7
Girls are able to take textbooks/materials home at night	33.3	26.7	N/A	N/A	26.1	36.1
Not enough seats for every student in class	7.0	19.0	36.4	37.0	6.7	34.3
School has reliable electricity	77.8	57.8	N/A	N/A	75.2	86.1
School has water access within 1 km	33.3	53.3	100.0	N/A	65.8	63.9
School provides at least one meal for students	11.1	33.3	100.0	N/A	11.8	41.7
School has only cement floors (no dirt)	55.6	57.8	100.0	N/A	82.6	61.1
School has separate toilets for girls	66.7	68.9	N/A	N/A	75.2	77.8
School charges school fees	55.6	33.3	N/A	N/A	59.6	41.7
My teachers lessons move too fast for me	48.8	61.4	47.5	52.1	58.4	32.9
Teacher used corporal punishment in last week	23.3	23.4	9.1	8.2	17.1	29.4
No female teachers, either FT or PT	33.3	13.3	100.0	N/A	13.0	22.2
Girl does not feel safe traveling to school	N/A	3.3	2.0	2.7	2.3	1.4
Caregiver does not feel it is safe for girls to travel to the school	0.0	4.3	1.0	N/A	1.8	2.1
It is unsafe for to be in school	25.0	3.2	2.4	3.1	10.3	N/A
Child says teachers mistreat her at school	5.0	1.1	2.4	3.1	8.2	4.5
Girls does not feel she can attend school if she is menstruating	N/A	10.8	6.3	8.3	8.8	3.1
Number of observations	8	184	99	73	598	143

Even the schools into which marginalized girls enrol can be fundamentally different from those of their peers. Girls from minority clans are more likely than the typical girl in our sample to attend schools that have reliable access to electricity. It is important to interpret some of these results with caution, though, as they can reflect the urbanicity of the area or wider structural features of the area; they may also reflect the attendance in schools set up in IDP areas, typically by international organisations, which may have better infrastructure. Better indicators of a school that is well-resourced are the extent to which students are required to share textbooks and whether there are sufficient desks for all the students. Among girls who attend school, those from minority groups are more likely (41.7 versus 39.9 percent) to attend a school that requires them to share their textbooks, and they are much more likely (34.3 percent, compared to just 13.5 percent in the overall sample) to attend a school with too few desks for all the students. Girls from minority groups are less likely to have to pay school fees to attend school, but there may be a trade-off in terms of quality for this privilege, as minority girls attend schools where physical punishment is more commonly used by their teachers, and are more likely to attend schools that have no female teachers of any kind.

The results presented in the table make clear that many of the characteristics typically associated with political and social exclusion in Somalia – disability status, membership in an agricultural clan, and membership in a minority group – are not strictly vehicles for social exclusion. Rather, they are also correlated with additional barriers to girls' education and the wider opportunities and life chances of girls. Girls in these groups often face greater household poverty and economic insecurity, and these barriers trickle down to educational outcomes, from lower enrolment rates to, potentially, enrolment in less-resourced schools. Moreover, these girls are more likely to face discrimination or bullying, as the final results in the table above seem to suggest. We return to this theme when we discuss CECs and school management in Section 7.4, but note here that the evidence is suggestive regarding broader mistreatment of girls according to their heritage and membership in particular social or ethnic groups. To the extent that AGES seeks to benefit the most marginalized girls, where marginalization is a vector of overlapping and non-mutually exclusive characteristics and barriers, targeting girls from socially-excluded groups such as the Bantu minorities and the Digil-Mirifle, among others, is an effective selection strategy.

3.5 PROFILES OF COHORT GIRLS

The discussion throughout this introductory overview of the sample and its characteristics has occasionally hinted at the possibility that the three groups of cohort girls may be systematically different from one another. Given their differences in age and, especially, the fact that they were drawn from entirely different types of learning institutions, it is natural to think that the girls may have fundamentally different characteristics, household backgrounds, or life experiences leading up to this point. In this section, we briefly profile the three types of girls and their socio-demographic characteristics. Elsewhere in the report, we spend considerable time discussing the different respondent types and their learning scores and YLI scores. Our interest in this section is simply to describe the girls as they are at the project's outset, to highlight any major differences between them that may be useful for project targeting, for understanding the results later in this report, or for understanding the "typical" girl of a given respondent group.

In the table below, we report demographic and household characteristics for FE, ABE, and NFE girls. In most ways, the girls have similar backgrounds and face similar barriers, overall. For instance, the share of girls whose caregivers and head of household have no formal education – or no education at all – is not dramatically different, and experiences of hunger and other indicators of financial marginalization (such as going without water or a cash income) are similar across groups.

There are several notable differences between the girls, however. First, NFE girls are much more likely to be married and a much less likely to live with neither parent, which are likely related outcomes –

married girls are less likely to live at home – and are clearly a function of their older age. Similarly, NFE girls are much more likely to have a child.

TABLE 19: PROFILES OF COHORT GIRLS

Main characteristics (in %)	FE	ABE	NFE
Girl has only one living parent	10.7	10.1	17.5
Girl has no living parents	0.7	1.2	1.0
Girl does not live with either parent in her HH	3.6	5.0	8.5
Girl is currently married (not divorced/separated)	0.2	0.8	9.3
Girl has ever been married (even if divorced now)	0.2	1.4	17.7
Girl has a child	0.0	0.4	13.8
HoH belongs to pastoralist clan	52.7	38.8	50.7
HoH belongs to agricultural clan	38.2	48.3	39.4
HoH belongs to a marginalized group or occupational minority	8.6	11.6	9.9
Female-headed household	34.9	33.7	42.7
HoH has no education of any kind (no Quranic)	22.1	22.1	19.8
HoH has no formal education (may have Quranic)	80.8	82.9	81.7
Caregiver has no education of any kind (no Quranic)	29.0	27.9	24.5
Caregiver has no formal education (may have Quranic)	85.7	86.0	83.9
HH has poor roof (mud/thatch/cardboard/plastic)	18.1	22.1	19.6
Went to sleep hungry most/all nights, last 12 months	6.9	7.9	6.2
Went without water for home use most/all days, last 12 months	4.8	6.2	4.7
Went without medicines most/all days, last 12 months	14.5	16.5	18.8
Went without cash income most/all days, last 12 months	13.5	18.0	12.8
HH owns land, either solely or jointly	38.0	31.40	29.1
Caregiver has savings	2.6	1.70	1.6
Household owns a phone	84.8	86.40	88.2
Household owns a smartphone	20.7	13.20	20.2
HoH does not have an occupation or does not earn a wage	37.1	36.60	43.7
Household is engaged in pastoralism	1.4	2.7	1.2

Beyond these age-related factors, the starkest differences between the groups arise in terms of clan membership and land ownership. ABE girls are significantly more likely to be members of an agricultural clan (Digil-Mirifle) than either FE and NFE girls and are very slightly more likely to be members of

minority groups, though there is no obvious reason why this should be the case.⁷⁶ In terms of land ownership, FE girls are much more likely to live in households that own land; again, it is not entirely clear why this should be the case, though FE girls are the least likely to be members of minority groups, who tend to have the most tenuous land claims in the areas targeted by AGES.

Perhaps the most compelling finding from this brief profile is that there are relatively few systematic differences between FE, ABE, and NFE girls, outside of characteristics directly driven by their large age differences. In some ways, this suggests that girls who have been enrolled in formal schools through AGES (i.e. FE girls) would, in the absence of AGES programming, likely become equivalent in many ways – marrying early, remaining out of the school – to NFE girls. In this sense, FE, ABE, and NFE girls are arguably on the same life track, but at different stages – if this is the case, AGES has a clear opportunity to substantially change the life outcomes of FE and ABE girls, moving them one trajectory to another.

3.6 APPROPRIATENESS OF PROJECT ACTIVITIES

The discussion in this section addresses the project’s Theory of Change, its planned activities, and assesses whether they are well-suited for the context in which the project is being implemented and the goals the project has set for itself. In short, this section provides feedback to the project on whether its selected activities are appropriate. Note that this discussion often connects back to the analysis of marginalisation and cohort demographics in the previous three subsections, to shed light on whether the project is targeting the right *types* of girls, and whether the project is addressing barriers that actually predict lower educational attainment and performance in this cohort.

At the same time, this section also casts a broader net, in that we also provide feedback on project activities and whether they are likely to bring about the change the project desires, beyond simply whether they are targeted correctly. We assess whether those activities are appropriate and fit-for-purpose, in two senses: first, that the project activities and the Theory of Change that underlies them are logical solutions to the types of barriers that girls in the area face and can be expected to bring about the changes the project seeks to make; second, whether the project activities are oriented toward the particular groups of girls that are most marginalized or face the barriers in question most acutely. We use this analysis to, wherever

⁷⁶ It is worth noting that some differences between groups could be simply coincidental, i.e. false positives. Across many statistical comparisons between two groups, in the absence of any *true* difference between them, 5 percent of all comparisons will show a statistically significant difference. Therefore, given that there are many comparisons being made in this table, we should be cautious about drawing firm conclusions regarding differences between the respondent types, especially where they constitute one-off differences (e.g., a difference on one variable, but not on other, correlated variables that suggest a theme).

possible, make constructive suggestions regarding intervention design and targeting, to bring both into line with the goal of targeting the most marginalized girls in a cost-effective manner.

Beneficiary targeting

In our sample, a large proportion of girls came from agricultural clans and marginalized groups, while we have identified a number of married school-going girls.⁷⁷ Around 14% of the participants are single or double orphans and 6% are not living with their parents, and therefore more at risk of having lower educational opportunities. Besides, 48% of the sampled girls is facing at least one of the following forms of exclusion: belonging to a minority; living in a food-insecure household; being an orphan; having a form of disability⁷⁸; being married or divorced. 17% of the girls face two or more of those barriers, therefore experiencing more extreme forms of marginalisation.

In the preceding sections, the analysis pointed at specific characteristics indicating higher educational marginalization, such as girls living with disabilities, married and divorced girls, girls engaged in child labour and girls from marginalized and minority groups and clans. Our research also identified specific barriers to education that different groups of marginalized girls are more likely to face. For instance, girls from minority clans are more likely to live in economically more insecure households and married girls and girls who are mothers are more likely to face higher chores burden, increasing opportunity costs of pursuing further education. Moreover, the previous results on the intersectionality between disability and key sub-groups indicate that girls with disabilities are more likely to face higher economic vulnerability and food insecurity.

Disaggregation by state shows that cohort girls from agricultural clans were mostly located in South West State, representing 58% of the girls from agricultural clans. The results remain robust even after controlling for other regions and respondent groups, suggesting that girls from agricultural clans are significantly more likely to be identified in South West State. Girls in South West State in formal schools were 0.30 percentage points and NFE girls were 0.20 percentage points more likely to belong to agricultural clans, compared to ABE girls. While the biggest share of cohort girls from marginalized groups and occupational minorities was found in Banadir, there were no statistically significant regional differences found, but girls in formal schools were significantly more likely come from marginalized groups or occupational minorities. On the other hand, while the highest share of girls who are mothers and single orphans was observed in Banadir, regression analysis controlling for different cohort groups does not show any statistically significant across states.

In the previous sections, we have also identified girls with disabilities, especially girls suffering from depression and anxiety on a weekly or monthly bases as another possible indication of marginalization. Disaggregation by state suggests that girls in the sample are more likely to suffer from mental health

⁷⁷ To be clear, school-going in the context of our sample can include enrolment at ABE or NFE centres through the AGES programme. In practice, the married school-going girls in question are invariably in the older age range of our sample, and are enrolled in NFE; this certainly does not mean that younger girls are never married, nor does it mean that no married girls are enrolled in formal school or ABE. Rather, the sample selection criteria mean that NFE girls are the oldest cohort in our sample, and, naturally, the most likely to be married.

⁷⁸ Considering all levels of functionality loss - “some level of difficulty”, “a lot of difficulty” and “cannot do at all”.

problems in Jubaland, compared to other states. The regression analysis suggests that girls in Jubaland were significantly more likely to suffer from depression and anxiety on a weekly and monthly basis. In addition, while only girls in formal schools were significantly more likely to deal with mental health issues on a weekly basis in Jubaland, girls in all cohort groups were equally likely to face the same problem on a monthly basis in Jubaland.

These findings suggest that greater efforts should be invested in reaching specific groups of respondents most at risk to face educational marginalization in South West State and Jubaland. The programme has further space to expand coverage among marginalised ethnic groups and minority clans. To illustrate, around 40 per cent of girls in our sample do not come from any of the marginalized ethnic groups or minority clans. Thus, as previously suggested, further targeting Bantu minorities and the Digil-Mirifle, would be an effective strategy for selecting the most marginalized girls. Lastly, difficulty to identify girls with disabilities in their communities could indicate a significant degree of marginalization. The findings suggest that girls suffering from anxiety and depression could be another group to target specifically in Jubaland.

Girls with disabilities

Some of AGES' activities specifically focus on improving educational outcomes and opportunities of the girls with disabilities. The low proportion of girls with disabilities (GwD) other than mental health issues in our sample, and the difficulties to locate additional girls with disabilities when conducting the baseline, replicates previous findings from SOMGEP-T and suggest that (i) the actual proportion of GwDs in the population may be lower than expected, potentially as a result of the inability to access appropriate healthcare and the difficulties faced by people with physical disabilities to survive recurrent violence and displacement in such an extreme context. It may also reflect the difficulty to identify and reach girls with disabilities in communities. Thus, the activities aiming at identification and enrolment of GwD by engaging CEC and providing additional training to teachers on inclusive teaching practices is especially appropriate for better inclusion of the GwDs into program activities.

A higher percentage of GwDs in our sample seem to come from economically more disadvantaged households. Around 35 percent of GwD live the houses with very poor roofing. Their households are also food insecure or have more limited access to water for use at home, compared to other groups or marginalized girls. Also, 42 percent of GwDs live in households where the HoH did not earn a wage or did not have a stable occupation. In combination with their disability, the girls seem to have extremely limited access to school and educational opportunities, as their households are less likely to have sufficient financial resources to provide them with special equipment, care or ensure their transport to schools. Thus, despite being enrolled and sensitized on the needs of the GwDs, girls may still be prevented from actually regularly attending schools.

The most common types of disability recorded in the sample, by far, were those that can be broadly classed as “affect disabilities”, i.e. depression and anxiety. In total, 11.4 percent of caregivers reported that their girl experienced anxiety or depression daily or weekly, and this percentage jumped to 30.1 percent when a more liberal standard – including monthly bouts of anxiety or depression – were included. In addition to training and sensitisation for teachers around disabilities in general, and the rights of children with disabilities, focused training on mental health counselling may be a wise investment. These efforts could be targeted to the most conflict-affected areas, or they could be directed using data on mental health

disabilities.⁷⁹ Mental health support could also be integrated into the GEFs in some schools, even if this support was relatively limited.

Note from the Project: Status of the Broader GwD Population in the Targeted Area

In addition to assessing GwDs identified in the sample, the baseline survey also collected information from an expanded sample of GwDs identified in the same target areas, but not included in the original sample. This expanded sample was identified using a ‘snowball approach’ where sampled students and their caregivers were asked if there were any GwDs living in the area. The enumerators then proceeded to collect data from the households identified as having GwDs, and in turn asked them if they knew of any other GwDs living in the same area. The expanded sample identified an additional 135 households with GwDs, and 112 of them confirmed the presence of a GwD in the house. The expanded sample allows for a better understanding of the prevalence of disability in the area, as well as to obtain a snapshot of their ability to access education services.

The expanded sample includes a larger proportion of girls with mobility disabilities (68%) as well as of girls with cognitive disabilities (43% self-care, 46% learning, 35% remembering). It also includes a large proportion of girls facing anxiety and depression (50%). Surprisingly, the expanded sample contains a relatively small proportion of girls with visual (10%) and hearing (22%) disabilities, potentially suggesting that girls with these types of disability had already been targeted by the project. Overall, 82% of these girls face multiple disabilities, indicating that a large proportion of the remaining GwDs in these areas face more severe conditions.

In terms of access to education services, 32% of the GwDs in the additional sample had ever attended school and 26% are currently enrolled (22% in formal education; 4% in non-formal education). The main reasons for non-enrolment included:

-Disability-related factors: Having a condition that prevents them from attending school (60%); needing assistive devices (60%); needing assistance (56%); unable to use the toilet at school (42%) or moving around school (38%); lack of transportation (33%); girl does not learn anything (18%);

-Poverty: 57%

-Lack of adequate services: Lack of special needs education (46%) or accelerated education (26%); teachers do not know how to teach the girl (33%); refused entry (13%);

-Lack of support for formal/ Western education: Formal/ Western education not acceptable (21%); schooling is not important (22%); school does not lead to a job (18%); has had enough schooling (12%); girl is not interested in attending (17%); conflict with religious education (15%); girl has to work (15%);

-Safety issues: Unsafe to attend school (19%); mistreated by teachers (14%) or other students (18%);

-Marriage: 13%; had a child (14%)

⁷⁹ For instance, based on the evaluation’s data, anxiety and depression are most pronounced in Kismayo and Dhobley.

Limited financial resources

To improve girls' educational opportunities, AGES activities are also aimed at strengthening the financial stability of households of beneficiary girls, by providing scholarships for girls and encouraging parental involvement in SLGs. Financial costs seemed to be one of the main barriers to education, especially as 62 percent of caregivers agreed that the costs of education are an acceptable reason to not attend schools and 56 percent of caregivers – among those whose girls were out-of-school – indicated the inability to pay for girls' education as the main reason why their girls were not enrolled.

Also, a high proportion of girls are from financially-insecure households with greater vulnerability to negative external shocks. Around 42 percent of girls in our sample live in households where the HoH did not have wage or occupation and 98 percent of girls lived in households with no savings. Thus, despite the provision of a certain degree of financial support during the program activities, long-term economic instability and unemployment are important factors for girls' school attendance.

Household chore burden

Overall 65 percent of girls spent several hours a day completing their household chores, while 40 percent of the caregivers agreed that housework is an acceptable reason for girls not to attend school. Project activities address the barrier by providing access to a mobile learning platform; through sensitization campaigns on out-of-school children, implemented in cooperation with CECs and Teachers-Parent Associations; and by working with religious leaders to change the perceptions girls' roles and education in their communities. Our findings suggest that mothers play an important role in decisions concerning housework burden of girls and thus, should be included in the future discussions on how to address girls' barriers to education.

On the other hand, our findings suggest that higher girls' chore burden is associated with the households' economic status and educational levels of the HoH or caregiver. For instance, girls in female-headed households and girls who have lost one parent (single orphans) are more likely to stay at home caring for household and siblings, while HoH or caregivers would stay outside to find resources to address household needs. Further support and long-term solutions to improve households' financial situation would be necessary to reduce the burden of the housework delegated to girls. Also, we found that caregivers and HoH who did not have any exposure to informal or formal education have significantly lower perceptions of the value of educating' girls.

Moreover, nearly all girls who are married or mothers had to cope with a high housework burden, while a significant majority of other marginalized groups faced the same challenge. Thus, instead of targeting general audience for awareness-raising campaign, sensitization and behaviour-change activities should prioritize needs of the most at-risk girls and their households.

School environment

The current project activities aim to create a more suitable environment for learning. By training teachers on inclusive and gender-sensitive teaching practices, we expect the quality of teaching – especially the tailoring of lessons to the varied needs of different students – to improve, and girls' participation and learning outcomes to improve as well. At baseline, 46 percent of girls reported that their teacher's lessons move too fast, which suggests there is potential for training of this kind to improve girls' ability to learn.

Secondly, around 30 percent of girls reported that their school does not have separate toilets for girls and boys, which was one of the most frequently mentioned barriers to girls' school attendance. This finding highlights the vulnerability of these locations as the proportion of schools without separate toilets for girls

is much higher than the average for south central Somalia (13%)⁸⁰. As we will further discuss in this report, girls often identified school toilets as the least secure place on the school grounds, with many girls citing fear that anyone could enter at any time. In addition, inadequate hygiene of the facilities was also preventing girls from using the toilets, especially during their menstrual cycle. Teachers also mentioned that girls feel ashamed to go to toilets during the class and prefer to use toilet facilities outside the schools, causing girls to miss valuable instructional time during class.

Many of the most negative aspects of the prevailing school environment in the project's locations are beyond the scope of AGES – or almost any project – to impact. For instance, during discussions with girls, many mentioned very real concerns about their safety on the way to school and in their communities. Unfortunately, AGES programming is not suited to the types of widespread peacebuilding interventions that would be needed to improve safety in the most affected areas. Within the confines of the school itself, AGES interventions around child protection and gender-sensitive teaching should help to improve the school environment for girls. However, more could be done to address the harassment and abuse that girls face at the hands of their male peers, especially the fact that some girls cannot use the toilets at school out of fear of harassment or spying. Child protection interventions targeting teachers should improve this situation, but the project should also target boys directly, possibly through the influence of religious leaders, who could highlight the fact that spying and harassment violate religious principles.

Involvement of Religious Leaders

Religious leaders can be powerful catalysts for attitudinal change, especially in the realm of topics where religious norms play a role in shaping public opinion. Among the caregivers of girls who were previously out-of-school, 13.7 percent cited – as one reason their girl was not enrolled in school – the viewpoint that formal "Western" education to be incompatible with cultural or religious traditions. Among all caregivers, 51.1 percent say that it is acceptable to keep a girl out of school if the teachings conflict with religious teachings. AGES works with religious leaders to promote enrolment of girls in school, as well as for early identification of less visible disabilities such as dyslexia, given that nearly all children attend Quranic education before starting formal school. To the extent that religious leaders can validate formal education, in general, and for girls, in particular, it has the potential to increase enrolment rates and promote more frequent attendance. Findings from the SOMGEP project highlight the importance of the support of religious leaders as a factor driving girls' enrolment⁸¹.

Girls Empowerment Forums

Support for Girls' Empowerment Forums and Boys Empowerment Forums is a key ingredient in AGES programming, and mirrors other CARE educational interventions in Somalia. To the extent that GEFs contribute to develop girls' agency, they can have a number of direct and indirect benefits, including increased self-confidence and voice to enhance classroom participation and ability to negotiate for rights. The results elsewhere in this report suggest that higher leadership skills are associated with better learning scores (though it is important to note that this could be the result of reverse causation, in which girls who

⁸⁰ Consilient (2020) Education Sector Program Implementation Grant - Baseline Evaluation, p.34

⁸¹ Zerihun Associates and CARE (2016) Mid-Term Evaluation of the Somali Girls' Education Promotion Programme, p.82

do well in school gain self-esteem).⁸² Improvements in girls' empowerment can have direct effects on enrolment and learning outcomes.

Less directly, empowerment forums should provide a venue for girls to discuss their concerns with their peers and with female adults, developing a support network in the process. If girls are able to voice their concerns around topics like safety at the toilets, they may feel safer, develop organic solutions to the problem, or make the problem known to school administrators in a way that encourages change. Similarly, such a venue may help girls who feel they cannot attend school during their menstrual periods to feel more comfortable doing so. These are just two examples of the possible benefits from empowerment forums.

Transition Pathways and Interventions

One fact that comes through in much of the analysis in this report is that the diverse age ranges of the three cohorts the project will work with has real consequences for the planned interventions. For instance, despite their lack of schooling, NFE girls – who are the oldest cohort – achieved the highest learning scores of any group. They also have the highest performance in terms of leadership or life skills scores. This fact – and the differences between learning streams – complicates our assessment of whether the program's interventions are fit-for-purpose.

This is especially pertinent to transition, as an outcome, because the project does not expect NFE girls – despite their higher learning scores – to enrol in formal schools, due to their age at the time of recruitment into non-formal education. We agree with this assumption, given the context.

With regard to transition, the activities planned by the project are suitable and sufficient to meet its transition targets. We would suggest minor revisions to the program's design only:

NFE girls are entering the program with different learning levels already achieved, a finding which is most clearly presented in the subtask “foundational skills gap” analysis. NFE girls should be filtered into two separate learning streams based on their initial performance at intake, with the higher-performing stream focusing on higher-order numeracy skills (multiplication, division, applied mathematics problems relevant to finance and business operations), while the lower-performing group should start with basics of numeracy. The same logic can be applied to literacy learning.

Formal school (FE) girls are entering the program with minimal exposure to schooling but many have already achieved low levels of literacy. We expect these girls to be able to catch up relatively rapidly to age-appropriate achievement levels, while noting that Somali girls often enter school later than girls in other countries (in other words, this is a lower bar than it may seem).

ABE girls face perhaps the biggest challenge in terms of meeting the transition goals laid out for them. They enter with a moderate-to-low level of learning achievement, but the project has set a goal for at least some of these girls to transition into formal schooling. Given their age at baseline, these girls would need to achieve grade 4-level literacy within two years of ABE programming, then transition into a formal school at grade 5 when they are between 15 and 18 years of age. Our concern is that this may not be

⁸² In other CARE interventions in similar contexts (e.g., SOMGEP-T), the evidence seems to support the impact of GEFs on girls' leadership skills and self-esteem. In fact, these groups appear to be among the most impactful activities across a range of outcomes in the case of SOMGEP-T, and we have no reason to expect the AGES cohort or context to differ in this regard.

realistic, given the societal pressures to marry that will begin to weigh on ABE girls, and the additional psychosocial barrier that may prevent them from wishing to enrol in a formal school when they are much older than the other girls in their grade. A formal plan for easing their re-entry into formal school may be needed – for instance, girls from multiple ABE centres in the same city could be grouped together into a single cohort that enters formal school together.

Beyond these broad comments on cohorts and age groups, the transition pathways appear to be well-designed for most or all of the girls targeted by AGES.

4. LEARNING OUTCOME

AGES targets two primary learning outcomes: learning, in terms of numeracy and Somali literacy, as well as transition of in-school girls from primary school to secondary school and transition of out-of-school girls to formal education institutions or alternative learning programs. This section presents key findings on learning outcomes, with emphasis on the identification of learning gaps and barriers, along with the discovery of sub-populations that tend to have the lowest learning outcomes. The section begins with a brief summary of the learning tests used and the scoring methods for those examinations. For literacy and numeracy assessments, a list of subtasks administered is provided for reference. Score distributions are explored for floor effects, and aggregate scores are presented by grade and by intervention versus comparison group.

4.1 ASSESSMENT DESIGN AND VALIDITY

Assessment Design

Numeracy and literacy are the key learning outcomes of the AGES project. Literacy is assessed for the Somali (Af-Mahatiri) language, which is the official language of instruction and the language of instruction for most intervention schools. The assessments are adapted versions of the Early Grade Reading Assessment (EGRA) and the Early Grade Math Assessment (EGMA), designed by CARE to suit the environment of Somalia and the specific context of the intervention. The assessments test the following general skills:⁸³

Somali Literacy

- Sub-task 1: Letter Sound Identification
- Sub-task 2: Words Commonly Used by Children (Timed)
- Sub-task 3: Reading Fluency (Timed)
- Sub-task 4: Reading Comprehension 1
- Sub-task 5: Reading Comprehension 2
- Sub-task 6: Reading Comprehension 3

Numeracy

- Sub-task 1: Number Identification
-

⁸³ For full text versions of the assessments, please refer to Annex 16.

- Sub-task 2: Quantity Discrimination
- Sub-task 3: Missing Numbers
- Sub-task 4: Addition (Single digit)
- Sub-task 5: Addition (Two digits)
- Sub-task 6: Subtraction (Single digit)
- Sub-task 7: Subtraction (Two digits)
- Sub-task 8: Word Problem (addition and subtraction)

Each subtask comprised a set of individual items, ranging from one to ten per subtask.

The scoring methodology ensured that each subtask was weighted equally in the final aggregate score. Specifically, each subtask was scored as the percentage of *items* correct out of the total number of items in that subtask (hence ranging from 0 to 100). The aggregate score is the mean of the subtask scores. As an example, a subtask with 5 individual test items, each of which results in a binary correct or incorrect scoring, would be scored as the percentage of items answered correctly (i.e. four out of five items results in a subtask score of 80 per cent). These subtask-specific scores are then averaged to arrive at a final aggregate score for the assessment. This ensures that subtasks with many questions are not over-weighted in the final score.

On language of administering the learning assessments

It is important to note that the guidance (instructions) were provided in the mother tongue of the child (Af-Maay or Af-Mahatiri variants of the Somali language). Hence, for numeracy, we do not expect language to affect the assessment scores negatively. Later sections of the report will show that there is no effect of the mother tongue on numeracy acquisition. In the case of literacy assessments – the instructions are provided in the mother tongue, but the tasks have to be conducted in the language of instruction. This naturally puts the Af-Maay speakers at a disadvantage as the official language of instruction is Af-Mahatiri. It would be possible to conduct reading comprehension assessments in both languages with a subsample to demonstrate the effect, but doing so with the entire Af-Maay sample may be quite expensive. More broadly, there may be a practical reason for maintaining the literacy assessment in af-Mahatiri: very few books or other documents are printed in af-Maay, even in areas where af-Maay is widely spoken. Even children’s books written in af-Maay are exceedingly rare. Therefore, from the perspective of achieving functional literacy, measurement of literacy rates in af-Mahatiri may be preferable.

Assessment Validity

Age and Assessment scores

The learning assessments for AGES appear to be generally fit for purpose. Firstly, average learning scores correlate very closely with program tracks: Formal Education (FE), Accelerated Basic Education (ABE) and Non-formal Education or Life Skills (NFE), with FE cohort girls scoring the lowest on both assessments, on average, and NFE beneficiaries scoring the highest.

TABLE 20: DISTRIBUTION OF ASSESSMENT SCORES BY PROGRAM TRACK

Program track	Numeracy	Somali Literacy	Frequency
FE	58.8	33.8	421

ABE	67.0	41.0	484
NFE	76.6	57.9	515

This is intuitive as cohort girls are put into different program tracks based on their age, as we noted and documented previously in our discussion of sample characteristics. To illustrate, the mean age of formal school, ABE, and NFE girls is 11.1, 14.1 and 17.9 years, respectively. Most of the cohort girls selected for the intervention will have had little or no formal education in the past, so differences in foundational reading and arithmetic skills may reflect differences in age and experience gained from a variety of sources. Indeed, assessment scores correlate strongly with the age of the cohort girls, with scores increasing almost monotonically with age.⁸⁴ The table below collates the aggregate learning scores for each age group in the cohort.

Older beneficiaries have had more time to be exposed to alternative sources of education as compared to their younger counterparts; additionally, it is likely that they might have developed some numeracy skills due to engagement in small scale business. It is also the case that ABE and NFE participants had already been exposed to 4-5 months of educational programming at the time of baseline data collection, which may partially explain their better performance vis-à-vis formal school girls. In fact, girls who scored well on the learning assessments were asked directly where they learned to read or do mathematics, and many cited education through previous programs, but the vast majority of those who cited education programs were referring to CARE and the AGES project.⁸⁵ It is also possible that a proportion of the enrolled girls had attended formal education in the past (as discussed in the section above) particularly through EiE programming. As such, the correlation between program tracks and learning scores is most likely a spurious relationship attributable to: the differential average age between the program tracks; differential exposure to learning opportunities, including education projects and practical learning via avenues such as transactions in the market; and differential exposure to AGES programming, depending on program track.

TABLE 21: DISTRIBUTION OF NUMERACY AND READING SCORES BY AGE

Age	Numeracy	Somali Literacy	Frequency
-----	----------	-----------------	-----------

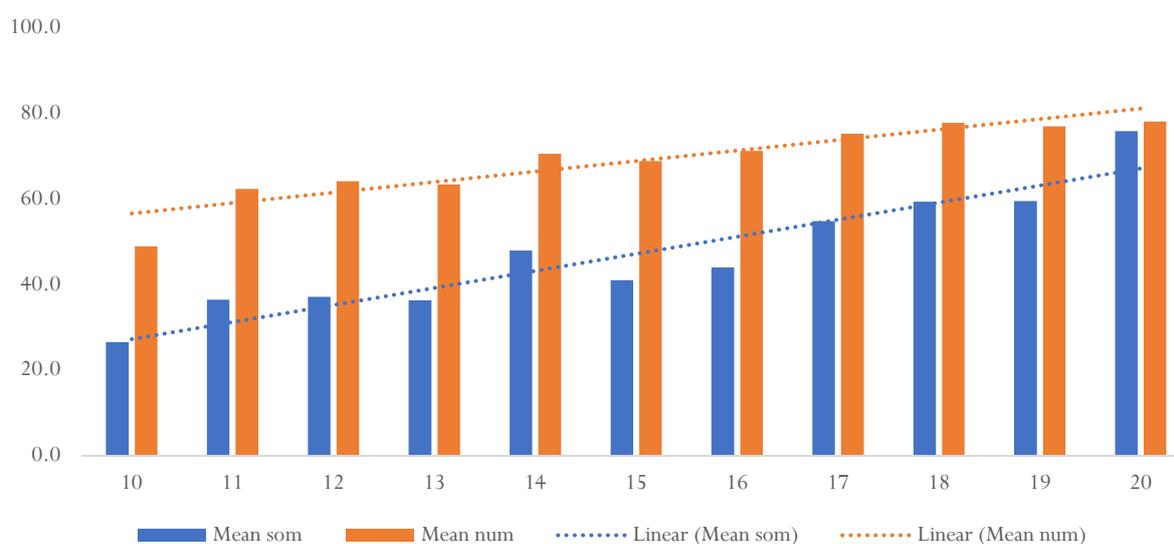
⁸⁴ There is a slight drop in learning scores between the ages of 14 and 15, but otherwise the scores increase neatly as the age of the beneficiary increases.

⁸⁵ This is consistent with the available evidence. Girls who scored above an arbitrary threshold on the learning assessment were asked where they had learned to read and do mathematics. Among these girls, 67.9 percent of ABE and NFE girls cited education during a project of some kind and a large share of them – in free-form responses – appear to be referring to CARE and the Girls' Education Challenge program. In contrast, only 25.0 percent of formal school girls cited exposure to education during a project, which is consistent with the fact that the AGES formal school intervention was just beginning at the time of the baseline, while ABE and NFE interventions had already been in progress for 4-5 months.

10	26.5	48.9	145
11	36.4	62.3	93
12	37.1	64.1	171
13	36.3	63.3	224
14	47.9	70.5	113
15	41.0	68.8	81
16	44.0	71.2	65
17	54.8	75.2	226
18	59.4	77.7	166
19	59.5	76.9	123
20	75.8	78.0	13
Overall	45.3	68.3	1420

Despite the slight dip in the learning scores moving from 12 years to 13 years of age, and an spike in learning scores at the age of 14, age is a good predictor of learning scores as scores by and large increase with an increase in the age of the beneficiary assessed. The graph below illustrates this relationship.⁸⁶

FIGURE 4: DISTRIBUTION OF LEARNING SCORES BY AGE



⁸⁶ Furthermore, the relationship between age and both learning assessments is found to be statistically significant at the 99% confidence level (P value for clustered and weighted regression 0.0000 for both)

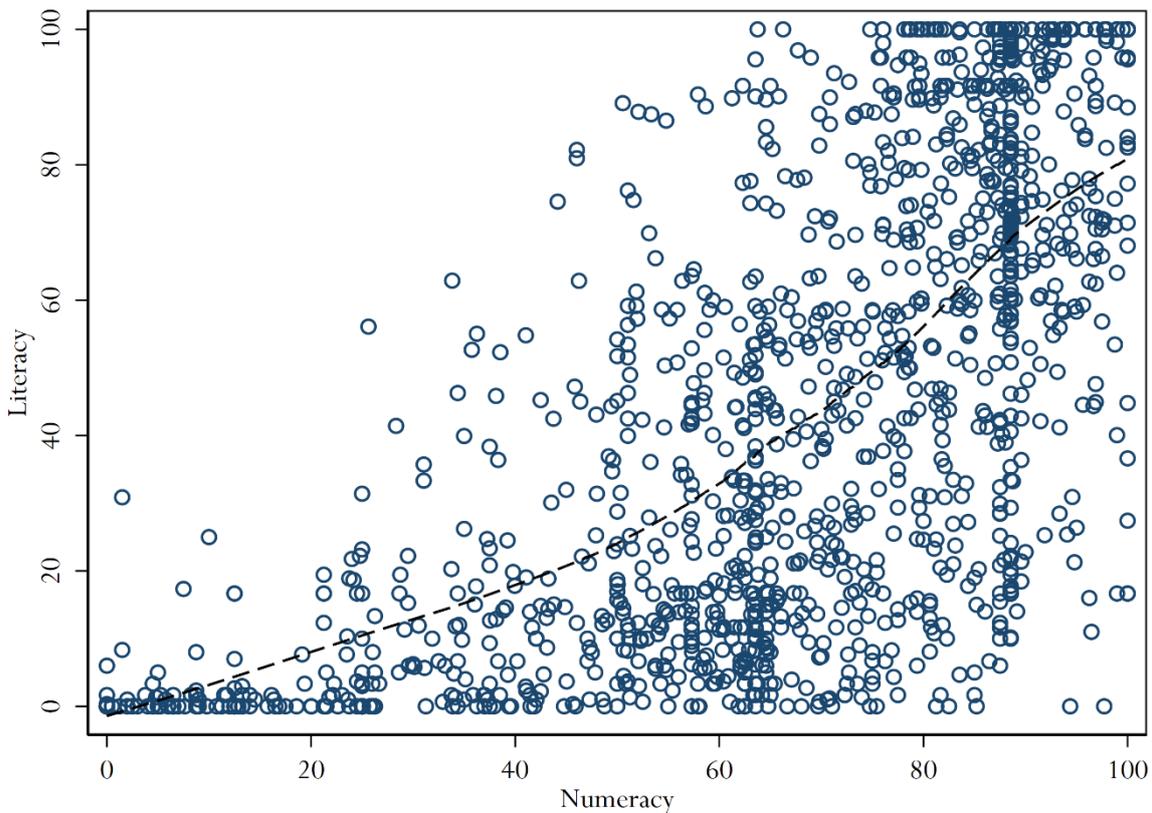
Relationship between Somali and Numeracy Scores

Secondly, the learning assessments for literacy and numeracy are themselves clearly positively correlated with increases in one predicting increases in the other. The following figure plots numeracy (X-axis) and literacy (Y-axis) scores among cohort girls ($n = 1,420$), where the fitted curve plots the association between the two scores.⁸⁷ The results exhibit a strong correlation between the assessments⁸⁸ whereby a student who scores well on their numeracy assessment is likely to score well on their Somali assessment, and vice-versa. Of course, a student might do well in literacy and not do well in numeracy and vice-versa, which would imply a failure of this test of validity. Assuming that most learners are exposed to both literacy and numeracy teaching at the same time, it is very encouraging to see that the scores from the assessments correlate in an almost linear fashion.

⁸⁷ The line in the graph is the locally weighted smoothing (LOESS) curve, which plots a local regression of the relationship between numeracy and literacy at each level of literacy.

⁸⁸ This relationship, again, is statistically significant at the 99% confidence level (weighted clustered regression P value: 0.0000). Moreover, the R-squared value for the regression is 0.4385 which is to say that 43.85% of the variance in Somali literacy scores can be explained by knowing the respondent's numeracy score.

FIGURE 5: SCATTERPLOT OF READING AND NUMERACY SCORES



Floor and ceiling effects of the assessments

Continuing with the inquiry into the validity of the learning assessments, we looked at possible floor and ceiling effects. The literacy assessment demonstrates both floor and ceiling effects. Approximately 8% of the students scored 0 percent in the literacy assessment and another 8% scored 100 in the same. This is positive in that the assessment is clearly able to distinguish learners in both ends of the spectrum. Yet, while the program will likely show significant increases in the low end of the scores with the approximately 20% of students who scored less than 10 percent,⁸⁹ the midline evaluation will have to find

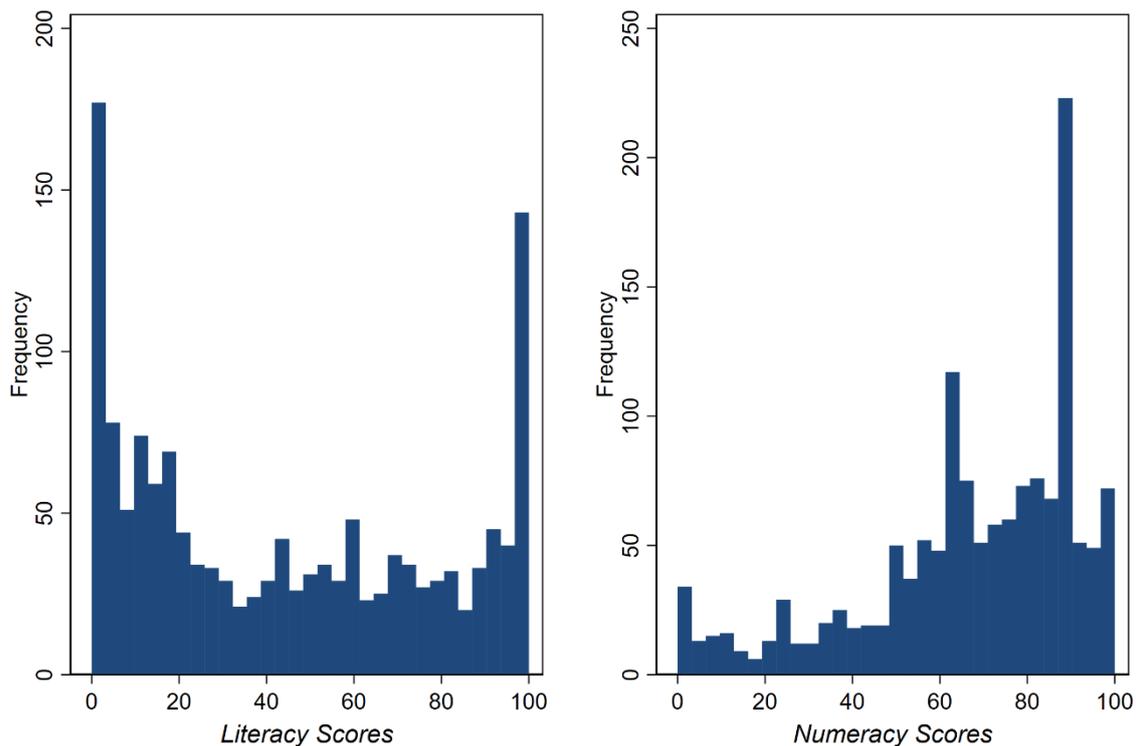
⁸⁹ The floor effects are also problematic, but mainly at baseline. For example, let us imagine that in reality the spectrum of learning results ranges from -50 to 150, but the assessment here is only able to distinguish performance on the range 0 to 100. Consequently, some students could theoretically improve by 50 points without showing any improvement in our test. However, considering that the assessment begins with the identification of letters, it cannot be much simpler for assessing literacy. It would also be doubtful if identifying more letters after the intervention should be counted as success when the aim is to actually create at least a level of functional literacy.

a way to address the likely ceiling effects with the 16% of the respondents who scored over 90 percent at baseline.

Meanwhile, floor effects for the numeracy assessment are practically non-existent, as only 1.6% of the students scored 0 percent. Similarly, at baseline only 2.5% of the students scored a full 100. Yet, the ceiling effect for the assessment will likely appear at midline as roughly 14% of the learners are currently at 90 percent or above.

FIGURE 6: DISTRIBUTION OF LEARNING SCORES

Distribution of learning scores



The consequences of ceiling effects can be quite severe, depending on the extent of ceiling effects at baseline, the expected "natural" gains in scores due to maturation over time in the absence of the intervention, and the expected impact – above and beyond the benchmark – of the intervention itself. That is, if there are large natural improvements from one year to another, ceiling effects will become more severe during each year of the project, even if the project has no discernible impact on learning scores. If combined with improved learning scores as a result of the project itself, ceiling effects could rapidly increase in magnitude to the point that it is impossible to discern any impact at all.

To illustrate the risks of this occurring, we performed a simulation of gains in learning scores over time, which we describe in more detail in Annex 13. Briefly, we started from the baseline distribution of Somali literacy scores, and calculated differences in scores across age cohorts, which are the "natural" expected improvement in learning from one year to another, in the absence of impact from the intervention. For

each girl, we added the appropriate amount of year-on-year improvement to her Somali literacy score.⁹⁰ Finally, we assumed a 5-point gain in learning scores due to the project, above and beyond the gains we assumed would occur naturally from maturation. Specifically, we assumed a 5-point "true effect" of the project on learning scores, and randomly simulated those gains in our sample of 1,420 girls. Truncating the results at 0 and 100 points, as would occur in practice, we calculated the "observed effect" of the project, taking into account benchmarking, to study the impact of ceiling effects on observed impact size. We repeated this simulation 1,000 times, and subsequently altered the true effect size of interest, to understand more fully the deleterious consequences of ceiling effects.

Overall, our simulations show that ceiling effects on the order of those observed at baseline will bias estimated midline effect sizes by between 30 and 60 percent, depending on the true impact of the project. For instance, if the project has a *true* impact of 4 percentage points on learning scores, we would expect to observe 2.8 point gain in learning scores in our sample, with learning scores truncated on the 0-100 point scale, a reduction in effect size of 30 percent. While this simulation is necessarily theoretical, it is tied very closely to empirical reality, insofar as it takes true baseline learning scores and benchmarked maturation rates as its starting point.

Without significant adjustments to the difficulty of the learning assessments, the potential for bias in measuring project impact is severe. As such, it is recommended that proactive changes are considered for both assessments to deal with likely ceiling effects at midline and endline. Otherwise ceiling effects could pose problems in the midline and endline; as student learning increases, there may not be sufficient room for high-level learners at the baseline to exhibit improvement.

Item Discrimination Analysis

Beyond floor and ceiling effects, we also evaluated each assessment in a classic item analysis framework. Specifically, we reviewed individual test items for their ability to discriminate between low- and high-performing students, calculating a discrimination index for each test item. The discrimination index measures the correlation between providing a correct answer on an individual test item and achieving a higher overall score on the exam.⁹¹ The fundamental idea is that a test question which elicits correct answers from otherwise high-performing students and elicits incorrect answers from low-performing students is able to *discriminate* between the two groups – the ultimate purpose of assessment. An effective test item should distinguish low- and high-performing learners by allowing correct responses from high-performers and forcing incorrect responses from low-performers, on average. That is, an effective test item is one that many low-performing students answer incorrectly, and one that many high-performing students answer correctly.

The following table summarizes these findings by subtask. A common metric for the type of discrimination index presented in the figure is that an item is effective if its index score is above 0.3. As the table below demonstrates, only one of the test items, namely Numeracy Subtask 3, does not pass this threshold, scoring just below 0.3. The scores for this item were overall very low compared to the other

⁹⁰ Note that our simulations include only girls enrolled in formal schools at baseline, as this is the cohort that will be tracked for learning outcomes over the entire duration of AGES programming. The bias that can be caused as a result of ceiling effects is marginally larger when ABE and NFE girls are also included in the simulation.

⁹¹ See Annex 9 for more details on the calculation of the discrimination index.

Subtasks. The Subtask requires the learners to fill in missing numbers in a sequence that can be based on subtraction or addition and is not explained to the student. This represents a somewhat applied mathematical skill that may not be a focus of the current teaching. Considering that this subtask focuses on logical reasoning, this is not a surprising result, given the rote learning practices often employed in basic numeracy teaching. The result mirrors previous findings observed in the SOMGEP-T baseline.⁹²

TABLE 22: ITEM DISCRIMINATION ANALYSIS

Numeracy Subtask	D Index	Literacy Subtask	D Index
1	0.52	1	0.57
2	0.59	2	0.59
3	0.29	3	0.57
4	0.61	4	0.78
5	0.59	5	0.77
6	0.65	6	0.64
7	0.55		
8	0.58		

The rest of the Subtasks are all well above the threshold. In many cases, d is greater than 0.6, representing well-designed test items, in terms of discriminating power as success in these test items correlates with overall success in the assessment. Individual subtasks of the learning assessments with high values on a discrimination index suggests that they are able to distinguish between low- and high-performing students. We can thus conclude that the test items are well designed for the purpose of the evaluation.

It should be noted, however, that the assessments are quite short and as such more likely to pass this test as the fewer the sub-tasks in the assessment, the more the overall score is composed by each sub-task. Consequently the standard for shorter tests should be higher. As such, it is encouraging to note that apart from Numeracy Subtask 3, all tasks are well above 0.5.

Logical Flow of Sub-Tasks

Continuing in looking at the individual test items, the order of the Subtasks does not imply a direction of increasing difficulty. For example, on average the scores for literacy Subtask 3 are lower than the following two Subtasks. While the Subtasks arguably measures an easier skill, reading fluency, the fact that the item is timed most likely explains the lower scores. Otherwise in the literacy the scores change logically, i.e. the more advanced reading comprehension items have an overall lower score than the preceding, simpler, Subtasks.

TABLE 23: DISTRIBUTION OF NUMERACY AND LITERACY SCORES BY SUB-TASK

Numeracy	Mean	Literacy	Mean
Sub-task 1: Number Identification	93.3	Sub-task 1: Letter Sound Identification	70.3

⁹² Ha, S. and Forney, J. (2018) Somali Girls' Education Promotion Project – Transition – Baseline Study Report, p.100

Sub-task 2: Quantity Discrimination	89.6	Sub-task 2: Words Commonly Used by Children (Timed)	44.4
Sub-task 3: Missing Numbers	26.9	Sub-task 3: Reading Fluency (Timed)	36.3
Sub-task 4: Addition (Single digit)	84.1	Sub-task 4: Reading Comprehension 1	44.4
Sub-task 5: Addition (Two digits)	53.1	Sub-task 5: Reading Comprehension 2	43.5
Sub-task 6: Subtraction (Single digit)	74.5	Sub-task 6: Reading Comprehension 3	35.5
Sub-task 7: Subtraction (Two digits)	45.6		
Sub-task 8: Word Problem	78.9		
Overall	68.3	Overall	45.3

Meanwhile, Numeracy Subtask 3 seems out of place in the assessment as it has the lowest overall score of all test items, at 26.9 while the second lowest Subtask score, for two-digit subtraction in the numeracy assessment, has a significantly higher average score, at 45.6. Considering the potential off-putting effect of a difficult task at the early part of the assessment, moving the missing numbers item at the end of the assessment should be considered. This is particularly pertinent as the exercise requires the use of all the skills covered by the other sub-tasks, including the most advanced calculations of double-digit subtraction. In addition, program activities should focus on improving the teaching of logical problems in mathematics as the score for the sub-task measuring the most difficult skill required to resolve the missing number problems, namely double-digit subtraction were significantly higher. Students clearly had trouble transferring the skills to a sub-task where they would need to be applied a logical problem. The results also suggest that the first two numeracy subtasks – number identification and quantity discrimination - could be removed given the extremely high scores.

Word problems, conversely, seem not to pose too much difficulty to the learners who average 78.9 percent for the sub-task. The scores for numeracy Subtask however are comparatively logical – scores for single digit subtraction and addition are much higher than the scores for similar exercises with two digits. Moreover, as corresponds with the above findings in the item discrimination index, all tasks had learners scoring both 0 and 100, which indicates that the items are able to capture the full range of learning.

Assessment Administration

Many measures were taken to ensure correct administration of the learning assessments. Enumerators for the exercise were given ample time to practice the administration of the learning assessment. The tests were first carefully explained and walked through in group in the training. Then a demonstration of the assessment was carried out by experienced fieldwork researchers. The enumerators then carried out mock assessments in pairs for two days. Training culminated in a pilot carried out at primary schools in Hargeisa where the enumerators had the opportunity to practice administering the assessment with girls of the same age group as the cohort girls. Finally, at the first sample point of the fieldwork, the team leaders were asked to spend time with each enumerator to observe their administration of the learning assessments and correct any lingering issues.

Beyond the discrimination index, a number of checks were conducted on the quality of the data. While the discrimination index focuses on the quality of assessment design, we also checked the quality of data collection efforts. Assessments can be sensitive to the performance of individual enumerators for any

number of reasons. For instance, enumerators might not follow instructions precisely, allowing students additional time, or confusing students with unclear instructions. The baseline data provide a strong test for such enumerator effects, because most enumerators completed assessments with 48 or more students. Such a large sample would allow us to detect if students being assessed by a single enumerator consistently under- or over-performed.

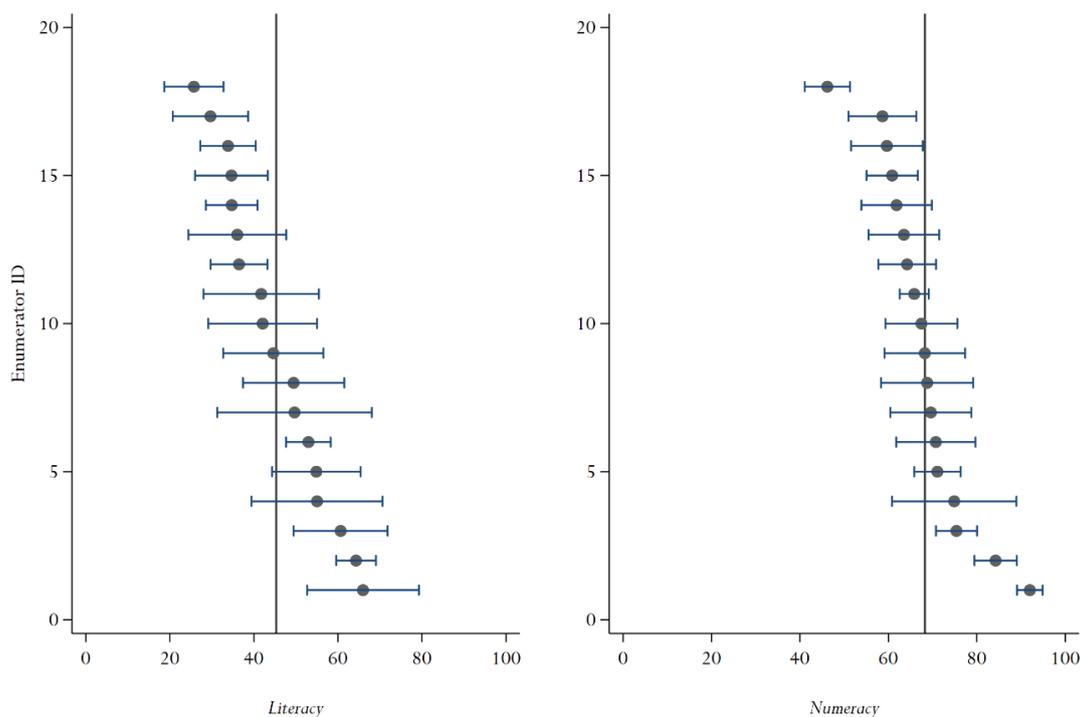
We calculated mean scores for each enumerator on each subject assessment, limiting the sample to those enumerators who assessed at least 10 students.⁹³ The figure above plots each enumerator's mean numeracy scores, along with 95 per cent confidence intervals. As the Figure shows, enumerators clearly vary in terms of their students' performance, with mean scores ranging from 25.4 to 66.6 percent in the literacy assessment. Importantly, variation by enumerators should be expected, because enumerators were assigned to regions with varying levels of student performance, and scores should be correlated within schools and within districts. However, the 95 per cent confidence intervals show that no single enumerator produces particularly extreme scores. That is, even in the case of the lowest- and highest-scoring enumerators, other enumerators obtained similar average scores. In this analysis, no enumerator stands out, suggesting that no individual enumerators were dramatically influencing student scores.⁹⁴

⁹³ This sample includes 18 enumerators, who account for 99.2 percent of all assessments included in the baseline. All of the 18 enumerators assessed at least 48 students.

⁹⁴ For the final dataset, observations collected by two enumerators were dropped because the assessment scores of the children assessed by the two enumerators differed from other enumerators even while controlling for sample point, region, type of respondent (FE, ABE, NFE) and age. The case of these two enumerators has been discussed in more detail in the method section of the report.

FIGURE 7: SCORES DISTRIBUTION BY ENUMERATOR

Learning scores by enumerator with 95% confidence interval



Design Recommendations

Going into the midline evaluation, it is pertinent to proactively address the potential ceiling effects of primarily the numeracy assessment and, to a lesser extent, the literacy assessment as well. Ceiling effects can worsen over time in longitudinal studies as the repeat respondents become familiar with the assessment. As such, it is recommended that a two-part assessment be developed into the subsequent phases of the evaluation. One part of the assessment should correspond in style, content and difficulty with the baseline to allow for comparisons over time. Then a more difficult second part should be developed. In this way the at the midline, comparisons to the baseline can be made using only the first portion of the learning assessment, while it will be possible to counter further ceiling effects in the later evaluations by comparing the full assessment in the two subsequent waves of research. While this approach will not completely obviate concern regarding ceiling effects, it preserves the ability to make valid, like-for-like comparisons at each stage of the evaluation.

In order to ensure comparability of assessments, two learning assessments of equal difficulty should be developed prior to midline. At midline, students should then be randomized into receiving one of the two assessment and the other at endline. The concern is that, without a comparison group, differences in the difficulty of the assessment from one round to another cannot be detected and any improvements in learning from one round to another could be a function of the test getting easier.⁹⁵

⁹⁵ This is further discussed in the limitations section of the report.

4.2 AGGREGATE LEARNING OUTCOMES

The girls learning cohort at the baseline consists of 1,420 girls in 108 learning centres. Distribution across different education tracks is somewhat even, ranging from 421 girls in 38 formal education centres to 515 girls in 36 NFE locations. For the establishment of baseline values, we employ sampling weights – as described in the methodological in the section above – to ensure that each school is equally weighted in the analysis.⁹⁶

TABLE 24: NUMBER OF INSTITUTION TYPES IN THE SAMPLE, BY STATE

Learning centres	BDR	JSS	SWS	Total
Formal education	16	13	9	38
Accelerated Basic Education	14	18	12	34
Non-formal education	17	9	10	36
Aggregate	47	40	31	108

Aggregate learning outcomes for the numeracy and literacy are presented in the table below. The table is divided into three panels, reporting scores for each of the three distinct learning tracks (FE, ABE, and NFE girls). Within each panel, we report mean Somali literacy and numeracy scores – and their standard deviations – for each state, and a program-wide aggregate. The evaluation design for AGES does not include a comparison or control group, nor are assessments with boys conducted at this baseline. As such, no other scores are presented.

As the table shows, there is some regional variation in average scores within particular learning tracks, but these differences are generally not dramatic. ABE girls are most similar to each other across regions, with mean scores within 1.5 points of each other for both types of learning assessment. FE and NFE girls exhibit much significant regional variation – formal school girls in the context of numeracy, specifically, while NFE girls show variation in literacy scores. Finally, as discussed above, girls in NFE score much higher, on average, than girls in ABE, who in turn score higher than the FE cohort girls. This is most likely an age-related factor as all cohort girls have previously had very little exposure to formal education.

TABLE 25: AGGREGATE LEARNING OUTCOMES BY LEARNING TRACK

Region	Mean Literacy Score	Standard deviation	Mean Numeracy score	Standard Deviation	Number of Observations
Formal School Girls					
Aggregate	33.8	31.2	58.8	26.2	421
NFE Girls					
Aggregate	57.9	33.9	76.6	19.6	515
ABE Girls					

⁹⁶ In analyses that report standard errors, confidence intervals, or p-values for formal hypothesis tests, we generally cluster standard errors at the school-level.

Aggregate	41.0	33.6	67.0	25.1	484
------------------	------	------	------	------	-----

Earlier in our discussion of learning scores, we analysed average scores for individual subtasks for both literacy and numeracy. As the results there showed, the first two subtasks on the numeracy assessment are markedly easier than all other subtasks. In principle, this is not surprising, as the test is generally designed to build from easier to harder test items. Moreover, these test items still effectively distinguish high- and low-performing girls, according to the discrimination analysis performed previously. On the other hand, discrimination analysis with less than 10 subtasks is not as effective as the same analysis on longer tests, with many more individual items. Given the significantly higher scores girls achieved on these subtasks, we calculated alternative numeracy scores, replicating the table above, but removing the first two subtasks from the scores. The results for this alternative score are presented below.

TABLE 26: ALTERNATIVE SCORING OF NUMERACY, BY LEARNING TRACK

Region	Mean Numeracy, Truncated	Mean Numeracy, Regular	Difference in Scores
Formal School Girls			
Aggregate	49.8	58.8	9.0
NFE Girls			
Aggregate	70.1	76.6	6.4
ABE Girls			
Aggregate	58.9	67.0	8.1

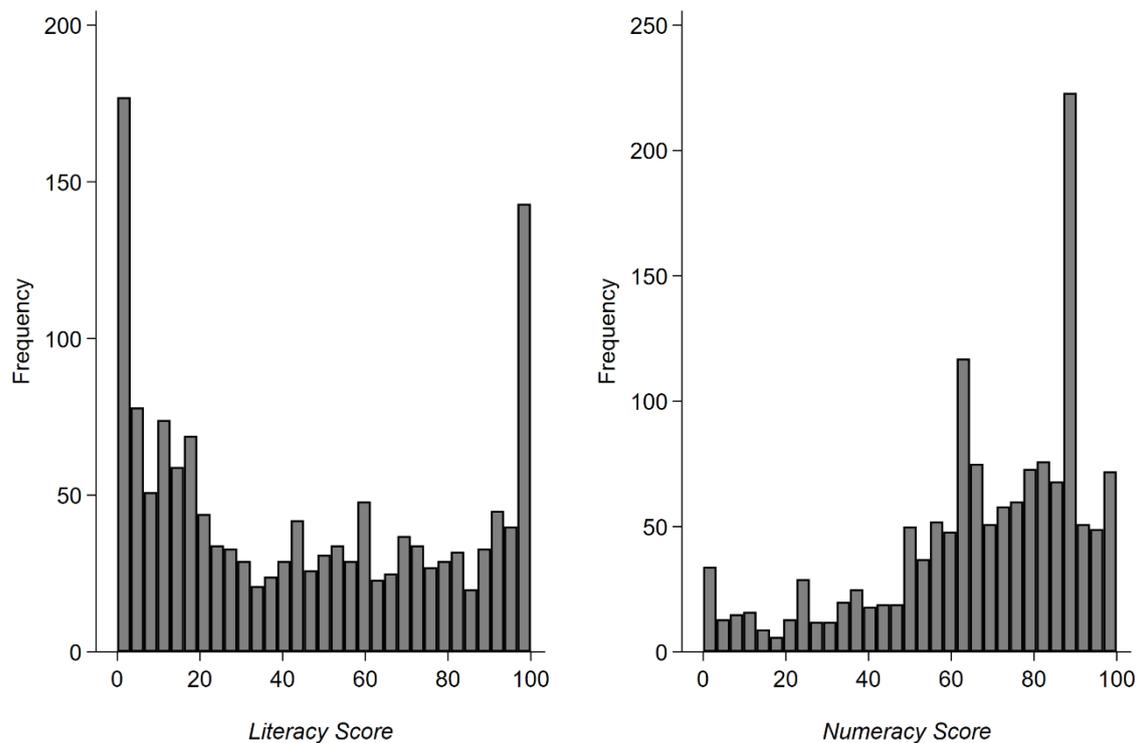
The truncated scores reported in the table are, unsurprisingly, lower, because these scores do not include the easiest two subtasks of the assessment. With respect to learning track, the truncated scores have the largest effect on formal school girls and the smallest effect on NFE girls, though the difference in impacts *between* learning tracks is not particularly large. Overall, removing the first two subtasks produces a net decrease in numeracy scores of 7.9 points.

When looking at learning scores by region and by education tracks, the standard deviations in numeracy and literacy scores do not differ dramatically. Overall, there is more variation in learners' literacy scores, however (in the aggregate, across all states and learning tracks, the standard deviation for literacy scores is 34.6 vs. 24.6 for numeracy). This indicates that there is slightly more bifurcation when it comes to literacy skills.

As discussed at length in the previous section, the learning scores have distinctly different distributions. The following graph shows the distribution of the learning scores. The illustrations of the distribution further highlight the issue of ceiling effects that characterize both assessments and will require remedial action going into midline.

FIGURE 8: DISTRIBUTION OF LEARNING SCORES

Distribution of Learning Scores



Indeed, when it comes to literacy, the distribution is much more bimodal, although with a slightly negative skew. Approximately 8 percent of the learners received no points for the assessment while another 8 percent received the full 100 percent. Note that, in line with our discussion elsewhere in this report, we find less reason to be concerned about floor effects, as girls who scored 0 percent are likely to reach the lower threshold of the assessment over time, due to maturation effects.⁹⁷ Numeracy, meanwhile, is a clearly skewed in the positive direction, yet with currently very little floor or ceiling effect. Nevertheless, already nearly 15 % of the learners scored 90 percent and over in the numeracy assessment. This indicates that there is currently very little room for this subgroup of top learners to show improvement in the coming years of the program.

Identifying Foundational Learning Gaps

One of the objectives of the baseline assessment is to investigate specific shortcomings in learning. This will allow for CARE to better target their teacher trainings and pedagogical focus. The analysis in this section is divided by learning track or institution type (i.e. formal school, ABE centres, and NFE centres). Girls in distinct learning tracks have different experiences of education prior to the baseline, especially

⁹⁷ Although such a process still has the potential to produce bias toward a null result, any bias is likely to be small in magnitude compared to the potential bias of ceiling effects, in this context.

because ABE and NFE centres began operation approximately 4-5 months prior to the baseline. Moreover, older girls like those enrolled in ABE and NFE programming may have learned basic reading and maths skills in the market, from family members, or through participation in other education programmes. Therefore, a full understanding of common skills gaps requires splitting the sample by learning track.

Before turning to analysis of each learning track, it is useful to consider the distribution of scores on individual subtasks, in general. Average scores, both in terms of an aggregate numeracy score or in terms of an average score on a single subtask, obscure important patterns how children actually complete subtasks. In the two figures below, we plot the distribution of scores for each individual subtask on the literacy and numeracy assessments. Note that our purpose in this analysis is not to identify skills gaps, but to highlight the fact that numeracy subtasks tend to exhibit bifurcated scores – the majority of children score either 0 or 100 percent on all but one of the eight numeracy subtasks. For instance, on the numeracy subtask 6, 18.6 percent of girls received a score of 0 percent, and 60.4 percent achieved a score of 100 percent; just 21.1 percent of girls are somewhere in the middle of the score distribution. The first figure below plots numeracy subtasks, and highlights this point.

Literacy scores on individual subtasks are similarly bifurcated, but with two nuances that are important to note. First, the first three Somali literacy subtasks have fewer extreme scores than other subtasks, with a wide variety of scores occurring between 0 and 100, exclusive. However, this is not an indication of a fundamental difference between the skills tested in these subtasks; rather, it is likely a consequence of the structure of the subtasks, which require girls to read a set of many letters, words, or a short passage. Subtasks such as these allow for a wider range of scores than those with just four yes/no questions. In fact, considering the nature of these subtasks, the number of girls who achieved either 0 or 100 percent is noteworthy: 32.5 percent of girls were able to identify all 50 letters successfully, while 8.2 percent were not able to identify a single letter correctly.

Second, the extent to which the last three subtasks of Somali literacy are divided into extreme ends of the scoring spectrum is still slightly less – perhaps only qualitatively so – than for numeracy. On nearly every numeracy subtask, at least 75 percent of girls achieved a score of 0 or 100 percent. But on the last three literacy subtasks, the equivalent rates are 69.2, 73.2, and 67.3 percent, respectively. This is certainly not a dramatic shift, but it may suggest something about the nature of literacy learning or literacy assessments that there is marginally fewer extreme scores than there are on numeracy tasks. Regardless of these nuances, however, the major takeaway point of the figures below is that scoring on most subtasks divides girls sharply into one of two camps, with relatively fewer girls in the middle of the score distribution.

FIGURE 9: NUMERACY SUB-TASK SCORES

Numeracy sub-task scores

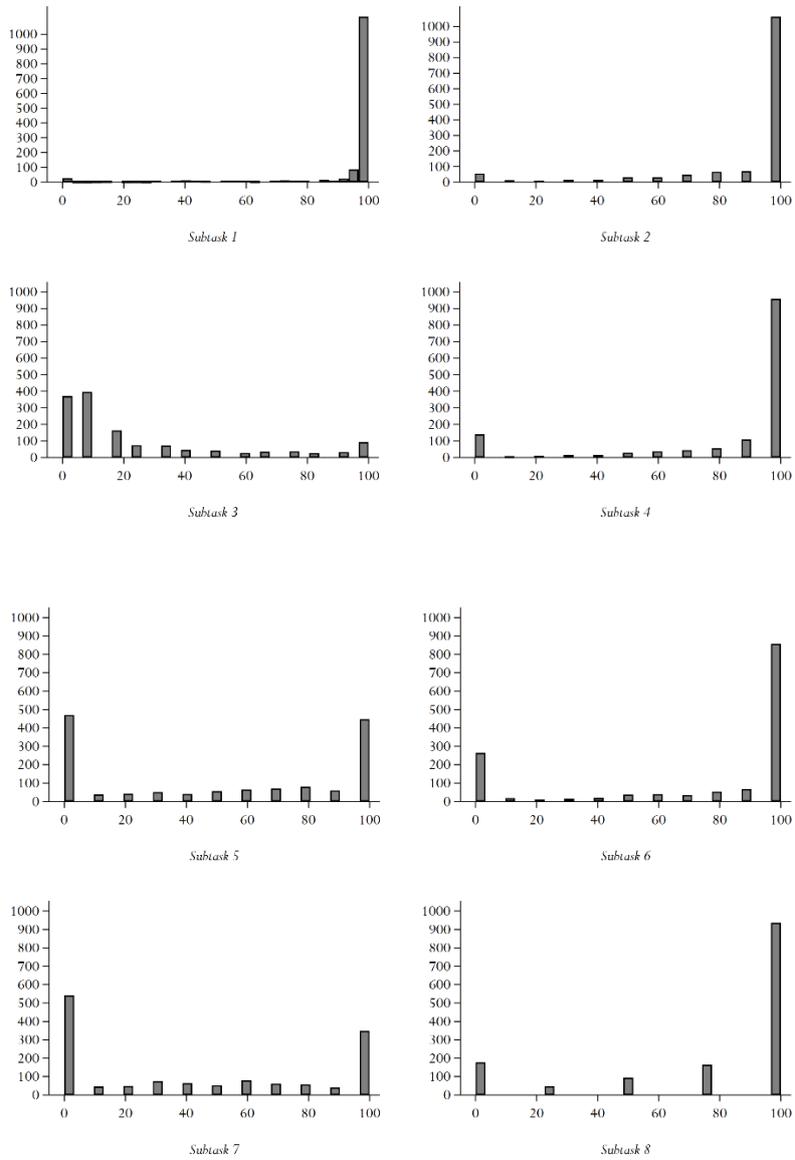
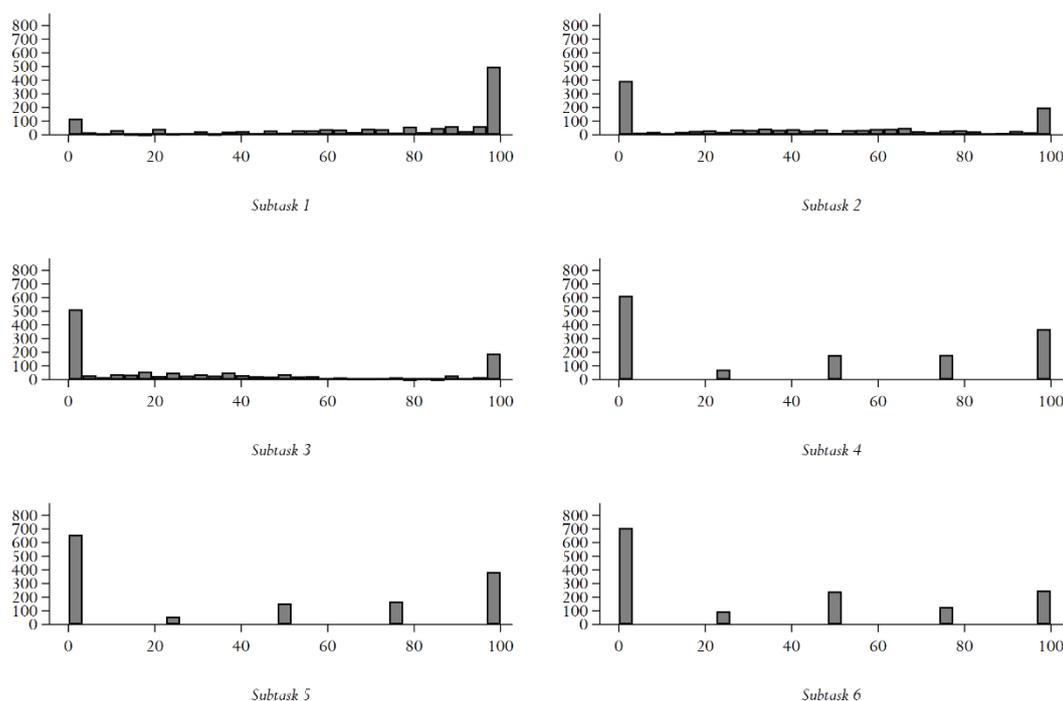


FIGURE 10: LITERACY SUB-TASK SCORES

Literacy sub-task scores



Our analysis now turns to individual cohorts of girls, with the goal of identifying patterns in their performance across subtasks that can be used to guide programming, by identifying learning plateaus, and, potentially, opportunities for jumpstarting improvements in learning outcomes.

In the sections that follow, we report performance on individual numeracy subtasks for each cohort of girls. The tables divide girls into four groups, based on their performance on a given subtask:

- Non-learners are those who do not complete any portion of the subtask correctly. These children clearly lack whatever skills are being assessed in the subtask.
- Emergent learners score between 1 and 40 percent on the subtask. This group encompasses a range of children, from those who may have the rudimentary skills necessary for the subtask, but only a minimal grasp of those skills at the moment, to those who may have guessed correctly on one or more of the subtasks individual test items.
- Established learners are those who score between 41 and 80 percent. These children have not entirely solidified the skills assessed in the subtask, but are performing sufficiently well that their performance cannot be based on guessing, and must reflect some degree of understanding or acquisition of the necessary skills.
- Proficient learners are those who score between 81 and 100 percent, and who have generally mastered the skills necessary for subtask.

Formal School Girls

The tables below report the proficiency of formal school girls on individual numeracy and literacy subtasks, respectively. When looking at the distribution of learning scores in numeracy it becomes apparent that the fundamental skills of identifying numbers and telling which one of a pair of numbers is bigger are skills most girls in the formal school cohort have acquired. Within the proficient group, the vast majority of girls (in fact, 71.0

percent of the entire set of formal school girls) are able to identify every number correctly in Subtask 1, while less than 3 percent of them do not recognize a single number. The number of proficient learners in Subtask 2 – quantity discrimination – is somewhat lower, but still high overall – 70 percent can choose the bigger one of every pair consistently.

It is in numeracy subtasks 4 and 5 that we begin to see a sharp decline in the performance of formal school girls – a trend that we will also observe in the other two cohorts studied below. In single-digit addition, 58.4 percent of girls achieved perfect scores, and approximately two-thirds of girls were considered proficient. But when the addition problems are extended to two digits, the share of girls who achieve perfect scores declines to just 19.0 percent. More telling, perhaps, is that suddenly nearly half (48.0 percent) of girls are unable to complete a single two-digit addition problem.

The results suggest that two-digit problems test a specific skill that many learners have not acquired. Such problems require the ability to "carry" digits from one column of the problem to another, which is the type of arithmetic rule that requires some degree of instruction to grasp – as opposed to single-digit addition, which many children can intuit. In addition, single-digit addition may be something children use more often in their daily lives while playing games or around the house, and they may have gained this skill with greater proficiency, even in the absence of formal schooling, as a result.

Most interesting are the 24.7 percent of girls (n = 104 in this cohort) who are classified as proficient in single-digit addition, but fail to correctly complete any double-digit addition problems. A similar trend, slightly less stark in some respects, is observed in single- versus double-digit subtraction. In our view, these girls illustrate plainly the possibility that knowledge of a specific rule separates girls specifically at the stage of double-digit arithmetic.

TABLE 27: FOUNDATIONAL GAPS IN NUMERACY, FE GIRLS

	Subtask 1	Subtask 2	Subtask 3	Subtask 4	Subtask 5	Subtask 6	Subtask 7	Subtask 8
	Number identification	Number discrimination	Missing numbers	Addition (single digit)	Addition (two digits)	Subtraction (single digit)	Subtraction (two digits)	Word Problems
Non-learner 0%	3%	6%	32%	13%	48%	27%	53%	19%
Emergent learner 1%-40%	6%	7%	47%	5%	14%	6%	15%	6%
Established learner 41%-80%	11%	18%	12%	16%	15%	15%	15%	20%
Proficient learner 81%-100%	80%	70%	9%	66%	23%	52%	16%	55%
	100%	100%	100%	100%	100%	100%	100%	100%

Trends in performance in literacy as the assessment becomes more difficult are somewhat similar to those observed in numeracy. In literacy, there is a relatively gradual decline in performance over the course of six subtasks, with the sharpest break between subtasks 1 and 2. This is not surprising, as the two subtasks are objectively different in terms of their difficulty – the former requires identifying letters, while the latter requires identifying words formed from those letters, which is a large leap in understanding.

In the literacy assessment, girls sort into two broad camps during the last three subtasks, especially. Briefly put, girls' scores on these subtasks are highly correlated: many girls are either not able to read at all or are able to read quite well. For instance, 58.4 percent of formal school girls were classified as either a non-learner or an emergent learner on all three reading comprehension subtasks, and nearly all of them achieved scores of 0 percent across all three subtasks. In contrast, 29.9 percent of girls were either established or proficient learners on all three reading comprehension subtasks. From this split, we can tentatively conclude that the subtasks identify girls who are either able to read well or are almost entirely unable to read, with a minority of girls falling in between.

TABLE 28: FOUNDATIONAL GAPS IN LITERACY, FE GIRLS

	Subtask 1	Subtask 2	Subtask 3	Subtask 4	Subtask 5	Subtask 6
	Letter recognition	Common words	Reading fluency	Reading comprehension 1	Reading comprehension 2	Reading comprehension 3
Non-learner 0%	11%	35%	46%	55%	59%	61%
Emergent learner 1%-40%	19%	27%	33%	6%	4%	5%
Established learner 41%-80%	29%	27%	13%	24%	21%	24%
Proficient learner 81%-100%	40%	11%	8%	15%	16%	10%
Total	100%	100%	100%	100%	100%	100%

ABE Girls

The results presented in this section for ABE girls illustrate both the similarity across types of girls in how learning skills relate to one another and the need to differentiate between different learning tracks in terms of the overall set of skills they have. The latter point is illustrated by most simply by the fact that ABE girls outperform formal school girls across every subtask in both learning assessments, with higher rates of proficiency and lower shares of girls classified as non-learners. The former point is slightly subtler, but is best shown by the fact that the trends identified among formal school girls – such as high performance on the first two numeracy subtasks, with a sharp decline at subtask 3, and the similarity in difficulties with two-digit mathematical operations – are substantively similar between FE and ABE girls.

TABLE 29: FOUNDATIONAL GAPS IN NUMERACY, ABE GIRLS

	Subtask 1	Subtask 2	Subtask 3	Subtask 4	Subtask 5	Subtask 6	Subtask 7	Subtask 8
	Number identificati on	Number discriminat ion	Missin g numbe rs	Additi on (single digit)	Additi on (two digits)	Subtracti on (single digit)	Subtracti on (two digits)	Word Proble ms
Non- learner 0%	2%	5%	26%	12%	35%	20%	41%	14%
Emergen t learner 1%-40%	3%	4%	54%	3%	14%	5%	17%	2%
Establish ed learner 41%-80%	6%	11%	10%	10%	19%	9%	16%	20%
Proficie nt learner 81%- 100%	89%	80%	11%	76%	33%	66%	26%	64%
	100%	100%	100%	100%	100%	100%	100%	100%

As with formal school girls – and the NFE girls discussed in the next section – the most compelling finding regarding literacy among ABE girls is that they split into two or three broad groups when it comes to reading comprehension. A large share of girls score 0 percent on all three reading comprehension tasks – though, it is important to note, these uniformly low scores result in part from the fact that girls who performed poorly on the first two subtasks were not presented with the remaining literacy subtasks. Therefore, the number of girls who truly would have achieved multiple 0 percent scores in a row is undoubtedly overstated.

On the other side of the spectrum, ABE girls are divided into two groups with moderate to high levels of reading comprehension. These two groups are not defined sharply, but are broadly distinguished by their consistent relative levels of proficiency. The first group achieves 80 percent or higher scores on all three reading comprehension subtasks, while the second group consistently falls into the "established learner" category. Of girls who achieve the highest level on Subtask 4, for instance, 63.2 percent also achieve the highest level of achievement on Subtask 6. Meanwhile, girls who fall into the established learner category on Subtask 4 do not typically jump up a level, nor do they typically regress a level in terms of achievement by Subtask 6 – the majority continue to fall into the established learner category even on Subtask 6.

TABLE 30: FOUNDATIONAL GAPS IN LITERACY, ABE GIRLS

Subtask 1	Subtask 2	Subtask 3	Subtask 4	Subtask 5	Subtask 6
--------------	--------------	--------------	--------------	--------------	--------------

	Letter recognition	Common words	Reading fluency	Reading comprehension 1	Reading comprehension 2	Reading comprehension 3
Non-learner 0%	10%	31%	40%	49%	52%	55%
Emergent learner 1%-40%	15%	25%	31%	6%	5%	6%
Established learner 41%-80%	26%	26%	14%	24%	20%	26%
Proficient learner 81%-100%	49%	17%	16%	22%	23%	13%
Total	100%	100%	100%	100%	100%	100%

NFE Girls

As the results of the aggregate learning analysis showed, NFE girls are the highest-achieving cohort in terms of both Somali literacy and numeracy. NFE girls are older than their FE and ABE counterparts, and may have gained skills through informal channels, including learning from siblings and learning-by-doing (by, for instance, completing basic arithmetic when shopping or playing with friends).

The general trend that NFE girls are the highest-performing learners is a difference in degree. However, NFE girls also show a difference of kind. That is, NFE girls do not simply exhibit identical *patterns* across subtasks as FE and ABE girls, but at a higher level of performance; rather, they show different patterns in subtask-to-subtask scores.

The most prominent such difference occurs in the numeracy assessment. NFE girls – like the other two cohorts of girls – show a tendency to decline in performance from single- to double-digit arithmetic operations. However, the magnitude of the decline is substantively different, and notable for how different it is. Whereas the share of FE girls who could not complete a single addition problem correctly jumped from 13 to 48 percent when moving from single-digit to double-digit problems, the same move upwards in difficulty results in a jump from 5 to 19 percent for NFE girls – still a marked increase, but far less sharp. The same trend holds true for subtraction, where NFE girls' scores do not decline as precipitously from subtask 6 to subtask 7.

TABLE 31: FOUNDATIONAL GAPS IN NUMERACY, NFE GIRLS

	Subtask 1	Subtask 2	Subtask 3	Subtask 4	Subtask 5	Subtask 6	Subtask 7	Subtask 8
	Number identification	Number discrimination	Missing numbers	Addition (single digit)	Addition (two digits)	Subtraction (single digit)	Subtraction (two digits)	Word Problems
Non-learner 0%	1%	2%	21%	5%	19%	10%	23%	6%

Emergent learner 1%-40%	1%	1%	48%	3%	9%	3%	17%	2%
Established learner 41%-80%	3%	10%	18%	10%	23%	11%	22%	15%
Proficient learner 81%-100%	95%	88%	12%	82%	49%	76%	38%	77%
	100%	100%	100%	100%	100%	100%	100%	100%

With respect to literacy, relatively few NFE girls lack the fundamental skill of identifying letters. In total, just 5 percent of girls are unable to identify a single letter, compared to 11 percent of formal school girls and even 10 percent of ABE girls, despite ABE and NFE girls' more similar ages. However, it is still important to emphasize the extent to which some girls lack all foundational reading skills – if 5 percent of NFE girls are entirely unable to identify common letters, this suggests that they will need to make very rapid progress from this point through recognizing words, onward to entire sentences, if they are expected to quickly reach a level of literacy that is considered functional.

Interestingly, girls' performance in identifying common words, reading fluency, and reading comprehension are all similar, at least with respect to the number of girls who are considered proficient and non-learners, respectively. One possible conclusion – though it should be considered cautiously – is that the transition from identifying letters to words is a very large barrier to literacy, while moving from identifying words to reading sentences is less significant of an obstacle.

TABLE 32: FOUNDATIONAL GAPS IN LITERACY, NFE GIRLS

	Subtask 1	Subtask 2	Subtask 3	Subtask 4	Subtask 5	Subtask 6
	Letter recognition	Common words	Reading fluency	Reading comprehension 1	Reading comprehension 2	Reading comprehension 3
Non-learner 0%	5%	17%	23%	29%	31%	35%
Emergent learner 1%-40%	9%	17%	31%	4%	3%	9%
Established learner 41%-80%	25%	35%	18%	28%	27%	28%
Proficient learner 81%-100%	61%	32%	28%	39%	39%	28%
Total	100%	100%	100%	100%	100%	100%

Grade level competency

Assessing the specific grade level competency cohort girls have achieved at baseline is not possible at baseline, for two reasons. First, the Federal Government of Somalia curriculum does not specify grade level targets. Instead, the somewhat vague learning targets are only set for lower primary and upper primary school levels. As such, it is not possible to assess specific grade level competencies achieved; only competency at the conclusion of multiple grade levels. Second, 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. The table below provides tentative findings on grade level competency at baseline for literacy and numeracy, broken down by learning track. In line with findings regarding aggregate learning scores, formal school girls performed worst on this metric, while NFE girls performed best. This is not surprising, as higher overall performance on the assessment will be correlated with achieving a minimum of 80 percent performance on each constituent subtask. Just 5.1 percent of formal school girls achieved lower primary competence, which is consistent with their grade levels and extremely limited previous exposure to educational opportunities. In contrast, 24.4 percent of NFE girls achieved partial lower primary competence in literacy; unfortunately, we cannot determine whether this is a function of several months of exposure to the NFE intervention itself, or due to their older age.

TABLE 33: SHARE OF GIRLS ACHIEVING LOWER PRIMARY COMPETENCY, BY RESPONDENT TYPE

Competency Level	FE Girls	ABE Girls	NFE Girls	Overall
Literacy, Lower Primary	5.1%	9.4%	24.4%	13.6%
Numeracy, Lower Primary	5.6%	7.2%	10.9%	6.8%

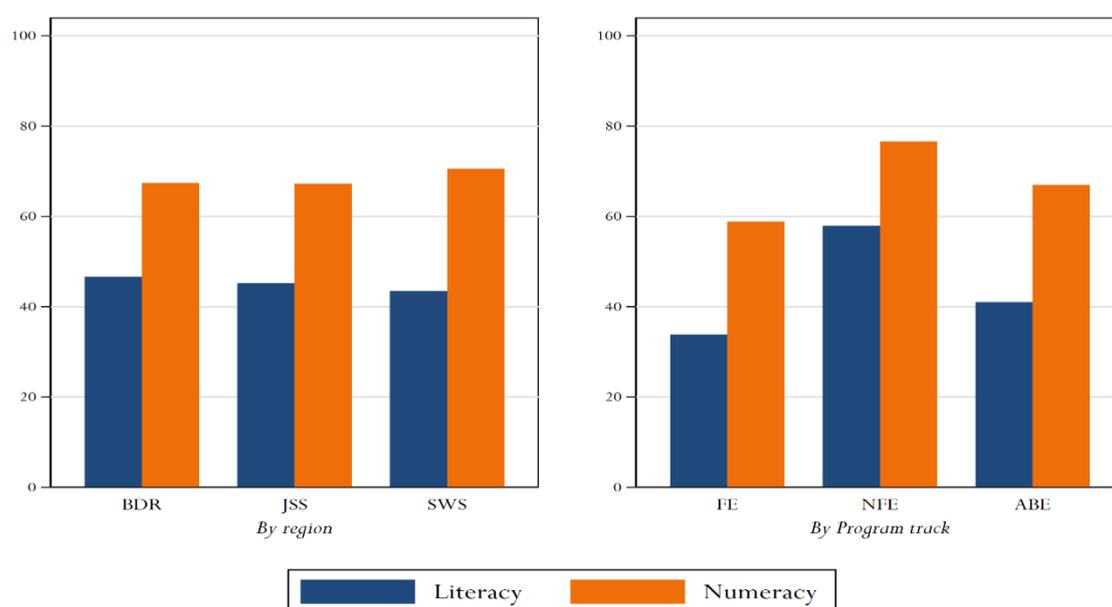
Relative to reading or literacy, the gaps between learning tracks in terms of mathematics competency are smaller. At the low end, 5.6 percent of formal school girls achieved partial lower primary competency, compared to 10.9 percent of NFE girls. This is somewhat surprising, as the gaps in mean scores between these groups were similarly large in numeracy as in Somali literacy. To clarify: girls in different learning tracks achieved significantly different mean numeracy scores, but they do not achieved significantly different rates of upper primary competency in mathematics. The explanation for this unusual result appears to be the third numeracy subtask, on which girls achieved a mean score of just 25.2 percent, far lower than any other numeracy subtask. When we define grade level competency after excluding subtask 3, there are two effects. First – and most obviously – because the most difficult subtask is being removed, the rate of girls achieving upper primary competency rises markedly. Second, a larger gap between formal

school, ABE and NFE girls opens, more in line with our expectations, based on mean numeracy scores among those groups.⁹⁸

4.3 SUB-GROUP LEARNING OUTCOMES

This section of the report explores learning outcomes for various different sub-groups of learners as well as for potential barriers. The section begins by looking at the relationship between geography and program track on learning. The section then looks at various types of disability, many of which are found to have a substantial and significant negative influence on levels of literacy and numeracy. The section then proceeds to looking at household and school resources and characteristics before looking at community attitudes and teaching quality. The results there are very varied. The following section looks at school governance. In the final part of the section brings together all the variables that were found to have a significant relationship with learning outcomes at baseline.

FIGURE 11: MEAN LEARNING SCORES BY PROGRAM TRACK AND REGION



⁹⁸ We do not focus on the results of this subsidiary analysis, because we are hesitant to alter our definition of competency to fit results we expect, especially since children really should achieve full competence across all subtasks in order to be so classified. For the sake of completeness, we report that this alternative approach yields 17.4 percent of formal school girls who achieved partial lower primary competency by this standard, compared to 42.1 percent of NFE girls.

Program track

To begin our inquiry into potential barriers for learning, we first wanted to control for any potential divergence in learning outcomes deriving from program track. The differences between beneficiaries in different program tracks are both substantial and significant. NFE respondents systematically score higher in both learning assessments than the ABE cohort who, in turn, score higher than the girls in formal education. Essentially the girls in the formal education cohort are systematically different from their non-formal education counterparts in that respondent type is heavily correlated with other factors that influence learning outcomes.⁹⁹

Given this substantive relationship between program track and learning, all the barrier relationships explored in this section will thus be controlled with program track. In other words, the regressions will be run with program variable in the model to see whether the relationship between the barrier and the learning outcomes persist when accounting for the respondents' program track. This will help in ruling out potential spurious relationships and collinearity. We, will however, report the results of both regressions, the simple regression and the controlled one. We, however, caution strongly against over-interpreting findings of the simple regressions and against the use of the 'naïve' results for the reasons discussed above and in the theory of change section of the report. Finally, before proceeding to the barrier analysis, the reader should know that in this section the analysis does not use weights. This is because the weights were designed to balance the aggregate results to reflect each cluster equally. They were not designed with the barriers in mind, and as such could significantly skew the results if applied for sub-group analysis where the sub-sample sizes can be very small.

The findings in this section are at times incoherent whereby some of the correlations that are found to be statistically significant affect learning in a way that is contrary to the assumptions. While we have sought to identify any issues of collinearity and omission of important variables, we can only assume that some of the sub-groups co-vary with factors that have not been accounted for. However, the findings will most likely change substantially going into midline. At baseline most of the cohort girls have had very little exposure to education, and as such many of the factors relating to specifically school resources as well as factors that may contribute to attendance and opportunities and time for learning have not yet intervened in the learning process.

The NFE girls score highest in both numeracy and literacy while ABE girls tend to score higher than girls in formal education, although this difference tends to be smaller than that between NFE and ABE girls. In sum, by and large in every region, the learning scores correlate similarly with the age of the cohort girls.

Disability

Once upon time, there was a man who said: "I want to kill a lion", and then he took a spear and shield. When he went to the forest there was a rabbit that jumped in front of him. And the man ran as fast as he can. Then people asked him "why are you running from the rabbit?". And he replied: "I wasn't ready for a rabbit; I didn't take that equipment, that's why." Therefore, as a committee, we can't do anything about children with disabilities because we

⁹⁹ Please refer to the section on the theory of change for more discussion on this topic.

don't have all things they need. For example, if we have a child who's deaf, how can we help them? So we need all the equipment that can help those children. – FGD with Community Education Committee

The first group of barriers to look at in this section are various disabilities. Respondents – both caregivers and the girls themselves – were asked if the cohort girl had difficulty with any tasks or activities. The below table reports the cases where the caregiver identifies the girl as having ‘some difficulty’ with the task or activity. The reason we are using this approach is because there is a lot of stigma around disability in the Somali context and in general many are reluctant to admit to anything that they perceive a weakness. Using the more relaxed definition by including those respondents who admitted to ‘some’ difficulty instead of ‘a lot’ will hopefully better capture those with disability, particularly as it relates to behavior and mental health where the stigma might be particularly strong. Moreover, using a definition which includes only those who report a lot of difficulty produces many sub-group samples of one individual which is not conducive to robust conclusions.

The following table presents the results for the disability variables. The table first shows the mean literacy scores for specific sub-groups as well as the p value for the regression between the sub-group and the literacy score variable as well as the number of observations affected. The same information is presented for numeracy. The last two columns, then, present the regression coefficient from the regression where the relationship is controlled with respondent type, if the relationship between the variable and the learning outcome is statistically significant. The p values for the controlled regression are also presented.

TABLE 34: ANALYSIS OF LEARNING SCORES BY DISABILITY TYPE

	Literacy			Numeracy		Controlled coefficient	
	Mean	P value	n	Mean	P value	Literacy	Numeracy
Overall	44.3		1420	67.1			
Disability: vision	53.9	0.260	27	73.8	0.103		
Disability: hearing	27.4	0.316	5	56.2	0.068		-15.5** 0.002
Disability: mobility	47.9	0.756	10	53.7	0.239		
Disability: arms	50.0	0.873	2	40.8	0.366		
Disability: communication	32.7	0.358	3	69.9	0.558		
Disability: cognitive	29.6	0.006	52	54.4	0.001	-15.0** 0.006	-13.1** 0.000
Disability: behaviour	33.5	0.082	40	58.9	0.042	-12.7* 0.023	-9.56936** 0.002
Disability: Mental health	38.0	0.015	162	64.3	0.168	-6.9* 0.028	

Disability: Any	38.7	0.016	251	64.0	0.048	- 7.377419 * 0.007	-4.210243* 0.018
------------------------	------	-------	-----	------	-------	-----------------------------	---------------------

With a sample of 251 cohort girls, it seems that having any disability is a negative predictor of skill level, even after controlling for respondent type. Those with any disability have a literacy score on average more than 7 percentage point lower than the rest of the cohort. In numeracy the difference is also negative with a different slightly over 4 percentage points.

Mental health disability is negatively correlated with literacy outcomes as are hearing difficulties with numeracy learning levels. Disabilities of cognition and behaviour, in turn, predict substantially lower levels in both literacy and numeracy. For example, a person whose caregiver reports some issues with cognition has on average 15 percentage points less in the literacy assessment than the rest of the cohort even when controlling for respondent type.

As a whole disability is a good predictor of learning levels at baseline as having any disability at all is statistically significant at the 99 percent confidence interval level. Importantly, all of the relationships are both substantial and running in the direction predicted, as they are associated with lower levels in the learning assessments.

If a girl has poor vision [or hearing], she should be given glasses, or a doctor for treatment. Then she should be given earphones which help her to hear, otherwise she will not learn anything. - FGD Mothers

As the above quote illustrates, disability affects learning. However, beyond questions of mobility, disability may affect attendance and thus retention and transition due to factors beyond the immediate ability to get to school and back:

If the girl is disabled, she will not go to school because she is being bullied. So she does not dare to go to school. - FGD Mothers

Demographics

Household and girl characteristics

Next we looked at general characteristics of the girls' household and herself. In terms of the girls living situation, we looked at whether she was partially orphaned or fully orphaned, or whether she is living without either one of her parents. These factors did not have a significant relationship with skill levels in a simple regression nor when controlling for respondent type. However, girls who come from a female-headed household have significantly lower learning levels when controlling for respondent type. Coming from a female-headed household lowers the literacy score by an average of 7.5 percentage and the numeracy score by 4.1 percentage points. Elsewhere in this report we have hypothesized that female-headed households are more likely to require the girl to take on more household and other chores. We have also seen that when parents are presented with the choice between chores that need doing and girl's education, many choose the former.

The obstacle is for the mother to do the house chores that the girl was doing. It won't be possible to tell her to go and learn, and still do the chores. The mother should do the chores, and the girl should go and study. – FGD CEC

TABLE 35: DISTRIBUTION OF LEARNING SCORES BY HOUSEHOLD AND GIRL CHARACTERISTICS

	Literacy			Numeracy			Coefficient after control	
	Mean	P value	n	Mean	P value	n	Lit.	Num.
Overall	44.3		1420	67.1		1420		
Orphan: mother deceased	45.2	0.772	184	69.0	0.353	184		
Orphan: both parents deceased	30.7	0.199	14	56.5	0.188	14		
Living without either parent	46.5	0.610	83	67.6	0.852	83		
Female HoH	40.6	0.006	530	65.2	0.048	530	-7.5** 0.000	-4.1** 0.003
Girl is currently married	49.9	0.306	53	72.9	0.081	53		
Girl is married or was married	48.1	0.409	99	70.7	0.161	99		
Girl is a mother	45.9	0.779	73	71.1	0.200	73	-12.1* 0.046	

In terms of girl's individual characteristics, we looked at marital status and whether girl is a mother. Most community members were of the opinion that once a girl marries, her education is over:

Every time when girl got married, she couldn't continue her education. When girls are married, 98% will stop their education. - FGD CEC

They may get married and take over their responsibilities. The husband may not allow her to go to school. - FGD Mothers

The quantitative data, however, did not find a relationship between learning outcomes and being married, as such. Yet, a girl being a mother

lowers the literacy score by more than 12 percentage points when keeping in mind the respondents program track.

In many ways the variables investigated here overlap with the indicators measuring poverty, in the following part of the section. Community members in qualitative interviews link marriage and poverty and marriage, for example:

In this village most underprivileged families give their daughters to rich men in order to get some money. So, in this village, minority or poor girls get married early for the sake of satisfying their families. – FGD CEC

If she gets married or her parents don't have any job, that will prevent her going from going to school – Risk Mapping

Household educational characteristics

If a parent is educated, they send their child to school whether they are a boy or a girl. But if they are ignorant, they send only boys to school while the sad girls stay at home to cook food. – FGD Mothers

TABLE 36: DISTRIBUTION OF LEARNING SCORES BY HOUSEHOLD EDUCATIONAL CHARACTERISTICS

	Literacy			Numeracy			Coefficient after control	
	Mean	P value	n	Mean	P value	n	Lit.	Num.
Overall	44.3		1420	67.1		1420		
HoH has no education	43.5	0.737	302	67.6	0.795	302		
HoH has no formal education	45.7	0.003	1162	67.8	0.051	1162	7.8** 0.003	
Caregiver has no education	45.9	0.468	383	68.0	0.573	383		
Caregiver has no formal education	45.7	0.001	1209	67.8	0.062	1209	10.4** 0.000	5.5* 0.031

When it comes to the educational background of the adults in the girls’ household, the results seem somewhat counter-intuitive. Namely, caregiver or head of household having no education at all has no relationship with learning outcomes in the data. Yet, caregiver and head of household having no formal

education¹⁰⁰, the relationship is in fact positive with learning outcomes. It would be more reasonable to hypothesize that the absence of formal education in the girls caregiver and head of household would be a negative predictor as it would make sense to assume that parents with formal education would be more likely to both transfer their skills to the girl as well as the desire to enrol.

Household economic characteristics

Today a poor person cannot develop. A person without money is not able to study. – FGD CEC

We try to bring them back, but their parents told us that they cannot let their daughter go to school, because they are busy trying to survive. Parents tell us that if they send their daughter to school, they cannot get one meal per day. – FGD CEC

TABLE 37: DISTRIBUTION OF LEARNING SCORES BY HOUSEHOLD EDUCATIONAL CHARACTERISTICS

	Literacy			Numeracy			Coefficient after control	
	Mean	P value	n	Mean	P value	n	Lit.	Num.
Overall	44.3		1420	67.1		1420		
HoH has no wage-earning occupation	48.7	0.002	558	70.2	0.004	558	6.0** 0.007	4.3** 0.010
HoH in pastoralism	51.6	0.177	25	84.5	0.000	25		18.6** 0.000
House is made of poor materials (roof)	40.6	0.188	284	68.2	0.517	284		
Gone to sleep hungry	42.7	0.002	1130	66.3	0.064	1130	-7.6** 0.002	-3.9 0.057
Gone without clean water	43.5	0.100	1144	66.4	0.099	1144	-4.6* 0.042	-4.0 0.062
Gone without medicines most days	50.6	0.018	238	70.7	0.068	238	6.4* 0.016	
Gone without cash income most days	43.3	0.719	210	69.1	0.323	210		
Owns land	48.5	0.018	462	69.8	0.036	462	7.7** 0.001	5.1** 0.002

¹⁰⁰ i.e. they may have qur'anic or other informal education but no formal education

Has savings	39.3	0.498	27	67.7	0.882	27		
Owns mobile phone	43.2	0.006	1229	66.3	0.017	1229	-9.3** 0.000	-7.0** 0.003
Owns smartphone	47.8	0.091	255	68.8	0.281	255		
Household has access to water	47.2	0.044	580	69.0	0.040	580		

To capture the economic status of the cohort girl’s household, we looked at a number of variables. The assumption here is that the more socio-economically advantaged household, the higher the girl’s learning baseline. The variables can be grouped into two categories, positive and negative predictors.

We begin with the negative predictors where some contradictory findings emerge. Namely, the head of household not having a wage-earning profession seems to be in fact positively correlated with learning outcomes. Similarly, contradictory is the fact that girls’ head of household being in pastoral livelihoods is positively correlated with numeracy outcomes but not literacy levels. Both of these variables were devised as proxies for poverty which was assumed to be negatively correlated with learning outcomes.¹⁰¹ In both cases, this may be a spurious finding if such households are IDPs; the availability of temporary learning options through education in emergencies may have had a positive influence in the acquisition of some skills. Additionally, receiving cash transfers may have allowed for temporary enrolment in the past (72% of the households where the head does not have an occupation had received cash transfers in the past). The difference in learning scores suggests that the difference is marginal in relation to the average (again suggesting potential exposure to temporary learning opportunities available to IDPs).

As it pertains to the pastoralists, it could be speculated that children of pastoralists are more exposed to mathematics as they might accompany their parents to the livestock market, but are less exposed to reading, spending long times outside of permanent settlements.

While the house having poor roof material – a proxy for overall economic status – was not found to significantly correlate with learning outcomes, going without clean water or food were found to be significant negative predictors of both numeracy and literacy skills. A respondent identifying that they did not have clean water most days had an average literacy score 4.6 percent lower than the rest of the cohort when considering the variance due to respondent type. In numeracy, likewise, those without clean water scored on average 4.0 percent lower. Meanwhile, the respondent having gone to be hungry was associated with a 7.6 percent lower score in literacy when taking into account learning track. In numeracy the effect

¹⁰¹ We conducted a further check by excluding ABE and NFE girls altogether as a substantial part (71%) of those girls whose HoHs do not have an occupation are in the two program tracks. Given that the intervention started earlier with ABE and NFE, it is possible that this could have positively impacted the learning outcomes. However, the results remain largely unaltered even when these girls are excluded. However, 74% of those ABE and NFE girls whose HoH does not have an occupation had received scholarships or cash grants in the past (most likely grants) which suggests that they had been able to access school (although temporarily). This might explain this outcome.

is 3.9 percent. Indeed, food insecurity can have a negative effect on attentiveness resulting in difficulty in learning. Likewise, not having enough water to drink can also mean that the respondent does not have sufficient water to take care of their personal impact. This can have a negative impact on health, which, in turn, can have a knock-on effect on learning.

When it comes to indicators that we assume would have a positive relationship with learning levels, household owning land is indeed correlated with learning levels, with cohort girls from these families scoring on average 7.7 and 5.1 percentage points higher in literacy and numeracy, respectively. Household having savings was not significantly correlated nor was household having access to water.

Finally, having a smart phone in the household does not seem to have a bearing on learning levels. But household having a mobile phone is negatively correlated with baseline literacy and numeracy skills which seems counter-intuitive.

Overall, in terms of the indicators capturing household economic status the findings are somewhat ambiguous. On the one hand household having land seems to be a positive predictor of baseline learning levels, and not having access to clean water is a negative one. These are intuitive as the relationship can be explained to essentially measure the relationship with the household's overall economic status whereby the assumption is that households with relatively more economic ease would result in girls having had more opportunity for learning. However, some of the other indicators do not 'move the right way'. For example HoH having no work is related with higher levels of learning instead of lower ones that would fit with the above theory (despite the potential caveat of IDP households having had access to cash grants and temporary learning opportunities through emergency programming, as noted above). Moreover, a household having a mobile phone, similarly correlates with lower levels of learning and not with higher ones which would fit the pattern. No systemic differences between respondents in these categories could be found through controlling for region and various other factors that could have caused potential omitted variable bias.

With the above in mind, there are two possible explanations at this juncture. Firstly, it is important to remember that all the cohort girls have their individual stories which are impossible to ever fully capture through quantitative analysis. However, what they do have in common is that most of them have had either very little or no exposure to formal education in the past. It is possible that as the intervention begins to gather speed, the different indicators capturing household economic status become more salient and more intuitive as they begin to interact with the program activities and play a role in retention, attendance and learning. Essentially the argument here is that, given the counter-intuitive results above, economic status does not affect learning in the absence of education. Secondly, it is possible that the indicators do not adequately capture levels of household status. Going into Midline it is recommendable to re-evaluate the indicators and consider adding new questions should that be deemed necessary.

Marginalisation

Also, as my friend mentioned, these minority clans don't go to school because of two reasons. First, they are afraid that if they go to the school they may not be enrolled. And, second, even if they go, they don't have the necessary equipment for school. Therefore, they need support and awareness to make them feel they are part of this community. – FGD CEC

TABLE 38: DISTRIBUTION OF LEARNING SCORES BY MARGINALIZED GROUPS

	Literacy			Numeracy			Coefficient after control	
	Mean	P value	n	Mean	P value	n	Lit.	Num.
Overall	44.3		1420	67.1		1420		
Dominant pastoralist lineages (Darood, Hawiye, Dir)	48.5	0.006	671	67.6	0.63	671	7.7** 0.003	
Agricultural lineage (Digil-Mirifle)	43.7	0.689	598	68.1	0.37	598		
Self-identified as marginalised group or minority	28.1	0	143	61.1	0.005	143	- 18.2* *	-7.0** 0.005
Wider definition of minority	40.4	0.015	653	63.9	0.004	653		-4.6* 0.02
Girl speaks minority language	39.1	0.027	460	63.8	0.035	460	-6.8* 0.029	-4.3* 0.036
Girl speaks a different language from LOI	23.5	0	95	56.1	0.01	95	- 13.2* *	
Language of instruction English	26.1	0	64	55.6	0.049	64	-8.9* 0.049	

In Somalia economic and social privilege or lack thereof is traditionally very linked to one’s clan lineage. The above quote demonstrates some of the dynamics in play. As such, in terms of marginalisation, various demographic things are under focus. All variables here behave in the predicted manner, and many have a statistically significant relationship with baseline learning levels. Beginning with respondents belonging to one of the traditionally dominant pastoralist lineages (Hawiye, Darood, Dir), respondents have overall higher levels of literacy when controlling for respondent type. Meanwhile, predictably, those self-identifying as a member of a marginalised group or a minority score on average much lower on literacy than the rest of the cohort, with their mean literacy score being more than 18 percentage points lower than the rest of the cohort when accounting for respondent type. For numeracy the effect is slightly less dramatic 7 percentage points lower (potentially due to previous experience in small business/ child labour), but still statistically significant at the 99 percent confidence interval.

Further, we looked at those who would fit a wider definition of minority by including those respondents who either self-identified as a member of a marginalised group as well as those who belong to a agricultural Digil-Mirifle lineage but reside in areas viewed as controlled by pastoralist lineages. This variable has a

significant and negative relationship with numeracy outcomes.

Next, we looked at the impact of language. The first group we looked at is learners whose first language is not the dominant language in the area in which they reside. As the table above shows, girl speaking a minority language negatively affects both learning assessment scores. The effect is more substantial for literacy, where speaking a minority language is associated with a score 6.8 percentage points lower when accounting for program learning track. For numeracy the same effect is 4.3 percentage points. Similarly, girls who speak a language other than the language of instruction at their learning centre score systematically lower in the Somali language assessment than the rest of the sample. A similar effect was found among girls who attend a learning facility where the language of instruction is English.

School resources and infrastructure, safety

TABLE 39: DISTRIBUTION OF LEARNING SCORES BY SCHOOL RESOURCES, INFRASTRUCTURE AND SAFETY SUB-GROUPS

	Literacy			Numeracy			Coefficient after control	
	Mean	P value	n	Mean	P value	n	Lit.	Num.
Overall	44.3		1420	67.1		1420		
Girl won't use drinking water facilities	41.9	0.337	317	66.5	0.738	317		
Girl won't use toilets at school	43.4	0.745	217	64.6	0.210	217		-6.3** 0.008
No computers available for use at school	44.5	0.351	1359	67.2	0.273	1359		
Girl cannot use books/learning materials at school	39.2	0.022	372	64.9	0.182	372	-6.4* 0.018	
Cannot take materials home at night	29.7	0.240	146	51.8	0.037	146		- 10.1** 0.037
Not enough seats for every student in class	43.3	0.748	191	64.7	0.277	191		-5.9** 0.010
Girl does not feel safe traveling to school	56.9	0.224	25	79.8	0.005	25		12.3** 0.002
Caregiver does not feel it is safe for girls to travel to the school	44.0	0.979	23	68.2	0.867	23		

The above table outlines the findings as it pertains to variables designed to measure overall quality of school resources and infrastructure. All of the variables that are found to have a statistically significant relationship with learning behave as they were expected. Girls who do not feel comfortable using toilets

at school have on average lower numeracy skills. Similarly, girls going to schools where they cannot use materials at school have lower literacy levels, while in schools where they cannot take materials home at night the numeracy levels are significantly lower. Finally, girls in schools where there are not enough seats for children in class have score fewer points in the numeracy assessment when controlling for the type of respondent.

It is important to avoid overanalysing the results since school infrastructure and resources that affect the learning environment should not play too big a role at the baseline as the girls have just enrolled at the learning program. However, all the variables in this category that correlate with learning in a statistically significant fashion which encourages to think that the indicators are fit for purpose. Yet, no measure predicts both literacy and numeracy levels.

Moreover, it should be considered that at baseline the indicators reflecting resources at school can also proxy the overall socio-economic levels in the neighbourhood or village that the girl resides in rather than a condition that will directly explain their baseline levels through affecting her learning experience, as exposure to education at baseline should have been minimal for formal school students.

In addition, both caregivers and girls were asked if they felt safe travelling to school. Girl not feeling safe to go to school in fact has a positive relationship, with this relationship being statistically significant with numeracy whereby those girls who felt it was unsafe to travel to school scored on average more than 12 points higher in numeracy than other girls. This may be a spurious finding as the majority of the respondents who feel unsafe traveling to/from school are ABE and NFE learners who have already been exposed to the program for 4-5 months, potentially resulting in higher learning scores. The qualitative interviews and risk mapping revealed that in fact most girls felt the journey to school was not safe when probed appropriately.

Parents are very worried about their daughters when they are at school. And they think too much about whether their daughter will come home safely or something else will happen. Therefore, because of these reasons, parents prefer that their daughters should stay at home. – FGD Mothers

When girls need to go outside without her brother, some boys make trouble and keep an eye on her. They terrorise girls that walk the street. Girls usually try run to save themselves. - Risk Mapping

Girls can get raped, but boys cannot. Also, girls are very weak compared to boys. So, we cannot even fight back. – Risk mapping

Safety was identified as one of the key issues for attending school. Yet, in the quantitative survey, only 25 out of the 1,420 girls interviewed admitted to the journey being unsafe. It is thus possible that these 25 girls are in fact the most critical and outspoken of their cohort, qualities that could very well relate to skill levels. However, given the findings it is perhaps recommendable that school safety at subsequent phases of evaluation are captured through another measure, for example by creating a score for each school based on the risk mappings.

School Characteristics - formal schools only

TABLE 40: DISTRIBUTION OF LEARNING SCORES BY FORMAL SCHOOL CHARACTERISTICS

	Literacy			Numeracy			Coefficients	
	Mean	P value	n	Mean	P value	n	Lit.	Num.
Overall	44.3		1420	67.1		1420		
Girl attends a private school	28	0.08	156	53.2	0.079	156		
Girl attends a public school	36	0.275	237	60.3	0.337	237		
Girl attends a community school	46.2	0.016	28	71.4	0.001	28	13.3*	14.0**
							0.016	0.001
School has reliable electricity	32.9	0.589	281	56.9	0.303	281		
School has water access within 1 km	27.4	0.001	230	54.7	0.062	230	-13.9	
							**	
							0.001	
School provides at least one meal for students	33.1	0.823	120	55.9	0.436	120		
School distributes sanitary towels	30.2	0.305	118	52.6	0.098	118		
School has only cement floors (no dirt)	31.9	0.275	285	56.7	0.195	285		
School has separate toilets for girls	34.6	0.522	301	59.1	0.558	301		
No female teachers, either FT or PT	20.8	0.004	64	51.4	0.125	64	-15.3	
							**	
							0.004	
Fewer than 5 hours of instruction per day	41	0.043	141	64.5	0.058	141	11.0*	
							0.043	
Teachers miss 1-2 days per week, on average	40.7	0.062	68	61.1	0.453	68		
Agree a little/lot: my teachers are often absent	38.6	0.01	406	64.3	0.061	406	-7.7**	-3.7
							0.005	0.06
Textbooks are shared between students	33.2	0.843	168	59.8	0.566	168		
School charges school fees	31.1	0.229	218	58.1	0.88	218		

This section looks at a sub-section of variables relating to school characteristics. The reader should note that these variables come from the survey with the head teacher and as such they only apply to the sub-sample of cohort girls enrolled in formal schools.

First, looking at the type of school it seems that whether girl goes to a community school or not affects her likelihood of scoring higher in both learning assessments. In both numeracy and literacy, girls attending a community school score 13 percentage and 14 percentage points higher than the rest of the sample of those attending formal schools. As discussed above in relation to school characteristics in general, it should be noted, however, that at this point, this is more likely a proxy measure capturing the kinds of places that have community schools rather than the quality of education in community schools. Indeed, when looking at all the other variables capturing school resources and characteristics – whether school provides meals or sanitary pads, has cement floors, or separate bathrooms for girls, or whether teachers tend to miss school, or if students need to share textbooks or pay attendance fees – none of these variables have a statistically significant relationship with learning levels at baseline.

However, when we look at teacher absenteeism reported by the girls in the surveys instead of that reported by the head teachers, we see that those who report that their teachers often miss school do systematically score lower in both assessments. The effect, as is often the case, seems to be more substantial for literacy where those with absentee teachers score on average 7.7 percentage points lower than their counterparts.

Nevertheless, the school not having female teachers is negatively correlated with literacy levels whereby girls in those schools have an average Somali literacy score 15 points lower than the rest of those girls who attend formal school. This is perhaps intuitive, but again at baseline we would caution against reading too much into this finding as school-related factors should not affect skill levels too much at this point given the very limited time that schools will have had to actually influence the girls' learning levels.

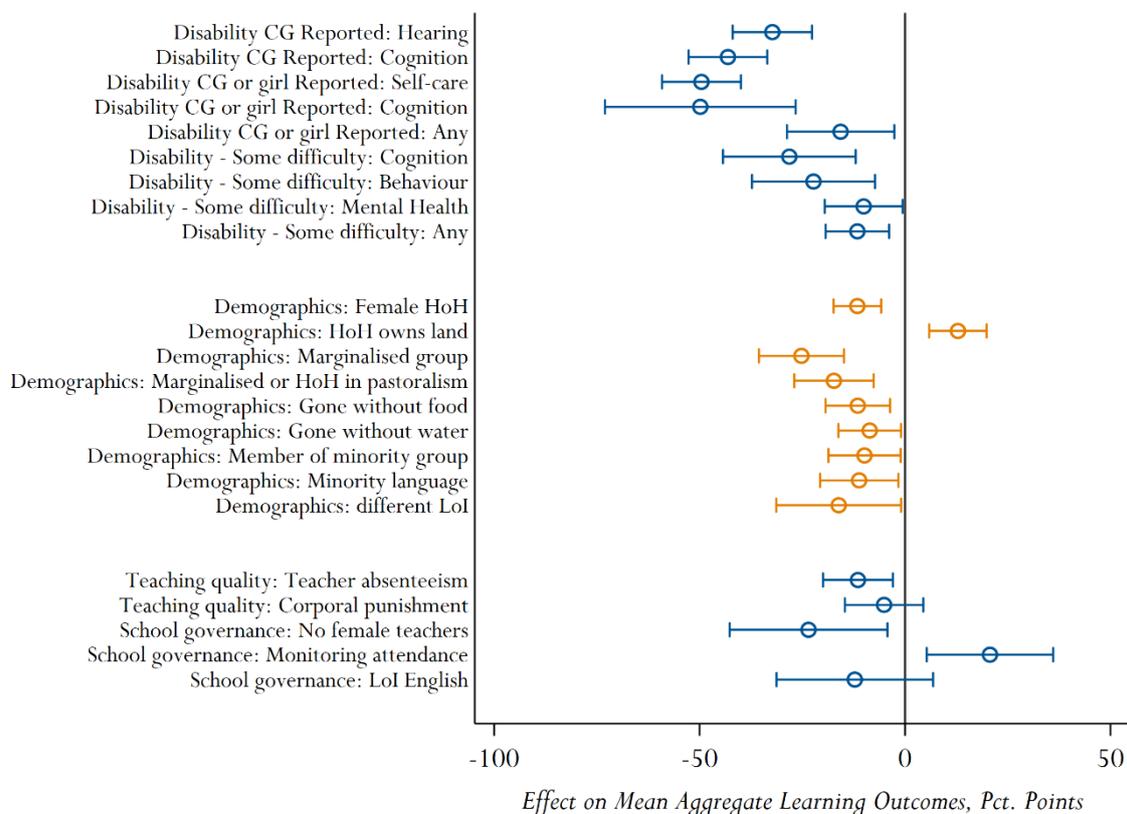
Finally, it should be noted that safety related concerns discussed in the qualitative interviews extend to the school environment:

When she comes to the school, she can go the toilet and she can get raped. She could feel scared and this may make her leave the school. – Vignette qualitative exercise

I sometimes feel scared when I get to school because there are many guys sitting in front of the class, and they insult and harass us. – Vignette qualitative exercise

The factors relating to school safety and comfort thus undoubtedly will play a big role in determining girls' attendance, retention and transition, and ultimately their learning outcomes.

FIGURE 12: SUBGROUP EFFECT ON MEAN AGGREGATE LEARNING OUTCOMES



This section brings together a number of variables that have been discussed in the preceding sections. These are all the different factors that have been found to have a statistically significant or marginally significant effect on learning at this Baseline.

As can be seen, various disability indicators correspond with lower aggregate learning outcomes. Similarly, many indicators on marginalisation in the demography category also correspond with lower learning outcomes, with a girl being from a marginalised group being the most negative predictor of learning levels at baseline of all variables that have been investigated. Similarly, respondents who have experienced lack of food or clean water also score significantly lower than others as do those whose parents are in pastoralist occupation. Meanwhile those whose families own land score higher overall. In turn, girls from female-headed households had overall lower levels in the learning assessments.

In terms of school characteristics for formal schools, very few factors were found to be significant at this juncture, and some of the variables produced counter-intuitive results. It can be assumed that since the program has had very limited time to work with the cohort girls, the factors relating to school resources have not had time to affect the learning outcomes. It is also possible that some of the factors when significant are that way because they co-vary with characteristics relating to the overall environment in the village or neighbourhood where the girl lives rather than due to their direct impact on her performance. At this Baseline, however, CEC monitoring attendance was positively correlated with learning outcomes, while teacher absenteeism and the lack of female teachers correlated negatively with assessment scores.

5. TRANSITION OUTCOME

In this section, we discuss the second core outcome for AGES, transition. Transition is a broad outcome, insofar as it encompasses multiple distinct pathways, and pathways are defined slightly differently for different types of girls. For example, for ABE girls, transition into an NFE center after finishing the ABE program is considered a positive transition outcome, while NFE girls who transition into gainful employment is also considered a positive transition outcome. However, for formal school girls – who are younger than ABE or NFE girls – transition into employment is not considered a positive outcome, as they should still be in education at their age.

The logic of AGES is based heavily on encouraging girls to transition successfully within their learning institutions and from one type of institution into another. The program primarily targets girls who have never been enrolled in school. In general, this means they are less likely to stay in school in the long-term, as girls who are starting school at an older age are less likely to complete primary education. Age is a major contributing factor to drop-out rates in Somalia, as pressure to marry and start a family intensifies in the later teenage years. Therefore, girls who are starting primary school at age 11 or 12 are at a significant disadvantage, as they will only be able to finish primary school before the common marriageable age if they stay in school continuously and advance without interruption from grade to grade. Unfortunately, girls who have not been in school at a younger age face a number of additional challenges: generally, their families are likely to undervalue education; they may face low expectations for the value of education, as completing a few years of primary education is not typically associated with significant job opportunities; and they may lack the basic skills and behavioral characteristics necessary to succeed in school.

The table below outlines the main transition pathways that the program considers a successful outcome. Note that the final row of the table identifies transition pathways for out-of-school girls, as AGES seeks to make ABE and NFE learning available to wide variety of girls and encourage their enrolment in such programs. However, these girls are not included in our sample at all, as the cohort sample was drawn from within formal schools, ABE centers, and NFE centers. As no cohort girls are out-of-school at the baseline, this transition pathway is not directly relevant to the baseline evaluation of AGES, but may become pertinent in future rounds.

Institution	Baseline point	Successful Transition	Unsuccessful Transition
Formal School	Enrolled in Grades 1-2	<ul style="list-style-type: none"> In-school progression to next grade level, including secondary school 	<ul style="list-style-type: none"> Drops out of school Remains in same grade
ABE	Enrolled in any stream of ABE	<ul style="list-style-type: none"> Progression within centre into next stream/level Transition into formal school Transition into paid employment or self-employment 	<ul style="list-style-type: none"> Drops out of ABE centre Remains in same level or does not progress
NFE	Enrolled in NFE	<ul style="list-style-type: none"> Enrolment in formal school 	<ul style="list-style-type: none"> Drop out

		<ul style="list-style-type: none"> • Transition into paid employment or self-employment 	
Out-of-School	Not enrolled	<ul style="list-style-type: none"> • Enrolment in formal school, ABE, or NFE 	<ul style="list-style-type: none"> • Continuing as out-of-school

The main transition pathways involve progression within school or other learning opportunities – upward movement from grade to grade, and transition from ABE into the formal school system, among others. The predominant forms of unsuccessful transition involve dropping out of the education system altogether, or failing to progress year-on-year across grade levels. Finally, though it is not extremely common, transition into paid employment or self-employment is considered a successful transition outcome.

As discussed in the methodology section previously, transition rates will be calculated at midline for the cohort girls. These rates will be compared to benchmarks established during both the baseline and midline, for different cohorts of girls. Our analysis at baseline focuses on the transition benchmark girls intended for comparison with NFE girls at midline and endline. We do not analyze the transition rates of cohort girls, because transition rates among this sample is 100 percent – in practice, all girls in the sample were recruited into the AGES program, and transitioned from being out-of-school into enrolment in either a formal school, ABE center, or NFE center. All girls "successfully transitioned" at the time of the baseline, making analysis of their transition outcome generally uninformative. Instead, our analysis in this section serves two goals: first, to establish the rate of transition among the benchmark sample; second, to analyze the nature and correlates of successful transition, to shed light on possible barriers to transition and opportunities for improved project design.

The women in the transition benchmark were recruited from the households of cohort girls. After recruitment at learning centers, enumerators followed up at the households of cohort girls, completing a household survey with each. As part of the interview, the head of the household was asked to enumerate the women present in the household who fall between the ages 20 and 24 years. After obtaining consent and completing the rest of the household survey, the selected woman was recruited into the benchmark transition sample; where households had more than one such woman, one woman was selected for interviewing randomly, using a random number generator pre-programmed into the survey script.

In total, the achieved sample of benchmark transition women was smaller than expected, with just 276 women interviewed (from a target of 314). Unfortunately, the targeted women were difficult to locate, as most women in the target age range no longer live with their parents. Girls in the regions targeted tend to marry at or before age 20, and most married women live with their husbands. While women living with their spouse would still be eligible for inclusion in the transition benchmark – indeed, the sample

includes many who are married and some who are mothers – women who live with their husbands generally do not simultaneously live with girls who were recruited into AGES programming.¹⁰²

The age breakdown of the sample is heavily skewed toward younger women: the mean age is 21.1 years, and 52.5 percent of the sample is 20 years old. Just 8.0 percent of the sample is aged 24 years. In terms of geographic distribution, 43.5 percent of the sample is drawn from Banadir, compared to 32.3 percent from Jubaland and 24.3 percent from South West state.

Women in the sample were not assessed for transition, in a precise sense, because transition is defined by one's situation at two points in time. For instance, when transition rates for cohort girls are assessed at midline, their position at that time will be compared to their position at baseline, to determine whether they have advanced a grade, transitioned into a new learning institution, and so forth. If a girl is enrolled in formal school at the second point in time, successful transition, nonetheless, requires knowing her grade level at the first point in time. In the analysis throughout this section, we are studying women at a specific point in time – a snapshot of their enrolment status, grade, and employment status at this moment. In this context, "transition outcomes" are a measure of what the woman in question is doing at that moment, whether it is working in formal employment, domestic work paid in-kind, or various types of schooling. In the following two sections, we report aggregate transition outcomes, and then investigate subgroup-specific outcomes, with the goal of understanding specific characteristics that impede enrolment and successful retention in school.

AGGREGATE TRANSITION OUTCOMES

The table below reports detailed transition outcomes by state, and in the aggregate, among the benchmark sample. The table makes clear that the vast majority of women aged 20-24 years are not engaged in any specific educational or employment activity – in total, 85.5 percent of girls are neither employed nor enrolled in school or alternative learning options. Just 7.6 percent of women are enrolled in formal school, with a further 5.8 percent in NFE or ABE programs, with a much heavier emphasis on NFE programs. Employment of all kinds is also very limited, even among this older cohort of women.

TABLE 41: TRANSITION OUTCOMES

Transition Outcome	Overall
No current activity	85.5%
In formal school	7.6%
NFE	5.1%

¹⁰² Specifically, one must consider the type of household that would include both a married woman age 20-24 and a girl eligible for AGES programming: if a married woman eligible for the transition benchmark sample lived with her husband, she is too young to have a daughter eligible for AGES programming; such a household would need to be multi-family or the husband would need to have children from a previous marriage. Our expectation is that most respondents in the transition benchmark sample are women who live with their parents, whether or not they are married.

ABE	0.7%
Self-employment	0.7%
In-Kind HH Work	0.4%
Total	100.0% (n = 276)

Again, it is important to note that our sample includes relatively few (n = 22) women who are aged 24 years. However, this small subsample is nonetheless informative, because it continues a downward trend from those aged 22 years, in which older women have lower transition rates. It also confirms our prior expectations regarding the impact of age on transition rates. We can, therefore, infer that age is a significant factor in shaping transition outcomes, though we cannot say, based on this analysis alone, whether there is a differential impact of age on enrolment versus employment rates.

In practice, enrolment rates drop precipitously with age among this group of women, while other positive transition outcomes are not clearly impacted by age. Among women 20-22 years, 9.1 percent are enrolled in formal school, compared to 1.8 percent of women 23-24 years old. On the other hand, participation in ABE or NFE programs are not clearly distinguished by age: 5.9 percent of younger girls are enrolled in such programs, compared to 5.5 percent of older girls. Similarly, there is no tangible difference in employment rates as a function of age. However, given that only three women in the entire benchmark sample are employed in any form, this information is far from conclusive.

To understand further the nature of transition in this group of women, we analyze the relationship between a number of subgroups and transition outcomes. In the table below, we disaggregate the data according to a wide range of subgroups, most of which are identical to those applied in the learning section above. Specifically, we study the effect of state, disability, household economic status, household demographic characteristics, and household members' views on girls' education on transition outcomes for benchmark girls.¹⁰³

We report on three outcomes in the table. The first is the share of women who are enrolled in formal school. The second is the share of women who have *ever* been enrolled in formal school, since age 5. This

¹⁰³ Where we report on subgroups that are specific to a girl or woman, as opposed to her household or community (e.g., disability), we use data collected specifically about the benchmark woman. Where we report on subgroups defined by household characteristics (e.g., clan membership, household food security, etc.) we use data reported, typically, by the head of each household.

second outcome is not of obvious relevance to benchmarking transition rates. However, it is informative in terms of relative marginalization – it will allow us to analyze which *types* of girls are most marginalized in terms of educational attainment, and the types of households in which they grow up. While the rate at which girls have *ever* attended school is not useful for benchmarking, we can infer from the results which types of girls are most marginalized, in a context in which the analysis is somewhat less subject to sample size concerns.¹⁰⁴ The third outcome is the share of respondents who are engaged in *any* educational or employment activity, such as domestic work that is paid in-kind, non-formal education, as well as formal employment and enrolment in a formal school.

The results in the top panel of the table reflect the findings discussed earlier, insofar as current enrolment is heavily determined by state of residence. Our analysis of disability is even more complicated, though the results are more straightforward in many ways. We first report results in which women are categorized according to the standard Washington Group classification of disability, which focuses on tasks that an individual "cannot do at all" or can only do with "a lot of difficulty." As with the cohort sample, the share of women in this group who self-identify as having a disability, according to this standard, is vanishingly small – just 16 women, in total, and nearly all of those who self-identified in this way did so on the basis of mental health difficulties. We also report results from our more liberal coding of disability, which includes women who have "some difficulty" with the tasks from the Washington Group questions. Unfortunately, this group includes 23.9 percent of the total sample, an implausibly high rate of overall disability among this demographic group. This high rate is driven primarily by mental health disabilities; usefully, when we consider only specific types of disability – such as cognitive or behavioral difficulties, the results are suggestive of higher rates of current enrolment than the sample average.

TABLE 42: TRANSITION OUTCOMES BY SUBGROUP

Subgroup	Enrolment Rate	Ever-Enrolled Rate	Rate of any Productive Activity	Sample Size
Full Sample	7.6	42.4	14.5	276
Disability Subgroups				
Vision disability	0.0	0.0	0.0	1

¹⁰⁴ In total, 117 women (42.4 percent) in the sample have been enrolled at some point after age 5, compared to just 21 (7.6 percent) that are currently enrolled. The larger share in the first group makes conclusions drawn about the correlates of "ever-enrolled" stronger; in addition, the ever-enrolled outcome is not driven by the current age of the woman – in sharp contrast to current enrolment – so our analysis is less obscured by age and the manner in which it may or may not be correlated with other predictors of enrolment status (e.g., marriage). Put bluntly, it is difficult to separate the effect of age and state of residence on current enrolment, because they are closely correlated; it is much more straightforward to analyze the effect of state of both variables on whether a woman was ever enrolled in the past, even if this outcome is less proximate to our true interest.

Vision disability (Alt. 2)	50.0	50.0	50.0	2
Hearing disability	N/A	N/A	N/A	0
Hearing disability (Alt. 2)	N/A	N/A	N/A	0
Arms/Hands disability	N/A	N/A	N/A	0
Arms/hands disability (Alt. 2)	N/A	N/A	N/A	0
Mobility disability	N/A	N/A	N/A	0
Mobility disability (Alt. 2)	0.0	0.0	0.0	2
Communication disability	N/A	N/A	N/A	0
Communication disability (Alt. 2)	0.0	50.0	0.0	2
Cognitive disability	100.0	100.0	100.0	1
Cognitive disability (Alt. 2)	14.3	57.1	21.4	14
Behavioral disability	50.0	50.0	50.0	2
Behavioral disability (Alt. 2)	15.4	38.5	23.1	13
Affect disability	4.9	39.0	7.3	41
Affect disability (Alt. 2)	3.6	41.1	10.7	112
Any Physical disability	0.0	0.0	0.0	1
Any Physical disability (Alt. 2)	25.0	25.0	25.0	4
Any non-affect disability	33.3	33.3	33.3	3
Any non-affect disability (Alt. 2)	16.7	50.0	25.0	24
Any disability	6.8	38.6	9.1	44
Any disability (Alt. 2)	5.7	40.7	12.2	123
Household Characteristics				
Female-headed Household	7.0	44.7	14.9	114
Head of household has no education	3.4	41.4	10.3	58
Head of household has no formal education	7.1	43.8	13.4	224
Caregiver has no education	5.3	42.1	17.1	76

Caregiver has no formal education	7.9	43.9	15.1	239
Head of household does not earn a wage	8.8	43.0	14.9	114
Pastoralist household	25.0	25.0	25.0	4
Household Economic Status				
Household has roof of inferior materials	8.5	29.8	14.9	47
No food most days	11.8	35.3	17.7	17
No food many/most days	11.1	48.2	19.8	81
No food many/most/a few days	8.2	45.3	14.7	232
No water most days	5.9	47.1	11.8	17
No water many/most days	8.4	56.6	19.3	83
No water many/most/a few days	7.9	45.2	14.9	241
No medicine most days	10.0	32.5	12.5	40
No cash income most days	8.6	37.1	17.1	35
Household owns land	7.1	46.9	13.3	98
Caregiver has savings	16.7	33.3	16.7	6
Household owns a phone	8.0	42.9	16.0	238
Household owns a smartphone	17.6	54.9	29.4	51
Woman's Characteristics				
Woman is married	1.1	30.7	6.8	88
Woman was ever married	2.6	36.4	7.8	154
Woman is a mother	1.6	34.4	4.7	128
Safety in Area				
Journey to school is unsafe for cohort girl, per cohort girl	0.0	33.3	33.3	3
Journey to school is unsafe for cohort girl, per caregiver	0.0	33.3	0.0	3
Caregiver Views of Education				

Caregiver aspires to university education for cohort girl	6.4	39.4	12.7	251
Caregiver believes it is worthwhile to invest in girls' education	7.7	43.3	14.6	247
Caregiver believes schooling is more important for girl than work	8.0	45.1	13.3	113
Caregiver believes cost of schooling is acceptable reason for girl to be out of school	8.6	44.3	14.9	174

As with learning outcomes, household characteristics – and the characteristics of the adults in the household, especially – are determinative of enrolment rates. Women growing up in households in which either their caregiver or the head of their household has no education at all (including Quranic education) are at a distinct disadvantage in terms of current enrolment, but not in terms of enrolment over their lifetime. Specifically, 41.4 percent of women whose head of household has no education have *ever* been enrolled, slightly lower than the ever-enrolled rate of all women in the sample. However, just 3.4 percent of women whose head of household has no education are *currently* enrolled, less than half the rate in the overall sample.

Household economic status presents a confusing blend of findings. In households with a roof made of inferior materials, women in the benchmark sample are slightly more likely to be enrolled currently, but much *less* likely to have ever been enrolled in their lifetime. Women from households that have faced food insecurity over the past year have higher current *and* ever-enrolment rates, yet land ownership is not associated with similarly large increases in enrolment rates. Perhaps the most compelling findings regarding household conditions concern access to water and ownership of a smartphone, both of which are associated with higher enrolment rates.

On the other hand, the best predictors of non-enrolment is marriage and motherhood. Girls who are currently married are enrolled at rates of just 1.1 percent, compared to 4.6 percent for girls who were previously married but are now divorced, and 13.9 percent for girls who have never married. Of course, marriage rates are somewhat correlated with age, which may account for a portion of the association between marriage and enrolment status. However, even when we control for age in a linear regression, marriage strongly predicts being out of school.¹⁰⁵ Indeed, marriage is also associated with a reduced rate

¹⁰⁵ The model we estimate is a linear model predicting the binary outcome enrolment, as a function of binary variables that identify women who are currently married, married in the past, and each age level. Women who are married see a 9.3 percentage point reduction in enrolment rates versus non-married women; women who have ever been married – even if they are not now – see a 10.8 point reduction compared to who have never been married.

of enrolment in non-formal education, which is in line with theoretical expectations, given the nature of expectations placed on women in Somali society.

It is important to consider the quantitative results reported in this section with methodological caveats in mind. The sample size ($n = 276$), while not overly small, is sufficiently small that subgroup analysis should be considered cautiously. Moreover, the sample itself is not entirely straightforward, as it is divided across locations – not just states, but even towns – with substantially different contexts. To the extent that these contextual differences are correlated with each other, they may obfuscate important findings, or even cause spurious findings. Finally, the fact that the sample was drawn from households that have a girl enrolled in the AGES program may shape the types of women who were interviewed. We do not necessarily believe a random household sample would have been better from a methodological perspective, but there are questions around the characteristics of women who are present at age 20-24 years in households with girls aged 11-18. They may not be entirely typical of women in the locations studied, and it is hard to determine whether this has had any effect on the findings reported here.

Given the methodological caveats mentioned above, triangulating findings from the qualitative data is particularly important. This is especially true because many barriers to enrolment we expect to be critical, based on previous research, are present in only small numbers in the transition benchmark sample. For instance, we have theoretical reasons to expect girls in pastoralist households to face additional barriers to enrolment in school, but there are only four pastoralist girls in the transition benchmark sample. Similarly, we expect a girl's domestic chore burden to influence her ability to attend school and, by extension, remain in school. But data on chores is only available for cohort girls, not transition benchmark girls, so we can only infer that a household might have placed a heavy chore burden on a benchmark girl indirectly. We cannot actually observe whether benchmark girls faced this barrier personally.

Themes in the qualitative data around barriers to transition emerged fairly clearly. One of the most straightforward findings is that girls with disabilities face a unique set of challenges, with a heavy emphasis on their ability to actually reach the school in question. For instance, CEC members cited transportation difficulties for a specific girl and the need for a car to take her to school, and the fact that the problem became even worse during the rainy season.¹⁰⁶ Another CEC member described a girl who was given a wheelchair and was then able to stay in school.¹⁰⁷ Not all difficulties for girls with disabilities are restricted to transportation, of course. Disability tends to be stigmatized in Somalia; in practice, some families "hide" members who face mental health or other problems. While it is not clear, in our experience, whether this inclination to make family members with a disability less visible applies also to physical disabilities or strictly to mental health issues, this issue can certainly restrict access to schooling.

Beyond disability, four additional barriers came through consistently in the qualitative data:

- Marriage and motherhood
- Pastoralist livelihoods
- Economic or financial limitations of households
- The household's reliance on girls for domestic work.

¹⁰⁶ FGD with CEC members. Other CEC members cited transportation difficulties for girls with disabilities as well.

¹⁰⁷ FGD with CEC members.

The first factor seen as inhibiting enrolment and successful transition is also, arguably, the one that came through most consistently or was described most forcefully as a barrier by qualitative interviewees. Many qualitative interviewees, from teachers to CEC members, to girls themselves, indicated that most girls drop out of school when they get married, and that marriage is among the most important barriers girls face.¹⁰⁸ As one CEC member described it: "The main reason [to stop attending school] is marriage. If she gets married, she may drop out of the school. Or maybe she will be away from the school for so long."¹⁰⁹ This finding fits closely with the quantitative results presented earlier, in which marriage was associated with a precipitous drop in current enrolment rates. As noted previously, the correlation between marriage and enrolment rates remains strong even after controlling for age, state, and other potential confounding factors. Marriage is broadly viewed as the point at which education for a girl or woman stops.

The obstacles faced by girls in pastoralist households also come through in the qualitative data, though fewer participants specifically cited pastoralism in discussing enrolment and transition. Responses regarding pastoralism tended to focus on two issues: the fact that pastoralist families may be far from a school and lack the transportation necessary for their child to reach the school; and the fact that pastoralist families, by definition, move from place to place, uprooting their families.¹¹⁰ The latter issue either results in a girl switching schools repeatedly or – probably more often – dropping out of school or missing the remainder of the school year after the family migrates. One CEC member relayed a story of a girl who was performing well in school, but suddenly stopped attending – after some investigation, the CEC became aware that her family had moved to a new grazing area, which was very far from town, preventing her from attending school.¹¹¹ It is also possible that children in pastoralist households face discrimination at school from teachers and administrators or – more likely – teasing and bullying from other students. This is a theme we investigate in more detail in Section 7.4 (School Management), but it bears mentioning here as well.

The third barrier that emerges from the qualitative data concerns the financial limitations of households. Naturally, this barrier often overlaps with pastoralism, as pastoralist households are more likely to experience food insecurity, and lack household access to water.¹¹² Several qualitative interviewees emphasized the financial strain of education and the fact that their households lacked the financial

¹⁰⁸ FGD with CEC members; FGD with teachers; FGD with teachers; Vignette exercise with girls.

¹⁰⁹ FGD with CEC members

¹¹⁰ FGD with CEC members (3)

¹¹¹ FGD with CEC members.

¹¹² Although the sample size of pastoralist households is small, 40.0 percent report experiencing hunger many or most days in the past year, compared to 26.9 percent of non-pastoralist households. In terms of water, the differences are even more stark: 19.6 percent of non-pastoralist households were fully water-secure (i.e. they *never* went without household access to clean water). No pastoralist households were able to claim the same, and 48.0 percent stated that they had gone without access to clean water many or most days during the previous year.

resources necessary to educate their daughters.¹¹³ These arguments typically came from FGDs with mothers, which is consistent with other quantitative data – not reported directly in this section – which showed that a majority (62.3 percent) of caregivers believe that the cost of education is an acceptable reason for their daughter to not attend school, and that a majority (56.4 percent) of caregivers cite a lack of money as the *specific* reason their girl was not enrolled in school previously. As one mother stated, "If girls get enough income to cover their school fees, it would be easy for them to stay at school."¹¹⁴ While it is reasonable to accept these viewpoints, they sometimes appear to contradict other interviewees in the same communities. For instance, while the mother cited above makes it sound as if the only barrier is economic, teachers and CEC members in some of the same communities view marriage as the most important barrier to girls' continued education, and others highlighted the role of anti-education pressure from girls' families.¹¹⁵

Girls also seemed to have a different perspective, which tended to contradict an exclusive focus on financial constraints. Using a participatory exercise designed by CARE, girls in "vignette FGDs" were presented with realistic scenarios that girls in their community might face, and asked how they thought the story would turn out. For instance, girls were presented with a story about a hypothetical girl named Sabirin, who lives with her aunt and helps her aunt with her tea shop and caring for her aunt's four children. When girls were asked what would happen to Sabirin, and whether she would go to school, almost all girls fixated on the household and shop-related responsibilities she faced as a barrier to her education.¹¹⁶ When asked, one participant indicated that she knows a girl like Sabirin, whose mother requires her to complete a lot of chores at home, and the girl has never attended school.¹¹⁷

It is not entirely surprising that girls would focus on the role of chores and domestic responsibilities in the context of this vignette. After all, the story mentions Sabirin's responsibilities directly, and the story is framed in a way that will prime girls to focus on that aspect of Sabirin's life. But it is telling that few or no girls mentioned Sabirin's father's sickness – which prompted the family to send Sabirin to live with her aunt in the first place. We do not view this evidence as dispositive, by any means, given the structure of the story and its emphasis on Sabirin's domestic responsibilities.

It is worth noting that caregivers themselves cite this as a reason why their daughters were not enrolled in school prior to intervention by the AGES program: of caregivers who described the reasons why their girls were not enrolled, 22.7 percent said that their girl needed to work, earn money, or help with domestic chores at home. The rate at which caregivers cited this as a reason for keeping a girl out of school rose with the girl's age, but even in the case of 12-14 year old girls, 19.5 percent of caregivers cited this as one of the reasons their girl had not been enrolled previously. Caregivers are also surprisingly frank about the extent of domestic work their girls complete. Over one-third (36.9 percent) of cohort girls spend either half or all of a typical day completing various household work, such as fetching water, caring for siblings, cooking, and tending crops. Just 35.3 percent of caregivers report that their girls complete no such work or complete only about one hour per day, typically. The fact that household is so extensive,

¹¹³ FGD with mothers (2).

¹¹⁴ FGD with mothers.

¹¹⁵ FGD with CEC members (2).

¹¹⁶ Vignette exercise with girls (4).

¹¹⁷ Vignette exercise with girls.

and so openly reported upon by caregivers, suggests the importance of this barrier, lending credence to our interpretation of the qualitative data.

The findings that emerge from the qualitative data do not always agree with the subgroup analysis of transition rates performed above. For instance, disability and pastoralism, do not appear to be a significant impediment to enrolment among transition benchmark girls. However, as we have noted, the quantitative data used for the subgroup analysis is far from conclusive, especially given the very small number of transition benchmark girls who have a disability, or whose families are engaged in pastoralism. Similar caveats apply to other potentially relevant subgroups analyzed statistically above. For this reason, it is important to triangulate the qualitative data against alternative quantitative data sources, such as the information provided by caregivers about their cohort girls. Our analysis of why cohort girls were not enrolled prior to the AGES intervention, how caregivers view the relative priorities of household finances versus girls' education, and their description of their girls' chore burdens all confirm the information gleaned from qualitative interviews, that marriage, a household livelihood built around pastoralism, and domestic chores act as some of the biggest impediments to girls' enrolment and retention in school.

6. SUSTAINABILITY OUTCOME

6.1 INDICATOR 3: ECONOMIC STATUS OF PARENTS

Proportion of parents able to support costs of girls' education

For many children in Somalia, fees charged for attendance at school – and/or additional costs of education, such as uniforms and supplies – present a significant barrier to enrolment and staying in school. Indeed, issues relating to the cost of education were a very common theme in the focus group discussions. The following quote from a CEC member represents a typical point of view: “Some students don't get one meal per day and they cannot focus on academics, so they have to stay at home and try to find their survivals instead of pursuing education.” The members of the CEC in that school said they try to provide support for school fees for some of the girls “but the CEC cannot help every girl economically.”

Indeed, the issues of economic barriers to education is very well recognized among both community members as well as those who work at schools. There is also clearly support for providing the less privileged girls with support for fees, and this is being done through CECs in some places.

However, this issue will require a wider solution across the board. Consider the fact that more than half of the girls in the cohort go to schools that charge fees. Secondly consider that when the caregivers of girls in the transition sample were interviewed 55% of those who did not attend school cited lack of money as the reason and 23 % the fact that the girl needs to help at home or work. Finally, let us keep in mind the fact that less than 13% of the families in the sample never went without a daily income. Moreover, the majority of the parents (98%) lack any savings; furthermore, 46% of the caregivers stated that their families are unable to meet basic needs without charity. In a similar manner, 46% of the households lack cash income on a regular basis. In other words, most of the families in the cohort are financially unstable, and it is unlikely that such households would be able to meet the costs of school fees.

In this situation, there is a triple risk of girls not attending school for financial reasons. Firstly, in case of hardship, the fees are likely to be high in the list of costs to cut, given that these girls were not attending school before, presumably in many cases due to economic reasons. Secondly, in case of economic hardship, the girl's input at home and potentially at family business becomes ever more valuable. Thirdly, as discussed elsewhere in this report, economically impoverished families might marry off their daughters

for financial gain, and according to community members, married girls are much less likely to continue going to school.

Short of a fundamental restructuring of the economy in every locality, AGES can intervene by providing school fees for those most in need in the sample. Yet, given the above, a majority of the beneficiaries will likely require some form of support. In the long term, thus, a more sustainable plan would require the intervention to be linked into initiatives where the local government collects tax revenue to be used for public services. In some of the districts the UNDP/Ministry of Interior implemented Joint Programme on Local Governance project is seeking to improve the revenue collection capacity of the recently established district councils. In the short term, still, it is unlikely that local governments will be able to provide the funding for either school fees or teacher salaries (which is what the fees are mainly used for).

AGES program is involved in a number of actions relating to mitigating this risk. Firstly, savings groups are being established and ABE and NFE beneficiaries are encouraged to participate. Secondly, the program aims to provide free education to 15,000 beneficiaries. Another one of AGES activities – mobilisation of savings groups – intends to support parents to access capital for small businesses and build safety nets for families, thus increasing family capacity to meet basic needs and subsequently, to support the cost of education.

If these actions go ahead as planned, they will of course have a big potential for reducing the economic burden on the beneficiary families.

	Rating	Justification
Sustainability rating	2 emergent	<ul style="list-style-type: none"> - School fees constitute a significant economic barrier for girls' attendance - This barrier is well recognized by communities and teachers, and there is clear support for economic assistance for those in need. However, there are doubts as to how much capacity there is locally to support all of the ones who need it. - Some CECs are involved in supporting some of the most underprivileged children with the fees - CARE will support savings groups and schools in providing free tuition but it is unclear where these activities are at the moment - Ideally a long-term solution would include local revenue collection and support for teachers salaries in order for the solution to be systemic

6.2 INDICATOR 4: GIRLS WITH DISABILITIES

Proportion of GwDs who remain in school

AGES is specifically targeting disabled girls in the intervention. In that regard there were many questions in the research tools pertaining to identifying disability. In addition some questions were devised to capture attitudes towards disabled people's participation in education.

As it comes to the community level attitudes towards disability, people overwhelmingly support the idea that disabled children should also go to school. Almost 100 percent of those interviewed were of the opinion that disabled children have the right to attend school. However, qualitative interviews did talk of disabled girls staying at home because of stigma and bullying that they face at school. For many community members disability was seen as something relating to mobility, and many of them, when asked about disability, talked about the difficulty of disabled girls to make the journey to school and depending on others for transportation.

When it comes to the schools, 57 percent of teachers said that teachers in their school supported girls with disabilities a good deal. However, as it pertains to the practical arrangements to make disabled girls' education a reality, much is still wanting. Much of the qualitative interviews lamented the fact that girls

with disabilities are unable to attend school and if they attend they rarely get the support needed to learn effectively. The desire to improve this does seem to exist as teachers talk about the need to support girls with disabilities, but oftentimes this is accompanied with talk of getting equipment for that purpose. However, some teachers talked about making sure that they seat the girls with vision and hearing disabilities in the front row. However, some others talked about even teachers joining in on the bullying of the disabled children

Moreover, some teachers said that in their schools they had organized transport for those who have difficulty with mobility. But it was more often in the qualitative interviews that those with disability were seen as a group that could currently not be helped by the education system, and it was seen as something that would be the government’s responsibility to address. This systemic level of change cannot be assessed at this baseline as no interviews were conducted with representatives of government.

When looking at current enrolment levels of disabled girls, they are very low indeed. The field research team was asked to specifically seek out girls with disability in every community. The teams were able to locate and interview 125 girls’ caregivers. Out of the 125 girls, only 26 or approximately 20% were currently enrolled. A majority, 69%, had never attended school at all.

TABLE 43: SUSTAINABILITY INDICATOR 4

Rating		Justification
Sustainability rating	<u>1</u> latent	<ul style="list-style-type: none"> - A majority of the disabled girls who were contacted through the communities, 69% had never attended school. A third of those who had attended school at some point, had dropped out. - While community members practically all (99.7%) agreed that girls with disability have the right to attend school, qualitative interviews talked about disability being common. - Most community members linked the inability to attend school also to issues of mobility and to lack of equipment needed due to disability. - In some schools, some actions have been taken to address those things that can be addressed with little resources, such as transport support and seating girls with hearing and vision difficulty in the front row. - Teachers overall are supportive of improving the situation of the disabled learners, yet the lack of resources and equipment to do that is always cited when talking about the question. - Schools tend to see the question of disability as something that the state government should address but the state governments most likely do not have the resources to do that at this point - The systemic level of this indicator cannot be assessed at this baseline because no interviews were conducted with government.

INDICATOR: ADOPTION OF A BROADER FRAMEWORK FOR ABE AS PART OF MOECHE’S NON-FORMAL EDUCATION POLICY, CATERING TO THE NEEDS OF MARGINALISED GIRLS

At the moment, there is no national framework for accelerated / alternative basic education in Somalia, or an official ABE curriculum. As of 2017, ABE had a limited reach (1,361 students in south central Somalia), despite the extraordinarily high demand for accelerated education options, as highlighted in the

Education Sector Strategic Plan¹¹⁸. The AGES curriculum has been officially approved by the MOECHE and State MoEs, and the project is currently collaborating with UNICEF and USAID on the development of a broader framework for ABE. As part of these efforts, AGES has shared emerging results to date with USAID as inputs to a national workshop on ABE, which took place in Mogadishu on February 4-6, 2020.

	System	Community	Learning space
Indicator 1:	MOECHE and State MOEs supporting a broader approach to ABE implementation	Self-replication rate of village savings and loans (VSL) groups	Proportion of umbrella schools adopting new methodologies
Sustainability rating:	N/A*	N/A	N/A
Indicator 2:	Adoption of a broader framework for ABE as part of MOECHE's non-formal education policy, catering to the needs of marginalised girls	Proportion of parents able to support costs of girls' education	Proportion of umbrella schools providing ABE
Sustainability rating:	N/A	2 - EMERGENT	N/A
Indicator 3:	OOS tracking mechanism incorporated in the EMIS	Parental support for girls' and boys' participation in GEFs and BEFs	Proportion of teachers implementing inclusive education strategies in class
Sustainability rating:	N/A	N/A	N/A
Indicator 4:	Strengthened evidence base on GwDs	Proportion of GwDs who remain in school	N/A
Sustainability rating:	N/A	1 – LATENT	N/A
Indicator 5:	N/A	Proportion of GEFs and BEFs implementing community actions to support attendance and retention	N/A
Sustainability rating:	N/A	N/A	N/A

¹¹⁸ Ministry of Education, Culture and Higher Education (2017), Education Sector Strategic Plan 2018-2020, p.103

Baseline Sustainability Score (0-4)	Not assessed at baseline	1.5	Not assessed at baseline
Overall Sustainability Score (0-4, average of the three level scores)		1.5	

TABLE 44: SUSTAINABILITY INDICATORS

**a variety of indicators were not assessed by the external evaluator as the data available did not lend itself to the analysis at this time. These indicators are marked with 'N/A'*

Overall, the sustainability indicators seem fit for purpose. However, as it pertains to systemic sustainability, much depends on the MOECHEs' resources. While this might not be the program's direct responsibility, it would be advisable to investigate any potential bottlenecks there and seek solutions that link other programs to the MOECHE staff who would require support in this regard.

As per community level indicators, they are apt for capturing the key issues, i.e. attitudes and resources. The more rural areas where CECs are unable to levy financial support and where the VSLAs might have limited success due to the community's poor resource base, CARE will need to supplement funding in the intermediate term. However, given the likely challenges for CEC and VSLA fund raising in those areas, a longer-term strategy might be required.

As regards the learning space, an issue that is not adequately outlined in the sustainability plan has to do with teacher salaries which is consistently viewed as a key challenge for retaining teachers, particularly in rural areas where issues relating to learning tend to compound. Explicitly outlining a strategy for securing teachers' pay in the sustainability plan will help address it.

7. INTERMEDIATE OUTCOMES

7.1 ATTENDANCE

The Intermediate Outcome Indicator 1 of the LNGB project – **“Improvement in marginalised girls' attendance in schools, ABEs, flexible learning courses throughout the life of the project”** fulfils a twofold objective in achieving improved learning outcomes of the marginalized beneficiary girls, such as literacy, numeracy, financial literacy and knowledge on menstrual hygiene management (MHM), and contributes to transition of these girls to (non-) formal education or employment. The main assumptions for successful project implementation rely on several expectations:

- 1) improved learning opportunities, along with the positive community and parental attitudes to girls' education will positively contribute to the attendance rates of the marginalized girls;
- 2) improved learning environment along with the increasing attendance rates will lead to improved learning outcome in the specific subject/areas of interventions of the marginalized girls;
- 3) 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.

The overall target for the **Intermediate Outcome Indicator 1 is 80% of beneficiary girls have improved their attendance rates** compared to the baseline values, with a midline target of 50% and an endline target of 70%. Consequently, this section will firstly explore the main properties of the attendance rates, their distribution across cohort groups and establish the baseline values. Secondly, the quality and reliability of the attendance scores will be further examined. Thirdly, further analysis will explore the statistical relationship with the key selected sub-groups and identify the main predictors of the attendance scores for each of the cohort groups. The analysis will be based on quantitative and qualitative data.

Analysis of attendance records

While visiting the schools, we collected information about the attendance from multiple sources, such as school records, teachers' or class attendance records and headcounts conducted by our team leaders. In the analysis, we will mostly rely on the headcount-based measure of attendance rates. The head counting is in general considered as a more reliable method to measure the attendance rates, compared to attendance based on school records, especially due to the lower quality, neatness and organization of attendance records in schools.

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 randomly selected one class for the administration of the headcount survey. To measure the attendance, an indicator was calculated as a percentage score based on a number of children counted as present in the class by the team leaders and a number of children enrolled in the class. Therefore, one observation in the dataset corresponds to one class in the formal schools or one stream in the ABE/NFE centres. As we are using the class-level data, the term 'cohort group' refers to the classes with FE, ABE or NFE students, rather than the individual students.

The next section will additionally explore the main properties of the attendance scores, quality of the records and explore further statistical linkages between the variety of indicators and attendance scores. As the enrolment rates may not be regularly updated, and there is a risk of artificially increased attendance rates due to presence of our teams, we will also assess the internal validity of attendance scores, by inspecting enrolment rates and by comparing them with the teachers attendance marked for the day of the visit and the day before. Moreover, we will also assess the overall quality of the record-keeping based on the observations of our teams.

School attendance rates

The main findings presented in this section suggest that there is a high concentration of the maximum attendance rates across all the cohort groups, accounting for 34% in the sample¹¹⁹. **The overall mean**

¹¹⁹ The headcounts and the attendance based on the teachers' counts were not collected in the case when the attendance/school records were not available or accessible. Consequently, a total of 111 out of 635 observations (representing the specific classes visited during the headcount) were excluded from the further analysis.

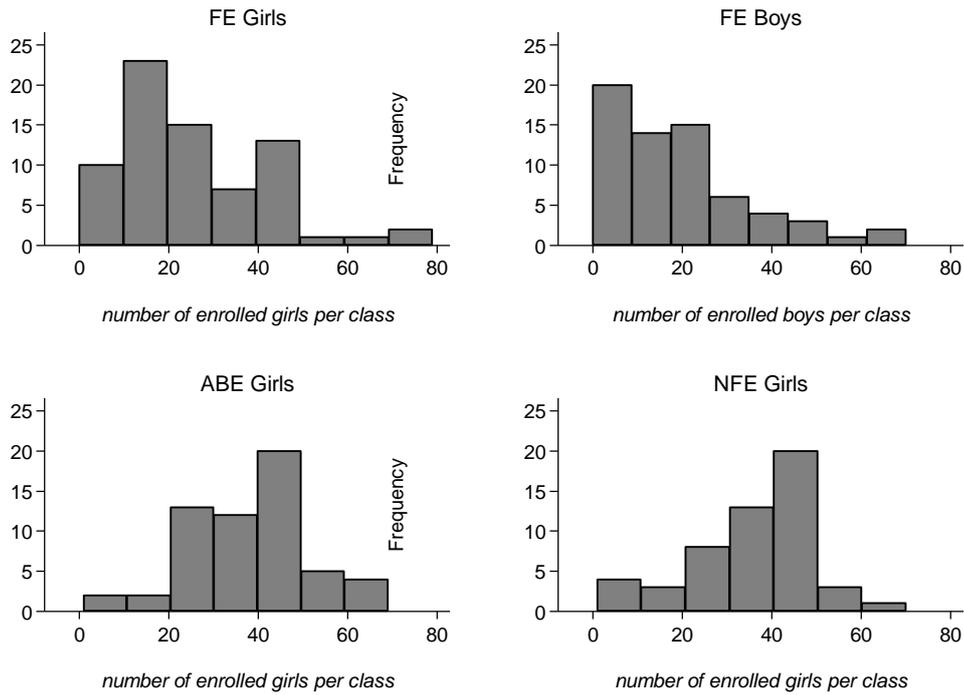
attendance rate is 86% of children present in the class at the time of the visit. In the following paragraphs, we will first assess the attendance rates disaggregated for different cohort groups, geographic locations and grades. First, we will analyse the internal validity of the attendance rates. Secondly, we will inspect the properties and distribution of the attendance rates. Thirdly, we will further assess the distribution and the differences in the attendance mean rates for different cohort groups.

As the measurement of attendance is based on the total enrolment numbers, the validity of the attendance rates is dependent on the reliability of enrolment records. The graph below illustrates the distribution of the number of children enrolled per class for each of the cohort groups. The results indicate that the majority of the class sizes in the formal schools in grades 1 and 2 is clustered around the values below 20 children enrolled per class. Specifically, for the girls enrolled in the formal schools there are few outliers with values above 70 girls enrolled in the class. On the other hand, class sizes for the ABE and NFE girls seem to be clustered around higher values, more precisely around 30 and 45 girls enrolled per class in the ABE classes and around 50 girls enrolled per class in the NFE classes¹²⁰.

Moreover, due to the errors related to the administration of the headcount surveys, when a value of 0 was entered as a number of present girls present in the class during the visit, the attendance scores were excluded from the subsequent analysis, which was the case for 3 observation among the ABE cohort groups and for 2 observations among the NFE girls (two headcounts in ABE centre).

¹²⁰ The original enrolment considers a maximum of 45 girls per class; larger class sizes reflect oversubscription and community pressure for additional enrolment.

FIGURE 13: DISTRIBUTION OF ENROLMENT NUMBERS PER COHORT GROUP¹²¹



In addition, certain enrolment numbers in formal schools were excluded from further analysis due to excessively high numbers of children enrolled in the class¹²², while no such cases were observed for the

¹²¹ The boys enrolled in formal education classes were not identified or enrolled through AGES, and are not part of the cohort, although they will benefit from AGES interventions in that school.

¹²² For the girls enrolled in the formal school, the values of the enrolment rates and the calculated attendance scores were excluded from the further analysis in one case, when the number of the enrolled girls in grade 5 was equal to 64 girls in a school, which was considered as excessively high compared to the enrolment trends for the given school and the average enrolment rates for the grade 5 for the cohort group. For the boys enrolled in the formal school, the same situation was considered in two separate cases, where enrolment rates and calculated attendance scores were not included in the subsequent analysis, as 82 boys were enrolled in the grade 7 in one school and 52 boys were enrolled in grade 7 in another one.

ABE and NFE cohort groups¹²³. Overall, the results suggest that the total numbers of children enrolled in such classes can be considered sufficiently reliable for the calculation of attendance, without compromising the overall results.

In the following paragraphs, we will continue to examine the distribution of the attendance rates. The graph below illustrates the distribution of the attendance rates. The results suggest that the attendance rates are clustered around the maximum values, with a long left tail¹²⁴. The maximum attendance rates were found in 29% of the classes.

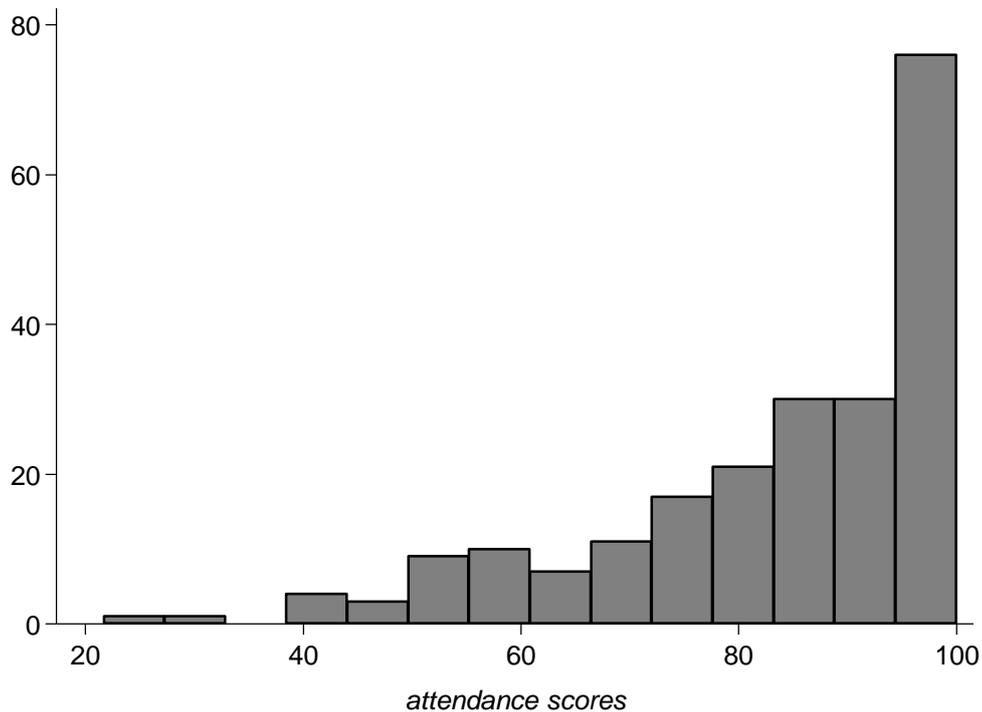
The overall mean attendance is 84%¹²⁵. Besides, in only 24% of the classes, the attendance rates were lower than 75%. To better understand the characteristics and distribution of the attendance rates, the following paragraphs will examine more in detail the attendance rates, disaggregated by cohort group.

¹²³ In three classes for the ABE cohort group the enrolment rates were ranging between 65-69 girls and in one ABE cohort group case the enrolment was 70 girls in one class. However, based on our observation during the fieldwork, these numbers were still considered relative plausible/acceptable.

¹²⁴ While inspecting the outliers and extreme values of the overall attendance rates, a number of observations with extremely high attendance rates (>100%) were excluded from the further analysis. This was the case of 2 headcounts for the girls and 3 headcounts for the boys in the formals school, 4 headcounts for the ABE cohort and 3 headcounts for the NFE cohort group.

¹²⁵ After including the weights adjusted for the unequal sample size at the school level, the overall percentage mean score remained the same., Thus weights will not be applied while we will compare distribution patterns In all the regression models we will control for the unequal cluster size.

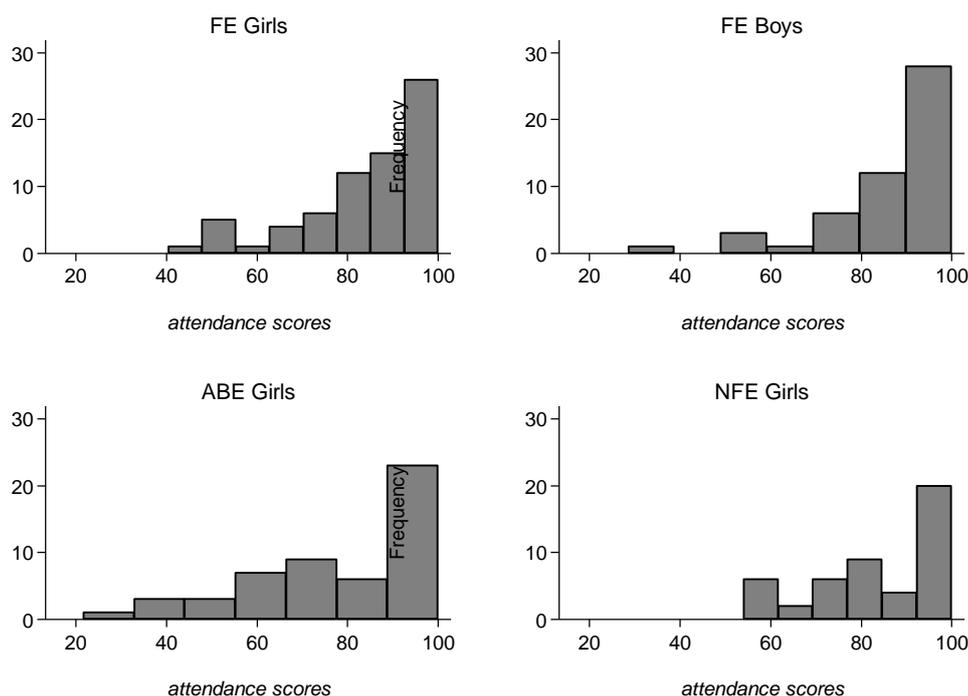
FIGURE 14: DISTRIBUTION OF THE OVERALL ATTENDANCE RATES



Overall, the attendance rates remain high across all cohort groups. The visualization of the attendance scores for each group is depicted in the graph below. For the girls in the formal school, in 16% (36) of the classes all enrolled girls were present at the time of the visit, all the enrolled boys were present in 13% (28) of the classes, all the enrolled ABE girls were present in 29% (16) of the classes and all the enrolled NFE girls were present in 30% (15) of the classes¹²⁶.

¹²⁶ Extremely low attendance rates (less than 40% attendance rates) were observed in only one class for FE boys and one ABE class.

FIGURE 15: DISTRIBUTION OF THE ATTENDANCE RATES BY COHORT GROUP



The table below summarizes the key statistic for each of the cohort groups. The mean attendance rates are higher than 75% across all the cohort groups. The highest mean values were observed for the girls and boys enrolled in the formal school. The lowest mean value was observed for the ABE girls. On the other hand, the minimum values of the attendance rates were higher for NFE girls, compared with other groups.

TABLE 45: DISTRIBUTION OF THE ATTENDANCE RATES BY COHORT GROUP

Cohort group	Obs.	Mean	Min	Max
FE girls	140	84.77	40.43	100.00
FE boys	102	87.03	28.57	100.00
ABE girls	52	79.59	21.67	100.00
NFE girls	47	83.99	54.00	100.00

The results of the regression analysis¹²⁷ in the table below indicate that boys and girls in the formal schools are more likely to have higher attendance rates. The difference in the mean rates is significant at 90% CI for boys. On the other hand, ABE girls are more likely to have lower attendance rates. The ABE cohort

¹²⁷ The results from the table 3 come from 4 different regression models, where each cohort group binary variable was examined separately, against other cohort groups.

group has a substantial and negative effect, decreasing the attendance by 5.65 percentage points. No difference was observed for the NFE girls.

TABLE 46: ATTENDANCE REGRESSION ANALYSIS PER COHORT GROUP

Cohort group	Coef.	Standard error
FE girls	1.27	2.25
FE boys	4.06*	2.36
ABE girls	-5.65	4.12
NFE girls	0.11	3.19

*significant at 90% confidence interval, ** significant at 95% confidence interval, *** significant at 99% confidence interval

The analysis of attendance rates disaggregated by cohort groups suggests that the attendance rates remain high across all the cohort groups. The children in the formal schools seem to have better attendance rates, compared to other cohort groups. The non-formal educational programs seem to have lower attendance rates, with the ABE girls more likely to have lower attendance rates. While we have included only children in the grades 1 and 2, in the following paragraphs we will briefly examine attendance rates disaggregated by grades in the formal schools.

To conclude this section, the results suggest that there is a significant share of classes with perfect attendance rates and the overall attendance rates are high. Girls in ABE and NFE classes have worse attendance compared to girls and boys in formal schools. In the next section, we will assess the reliability and validity of the attendance scores, by comparing the attendance based on headcounts with attendance estimations from alternative sources – teacher/school attendance records. Also, we will inspect the quality and availability of attendance records in formal schools.

Quality of attendance records & measurement

In this section, we will first assess the quality of the teachers' attendance records in comparison to the attendance from the headcount surveys. Secondly, we will examine the reliability of the attendance from the headcount surveys in comparison to the attendance marked the day before the visit. Thirdly, we will assess the quality of the school records.

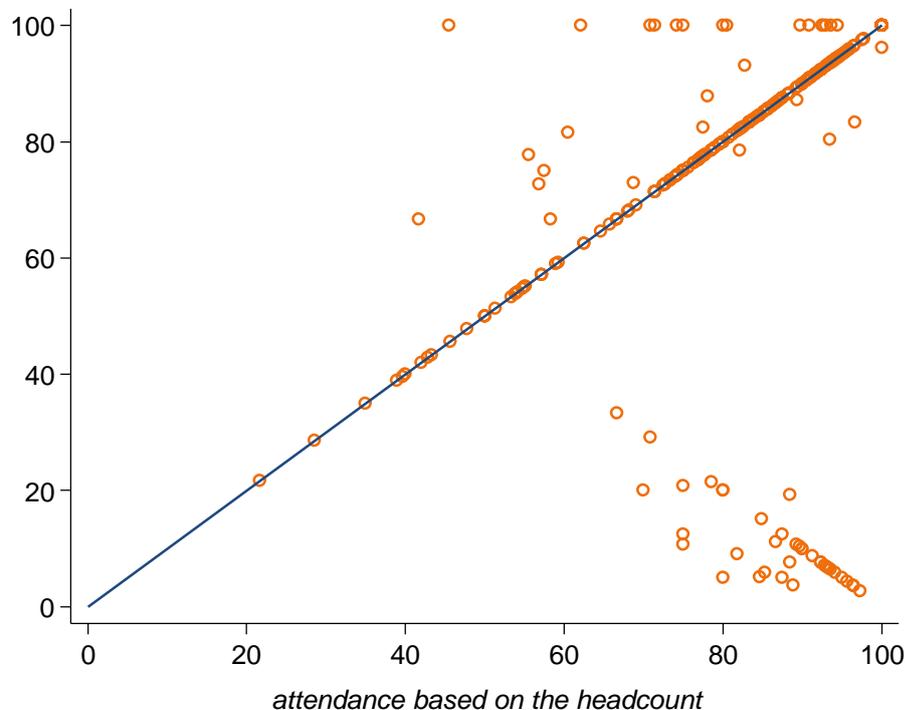
The graph below plots the percentage attendance measure (*x-axis*) against the percentage attendance calculated as a ratio of the attendance marked by the teacher for the same day over the number of enrolled students. The results suggest that the distribution of teachers' attendance mostly corresponds to the distribution of attendance from the headcounts. In around 62% of the cases, the attendance marked by the teachers is the same as the headcount.

It is possible that in some cases the teacher would rather mark the number of students missing rather than the number of students present in the class, which could partly explain the inverse relationship with the teachers' attendance and headcount observed in the lower right corner of the graph below. Only in around 8% of the cases, the attendance marked by the teachers is lower than the headcount.

On the other hand, in around 28% of the cases, the teachers seem to have indicated higher numbers of the students present in the class compared to the numbers present at the time of the visit. The smaller discrepancy could be explained by the fact that the number of students present in the class in the morning, when the attendance was taken by the teacher, might have changed by the time our teams visited the class. However, the very big positive differences between the teachers' attendance records and the headcount

might suggest higher inaccuracy of the records and a tendency to inflate the numbers of the present students. In around 25% of the cases, the attendance marked by the teachers is higher by 20% or more, than the headcount.

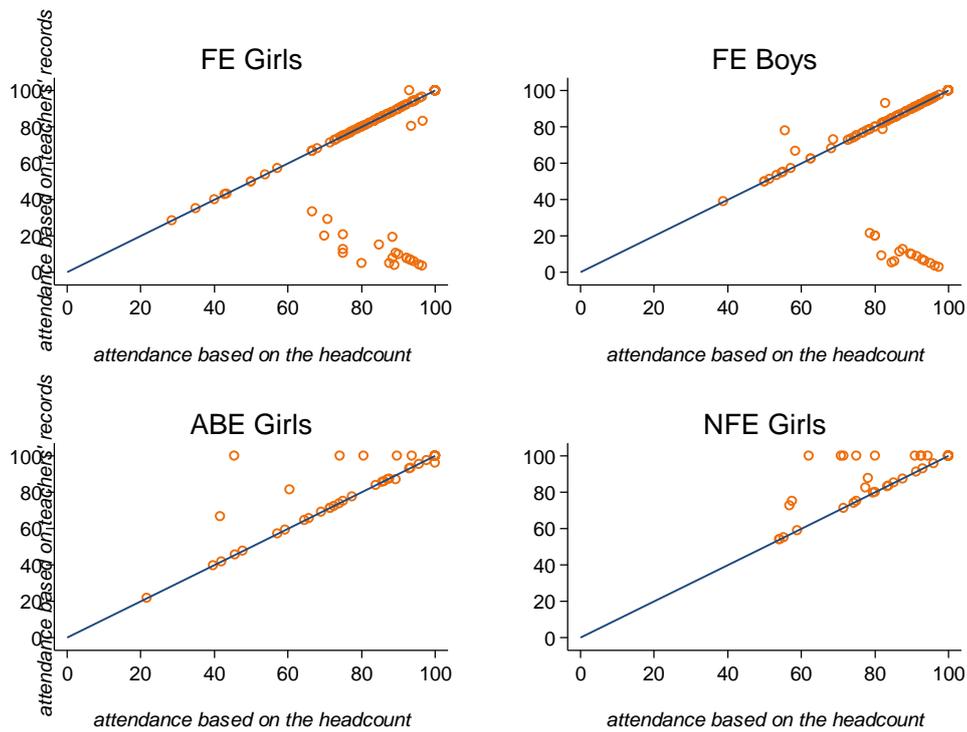
FIGURE 16: COMPARISON OF ATTENDANCE RATES



The graph below further details the breakdown of attendance records and headcount for each of the cohort groups. While the attendance records in the formal schools seem to better fit the headcount, there is also a reverse relationship between the attendance from the records and the attendance from the headcounts, in the lower right corner of the graph. For girls in the schools, around 60% of the attendance records fit the headcounts, while 31% of the attendance records were higher than the headcount values. Similar results were observed for the boys in the formal schools, with 65% of the attendance records fitting the headcounts and 29% of attendance records higher than the headcounts.

For the ABE and NFE girls, there seems to be a bigger discrepancy between the attendance record and the headcounts. Besides, for both the ABE and NFE cohort groups, there is a bigger share of attendance records that are higher than headcounts values. Around 70% of the ABE attendance records correspond to the attendance from the headcounts in the ABE classes, while in 16% of the classes, the attendance records were higher than the headcount values. In around 65% of the NFE classes the attendance records are the same as the attendance from the headcounts, while in 31% of the NFE classes, the attendance records were higher than the headcount values.

FIGURE 17: COMPARISON OF ATTENDANCE RATES BY COHORT GROUP



As indicated in the table below, the results indicate that in the NFE classes girls’ attendance records seem to be significantly more likely to be higher, suggesting lower accuracy of the attendance records. The attendance records from the headcounts in the ABE classes seem to be significantly different and lower, compared to the attendance records, suggesting the attendance records in the ABE classes may be less accurate, compared to the attendance taken from the headcount surveys.

TABLE 47: ATTENDANCE SOURCE BY COHORT GROUP¹²⁸

Cohort group	FE Girls		ABE Girls		NFE Girls	
	Headcount	Records	Headcount	Records	Headcount	Records
Coef.	-0.19	0.07	-7.16**	4.19	-2.76	10.13***

¹²⁸ When we have compared the headcount attendance and attendance marked by teachers for girls in grades 1 & 2 only, results remained statistically insignificant, with coefficients -2.56 for the headcount attendance and -2.47 for the attendance marked by the teacher.

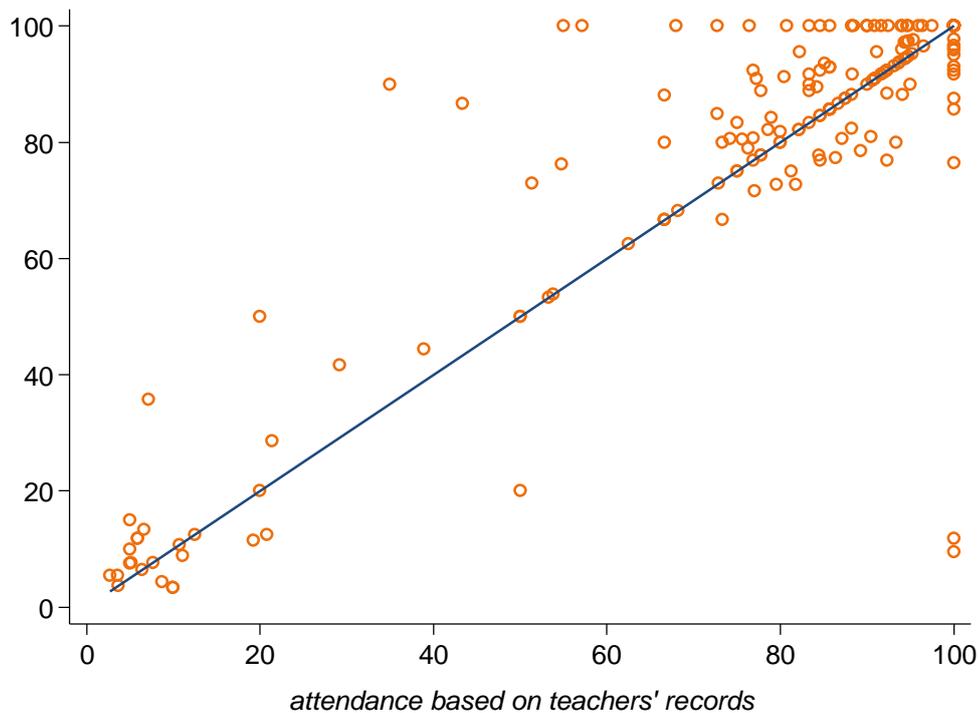
Moreover, the graphical visualization of the distribution of attendance rates when counting only children in grades 1 and 2 (as in figures 19 and 20) is fairly similar, with absent observations depicted in the lower right corner of the present graphs.

Standard error	1.07	1.76	4.15	5.63	3.11	4.74
-----------------------	------	------	------	------	------	------

*significant at 95% confidence interval, ** significant at 90% confidence interval, *** significant at 95% confidence interval

The graph below plots the distributions of the attendance records from the day of the visit and the previous day¹²⁹. While comparing the attendance records in the formal schools, as illustrated in the graph below, more observations are located above the 45° line, suggesting that the attendance records from the day before the visit tended to be higher. Moreover, 44% of the attendance records from the day before the visit correspond to the attendance records from the day of the visit. Additional 44% of the records from the day before the visit are higher than the records from the day of the visit and only 10% are lower. These results suggest it is less likely that the attendance rates would be higher because of the school visits by our teams. On the other hand, it is possible that the headcounts were conducted on the second day of the school visit and thus, the attendance records marked for the day before the headcount was taken would be already higher.

FIGURE 18: COMPARISON OF YESTERDAY’S AND TODAY’S ATTENDANCE RATES



To gather additional information about the quality of the attendance records, we have collected additional information based on our observation of the quality of the record-keeping while visiting the formal schools. Team leaders were required to indicate the availability of the attendance, record the number of

¹²⁹ This question was asked only in the formal schools.

days when attendance was marked during the past 5 days and assess their completeness, neatness and organization, on the 4-point Likert scale. The table below summarized the questions and answer options¹³⁰.

Overall, in 76 % of classes, we were able to access the attendance register. In case when we were able to access the records of the formal schools, in 64% of the classes the attendance was recorded for the last 5 days, in 14% of the classes for the last 4 days and in 30% of the classes, the attendance was recorded for 3 days or less. In 58% of the cases, all records were completed and in 38% of the cases, only some of the school records were completed.

TABLE 48: STATUS OF ATTENDANCE RECORDS WHEN YOU ARRIVE

Status (in %)	Obs.	Overall
All completed	115	58.10
Some completed	76	38.40
None completed	7	3.50

In addition, when we were able to access the school records, in 18% of the classes visited in the formal schools the attendance records were considered as somewhat incomplete and in 8% of the classes as extremely incomplete, while 44% of the records were mostly and 29% extremely complete.

TABLE 49: COMPLETENESS OF THE OVERALL ATTENDANCE RECORDS

Completeness (in %)	Obs.	Overall
Extremely incomplete	16	8.00
Somewhat incomplete	35	18.00
Mostly complete	84	44.00
Extremely complete	56	29.00

Similarly, as the table below suggests, when considering the neatness and organization of the attendance records, in 17% of the classes the attendance records were considered as somewhat disorganized and in only 3% of the classes as extremely disorganized.

TABLE 50: NEATNESS AND ORGANIZATION OF THE OVERALL ATTENDANCE RECORDS

Neatness and organization (in %)	Obs.	Overall
Extremely disorganized	6	3.00
Somewhat disorganized	32	17.00

¹³⁰ Completeness of the overall attendance records: Extremely incomplete; Somewhat incomplete; Mostly complete; Extremely complete; Neatness and organization of the overall attendance records: Extremely disorganized; Somewhat disorganized; Mostly organized; Extremely well-organized;

Mostly organized	91	48.00
Extremely well-organized	62	33.00

While the headcount-based attendance scores seem to be reliable, especially compared with the teachers' attendance records for the day of the visit and the day before the visit, more notable discrepancies between the headcount and attendance records were observed in the non-formal educational programs. In one of the previous sections, we have noted that the headcounts in the non-formal educational programs were more likely to be lower compared to the headcount-based attendance rates in the formal institutions.

Sub-group analysis

The present section will focus on the more nuanced analysis of the attendance rates, building multivariate regression models, and identify factors that are likely to impact the measured attendance and explain the previously identified differences across cohort groups and geographic locations.

A set of variables from different sources will be used for the analysis, such as individual-level data from the household survey and the teacher survey and the school-level data from the school survey. First, we will use the household survey data, aggregated at the school or centre level to analyse the overall attendance scores. Then we will focus on the formal schools and ABE and NFE centres specific variables and their statistical relationship with the attendance scores. Lastly, we will combine the most significant variables for multivariate regression analysis. To better inform our analysis, we will also reply on the main findings from the qualitative data analysis.

Overall attendance & sub-group analysis

The school-level data for the selected subgroups was originally drawn from the teacher survey¹³¹ and household survey, where the variables originally coded as dummy variables for individual-level data were aggregated at the school level¹³². Thus, the dummy variables transformed into the continuous variables at the school-level will be interpreted as the (probability) that the attendance rates increase or decrease with a higher or smaller percentage of girls corresponding to a specific category at the school or centre level. Only attendance rates for girls in grades 1 and 2 were included for the girls in formal schools.

¹³¹ The only two variables coming from the teacher surveys are: 1) Barriers to remain in school: lack of parental support for education and 2) Girls are more likely to miss school.

¹³² During the process of aggregating the data, the mean values of the dummy variables were calculated by the school level data, while the values of the transformed variables represented the percentage of the responses with the value 1 per school. Thus, the individual level dummy variables were transformed into continuous variables at the school level. The only exception for this process was the YLI variable, which as already a continuous variable, was aggregated by its mean values at the school level.

As indicated in the table below, the initial simple regression analysis of the subgroups indicates that variables related to the lack of mental health disability (a), land ownership (b), several hours a day spent on household chores a day (c), caregivers prioritizing housework over girls' education (d) and teachers miss 1-2 days a week on average (e) have statistically significant and negative impact on the attendance rates.

- a) The attendance rates also tended to be 0.09 percentage points lower, with a higher proportion of girls who reported they had a mental health disability¹³³. There has been a consensus among all the respondents that girls with disabilities were significantly less likely to not to go to school, especially when the school was missing supportive devices, teachers were not qualified to teach disabled girls or girls with reduced mobility lacked means to access the school.
- b) A higher number of the households in the school community who own land would additionally decrease the attendance rates in the school by 0.16 percentage points. Some respondents explained that girls in the poorer households with land and farming activities were more likely to stay at home to work rather than to attend the school, especially during the rainy season and in less accessible locations.
- c) The fact that girls at school were more likely to spend several hours a day on HH chores further decreased the attendance rates by 0.17 percentage points. Similarly, during the FGDs the traditional perceptions of girls' roles at home were the most frequently cited reason for girls' absence in schools. Also, mothers often expressed an opinion that girls' caretakers (mostly mothers) decide if the girls are going to miss school because of the housework.
- d) In the same way, girls were significantly less likely to attend school when caregivers were more likely to believe that housework is an acceptable reason for girls to not attend school, representing a negative change of 0.14 percentage points.

"Boys can't keep children, they can't manage families, they don't wash clothes. A girl has a lot to do at home. Parents then sent boys to schools for education and the girls to do their housework." (Mother FGD)

- e) When the teacher was likely to miss 1-2 days a week, the attendance rates were significantly likely to drop by 0.17 percentage points, suggesting that the effective supervision not only of the student but also teacher attendance could significantly improve the attendance rates in the schools.

¹³³ Due to the small sample size for the girls with disabilities, alternative and more liberal coding was used for the purposes of the analysis. The dummy variables for each type of the disability were based on the disability levels reported by girls, also including cases when girls responded to have experienced some level of disability, as opposed to stricter coding, then only cases when girls reported to have experienced 'a lot of difficulty' or 'cannot do at all'.

In addition, other variables with were not statistically significant but had a substantial negative effect on the attendance rates, such as when girls were mothers (0.07 percentage points change) and when teachers did not sufficiently explain the usefulness of subjects (0.11 percentage points change). The low number of observations could explain the absence of a statistically significant change, and for the same reason, these variables will be excluded from further analysis.

On the other hand, several factors have positive and statistically significant effect on the attendance rates, such as when HoH had no formal education or when the household had limited access to food, water, medication and cash income (a); pastoral household (b); household ownership of the phone (d); reliable electricity at school (e); higher YLI scores (f); and caregivers' aspirations for university education (g) have positive and significant impact on the attendance rates.

- a) A higher share of households, where HoH have no education at all, is associated with a positive 0.08 percentage points change in the attendance rates. Also, when more girls are coming from more vulnerable households, that are food insecure or have limited access to water, medication or cash income, the attendance rates are more likely to increase by 0.27 percentage points for food insecure households, 0.34 percentage points when there is limited access to water and 0.21 percentage points when there is limited access to medication or cash income. Possible explanations of these results are a) the data was collected in an urban or peri-urban areas where girls are in general more likely attend schools, and b) households are more likely to have lower economic and educational status, and c) in these communities there is a likelihood there are already existing interventions to support girls' education and sensitize parents ¹³⁴.
- b) The fact that the children are more likely to come from pastoralist households is associated with a positive change of 0.23 percentage points in the attendance rates. Same factors as detailed above could partly explain this finding. Moreover, due to the large effect on the attendance rates, we will include the variable in the multivariate regression analysis to assess whether its impact on the attendance rates is not driven by additional factors.
- c) Household phone ownership increases the attendance rates by 0.17 percentage points, which could be associated with the fact that phone ownership potentially improves and facilitates the communication between parents and the school/teachers, which could have a positive impact on the attendance rates. This claim was often supported during the FGDs with the CEC members, where the fact of directly contacting parents was often mentioned as a way to improve girls' attendance in schools.
- d) The fact that the school has reliable electricity seems to have a positive impact on the attendance rates of the girls, with a positive change of 0.11 percentage points. The finding could be linked to the previous assumption that improved learning environment has a positive effect on the girls' attendance.

¹³⁴ After controlling for the cohort group, geographic location and age are, for each of these variables, the results suggest that only the fact that the household has limited access to medication has a statistically significant impact on the attendance scores, representing a change of 17 percentage points, significant at a 95% CI,

- e) Girls with higher YLI scores were (statistically) significantly more likely have better attendance, with a 0.21 percentage points change. This finding may suggest that a) girls are less likely to miss school when they are able to advocate for their educational needs and express their opinions or b) girls have higher self-esteem in the household that the girls' education is more valued.
- f) Girls were significantly more likely to attend school when caregivers did not believe it was safe for girl to attend school, representing a positive change of 0.27 percentage points. While we do not have a direct explanation for the association, the low number of observations and prolonged insecurity in the locations where we have collected data may partly explain he results.
- g) Girls were significantly more likely to attend school or centre when caregivers of the girls were more likely to aspire for university education of girls, representing an increase of 0.19 percentage points.

Not significant, but substantial and positive change was observed when girls did not believe they could attend the school while menstruating, with 0.21 percentage points increase in the attendance scores. The counter-intuitive and insignificant results could be partly explained by a low number of observations.

One of the outcomes of the qualitative analysis suggests that menstruation is one of the main barriers for attending schools, mentioned by several mothers in 9 out of 10 FGDs with mothers. Respondents also mentioned pain, other negative health symptoms and limited access to menstrual pads and material kits as the main reasons why menstruation would prevent girls from attending schools.

Several variables seemed to have no statistically significant impact on the attendance rates, such as absence of the female teachers in the school; parental perceptions of the value of the girls education, in comparison to the cost of the education; and access to the school learning materials, such as textbooks.

TABLE 51: OVERALL ATTENDANCE AND SUB-GROUP ANALYSIS

	Sub-group	Obs.	Coef.
Barriers to remain in school: lack of parental support for education		99	-0.07
Mental health disability		195	-0.09**
Single orphan		150	0.00
HoH has no education		150	0.08*
Caregiver has no education		168	0.08
HoH does not have an occupation or does not earn income		236	0.0.09
Pastoral household		51	0.23***
Went to sleep hungry most/all nights, last 12 months		91	0.27**
Went to sleep hungry many days (more than 10), last 12 months		220	0.02
Went without water for home use most/all days, last 12 months		74	0.34***
Went without medicines most/all days, last 12 months		127	0.21***
Went without cash income mostly/all days, last 12 months		134	0.21**
Household owns land		233	-0.16***
Caregiver has savings		47	0.10
Owns a phone		255	0.17*
Girl is or was married		46	-0.01

Girl is a mother	33	-0.07
Girl does not use drinking water facilities at school	126	-0.02
Not enough seats for every student in a class	82	-0.07
School has reliable electricity	96	0.11***
School has separate toilets for girls	104	0.07
No female teachers, either PT or FT	24	0.01
Caregiver does not feel it is safe for girls to travel to school	31	0.27**
Girl spends few hours or more per day doing HH chores	251	-0.17***
Caregiver aspires to university education for girl	255	0.19*
Caregiver believes education worthwhile even if funds are limited	255	0.13
Caregiver believes work or HH chores are an acceptable reason to not to attend school	239	-0.14***
Caregiver believes the cost of education is an acceptable reason to not attend school	255	0.07
Teacher explains how the topic is useful for students' lives	62	-0.41
Physical punishment witnessed last week	216	-0.10
Girl does not believe she can attend school if menstruating	37	0.21
Teachers miss 1-2 days per week on average	22	-0.17**
Textbooks are shared between students	60	-0.04
Girls can take textbooks/materials home at night	46	0.02
Girl speaks Af-Maay at home	143	-0.05
School language of instruction is English	2	0.04
Girl speaks a different language than the language of instruction at a school	58	0.03
Youth leadership index	255	0.21*

*significant at 90% confidence interval, ** significant at 95% confidence interval, *** significant at 99% confidence interval

In the following sections, we will explore in the same way the statistical relationships for additional variables specific for the formal schools and ABE/NFE centres, and subsequently create more advanced regression models to further the analysis and identify the main predictors of the attendance rates.

Attendance & formal schools' specific sub-groups

This section analyses the variables selected from the school survey, which were initially collected at the school level data¹³⁵. The aim is to provide complementary information on the factors influencing the attendance rates, in addition to the household and teacher survey data in the previous section. In this case,

¹³⁵ As the data was collected at the school level, to ensure the results of the analysis are robust and to keep higher number of observations, we will include attendance rates from all the grades in the formal schools.

the analysis concerns only the attendance rates of girls and boys in formal schools, as the school survey was primarily targeting the formal schools only.

The table below summarizes the main findings from the single regression models. The attendance rates decrease by 0.19 percentage points with a higher number of students per teacher, at a 99% CI. Other variables, such as the fact that children were attending either public or community schools, seem to have a substantial and negative but statistically insignificant impact on the attendance rates.

Other subgroups related to WASH knowledge and the provision of the scholarships to girls had insignificant but negative and relatively large coefficients. These results could be partially explained by the fact that the interventions related to WASH and scholarships were implemented in the schools which were already more likely to have lower attendance rates, due to the poor hygiene or limited funds to support girls education.

The most significant and positive impact on the school attendance was observed when the school had a functioning CEC and when a higher number of teachers were trained on gender-sensitive teaching methods, while other non-significant variables with high and positive coefficients were the fact of attending the private school and the use of new learning materials at the school.

The fact that the school has a functioning and established CEC committee, was associated with 0.20 percentage points increase in the attendance rates, representing a substantial change in the outcome indicator. Also, CEC members often mentioned that one of the main responsibilities of the CEC committee is to supervise the attendance of students, communicate with parents and in the same way, supervise the teachers' presence in the classes, which could have a positive impact on the attendance rates.

TABLE 52: FORMAL SCHOOL-SPECIFIC SUB-GROUPS AND THE ATTENDANCE SCORES

Sub-group	Obs.	Coef.
Public school	326	-1.00
Private school	194	1.45
Community school	22	-2.82
School provides school meal	524	0.34
WASH-related education at school	524	-3.53
WASH-related education at school: handwashing	270	1.21
WASH-related education at school: use of latrines	250	-3.06
WASH-related education at school: use of clean water	224	-3.72
WASH-related education at school: causes of diarrhoea	120	-0.96
Number of students per teacher	524	-0.19***
Number of teachers trained on gender-sensitive teaching	144	0.28*
Use of new learning materials at school (curriculum etc.)	524	1.42
School provides scholarships for girls	524	-2.12
Number of toilets available in the school per girl	430	0.17
School has a CEC	494	0.20***

*significant at 90% confidence interval, ** significant at 95% confidence interval,

*** significant at 99% confidence interval

Attendance & NFE/ABE specific sub-groups

This section will additionally focus on the analysis of subgroups selected from the headcount survey, which were asked in the ABE/NFE centres only. This section aims to better understand how the surroundings and availability of the facilities at the ABE/NFE centres affect the attendance rates.

The table below summarizing the main findings indicates that only the fact of having a handwashing station in the centre has a positive and significant effect on the girls' attendance rates. However, the overall low significance levels of other sub-groups could be partly explained by a low number of observations. Other variables, with substantially big coefficients, are (a) the fact of having reliable electricity at the centre and (b) there is soap at the handwashing station near the girls' toilets¹³⁶.

TABLE 53: ABE AND NFE SPECIFIC SUBGROUPS AND ATTENDANCE SCORES

Sub-group	Obs.	Coef.
Center has access to water	81	-0.84
Centre has reliable electricity	48	6.33
Number of toilets available for girls	96	-0.03
Number of toilets separated by walls	79	-0.38
Handwashing station in the centre	22	9.68**
Handwashing station near girls' latrines	18	-2.18
Soap at the handwashing station near girls' latrines	22	2.05

*significant at 90% confidence interval, ** significant at 95% confidence interval,
*** significant at 99% confidence interval

In the next section, we will build a multivariate regression model, including the most significant sub-groups with the most meaningful impact and compare the results of the regression by cohort group and state.

Further analysis of sub-groups

This section will further explore the statistical relationship of the selected most relevant and significant subgroups and the attendance scores, through multivariate regression analysis, including the disaggregation by the cohort groups and by geographic locations. Based on the findings in the previous section, 7 sub-groups aggregated at the school level¹³⁷ and applicable to both formal schools and the ABE/NFE centres were selected for the analysis.

¹³⁶ The results in the table are partly counter-intuitive as the fact of having 'handwashing station near girls' latrines' has a non-significant but high and negative coefficient. While we do not know a possible explanation of these results, due to the low significance levels and the number of observations, the results will not be included in the further analysis.

¹³⁷ Including only children in grades 1 and 2.

The list of the selected variables is indicated below:

1. Girl comes from a pastoral household
2. Youth leadership index
3. Household owns land
4. School has reliable electricity
5. Girl spends a few hours or more per day doing HH chores
6. Teachers miss 1-2 days per week on average
7. School has a CEC

The table below details the results of the analysis. The above-mentioned sub-groups were included in the regression model (column 1), while the regression was run separately for the cohort groups (columns 2-4) and the state (columns 4-8)¹³⁸. The results suggest that the most significant variables for the overall attendance rates are related to household/girl characteristics (pastoral household and land ownership) and school management (teacher absent 1-2 days per week, established CEC committee and electricity at the centre), which could be considered as fairly important predictors of the attendance rates.

The variable 'pastoral household' had the strongest and the most significant impact on the overall attendance rates, suggesting that the attendance rates are 0.48 percentage points more likely to increase with a higher percentage of girls coming from the pastoral households, who are also attending the school. While looking across different cohort groups, the variable is especially significant with a very high coefficient for the ABE cohort group, representing 0.47 percentage points change, and for the NFE cohort group, representing 0.59 percentage points change. The coefficients for girls in the formal schools have also a substantial and statistically significant effect on the attendance rates, representing 0.49 percentage points change. While we do not have a straightforward explanation why the higher share of pastoralist school-going children may have a positive impact on the attendance rate at the school level, it is possible that other factors, such as better school management, CEC involvement and supervision, may improve both, the overall attendance rates and the enrolment and attendance of children from the pastoralist communities. It is also possible that the majority of the (few) pastoralist households in this study are actually IDPs following the 2016-2017 drought, and therefore not actually engaged in pastoralism as a primary livelihood; 74% of those respondents are currently depending on charity, 70% do not own any camels and 59% do not own any medium-sized livestock. Also, respondents in two different FGDs with the CEC mentioned that children in pastoral households face in general bigger challenges to attending school regularly, especially due to displacement and parental attitudes.

¹³⁸ Specific cells containing “-“ indicate the variable was omitted from the analysis due to multicollinearity. In addition, empty variables suggest that the variable does not have observations recorded for the given cohort/state group. In case of the 'reliable electricity' with missing observation for the ABE/NFE centres in the household survey, an alternative substitution variable was identified in the headcount survey and additionally included in the regression as 'centre has a reliable electricity'.

“Nomadic students face tremendous challenges. They sometimes don’t show up at school for three consecutive days.” (CEC FGD)

The second most significant variable was the probability that the household of the school-going children or ABE and NFE participants owns land. The variable has a negative, statistically significant impact on the attendance rates, representing a change of 0.17 percentage points. The effect was also substantial for older girls in the ABE and NFE groups, representing a negative change of 0.22 percentage points and 0.13 percentage points, respectively. Besides, respondents during the FGD sessions also expressed an opinion that it is more difficult for ABE and NFE girls to attend the school due to fact that they may be caring for children, working or be responsible for the housework during the day, while more time-flexible classes could improve girls attendance of these courses. The third most significant variable was teachers’ absence in classes, with 0.11 percentage points negative change and significant only for boys in the formal school. The fourth most impactful variable is the fact that the school has a CEC¹³⁹, increasing the attendance rates by 0.07 percentage points, and suggesting that the monitoring activities of the attendance and teachers’ absenteeism improve the overall attendance rates in the formal schools. For the attendance rates of the girls in the formal schools, the attendance rates seemed to increase by 0.04 percentage points. The CEC seems to matter most for the attendance of the boys, increasing their attendance rates by 0.14 percentage points, at a 95% CI. It is possible that other barriers are more likely to prevent girls from attending the schools and the CEC has only a limited impact on their attendance rates.

While other variables do seem to have a strong of significant impact on the attendance rates, girls in the NFE cohort group are significantly more likely to have higher attendance scores if they more likely to achieve higher YLI scores, increasing their attendance by 0.54 percentage points. The same results were found when the NFE centre had reliable electricity, increasing their attendance by 11.14 percentage points. In the other hand, while the higher chores burden did not seem to have any impact on the attendance rates overall and for children in the formals schools and ABE centres, NFE girls were significantly less likely to attend classes if they spend several hours a day on housework. These findings may suggest that as NFE girls are older, they are more likely to assume higher responsibilities in their homes and have less time available to attend courses. This was also frequently mentioned as a barrier to attendance of the girls in the life skills courses during the qualitative interviews.

TABLE 54: MULTIVARIATE REGRESSION ANALYSIS

Variables	(1) All	(2) FE Girls	(3) FE Boys	(4) ABE Girls	(5) NFE Girls
Girl comes from a pastoral household	0.48***	0.49**	0.29	0.47***	0.59***
Youth leadership index	0.09	0.05	0.13	0.36	0.54***
Household owns land	-0.17**	-0.20*	-0.10	-0.22	-0.13

¹³⁹ The variable was variable only for the formal schools.

School has reliable electricity	0.05	0.07*	0.01		
Girl spends few hours or more per day doing HH chores	0.07	0.07	0.02	0.05	-0.20*
Teachers miss 1-2 days per week on average	-0.11*	-0.07	-0.14*		
School has a CEC	0.07	0.0*	0.14**		
Number of students per teacher		0.14	0.15		
Center has a reliable electricity				-1.44	11.14**
Constant	74.24***	66.74***	66.11***	65.23***	68.27***
Observations	121	70	51	52	47
R-squared	0.30	0.31	0.35	0.11	0.50

To conclude, the main finding of this section suggests that the attendance rates, in general, tend to be relatively high. In our sample, one-third of the classes where we conducted headcounts have perfect attendance rates. Comparison with the attendance records also showed that teachers were more likely to mark higher attendance for the day of the visit in the ABE and NFE classes, compared to the attendance of children in the formal schools. On the other hand, while considering the quality of the records, in the majority of the classes the attendance records seemed to be complete and well organized.

The main findings indicate that the attendance rates are significantly lower for the ABE groups. Also, the results of the sub-group analysis indicate that the main barriers to the girls' attendance are: land ownership; high homework and teachers absence in the classes. In addition, the findings from the qualitative data analysis suggest that menstruation and lack of financial resources are important barriers preventing girls from regularly attending classes.

On the other hand, the main predictors of the higher attendance rates were: higher share of girls coming from the pastoral household; the fact that household owned a phone; reliable electricity at school; and established and functioning CEC committee at the school. The additional analysis combining several of the most predictive subgroups revealed that a girl comes from a pastoral household and that the household owns a land are very strong predictors of the negative attendance rates.

7.2 TEACHING QUALITY AND PRACTICES

This section of the report focuses on the intermediate outcome of teaching quality. The section is divided into three parts. The first part looks at the overall professionalism and gender equity in teaching practices. The second part investigates the use of physical punishment by teachers. In the third part the focus is on the pedagogical prowess of the teachers.

Note that when classroom observations are being referred to, they were only conducted with teachers in formal schools. However, other sources of data (teacher surveys, girl surveys) cover the other types of learning institutions as well.

For an education intervention a focus on teaching quality is an obvious choice. Teaching has been deemed of poor quality in Somalia through previous interventions, and it has been linked to both learning and transition outcomes.

A girl will likely learn more when the teacher has teaching methods and practices that are high quality. High quality teaching may also be an incentive to the girl to stay in school as she feels like she is learning and that school is worth her time and money. However, the opposite is true too – a girl may be more inclined to drop out of school if the teacher is often absent or not prepared and she will not learn as much if the teacher is not engaging both boys and girls or has a poor teaching approach.

- There are variety of indicators and data sources that have been collected in order to assess teaching quality
- Questions in the girl survey on teachers’ behaviour in class and with the students
- Teacher survey to look at the teachers’ attitudes
- Classroom observation to look at their teaching practices. This however, was only conducted in formal schools

Teacher professionalism

It is a reasonable assumption that feeling welcome in school and having a teacher present in class are positive factors for both learning as well as staying at school. The first category of teaching quality thus looks at responses in the school survey to two questions:

- 1) Do you agree or disagree with the following statement: “My teachers make me feel welcome in the classroom”?
- 2) Do you agree or disagree with the following statement: “My teachers are often absent for class”?

The findings presented below are the percentage share of those who agree somewhat or strongly with those statements.

When cohort girls were asked if their teachers made them feel welcome in class, the responses were overwhelmingly positive with 97% on average agreeing with the statement.¹⁴⁰ Importantly, there was hardly any variation over program track or region. This implies that the teachers were overall effective at creating a welcoming environment for learning.

TABLE 55: TEACHER PROFESSIONALISM BY EDUCATIONAL TRACK

	By education track			
	FE	NFE	ABE	Total
Agree: My teachers make me feel welcome in the classroom	97%	97%	96%	97%
Agree: My teachers are often absent for class	34%	31%	31%	32%

¹⁴⁰ This datum aggregates the answers ‘agree’ and ‘strongly agree’. While there is little variance in this compounded measure, there is some variance in the share of those who strongly agree with the measure.

However, when looking at teacher absenteeism as reported by cohort girls, the results vary somewhat more. An average of 32% agreed or agreed strongly with the statement, indicating that teacher absenteeism may be an issue in some schools. There was little variation in responses to the question when comparing different types of learning institutions.

Classroom gender equity

Pertinent to the intervention's focus on girls, we look at gender equity in the classroom. Much of the approach here revolves around the question of whether teachers treat boys and girls in the classroom in the same manner.¹⁴¹ The approach poses one significant problem. Namely, without a full contextual understanding of the specific classroom, group of students and the teacher themselves, the approach cannot detect cases where the teacher is in fact treating girl and male students differently in order to break down, rather than create, barriers to girls' participation.

With this caveat in mind, we feel that these questions are suitable proxies for gender-sensitivity in teaching. Even if teachers treat students differently in an effort to help girls perform better in school, it is not desirable for female (or male) students to feel that they are being treated differently.

Data for this indicator was collected through the girl's survey, the classroom observation as well as the survey with the teachers themselves. In order to analyse gender equity, we looked at the following questions:

Girl's survey:

- To what extent do you agree with the following statement: "My teachers treat boys and girls differently in the classroom.?"

Classroom observation:

- Observed: Girls and boys have equal access to learning materials
- Not observed: Teacher directed questions to one gender, not both
- Observed: Teacher asked boys and girls questions of equal difficulty
- Observed: Teacher gave the same amount of time to boys and girls to answer questions

Teacher survey:

- Who is able to solve problems?
- Who is good at school?
- Who acts as a leader in class?
- Who should be prepared for a professional career?
- For whom is it more culturally acceptable to attend school regularly?

The table below presents the results. However, it should be noted that all the answers presented have been converted to ensure that they run in the same direction. For example, as it pertains to the girl's survey question "My teachers treat boys and girls differently in the classroom", what is presented is the percentage of those who disagreed with the statement (either a lot or a little). This way, the higher the percentage, the greater the equality in treatment of boys and girls. Similarly, the teacher survey findings

¹⁴¹ It should be noted, however, that with this approach it is not possible to detect whether some of the differences in treatment are in fact 'positive discrimination'.

present data coming from a set of questions where the answer options for the teachers interviewed were: 1) boys and girls equally; 2) girls more than boys; 3) boys more than girls; 4) neither boys nor girls. The results presented here show where the teachers answered ‘boys and girls equally’ to each question.

It is important to note two philosophical and methodological shortcomings of this approach. First, the questions in this section are measures of gender equality, rather than equity. The questions assess whether girls and boys are treated equally; equitable treatment, on the other hand, might imply that girls should be prioritized over boys by their teachers, in order to counter broader inequality. Second, even where we can identify teachers who favour one gender over the other in the classroom – which is possible in the context of the teacher survey questions – we have, nonetheless, opted to focus on equal treatment. In practice, many teachers who report unequal treatment actually report favouring female students. For instance, when asked “who is good at school”, a plurality of teachers said that boys and girls are equally good at school; crucially, *among those who chose one gender over the other*, 68.8 percent of teachers believed girls were better at school than boys. In the coding employed here, such an attitude represents gender inequality.

It should also be noted that classroom observations were only conducted in the formal schools. As such, where disaggregation by program track (FE, NFE, ABE) are presented, this data source is naturally excluded (furthermore, only FE classes include girls as well as boys). In addition, as NFE and ABE programs do not have boys, the data presented for the girl’s survey includes responses only from formal school girls as they are the only part of the cohort where gender differentiation in class is relevant.

Further, facilitators and teachers involved in ABE and NFE programs often teach more than one level of education. Teacher survey asked them to indicate every grade or level that they teach in a multiple choice question. Consequently, the FE, NFE and ABE categories for teacher survey responses are not necessarily mutually exclusive, i.e. the same respondent can belong to more than one category. However, when looking at responses for any category alone, all respondents do teach that level.

Looking at the girl’s perceptions of gender equity, the responses are perhaps quite encouraging in that although overall slightly less than 60% of the girls feel that teachers treat boys and girls in the same manner in the classroom, there is very little variation between different regions. Yet, of course there seems to be room for improving gender equity in the classroom as nearly 40% of the respondents did agree that girls and boys were treated differently.

When it comes to learning materials, very little difference across regions can be detected. The responses are overwhelmingly positive, with 96% of classroom observations reporting girls and boys having equal access to learning materials.

A teacher outlines the logic of differential gender treatment in class:

When girls and boys are in the same class, we often ask questions to boys. Because girls are very shy in the class, we don’t ask more questions to them. And this is deeply rooted in our Somali culture. Girls don’t speak up when they are with boys. Even the ones who try to answer a question answer it very slowly, whereby we as teachers cannot even hear it. – Teacher

Meanwhile, many teachers in qualitative discussions talked about balancing the genders in class:

I do the so-called gender balance. When I ask a boy a question, the next question is given to a girl. For example, when I ask for an introduction to yesterday’s lesson I like to start with girls, and when the girl answers the question, the next question is given to a boy. – Teacher

TABLE 56: CLASSROOM GENDER EQUALITY

Gender equity by State – mean of responses	Overall mean	n	Data source
Disagree little/lot: My teacher treats boys/girls differently in classroom	60%	421	Girl Survey
Sub-score	60%		
Observed: Girls and boys have equal access to learning materials	96%	70	Classroom observation
Not observed: Teacher directed questions to one gender, not both	51%		
Observed: Teacher asked boys and girls questions of equal difficulty	64%		
Observed: Teacher gave the same amount of time to boys and girls to answer questions	83%		
Sub-score	74%		
Boys & Girls equally: Who is able to solve problems?	39%	310	Teacher survey
Boys & Girls equally: Who is good at school?	35%		
Boys & Girls equally: Who acts as a leader in class?	32%		
Boys & Girls equally: Who should be prepared for a professional career?	38%		
Boys & Girls equally: For whom is it more culturally acceptable to attend school regularly?	30%		
Sub-score	35%		
Average of all data sources	56%		

When disaggregating by program track, there is little difference between the perceptions of the cohort girls themselves. However, their teachers or facilitators exhibit substantial differences in their answers to the questions looking at gender equity. When looking at the aggregate score between teachers or moderators across the three tracks, we can see that the formal education teachers overall exhibit much more positive attitudes on gender equity, at 44%. While this is still relatively low, it is much higher than for those teaching ABE, at 26%, and NFE especially, at 17%.

TABLE 57: GENDER EQUITY BY PROGRAM TRACK

Gender equity by program track – mean of responses	FE	NFE	ABE	Overall mean	N	Data source
Disagree little/lot: My teacher treats boys/girls differently in classroom	60%					Girl Survey
Sub-score	60%					
Boys & Girls equally: Who is able to solve problems?	51%	17%	21%	39%	310	Teacher survey
Boys & Girls equally: Who is good at school?	44%	13%	25%	35%		
Boys & Girls equally: Who acts as a leader in class?	43%	13%	18%	32%		
Boys & Girls equally: Who should be prepared for a professional career?	45%	25%	36%	38%		
Boys & Girls equally: For whom is it more culturally acceptable to attend school regularly?	37%	15%	28%	30%		
Sub-score	44%	17%	26%	35%		
Average of all data sources	52%	37%	44%	47%		

Given that the girls in these program tracks have been selected in due to their age, this can be a reflection of the pragmatism relating to age-related concerns. This could perhaps explain the more negative responses to the question on who should be prepared for a professional career, as the teacher may have as their specific reference point, the girls in their program who can be 18 years of age with no formal education. However, as it pertains to the question on for whom it is more culturally acceptable to attend school the same explanation does not hold as much water. As such, it does seem that the teachers involved in NFE and ABE programs are exhibiting opinions that do not reflect gender equity. This may become a problem for retaining some of the students as undoubtedly these kinds of attitudes can partially explain why these girls have not enrolled or stayed in education in the past.

Disciplinary practices

To look at disciplinary practices, this section focuses on two main questions. Firstly, girls were asked if their teachers punish students for wrong answers. It is assumed that punishing students for wrong answers might have a negative impact on learning as it prevents participation as students are afraid to take part in case their answer is wrong. Secondly, the use of corporal punishment was looked at through the girl’s survey as well as classroom observation. This is an important focus as the use of corporal punishment may deter students from coming or staying at school. A female community member recounts in an FGD:

“There was a girl who was my neighbour. She was clever but dropped out because a teacher rebuked her. I looked for her and went to their house. I met her father and asked about the whereabouts of the clever girl. He then informed me that she told him that a teacher beat her, so she does not want to go to school anymore.”

The table below presents the results. Again, as above, the answers are converted so that they share a directionality (i.e. higher numbers indicate positive attitudes). In total, less than a third of the girls said their teachers did not punish students for wrong answers.

TABLE 58: DISCIPLINARY PRACTICES

Disciplinary practices by region	Total		Data source
	Total	n	
No: Does your teacher/facilitator discipline or punish students who get things wrong?	32%	1420	Girl survey
No: Teacher/facilitator uses physical punishment to punish students.	73%	1416	
Sub-score	53%		
Not observed: Boys are disciplined physically in class	61%	51 ¹⁴²	Classroom observation
Not observed: Girls are disciplined physically in class	61%	70	
Sub-score	61%		
Average of all data sources	57%		

Some potential explanations exist for the incongruence between the negative findings in the girls’ survey data and the classroom observations. It is possible that the teachers changed their practices during the observation, knowing that the program seeks to eradicate the use of corporal punishment in the intervention schools. Moreover, the number of classroom observations is rather small. It is recommendable that at midline observations are conducted also with ABE and NFE moderators to systematically compare their practices to those in formal schools. The qualitative data did not reveal any fundamental differences between the three regions when it comes to physical punishment. It is thus unlikely that some locations would have eradicated corporal punishment in their schools, especially when a teacher discusses in this manner:

Yes, it exists [difference in how you treat girls vs. boys] in terms of discipline. The way teachers treat boys, they don't treat girls like that. Because you can beat boys in many places of his body, but the girls you cannot beat in many places of her body [...].

As per differences in practices of punishment, a clear trend exists. Formal education scores the lowest for the absence of physical punishment (39%), followed by ABE (49%) while NFE scores highest in this regard, at 65%. We interpret this to relate mostly to the age of the students; they younger the student

¹⁴² Please note that only 51 of the classrooms observed had boys present. Some schools are girls only.

the more acceptable it is viewed to use physical punishment. Additionally, ABE and NFE facilitators are explicitly monitored for the use of corporal punishment and are aware that cases will lead to dismissal.

Since they are young and won't listen, punishment is important. But it shouldn't be too much.
– Teacher

TABLE 59: DISCIPLINARY PRACTICES BY PROGRAM TRACK

Disciplinary practices by program track	Program			Total		Data source
	FE	NFE	ABE	Total	Freq.	
No: Does your teacher/facilitator discipline or punish students who get things wrong?	22%	41%	29%	32%	1420	Girl survey
No: Teacher/facilitator uses physical punishment to punish students.	56%	88%	69%	73%	1416	
Average of all data sources	39%	65%	49%	53%		

Qualitative data shows that, despite some variance, attitudes favouring the use of the physical punishment are still prevalent.

If the teacher is intimidating, the child will learn something. But if he is not, the child will not learn since he is not afraid of the teacher. – Mother

Meanwhile, an FGD with teachers saw a debate over the use of punishment whereby most teachers were of the opinion that students should be beaten to get some ‘character’ out of them. One teacher argued, conversely:

Students should be told great stories about education and given good motivation. Punishing students only cause to them to be a drop out. – Teacher

AGES will have some way to go in way of convincing the rest of the teachers to share this point of view. Across the entire cohort, only 32% of the girls surveyed said their teachers did not punish students who get things wrong.

Pedagogical practices

In terms of good teaching practices, we look at the use of good pedagogical tools and methods. The section first looks at the use of formative assessments, both reported by the teacher during the classroom observation and verified by the researcher by asking the teacher to show the formative assessments.

We then look at a number of teaching practices as observed by the cohort girls in the girl’s survey. The following questions are under focus:

TABLE 60: LIST OF QUESTION USED FOR THE ANALYSIS OF PEDAGOGICAL PRACTICES

Question	Answers reported
Does the teacher/facilitator explain how the things you are learning will be useful to you in your life?	Reporting those who answered 'often'
Does the teacher/facilitator give you ideas for how you can learn outside the learning center/class as well as inside it?	
Do you find the lessons go at a good speed for you?	Reporting those who answered 'just right'
If you don't understand something, do your teachers/educators use a different language to help you understand?	Reporting those who answered 'often'
Does your teacher/facilitator encourage students to participate during lessons, for example by answering questions?	
Does your teacher/facilitator suggest ways you can continue to study after school/at home?	

In the classroom observation, the researcher was asked to observe if the teacher used any of the following pedagogical practices:

- Teacher used student-centred activities or games
- Teacher allowed students to instruct each other
- Teacher asked open-ended questions that encourage thinking
- Teacher used questions that ask for student opinions
- Teacher tried to involve a student that was not participating
- Students worked together in groups

Use of formative assessments

For the use of formative assessments, the only data source is the classroom observation. The starting point is quite low. When averaging reported and recorded use of the assessments, approximately only a third of the teachers observed in formal schools make use of these tools to support their teaching.

TABLE 61: USE OF FORMATIVE ASSESSMENTS

Use of formative assessments	Total		Data source
	Mean	Frequency	
Teacher saying they use formative assessments	43%	70	Classroom observation
Records of formative assessments are shown	21%		

Use of participatory and student-centred practices

When looking at the specific participatory and student-centred practices employed by teachers by region, we have both the classroom observation and the girl survey data to refer to.

Overall, while students indicate teachers are good at suggesting ways to continue learning outside of school – overall 94% indicated as

such – many seemingly struggled with the pace of the lessons as only 40% of all cohort girls indicated that the pace of the lessons was ‘just right’. This is, perhaps, a natural result of oversized classes. However, it is concerning that across all interviewed, 45% found the speed of the lesson to be too fast and only 15% thought it was too slow. This implies that many teachers and facilitators might be opting to follow the pace of the most advanced students in the class rather than taking on the challenge to trying to cater the lessons for those who require more support.

TABLE 62: PEDAGOGICAL PRACTICES

Pedagogical practices - mean of responses	Mean	Freq	
Often: Does the teacher/facilitator explain how the things you are learning will be useful to you in your life?	76%	1420	Girl survey
Often: Does the teacher/facilitator give you ideas for how you can learn outside the learning centre/class as well as inside it?	57%		
Just right: Do you find the lessons go at a good speed for you?	40%		
Often: If you don't understand something, do your teachers/educators use a different language to help you understand?	52%		
Often: Does your teacher/facilitator encourage students to participate during lessons, for example by answering questions?	62%		
Often: Does your teacher/facilitator suggest ways you can continue to study after school/at home?	94%		
Sub-component score - Pedagogical Practices	64%		
Observed: Teacher used student-centred activities or games	44%	70	Classroom observation
Observed: Teacher allowed students to instruct each other	84%		
Observed: Teacher asked open-ended questions that encourage thinking	70%		
Observed: Teacher used questions that ask for student opinions	67%		

Observed: Teacher tried to involve a student that was not participating	81%		
Observed: Students worked together in groups	43%		
Sub-component score - Pedagogical Practices	65%		

Approximately half of the students across the board indicated that their teachers would use a different language to help a student who is struggling to understand. Disaggregation by type of learning centre reveals very little variation of the girl's survey findings. As such, there do not seem to be any significant differences in teaching quality based on the type of education institute. As discussed above, lesson speed seems to be perceived somewhat problematic by many of the respondents as is the teacher giving ideas for learning outside of class, but these challenges seem to differ only marginally between different education tracks.

TABLE 63: PEDAGOGICAL PRACTICES BY PROGRAM TRACK

Pedagogical practices by program track	Program			Total		Data source		
	FE	NFE	ABE	Mean	Freq.			
Often: Does the teacher/facilitator explain how the things you are learning will be useful to you in your life?	72%	78%	77%	76%	1420	Girl survey		
Often: Does the teacher/facilitator give you ideas for how you can learn outside the learning centre/class as well as inside it?	52%	61%	57%	57%				
Just right: Do you find the lessons go at a good speed for you?	40%	42%	37%	40%				
Often: If you don't understand something, do your teachers/educators use a different language to help you understand?	53%	55%	50%	52%				
Often: Does your teacher/facilitator encourage students to participate during lessons, for example by answering questions?	58%	68%	59%	62%				
Often: Does your teacher/facilitator suggest ways you can continue to study after school/at home?	94%	97%	93%	94%				
Sub-component score - Pedagogical Practices	62%	67%	62%	64%				

Teachers themselves called for more training, and a clear curriculum to guide them in their work. When teachers were asked what subjects they most struggled teaching, they talked about science and other subjects where practical instruction would be necessary but is difficult because of a lack of equipment. One teacher had a particularly practical suggestion in this regard:

Teachers of this school or region need equipment of education, especially equipment of science, because of students need to be shown in a practical way and teachers can't afford to buy such equipment. Therefore, it is good to support that side in order to provide lessons in practical way. Also, teachers have to have a back-up place if they meet any challenges, for example a library where they can find [responses for] what they don't understand and has internet and all of the equipment are needed. – Teacher

In this regard, the program could explore teaching science teachers in how to make their subject practical in the absence of additional equipment. For example some activities could focus on how to give practical examples of science problems that are relevant to the students. Moreover, teachers should be informed about existing learning resources that they can use for continued professional learning without new or additional resources.

Overall, the Baseline assessment revealed that most girls struggled with the pace of the lessons as only 40 percent found the pace of the lessons to be just right for them. The teacher rushing through a material could be a sign of their lack of confidence in their mastery of the subject but exploring trainings in this regard could be useful. In terms of other teaching practices that are less prevalent, the observations with teachers in formal schools rarely featured games or student-centred activities or group work. Teacher trainings on these practices would be relevant.

Discussion of findings

Teaching quality by education track

As it pertains to differences in program track measures for teaching quality have slightly different data sources. For formal schools there is most data as some of the teachers were observed in class.

As it comes to professionalism: little difference nearly all girls, regardless of type of education intervention they are part of, indicated that their teacher made them feel welcome. Approximately 30 percent of the girls across different types of education programs indicated that their teacher was often absent for class.

When looking at gender equity, while ABE and NFE programs only include girls, it is still useful to look at the gender attitudes of the moderators and facilitators as these might reflect on girls' attendance, learning and retention. Teachers' attitudes towards girls are seemingly most negative with NFE than ABE while teachers in the latter exhibit more negative attitudes than those in formal schools. For example, when asked who is good at school, 44 percent of formal school teachers indicated that both boys and girls equally were good. In the ABE centres 25 percent did, while in NFE the share was only 13 percent. Given the relationship between gender-sensitive teaching practices and learning, transition and retention, the program should perhaps focus on the teachers of NFE and ABE programs to address some of these attitudes as they might reflect negatively on their teaching practices and the overall learning environment.

Yet, in terms of actual teaching practices, only formal schools had classroom observations, so when contrasting findings by education track, we rely on cohort girl's accounts of their teachers' practices. No systematic differences by program track can be observed at this Baseline.

As per discipline, the cohort girl surveys indicated that physical punishment was most common in FE and least common in NFE. This is perhaps negatively correlated with age as physical punishment of the older girls in NFE courses is likely deemed less socially acceptable. Midline and endline evaluations should expand the observations to cover ABE and NFE learning institutions to maximize comparability of data.

Qualitative interviews revealed problematic attitudes especially as it pertains to differential treatment of girls and boys and corporal punishment though the views are not universally held. Meanwhile, teachers used to opportunity to call for more training, a clearer curriculum and added resources for practical guidance of subjects.

7.3 YOUTH LEADERSHIP INDEX

This section aims to further estimate which factors determine the level of the girls’ self-perceptions of leadership skills, by using the Youth Leadership Index (YLI), as the main measurement tool. As a part of the LNGB project’s theory of change, girls’ leadership skills, measured by a percentage of girls scoring 70% in YLI, are expected to contribute to the improvement of girls’ learning and transition outcomes and to ensure that the outcome change will be sustainable.

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 GEC initiative. The GEC thematic review of the self-esteem stipulated that the interventions focused on building the girls’ self-esteem lead to positive changes in the girls’ attitudes, more concretely “motivation to attend, new aspirations or a sense of school belonging”¹⁴³, which in turn have a positive impact on attendance, class participation and lead to overall improvement of the learning outcomes.

In the first section, we will explore the statistical properties of the distribution of the YLI index, disaggregated by the cohort group and state. In the second section, we will further assess the nature of the statistical relationship for selected sub-groups and the YLI scores.

The Youth Leadership Index (YLI) 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 is measuring 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)¹⁴⁴. The table below contains the list of the questions used for the construction of the YLI.

TABLE 64: LIST OF YLI QUESTIONS.

Questions
I like to try new activities that I may not know how to do.
My friends ask me for advice.

¹⁴³ 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

¹⁴⁴ CARE Education, *CARE’s Youth Leadership Index Toolkit* (2014), link: <https://www.care.org/cares-youth-leadership-index>

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.

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. The lower values indicate more negative outcomes and the higher values indicate 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%.

In the following section, we will examine more in detail the YLI distribution and the score's breakdown for different cohort groups, geographic locations, and selected sub-groups/characteristics. The aim is to assess and establish the baseline values for the YLI scores, disaggregated by cohort group and geographic location and determine the most relevant sub-groups.

YLI analysis by cohort group and geographic location

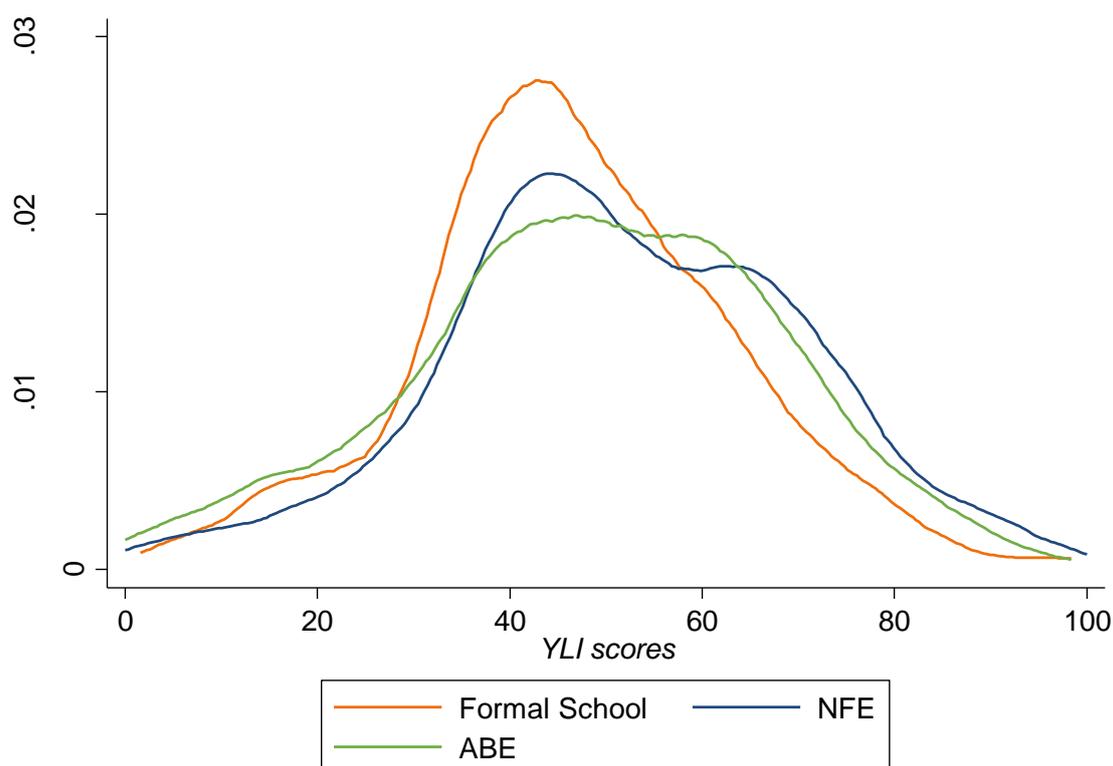
This section will further examine the properties and implications of the overall distribution of YLI scores, and disaggregated by the three main cohort groups (girls participating in formal education (FE), girls participating in the non-formal educational (NFE) and girls participating in the alternative basic education (ABE) programs).

Overall, **12% of girls (171) scored with YLI of 70% or higher**. The measured value seems to be significantly low, in comparison with the **expected endline target 80% of girls reaching a score of 70% in the YLI**. Around 54% of girls scored less than 50%, with the cohort groups equally distributed among the girls. Within the FE cohort group, only 8% of the girls scored 70% or higher in the YLI, while within the ABE cohort group 12% of the girls and within the NFE cohort group 16% of the girls scored 70% or higher in the YLI.

As indicated in the graph below, the distribution of the YLI scores approached normal distribution¹⁴⁵. The values of the YLI of the NFE girls are clustered around 45% and 65%. The YLI scores distribution for the FE girls peaks around scores of 45% and for the ABE girls the distribution peaks around 50%.

Consequently, the distribution of the scores indicates that the girls in the FE cohort group are more likely to score lower, compared to the NFE and ABE groups. The lower self-confidence levels for the FE cohort group could be a function of age, as younger girls can be less likely to be self-aware. In the course of the analysis, we will further explore the relationship between the age and the YLI, to establish a list of main predictors of the girls' self-perceptions of leadership skills.

FIGURE 19: YLI DISTRIBUTION – KERNEL DENSITY PLOT, BY COHORT GROUP



¹⁴⁵ The kurtosis of the overall YLI scores approaches the value of 3 (2.96), indicating the nearly normal distribution of the scores, while the skewness of -0.06 indicates the score's distribution is slightly skewed to the left. Similar results can be observed for the YLE distribution for each of the cohort groups.

As indicated in the table below, the overall YLI mean score is 49%. Also, the results disaggregated by the cohort groups indicate that the NFE girls reached the highest mean scores (52%), while the FE girls reached the lowest YLI mean scores (47%). Results of simple regression analysis of the YLI scores for each of the cohort groups suggest that the NFE girls are significantly more likely to score higher, compared to other cohort groups. The mean difference is large and significant at a 90% CI, even after accounting for clustering at the school level. On the other hand, the FE girls are significantly more likely to score lower, compared to other cohort groups. Similar to the NFE cohort group, the mean difference for the FE girls is equally large and significant at a 90% CI. After controlling for the ABE cohort group, the difference in the mean scores for both, NFE and ABE cohort groups, is even bigger and significant at a 95% confidence interval¹⁴⁶.

TABLE 65: YLI DISTRIBUTION, BY COHORT GROUP

Cohort group	Obs.	Mean (in %)	Coef.
Overall YLI scores	1420	49.29	-
NFE Girls	515	51.79	3.92*
ABE Girls	484	48.83	-0.69
FE Girls	421	46.76	-3.60*

*** significant at 99% confidence interval, ** significant at 95% confidence interval, * significant at 90% confidence interval

Older girls and girls in higher grades may be more likely to be more confident and to have developed higher voice and organisation skills. To further examine the impact of the FE and NFE cohort groups on the YLI scores, we run additional regression analysis of the YLI scores for the two cohort groups, controlling for age of the individual girls and grade, for girls who were in formal schools only. The regression results suggest that both FE and NFE groups do not have a statistically significant and meaningful impact on the YLI scores. On the other hand, the mean difference for the continuous age variable was small but significant at a 95% CI. While the results suggest that the difference in the YLI means scores is driven by age rather than by the cohort group, the cohort groups and age are closely correlated.

To conclude, the initial analysis of the YLI scores found that the composite indicator is well constructed as the distribution of the scores approached the normal distribution. The analysis results of the YLI scores disaggregated by the cohort groups could be partially driven by the difference in age.

In the next section, we will further explore additional factors that could influence the YLI scores, such as characteristics of girls and their households, learning environment, teachers' attitudes, parental attitudes,

¹⁴⁶ The coefficients of the simple regression of YLI for FE and NFE groups, controlling for the ABE group, age and grade are: 5.03 for the NFE girls and -5.03 for the FE girls. The standard deviation of the YLI scores is 18.23, suggesting that the difference in the means scores for FE and NFE cohort groups is bigger than a quarter of a standard deviation, taking the value of 0.27 standard deviations.

and disability. We will also disaggregate the analysis by cohort groups and state, to complete the results of the present section.

Sub-groups analysis

In this section, we will first assess the YLI mean difference for each of the variables selected for the sub-group analysis to determine the most relevant YLI predictors. The sub-groups were further divided into categories to facilitate the analysis: general characteristics, household characteristics, parental attitudes, learning environment, and disability. Subsequently, we will run multivariate regressions for the identified predictors, controlling for additional variables, such as age, cohort group grade, and state.

The table below suggests that among the girls' characteristics, only the fact of being a single orphan and not living with parents seems to have a significant and large impact on the YLI results, suggesting that these girls are more likely to have higher YLI scores.¹⁴⁷ Other characteristics had an insignificant effect on the YLI scores, which could be mostly explained by a low number of observations.

Variables such as age, higher household chore burden and the fact that the head of household is female might affect the scores, especially as older girls are more likely to live without their parents and orphaned girls and the girls not living with their parents will be more likely to assume more responsibilities at home, potentially increasing the importance of the role in the household. Also, for the girls who live in the household with female HoH, the fact that the main decision-maker in the household is female could have a positive impact on their self-confidence and voice, as girls could perceive the female HoH as a role model. After controlling for these variables, the difference in the mean scores for both sub-groups remains positive and statistically significant at a 90% CI¹⁴⁸, with a big effect on the scores when the girls do not live with her parents.

TABLE 66: SUB-GROUP ANALYSIS OF THE GIRLS' CHARACTERISTICS

Sub-group	Mean	Obs.	Significance
Girl is a single orphan	53.10	184	***
Girl does not live with her parents	54.94	83	***
Girl is currently married	50.22	53	-
Girl is married or was married	50.89	99	-
Girl is a mother	49.27	73	-
Girl speaks Af-Maay at home	50.45	398	-
Girl speaks a different language at home than LOI	56.85	64	-

*** significant at 99% confidence interval, ** significant at 95% confidence interval, * significant at 90% confidence interval

¹⁴⁷ The coefficient of the single orphaned girls is 4.37 points and the coefficient for the not living with their parents is 6.00 points.

¹⁴⁸ The coefficient for the single orphaned girls is 3.67 points and for the girls not living with their parents the coefficient is 4.98 points.

The results in the table below for the household characteristics suggest education, pastoralist households and limited access to food, medicine, and income have a statistically significant impact on the YLI scores. Due to the low number of observations, the fact that the girl comes from a pastoral household will be excluded from further analysis. Also, when an HoH belonged to a pastoralist clan or agricultural clan, the YLI scores were significantly lower for the former and higher for the latter. After controlling for age, cohort group and state, the fact of belonging to an agricultural clan had a positive and statistically significant effect on the YLI scores, but the change represented only 0.21 standard deviations.

The low education had a positive and substantial impact on the YLI mean scores¹⁴⁹. While considering the economic status of the household, the limited access to food, medication, and cash income have all strong and positive impacts on the YLI scores¹⁵⁰, significant at a 99% CI. After controlling for age and cohort groups, the results remain equally significant for each of these variables. While we do not have straightforward explanations for such results, the subgroups' impact on the YLI scores will be further examined in the multivariate regression analysis.

TABLE 67: SUB-GROUP ANALYSIS OF THE HOUSEHOLD'S CHARACTERISTICS

Household characteristics	YLI mean scores	Obs.	Significance
HoH has no education at all	53.65	302	***
HoH has no formal education	50.43	1162	***
Caregiver has no education at all	54.18	383	***
Caregiver has no formal education	50.91	1209	***
Female HoH	49.24	530	-
Pastoralist household	26.54	25	***
HoH has no wage-earning occupation	50.02	558	-
Gone to sleep hungry most days, past 12 months	56.99	99	***
Gone to sleep hungry many (more than 10) days or more, past 12 months	52.60	385	***
Gone without medicines most days, past 12 months	57.20	238	***
Gone without cash income most days, past 12 months	57.97	210	***
Household owns land	47.80	462	-
Household has savings	45.68	27	-
HoH belongs to a pastoral clan	47.73	671	**
HoH belongs to an agricultural clan	51.86	598	***

¹⁴⁹ The coefficient for the 'HoH has no education at all' is 5.53, for 'HH has no formal education' is 6.30, for 'caregiver has no education at all is 6.70' and for 'caregiver has no formal education is 10.93'.

¹⁵⁰ The coefficient for 'no food' variable is 8.27, for 'no medication' variable 9.50, for 'no cash income' variable is 10.18.

HoH belongs to a marginalized group or occupational minority	46.48	143	-
---	-------	-----	---

*** significant at 99% confidence interval, ** significant at 95% confidence interval, * significant at 90% confidence interval

The table below illustrates the effect of caregivers' attitudes on the girls' YLI scores. Only two variables seem to be significantly related to the YLI scores, which are the high chores burden reported by the girls, with a relatively strong and positive impact, and caregivers' beliefs that it is acceptable for the child not to attend the school due to the financial burden, also with strong positive impact¹⁵¹.

Even after controlling for cohort groups, age, and state, the results remain mostly unchanged¹⁵². While we do not have a clear explanation for these results, the positive impact of the higher chores burden seems to be in line with the above-mentioned assumption that the orphaned girls and the girls not living with their parents are more likely to take on more responsibilities in the household and to acquire new skills, with positive impact on the self-confidence levels, as measured by the YLI index.

TABLE 68: SUB-GROUP ANALYSIS OF PARENTAL ATTITUDES

Parental attitudes	YLI mean scores	Obs.	Significance
High chore burden	50.65	919	**
Caregiver wants a girl to attend university	49.67	1241	-
Caregiver does not believe girls' education is worth investment	47.00	174	-
Believes it is acceptable for a child to not attend school if they have to work or do housework	48.02	569	-
Believes it is acceptable for a child to not attend school if a school is too costly	52.03	884	***

*** significant at 99% confidence interval, ** significant at 95% confidence interval, * significant at 90% confidence interval

Several variables related to the learning environment seem to be significant for the YLI scores, as indicated in the table below. The limited access to the learning materials reported by girls seems to have a large and negative impact on the YLI scores, significant at a 99% CI. Among the variables describing teachers; attitudes in the class, the fact that the teacher is often absent, does not encourage participation and does not treat girls and boys equally in the class have a negative impact on the YLI scores and is significant at a 99% CI. Moreover, the lack of female teachers and the use of physical punishment seem to have also a strong and significant impact on the YLI scores.

During the FGDs with teachers, respondents often mentioned that the fact of having more female teachers in the schools could encourage girls to feel more comfortable and to participate more during the class.

¹⁵¹ The coefficient for 'high core burden' variable is 3.85 and the coefficient for 'acceptable for child to not attend school if school is too costly' is 7.26 points, implying the difference in the means scores above 0.25 of YLI standard deviation.

¹⁵² With the coefficient for the 'high core burden' reduced by 0.57 points and significant at 90% CI and the coefficient for the caregivers believed concerning the cost of education reduced by 0.7 points, significant at 99% CI.

However, due to the low number of observations and missing observations for ABE and NFE groups, the subgroup ‘no female teachers in school’ will be excluded from further analysis.

“The important encouragement for girls is to increase the number of female teachers.” (Teachers FGD)

After controlling for age, cohort group and state, the mean difference for all these variables remain negative and higher than a quarter of standard deviation¹⁵³. Also, all the variables are significant at a 99% CI. Overall, less adequate learning environment seems to have a negative impact on the girls' YLI scores, which is an important finding as formal and informal education institutions play an important role in the development of the girls' non-cognitive skills.

TABLE 69: SUB-GROUP ANALYSIS OF THE LEARNING ENVIRONMENT

Learning environment	YLI mean scores	Obs.	Sig.
No computers available for use at school	48.95	1359	***
Girl cannot use books/learning materials at school	43.32	372	***
My teacher makes me feel welcome	47.02	176	-
Teacher is often absent	44.98	406	***
Teacher does not encourage participation	39.13	144	***
Teachers treat girls/boys differently in the classroom	46.39	530	***
Physical punishment witnessed last week	44.31	299	***
No female teachers in school	38.62	64	***

*** significant at 99% confidence interval, ** significant at 95% confidence interval, * significant at 90% confidence interval

As indicated in the table below, only a few disability variables¹⁵⁴ were significantly related to the YLI scores, which can be partly explained by the low numbers of observations. Overall, the fact of living with any type of disability seems to have a negative impact on the YLI scores and is significant at a 95% CI. The results remain nearly unchanged when controlling for individual age, the type of cohort groups and state¹⁵⁵. Also, mental health disability and cognitive disability were the only significant disability-specific

¹⁵³ The coefficients resulting from the regression analysis are: ‘no computers available for use at school’ - -8.46; ‘girl cannot use books/learning materials at school’ - -8.13; ‘teacher is often absent’ - -5.94; ‘teacher does not encourage participation’ - -11.40; ‘teachers treat girls/boys differently in the classroom’ - -4.67; ‘physical punishment witnessed last week’ - -5.34;

¹⁵⁴ Other types of disability were not included in the analysis due to low number of observations.

¹⁵⁵ The coefficient for any type of disability is -4.00 points, representing 0.34 of YLI standard deviation.

variables. After controlling for the age, cohort group and state, only mental health disability remains statistically significant, with the coefficient of -6.32 points, which is a considerably large impact.

However, due to the liberal coding of the variable, as very few girls reported they experienced ‘a lot’ or even higher level of difficulties related to their disability, the disability variables will be excluded from the further analysis. But the statistically significant and negative impact on the YLI scores suggests that girls with disabilities as a group are more likely to have lower self-confidence, vision and voice.

TABLE 70: SUB-GROUP ANALYSIS OF THE DISABILITY¹⁵⁶

Disability	YLI mean scores	Obs.	Significance
Cognitive disability	45.5128	52	*
Mental health (anxiety/depression)	43.7096	162	***
Any type of disability	46.2025	251	**

*** significant at 99% confidence interval, ** significant at 95% confidence interval, * significant at 90% confidence interval

During the rest of this section, we will further analyze the effect of the selected sub-groups on the YLI by including the most impactful groups in the multivariate regression model. The analysis will be also informed by the main findings from the qualitative data analysis. Based on the results of previous subgroup analysis, we have identified the most relevant variables. The list below indicates which subgroups were identified:

1. Girl is a single orphan
2. Girl does not live with her parents
3. HoH has no formal education
4. The caregiver has no formal education
5. Gone to sleep hungry most days
6. Gone without medicines most days
7. Gone without cash income most days
8. High chores burden
9. Believes it is acceptable for a child to not attend school if a school is too costly
10. Girl cannot use books/learning materials at school
11. A teacher does not encourage participation
12. Teachers treat girls/boys differently in the classroom
13. Physical punishment witnessed last week

¹⁵⁶ Due to the small sample size for the girls with disabilities, alternative and more liberal coding was used for the purposes of the analysis. The dummy variables for each type of the disability were based on the disability levels reported by girls, also including cases when girls responded to have experienced some level of disability, as opposed to stricter coding, then only cases when girls reported to have experienced ‘a lot of difficulty’ or ‘cannot do at all’.

TABLE 71: MULTIVARIATE REGRESSION ANALYSIS OF SUB-GROUPS AND YLI

Variables	(1)	(2)	(3)	(4)
	All	FE Girls	ABE Girls	NFE Girls
Girls is a single orphan	1.81	0.25	0.74	2.89
Girls does not live with her parents	4.45**	2.31	7.71**	3.78
HoH has no formal education	-0.78	-1.57	1.97	-1.45
Caregiver has no formal education	12.88***	7.85**	14.55***	13.21***
Gone to sleep hungry many days	0.24	2.60	-4.32*	3.07
Gone without medicines	3.92***	6.45**	0.03	4.79*
Gone without cash income	7.12***	3.08	10.06***	7.90*
High chores burden	2.42*	3.43	0.88	3.20
Caregiver believes girls' education worthwhile, even if funds are limited	4.73***	4.89*	4.56*	4.61
Girl cannot use books/learning materials at school	-7.02***	-1.53	-12.17***	-7.11**
Teacher does not encourage participation	-10.53***	-4.53	-12.63***	-10.27**
Teachers treat girls/boys differently in the classroom	-4.10***	-2.84	-4.04*	-4.99**
Physical punishment witnessed	-5.06***	-7.40***	-4.44*	0.69
Age	0.48*	1.11	1.78***	-0.95
Constant	31.68***	27.90**	13.06	55.49**
Observations	1,420	421	484	515
R-squared	0.25	0.18	0.35	0.27

The table above summarizes the main results of the analysis. The same regression model, controlling for an additional age variable, was run separately for each of the cohort groups and state. In the following paragraphs, we will first assess each of the sub-groups individually and compare the differences across cohort groups.

As far as the girl characteristics are concerned, only the fact that the girl does not live with her parents is significant at 90% CI, representing a positive change of 0.24 standard deviations. The effect of the variable was substantially large and statistically

significant for the girls in the ABE cohort group. Thus, these findings suggest that girls less reliant on their caregivers are more likely to become independent, which could be partly explained by the fact that girls living with their relatives may face higher responsibilities in terms of housework and also may be more likely to take care of their own needs, especially when the household has lower economic status.

Among the household characteristics, an especially large and statistically significant effect was observed when the caregiver did not have formal education, representing 0.71 standard deviations increase in the YLI scores. The mean difference was very substantial and significant for the ABE and NFE girls. As mentioned in the previous analysis of these sub-groups, the households with lower educational levels of the HoH are more likely to have lower economic status, suggesting that the girl in these households are more likely to assume higher responsibilities for the housework. Their parents are under bigger financial

stress to address the household needs, which in turn could increase the importance of the roles assumed by girls within the households and help them develop additional organizational and communication skills.

The sub-groups related to the economic status of the household seem to be strong predictors of the YLI scores as well. The fact that the household had limited access to medication and cash income during the past 12 months has a large, positive and significant effect on the YLI scores, representing 0.21 and 0.39 standard deviations, respectively. These results confirm the previous assumption when we discussed the results of the girls' and households' characteristics.

The parental support for the girls' education, despite low financial resources, seems to have a positive, substantial and statistically significant impact on the girls' YLI scores. Also, parental support seems to be important only for the FE cohort group and has an especially large impact on the scores in South West State, increasing the YLI scores by 0.80 standard deviations. On the other hand higher HH chores burden seems to have a smaller effect on the YLI scores. Thus, the previously substantial and statistically significant effect on the higher HH chores burden of girls on the YLI seems to have been driven by other factors, such as the economic situation of the households and higher vulnerability to external economic shocks. Also, consistent with the findings, mothers during the FGDs often mentioned that a good relationship between mothers and daughters is essential for the development of the girls' self-confidence and decision-making capacities. In the same way, mothers investing in relationships with their daughters and mothers that are more involved in the girls' education and acknowledge the importance of educating their girls are more likely to support girls. This could additionally improve girls' perceptions of their self-worth.

“If you have friendship with your daughters, they will share their opinions also they can influence decisions about home, but if you do not become friends with your daughters, it is difficult for them to express themselves.” (Mother FGD)

“If Fadumo's mom motivates her and inspires her to go to the school and attend these free courses, there is no doubt that Fadumo will benefit from it.” (Girl during the risk mapping)

The results also suggest that positive and adequate learning environment is extremely important for the development of girls' leadership skills. The biggest decrease in the YLI scores was observed when the teacher was said to not encourage participation in the class. Statistically significant, the mean difference accounted for 0.57 standard deviations. The girls' limited access to learning materials seems to have a significant, negative and strong impact on the YLI levels. The fact that girls witnessed corporal punishment during the past week seems to have a strong and negative impact on the overall YLI scores. Last but not least, the differential treatment of the boys and girls had a negative impact on the YLI scores, decreasing the scores.

While considering the differences across different cohort groups for the learning environment subgroups, the YLI scores for the FE cohort groups were significantly affected only by the use of corporal punishment, representing a negative change in scores by 0.41 standard deviations. The limited access to learning material and teachers not encouraging participation seem to have a negative large effect on the YLI scores of ABE and NFE cohort groups.

Respondents during the FGDs with teachers often mentioned the importance of the teachers' encouragement in schools for the girls' ambitions and future expectations. While some of the teachers

frequently mentioned that they tried to encourage girls to pursue further education, others also stated that they tried to apply more gender-sensitive teaching practices, especially encouraging gender balance in the children's participation during the classes.

“When I'm with teachers, I get motivated. They always inspire us to be excellent students. Also, they help us understand the lesson and they always repeat the previous lessons for us.” (Girl during the risk mapping)

The majority of the teachers mentioned that girls were less likely to participate in the class because they were shyer than boys, feared to speak in front of the class and especially in front of the boys or in some cases in front of the teachers. In two instances teachers mentioned that girls felt especially ashamed to ask to go to the toilets during the class. On the other hand, in two cases teachers mentioned that the girls were as active and outspoken as boys and that there is no difference between them. Also, teachers often perceived girls as better students in schools and more diligent and disciplined, compared to the boys.

*“Yes, there is a difference between them, for example, boys participate more in the lessons because girls are shy, also about 80% of girls aren't able to face teachers.”
(Teacher FGD)*

During the vignette exercises girls often mentioned that if a girl is educated, there is nothing else that can stop her from achieving her ambitions. In the same way, few girls also stated that if a girl did not go to school, she would be less confident in her life. Moreover, some girls also mentioned that education allows girls to better understand themselves, become involved in decision-making, be listened to in their communities and uplift their communities. However, girls frequently mentioned that if the parents did not let girls go to school or there was no one to help girls with their housework, girls would not go to school.

“She can learn everything about herself. Also, the whole community will consider her as a real role model too.” (Girl during the vignette exercise, explaining what will change if a girl can overcome barriers to attend school)

To conclude, the main findings in this section suggest that girls in formal schools and younger girls are significantly less likely to have lower self-perceptions of leadership skills. Overall, the most significant predictors of the YLI scores are lower education levels of the HoH, lower economic status and the fact that the teachers do not encourage participation in the class. The main qualitative findings suggest that the girls' self-confidence levels are also affected by the mother-daughter relationships, the support and encouragement they receive from their teachers, and the fact of being able to overcome their barriers to education and learn. The common perceptions and expectations of the girls' behaviour, such as being 'shy' and 'quiet' in the class seem to have also a negative effect on their perceptions of leadership skills.

7.4 SCHOOL MANAGEMENT AND GOVERNANCE

The next intermediate outcome targeted by AGES concerns the quality of school management and governance. Proper school management has a number of downstream benefits for student learning, promoting enrolment, and encouraging retention, among other outcomes. Well-managed schools facilitate student learning by hiring better teachers, providing them with professional development opportunities, ensuring they are paid on time, and monitoring their teaching practices and attendance. Well-managed schools also improve community education levels by encouraging children to enrol in school, monitoring attendance rates, and promoting pro-education attitudes among the community. Finally, well-managed schools make good use of financial resources to improve infrastructure, design and implement effective school development plans, and develop and enforce standards of conduct that protect children and encourage inclusivity.

In the area of school governance, AGES is focused on the efforts of Community Education Committees (CECs) and the Ministry of Education officials that oversee schooling. The project is supporting CECs to increase their outreach to and capacity to assist the most marginalized girls – especially girls with disabilities (GWDs) – to enrol in school and remain enrolled over time. The project is also working with relevant government officials to promote inclusive education and special needs education, by supporting the finalization and implementation of a policy around these topics. AGES will also promote a review of the policy framework under which ABE programmes operate. Below the policymaking level, the project will support ministry of education officials through the development of simplified quality assurance procedures, alongside capacity-building to facilitate effective school-level monitoring by officials.

Given the project's focus on inclusive education and quality assurance procedures, the indicators tracked as part of this intermediate outcome are oriented toward the actions of CECs and MOE officials. 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

The baseline evaluation did not collect direct quantitative data on these indicators, but did collect data that is indicative of actions by CECs to promote enrolment and encourage inclusion of marginalized girls; moreover, the qualitative data provides insight into these topics in some detail.

CEC Actions to Promote Inclusive Education

The environment in which AGES is operating includes several axes of marginalization, some of which overlap with one another in complex patterns, as highlighted in our earlier discussion of beneficiary characteristics (Section 3.3). For instance, children of pastoralist families are disadvantaged due to their distance from urban centres and the schools that they are often expected to attend; moreover, their families tend to migrate seasonally and – during times of environmental or economic stress – migrate more broadly and more permanently. But marginalization is far from binary in this context: many of the most disadvantaged children in South Central Somalia are those in agricultural families, especially the minority groups in riverine areas along the Juba and Shabelle Rivers. Other families live in an environment of insecure property rights or as IDPs from conflict, drought, or floods, depending on their home areas. To be clear, there is no singularly disadvantaged group in the project locations – marginalization in this context is multilayered, and interacts with often extremely localized political and conflict dynamics.

The project's approach to school management is focused on promoting inclusion of marginalized girls, which can include pastoralists, members of agricultural clans, and minority groups, as well as girls with disabilities and others. According to the self-reflections of CEC members, many of them are actively reaching out to marginalized groups. CEC members who participated in qualitative interviews reported that they provide tangible benefits to girls from marginalized groups, such as children of pastoralist families, minority groups, IDPs, orphan girls, and girls with disabilities. For instance, many of the CEC members cite the fact that their school does not charge school fees as evidence that they are able to enrol and promote the inclusion of marginalized girls.¹⁵⁷ It is worth noting that just under half of formal schools in the sample (47.4 percent, n = 38) are free of charge, according to head teachers.

The provision of free primary education is an important step toward encouraging enrolment of marginalized girls, but it is not an intervention specifically targeted at these girls. Many CEC members did report more targeted actions they have taken, such as paying for uniforms, books, or other ancillary educational expenses for marginalized girls.¹⁵⁸ As one CEC member described, their committee helps children from pastoralist families buy school uniforms, books, and pens, and occasionally helps them pay their school fees.¹⁵⁹ A common theme in interviews with CEC members was that they support economically-marginalized families by buying school supplies and paying their schools; however, it is unclear how extensive financial support is or could be, in a context of widespread poverty. Some CEC members state that they fundraise from among the community to support marginalized children, which is likely a more common approach than the qualitative interviews indicated – whereas many interviewees reported that their committee supports girls financially, it is likely that this support is often generated from the broader community, through the fundraising efforts of the CEC, rather than directly from the CEC itself.¹⁶⁰ Financial support may also come from other sources: one CEC member reported that female teachers at their school donate part of their wages to pay school fees for female students.¹⁶¹

Aside from financial support, CECs can also encourage enrolment by visiting families at home, raising awareness among their community. They can also take action to prevent discrimination against marginalized children, or promote more inclusive attitudes among teachers, students, and community members. Compared to financial support, these possible actions to promote inclusion were cited less often by CEC members. In two separate FGDs, CEC members described going door-to-door to promote enrolment by female children; another CEC member reported that "they often visit different refugee camps and help nomadic kids to come to school."¹⁶² Unfortunately, there are also significant concerns – based primarily on qualitative interviews, but tentatively confirmed in the quantitative data as well – that students from minority groups, children of pastoralist households, and girls with disabilities face

¹⁵⁷ FGD with CEC members.

¹⁵⁸ FGD with CEC members (4).

¹⁵⁹ FGD with CEC members.

¹⁶⁰ One CEC member reported "knocking door-by-door" among the community to raise funds to support the education of pastoralist children in their area (FGD with CEC members).

¹⁶¹ FGD with CEC members.

¹⁶² FGD with CEC members (2).

discrimination in schools, and few CEC members participating in FGDs gave any indication they were able to effectively combat such issues, a point we return to in greater detail below.

According to data collected from head teachers, CECs are moderately active. Of the 38 formal schools sampled, nearly all (94.7 percent) had a functional CEC, according to head teachers. The most common activities CECs undertake are -- in order -- the promotion of enrolment by out-of-school (OOS) children, tracking teacher attendance, following up on cases in which a child has dropped out, and tracking student attendance. However, the rates at which CECs pursue these activities are fairly low: the most common activity – promoting enrolment of out-of-school children – is undertaken by CECs in just 60.5 percent of schools. Even fewer schools (44.7 percent of all schools sampled) have a CEC that actively tracks student attendance.

In general, the CECs appear to be orienting their activities in line with the project's goal of promoting enrolment and increasing transition, though it is not clear whether these efforts are specifically targeted at marginalized girls. Promoting enrolment of OOS children and following up on cases of dropouts are two direct ways CECs can encourage better transition rates, though, again, we emphasize that we cannot say whether these efforts are targeted at the most marginalized girls or simply reify any existing structures of inequality. Most importantly, it is clear that CECs overall are not as engaged as they should be in terms of pro-active management.

To illustrate this point, consider head teachers' reports of how often a member of the CEC has visited their school for monitoring purposes in the last year. Just 65.8 percent of head teachers report a monitoring visit by a CEC member over a lengthy one-year timeframe, which suggest that – even if CECs undertake monitoring efforts in theory – they are less likely to engage in it actively in practice. This question does not address monitoring a school for a specific outcome; rather, it captures *all* monitoring visits, of any kind, implying a relatively low level of CEC engagement.¹⁶³

Based on the qualitative data, CECs are actively engaged in promoting education – specifically, enrolment – of marginalized children. However, the quantitative findings described above beg the question of how engaged CECs are, in practice, in these activities. CEC members participating in qualitative interviews may indicate they support marginalized girls financially, but if a large minority of CECs do not monitor their schools on even an annual basis, it is unlikely they provide material financial support. One way to make sense of these seemingly contradictory results is to recall the sampling procedure for qualitative interviews at the baseline: data collection teams were advised to conduct qualitative interviews in areas with active CECs. As such, the disjuncture between qualitative findings of engaged CECs and quantitative findings of only low-moderate levels of CEC activity can be understood as a function of performing qualitative interviews at the schools with more obviously active CECs.

Bearing in mind the nature of qualitative sampling makes clear that the typical CEC is only marginally engaged in school management and promotion of education among marginalized girls. For instance, while several CEC members described tangible financial support they provided to in-need students and their families, relatively few teachers indicated that their CEC provided material support to their schools. When asked, 25.2 percent of teachers reported that their school's CEC supported their school financially in the last year, and just 31.6 percent reported *any* kind of support, either financial or in-kind. While it is possible that CECs support students and their families instead of the schools themselves, it seems unlikely that this would be a particularly common practice.

¹⁶³ This point should not be overstated.

Bias as a Barrier to Inclusion

The project activities relating to CEC management of schools are, probably rightfully, focused on encouraging enrolment and retention of marginalized girls. Our discussion above was, therefore, focused on CEC engagement along these lines, through financial support, enrolment campaigns, and monitoring of schools for student attendance and dropouts. Despite this focus, our analysis of the qualitative data revealed troubling evidence of less-tangible barriers to educational attainment by several of the marginalized groups targeted by AGES, in the form of discrimination and mistreatment at schools.

This mistreatment can take the form of institutionalized discrimination or mistreatment at the hands of teachers, which is especially problematic, but harassment and bullying by other students is perhaps more common. FGD participants intimated that students from minority groups, children of pastoralist households, and girls with disabilities can face discrimination in schools, and few participants gave any indication that they were able to effectively combat such issues. Participants said that "sometimes the children insult the nomadic kids with bad names"; another CEC member said that their neighbor, a boy with a hearing disability, attends school but the other students insult and mock him.¹⁶⁴ Some participants stated that "tribalism" and clan-based discrimination is not a problem in their communities, but others reported that children from minority groups may not bother attempting to enrol, because they are concerned they will be turned away and – even if they are able to enrol – they lack the necessary learning materials.¹⁶⁵

Recall that qualitative interviews were conducted in sample points with particularly active CECs, which we would assume mean the schools in question are, on average, better-managed. In light of these considerations, the qualitative findings that children from marginalized families and groups face bullying at school by other students and potentially experience discrimination at the hands of teachers or school administrators is even starker. Even in communities and schools with more active CECs, CEC members report that children with disabilities, and children from pastoralist families, face insults at school and may face more institutionalized discrimination within the school.¹⁶⁶ Other participants report openly that discrimination along clan lines is common in their communities – a problem which almost certainly manifests in schools, as well as other community institutions.¹⁶⁷

The quantitative data provide two avenues to confirm – in part – the potential issues faced by marginalized girls. The baseline evaluation included a survey specifically focused on girls with disabilities, who were recruited separately from the primary sample of girls and their households. In cases where the interviewed girl was not enrolled in school, their caregiver was asked why they were not enrolled. The answers given are illustrative: 13.8 percent of these caregivers reported that their girl was mistreated by her teachers at school, and 12.8 percent reported that their girl was flatly refused entry at the school.¹⁶⁸ The results suggest that girls with disabilities are, or at least perceive themselves to be, mistreated at school in a way that actively discourages their enrolment.

¹⁶⁴ FGD with CEC members; FGD with mothers.

¹⁶⁵ FGD with CEC members.

¹⁶⁶ FGD with mothers; FGD with CEC members (2).

¹⁶⁷ FGD with mothers.

¹⁶⁸ These two groups largely overlap.

The second avenue by which the quantitative can shed light on the nature of marginalization comes from interviews of caregivers whose girls have never been enrolled in school. Caregivers matching this description were asked the reasons why their girls had not been enrolled in school; the battery of available options, shown in the table below, included several answers that reflect a degree of discrimination or bullying at school. The table below reports the share of, for instance, pastoralist girls, whose caregiver says it is unsafe for their girl to be in school, among only those respondents whose girl has never been enrolled. We report this statistic for several subgroups who are likely to face discrimination – pastoralists, members of the Digil-Mirifle agricultural clan, members of minority groups, girls who speak Af-Maay at home, and girls with a disability. For comparison, we report the same share for caregivers of girls who do not fit any of the aforementioned subgroups.

TABLE 72: REASONS FOR NOT ENROLLING GIRLS TO SCHOOLS

Unsafe for girl to be in school	Share of caregivers in subgroup citing reason for girl to not be enrolled
Overall	1.6%
Pastoralist	10.0%
Agricultural clan	10.3%
Minority group	0.0%
Speaks Af-Maay at home	13.0%
Has disability	16.1%
Teachers mistreat girl at school	
Overall	1.0%
Pastoralist	0.0%
Agricultural clan	8.2%
Minority group	4.5%
Speaks Af-Maay at home	13.0%
Has disability	3.2%
Refused entry at school	
Overall	0.3%
Pastoralist	0.0%
Agricultural clan	7.5%
Minority group	4.5%
Speaks Af-Maay at home	12.4%
Has disability	6.5%
Other students bully/mistreat girl at school	
Overall	0.3%
Pastoralist	0.0%
Agricultural clan	6.8%
Minority group	0.0%
Speaks Af-Maay at home	10.7%
Has disability	3.2%

Reading down the rows in each panel of the graph makes fairly clear that girls from particular subgroups experience mistreatment at school. The findings focused on girls whose families speak Af-Maay at home are the starkest: of the 177 such girls in the baseline sample who are not enrolled, 12.4 percent report they were refused entry at the school and 13.0 percent report that teachers at school mistreat them and this was a reason they were not enrolled in the past. In comparison, girls who are not included in any of the specific disadvantaged groups analysed here are significantly less likely to cite these barriers, even among those who are also out-of-school.

These results speak to a more complex problem facing marginalized girls. While CECs and other organizations can focus on providing financial assistance and promoting education broadly, girls face bias that may be more difficult to uproot. CECs and their supporting organizations, such as CARE, may need to make special efforts to ensure schools are free of discrimination by teachers and students alike, and engage in wider behavioural and attitudinal change efforts.

MOE use of inclusive quality assurance procedures

The second indicator of inclusive school governance concerns the use of quality assurance procedures by government officials. The project will work with MOECHE and State MOE officials to implement inclusive quality assurance processes for use by MOECHE and State MoE officials and provide capacity-building to officials to facilitate their use.

Unfortunately, the baseline evaluation did not capture systematic data regarding MOE quality assurance procedures or, even more problematically, data on MOE engagement with schools in general, as these dimensions are captured through the monitoring system (joint visits and minutes of engagement meetings). Head teachers were not asked directly about monitoring visits from MOE officials. This means that our understanding of MOE engagement is impressionistic.

In general, the level of engagement in most schools appears to be limited. The federal government's role in education provision is relatively limited, although this situation is slowly shifting as the federal model implementation evolves. As we described at the outset of this report, education in Somalia is primarily administered not through government, but through umbrella associations of privately-run schools, as well as through independent private schools.

CEC members participating in FGDs described a limited, one-way, and very occasional level of engagement. One group of teachers stated that they receive no financial support from the government, and no CEC member specifically stated that they receive financial support.¹⁶⁹ Schools' primary avenue of engagement is the provision of the national education curriculum, and textbooks distributed as part of the curricular rollout.¹⁷⁰ The most expansive role for government was cited by CEC members at one school, who described the government's payment of teacher salaries, and provision of teacher training.¹⁷¹

Monitoring by government officials is extremely limited. It is important to remember that FGDs with CEC members were conducted in schools with particularly active CECs, implying that they would typically be aware of things occurring in their schools, such as visits by MOE officials. Despite this, one

¹⁶⁹ FGD with teachers; FGD with CEC members.

¹⁷⁰ FGD with CEC members.

¹⁷¹ FGD with CEC members.

CEC member stated explicitly that no official visits had occurred; another stated that monitoring occurred only during examination administration.¹⁷² Interestingly, two FGDs produced extremely tentative findings that might relate to securitization of the relationship between school management and the government: in these cases, when asked about the role of government in the school, participants highlighted the role of security officers, the potential for attacks at their school, and the role of the government in keeping the school safe.¹⁷³ On the other hand, assuring security is a legitimate need, as a lack of security reduces girls' enrolment and disrupts their regular attendance, and the government can and does play a role in reducing the risk of traveling to school.¹⁷⁴

In important ways, the qualitative evidence is mixed. While some interviewees described a total lack of government interaction or engagement, others occasionally reported monitoring visits – even frequent monitoring, among one set of teachers.¹⁷⁵ To some degree, the gaps may be the result of differing perspectives between CEC members and teachers: members of CECs tended to indicate limited engagement from the government, while teachers cited more frequent interaction, including monitoring, design of the curriculum, and provision of teacher training. Teachers, who are at the school daily, may be more reliable informants in this area. But the variability in responses almost certainly reflects regional differences as well: some participants citing the importance of security are in particularly conflict-affected areas, where localized security is critical to school functioning. And regional variation in the extent of the government's capacity almost certainly explains part of the variation in monitoring efforts, curriculum rollout, and even textbook provision.¹⁷⁶

Summary

The project's role in promoting school management takes place at several different levels. At the highest level, interventions are attempting to shape federal policy; at a meso level, the project seeks to influence monitoring efforts by ministry bureaucrats; finally, at the local level, the project's focus is on inclusivity in CEC efforts to promote enrolment and other outcomes. The results in this section emphasize that CECs exist and are broadly functional, but their level of active engagement is generally low. CEC members are cognizant of the need to encouragement enrolment and provide support to marginalized groups, but the extent of the support they provide, in practice, is unclear.

The government's role in education is more clearly limited, which fits with our prior understanding of the South Central Somalia educational context. The government does provide critical services to schools, such as developing a national curriculum, and helping to ensure security at schools and in surrounding communities. But their role in monitoring and provision of learning materials is extremely limited in frequency and scope.

¹⁷² FGD with CEC members (2).

¹⁷³ FGD with CEC members (2).

¹⁷⁴ FGD with CEC members; FGD with teachers.

¹⁷⁵ FGD with teachers.

¹⁷⁶ For instance, one group of teachers lamented the inability of the government to provide textbooks in line with the new curriculum, and linked it explicitly to the region in which they live and work (FGD with teachers).

The data in this section regarding government actions and – to a lesser extent – the role of CECs, is limited, which prevents us from drawing firm conclusions regarding school management. During preparations for the midline evaluation, efforts should be made to more specifically measure the outcomes of interest to the project: CEC actions that focus explicitly on inclusion of marginalized groups could be measured more directly via quantitative data collection, as could MOECHE/ State MOE engagement with school officials and teachers. Improved tool design will facilitate firmer conclusions regarding the performance of the project in terms of this intermediate outcome.

This section will specifically focus on the analysis of community attitudes towards girls' education. The GEC conceptualization of the link between the change in community attitudes and girls' education opportunities is based on the 'virtuous circle', where at the first stage, the positive change in the community perceptions of the value of girls' education will improve their attendance and enrolment in the short run. In the long term, increasing numbers of women gaining higher education will improve women's economic opportunities and leadership roles in society, challenging existing norms. In this way, the change achieved in the learning opportunities and inclusion for the girls will be sustainable.

From the implementation of Phase 1 of GEC initiatives, the main lessons learned indicated that 1) a positive change in the community attitudes is important for the learning and transition outcomes; and 2) the attitudes towards girls' education are complex and rather than the general lack of support, it is a function of the economic situation of the household, where the financial burden of girls' education defines how likely girls are to attend a school. This is a particularly important aspect in Somalia, where the clan system plays a particularly important role in society. Through marriage, girls become members of another clan, which may reduce incentives to support girls' education as it is not going to be beneficial for the household. Besides, as women's roles remain restricted to the domestic sphere, girls are less likely to use their education after the marriage, further reducing the perceptions of the value of girls' education.

The LNGB AGES project aims at achieving positive shifts in community practices towards girls' education to achieve 1) improved literacy, numeracy, financial literacy and knowledge of menstrual hygiene management; 2) transitions to formal/non-formal education and employment/ self-employment; and 3) overall sustainability of the interventions. The expected change will be measured through the following indicators:

1. Parents investing profits from income generation activities in expenses related to girls' education;
2. Changes in caregivers' attitudes and practices towards new roles for girls (reflected in changes in attendance, mobility, and early marriage).

We will approach the analysis by incorporating both quantitative and qualitative data. As the main indicators to better understand the community attitudes towards the girls' education, we will look at caregivers' aspirations for the girls' education, caregivers' attitudes towards the girls' education in relation to the household responsibilities and chores burden and caregivers' attitudes towards the cost of the girls' education.

In addition, the analysis will rely on the qualitative data from the FGDs with mothers and CECs at the schools. The community attitudes specific questions for CEC members were:

1. In what ways do families in the communities around the school support girls' education? Can you provide some examples or stories?;
2. How supportive are religious leaders in your community of girls' education? Can you provide some examples or stories? How do religious leaders work with the CEC to support the school?;

The community attitudes specific questions for mothers were: 1) Is it common to have as many adolescent girls enrolled as adolescent boys in this community? If not, why do you think some parents in this community might not enroll (or are not enrolling) their adolescent daughters in school?; and 2) What kind of factors do typical families consider when deciding to send their adolescent daughters to school?.

In addition, to complete the analysis, the information from the risk mapping and vignette exercises with girls will be incorporated. The vignette exercises will especially useful in the analysis of the community attitudes for a better understanding of how girls perceive community attitudes towards their education, as the tool relies on indirect questions, where girls are asked to complete different stories featuring girl characters facing different barriers to continue their education.

Consequently, in the following section, we will first analyze parents’ willingness and capacity to support the financial costs of girls’ education. The second section will then explore general attitudes of caregivers towards the girls’ education and their expectations towards the prioritization of their household responsibilities and early marriage, especially in the context of marginalization, negative economic situation, security risks, and violence.

7.5 COMMUNITY ATTITUDES

The aim of this section is to establish baseline values to measure parental and caregivers’ perceptions of the value of girls’ education. In the first section, we will first discuss the attitudes towards the cost of education and willingness to financially support girls’ education, despite households’ financial situation. Secondly, we will analyze caregivers’ aspirations for girls’ higher education and heir willingness to prioritize education over household chores. The analysis will mainly rely on the selected quantitative indicators and will be informed by the main findings from the qualitative data analysis. In addition, further analysis was done for the main quantities indicators and selected sub-groups to identify key predictors of caregivers’ attitudes.

Financial Burden of Girls' Education

The end line target value is an increase of 80% of parents investing their income profit in girls’ education related expenses. To establish the baseline values for the parental attitudes towards the cost girls’ education, this section will mostly focus on the analysis of the parents’ *willingness to invest* in the girls’ education vis-à-vis other household expenses.

The quantitative data analysis will be mostly centered around the two variables, both looking at the caregivers’ perceptions of the economic costs of the girls’ education. The first indicator, a dummy variable is specifically exploring the likelihood that the caregiver agrees with the statement that it is worth to invest in the girls’ education, even if financial resources are limited. The second indicator is looking at a similar question, asking caregivers whether the limited access to financial resources is an acceptable reason for girls to not attend the school.

Overall, as the below table shows, nearly 88% agreed that the girls’ education is worthwhile, even if the funds are limited. No differences in terms of the distribution and the statistical significance were observed when the analysis was disaggregated by the cohort group.

TABLE 73: CAREGIVER BELIEVES GIRLS' EDUCATION WORTHWHILE, BY COHORT GROUP

Caregiver believes girls' education worthwhile, even if funds are limited		Coef.	St. Err.
Overall	87.75%		
FE Girls	87.89%	0.00	0.03

NFE Girls	88.74%	0.02	0.03
ABE Girls	86.57%	-0.02	0.03

*** significant at 99% confidence interval, ** significant at 95% confidence interval, * significant at 90% confidence interval

Subsequently, we analyzed the effect on specific groups of variables describing the households' economic situation and education levels of the head of a household and the caregiver. Using simple regression analysis we have compared the difference in means of the caregivers' attitudes, a binary variable, and each of the sub-group variables also coded as binary variables. The results are indicated in the table below.

The results suggest that households with limited access to water are 0.16 percentage points and households that are food insecure or have limited access to education are 0.07 percentage points less likely to support girls' education. In the same way, land ownership seems to decrease the willingness of parents to invest in girls' education by 0.04 percentage points.

Thus, when a household has to decide between spending financial resources on girls' education and household needs, especially when the household is highly food insecure or has very limited access to water and medication, the household is less likely to invest in girls' education. Besides, girls in more insecure households under higher financial stress may be more likely to stay at home, caring for their siblings and household, while parents would work outside of the household. The burden of the household chores may be higher for girls when the household had limited access to the water supplies, requiring more time and effort to access the water source.

The qualitative data supports this interpretation. Mothers commonly expressed the opinion that in households with lower economic status, girls are more likely to stay at home to help. Also, the fact that the opportunity costs of investing in girls' education are perceived to be higher than the costs of investing in boys' education because girls were expected to move to other households once they are married.

“They (parents) mostly look at the economic side, because when they have boys and girls, they love to send boys to the school, because they are convinced that a boy can enrich and develop the family.” (Mother FGD)

Pastoral households are 0.28 percentage points less likely to support girls' education if their funds are limited, and livestock ownership seems to have a statistically significant and negative effect, representing a negative change of 0.08 percentage points. While pastoral households are generally considered to be economically less stable, the respondents during the focus group discussions also often mentioned that girls in the pastoral and nomadic communities are more likely to be married earlier. This is supported by the fact that in Somalia, pastoral communities are considered as very traditional, therefore more likely to hold negative attitudes to girls' education. Moreover, the financial advantages are likely to serve as an incentive for early marriage in these communities. This was also mentioned as a case for the IDP household and the marginalized communities.

“Some of the parents are forced to marry their girls due to economic reasons, for the internally displaced people early marriage is more (frequent)” (CEC member)

Moreover, the fact that the caregiver has no formal education¹⁷⁷ has a statistically significant and negative impact on the parents' willingness to support girls' educational costs. Meanwhile, the fact that the HoH does not earn a wage or has no occupation does not affect the parental willingness to support girls' education. When the caregiver lives in the household with savings parents are less likely to support girls' education (by 0.14 percentage points) but the difference in means is not statistically significant.

Only the fact that the caregiver lived in the female-headed household has a positive and statistically significant impact on the caregiver's support, representing a positive change in the mean scores by 0.06 percentage points. Mothers, in general, may have a more positive view of the value of investing in girls' education. Single mothers as possibly the only adult breadwinners in their households are more likely to understand the importance of building girls' independence and ability to take care of themselves, by gaining better education. Moreover, mothers or female relatives as the main decision-makers in the household could represent a positive role model for girls. One of the outcomes of the focus group discussions was that mothers tended to support the opinion that the girls' education is especially valuable for their future roles as mothers to make better decisions and to take better care of their children. Moreover, mothers often referred to the positive impact of the educated girls (and future mothers) for their communities.

“Education is very important for girls because girls are an important pillar in our community. An educated mother is just like an educated community; therefore girls deserve to be educated just like the boys.” (Mother FGD)

TABLE 74: CAREGIVER BELIEVES GIRLS' EDUCATION WORTHWHILE, BY SUBGROUP

Sub-group	Obs	Coef.	St. err
Female HoH	530	0.06** *	0.0 2
HoH has no education	302	-0.03	0.0 3
HoH has no formal education	1162	-0.03	0.0 2
Caregiver has no education	383	-0.01	0.0 3
Caregiver has no formal education	1209	- 0.08** *	0.0 2
HoH doesn't have wage/occupation	558	0.00	0.0 2

¹⁷⁷ The variable includes cases when the caregiver did not have any education at all or attended Madrasa or received some form of informal education.

Pastoral household	25	-0.28** *	0.09
Poor roof	284	0.05*	0.03
Went to sleep hungry many (more than 10) nights, last 12 months	385	-0.07**	0.03
Went without water for home use most/all days, last 12 months	74	-0.16**	0.07
Went without medicines most/all days, last 12 months	238	-0.11**	0.05
Went without cash income most/all days, last 12 months	210	-0.04	0.04
HH owns land	462	-0.04*	0.02
HH has savings	27	-0.14	0.10
HH owns phone	1229	0.03	0.04
HH owns livestock	253	-0.08** *	0.03
HH owns cattle/camels	30	0.00	0.00
HH owns middle livestock (goats/sheep)	106	-0.05	0.04
HH owns small livestock (chicken)	172	0.00	0.00

*** significant at 99% confidence interval, ** significant at 95% confidence interval, * significant at 90% confidence interval

The second indicator we will consider is whether a caregiver would consider the cost of education as an acceptable reason not to attend a school. The table below summarizes the distribution of the indicator, across different cohort groups and states. Most of the respondents (62%) agreed with the statement.

There is a small variation across different cohort groups, with only 59% of FE caregivers who agree with the statement, 63% of NFE caregivers and nearly 65% of the ABE caregivers who agree with the statement. Further regression analysis suggests that neither cohort group nor the specific states have a significant effect on the likelihood that the caregiver would accept girls' not attending the schools due to the cost of education.

TABLE 75: CAREGIVER ATTITUDES TOWARD COST OF EDUCATION, BY COHORT GROUP

Caregiver believes the cost of education is an acceptable reason to not attend school	Coef.	St. Err.
Overall	62.25%	

FE Girls	58.67%	-0.05	0.03
NFE Girls	62.72%	0.01	0.03
ABE Girls	64.88%	0.04	0.03

*** significant at 95% confidence interval, ** significant at 90% confidence interval, * significant at 85% confidence interval

Similar to the results of the first indicator, caregivers in the households with a less favorable economic situation are more likely to accept girls not going to the school, if their education is costly, as indicated in the table below. This is the case especially when the household is food insecure, representing a statistically significant 0.17 percentage points decrease in the mean scores, and when the household has limited access to cash income, representing a statistically significant change of 0.23 percentage points. On the other hand, the fact that the household has limited access to the medication or water seems to have a marginal effect on the likelihood that the caregiver would agree with the statement.

There also seems to be a difference between the fact that the HoH does not earn a wage or does not have an occupation, compared to the long-lasting financial insecurity, then the household did not have access to cash income during the most/all days of the past 12 months. The households that have been financially insecure for a long time seem to be more likely to agree with the statement, compared to the household when the HoH is currently unemployed. Thus, protracted long-term vulnerability is more predictive of the probability girls will not attend the school if the education is too costly.

On the other hand, pastoral households are significantly more likely to disagree with the statement, which a negative change of 0.31 percentage points in the mean scores, significant at 99% CI. Livestock ownership seems to have only a small negative effect on parental attitudes, significant only for the ownership of small livestock. Compared to findings in the previous section, where pastoral households were significantly more likely to disagree with the statement that it is worthwhile to invest in the girls' education, even if the funds are limited, the present results seem to be counter-intuitive.

The findings from the analysis of the attendance rates, in the previous section, suggest that schools with a higher share of girls from pastoral households are significantly more likely to have higher attendance rates. Thus, it is possible that as the pastoral households face bigger barriers to send their children to schools, the households who can send their girls to schools are more determined to support their education, despite limited financial resources.

Also, caregivers in female-headed households more likely to disagree with the statement. The results are very similar to the results for the previous indicator. The female-headed households are (statistically) significantly more likely to disagree with the statement, with the 0.8 percentage points change in the mean scores.

Moreover, in households where caregivers or the HoH received informal education¹⁷⁸, caregivers seem to be significantly more likely to disagree with the statement (by 0.21 and 0.14 percentage points, respectively). The results suggest that there is a difference in the parental support when the HoH or caregiver received informal education, compared to the households where HoH and caregivers received

¹⁷⁸ 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.

no education at all. This suggests that caregivers who had exposure to some informal education are likely to recognize the value of educating girls.

TABLE 76: CAREGIVER ATTITUDES TOWARD COST OF EDUCATION, BY SUB-GROUP

Sub-group	Obs	Coef.	St. err.
Female HoH	530	-0.08**	0.03
HoH has no education	302	0.00	0.03
HoH has no formal education	1162	-0.14***	0.03
Caregiver has no education	383	0.02	0.03
Caregiver has no formal education	1209	-0.21***	0.04
HoH doesn't have wage/occupation	558	-0.08***	0.03
Pastoral household	25	-0.31***	0.10
Poor roof	284	0.04	0.03
Went to sleep hungry many (more than 10) nights, last 12 months	385	0.17***	0.03
Went without water for home use most/all days, last 12 months	74	0.11	0.06
Went without medicines most/all days, last 12 months	238	0.08*	0.03
Went without cash income most/all days, last 12 months	210	0.23***	0.04
HH owns land	462	0.05	0.03
HH has savings	27	0.01	0.09
HH owns phone	1229	-0.01	0.04
HH owns livestock	253	0.06	0.04
HH owns cattle/camels	30	0.00*	0.00
HH owns middle livestock (goats/sheep)	106	0.01	0.05
HH owns small livestock (chicken)	172	0.01***	0.00

*** significant at 99% confidence interval, ** significant at 95% confidence interval, * significant at 90% confidence interval

In some cases, the results from the analysis of the two indicators in this section diverged, such as the 88% of caregivers who agreed it is worthwhile to support the girls' education even if the funds are limited, while almost 62% of respondents agreed that cost of the education is an acceptable reason to not to attend the school for girls.

The subgroup analysis showed that caregivers in female-headed households are more likely to invest in girls' education despite financial constraints. When the households were food insecure, had limited access to water, medication and cash wage, the caregivers were more likely to prioritize other expenses, over girls' education costs, especially when the household lacked food or cash income. On the other hand, different and sometimes contradictory findings were found for other subgroups, especially for pastoral households, livestock ownership, land ownership, HoH and caregiver (lack of) education and lack of regular wage or stable occupation of the HoH.

The way how the two indicators were answered may vary, which could partly explain the differences in the results. The first indicator could be perceived as less concrete and more hypothetical, while the formulation of the second question is more concrete and asked in the present tense and the caregiver could be more likely to refer to own experiences or past decisions. On the other hand, the first indicator is directly asking about the personal opinions of the caregiver. While responding to the second question, the caregiver may refer to general attitudes or opinions in a community, as the question is less targeted at personal experiences. Lastly, the formulation of the answer options¹⁷⁹ for the second question could be less intuitive and might influence the way respondents would answer.

Caregivers' attitudes and practices

The present section will mostly focus on the analysis of the caregivers' attitudes and practices towards the roles for girls and their expectations for the girls' education. The table below summarized the main indications related to parental attitudes¹⁸⁰. The results below indicate that around 87% of the caretakers aspire for the girls to pursue a further university education.

Similar results were observed when the indicator was broken down by cohort groups. Slightly lower percentages of caretakers who aspire for girls' university education (85%) were registered for the NFE cohort group. Additional regression analysis indicates that the difference in the mean scores is significant only for the NFE cohort group. These results suggest that caregivers of the NFE girls are less likely to aspire for their higher education, which could be related to the fact that the girls in the NFE groups are older (17 to 20 years) and were more likely to drop out of the school in the past or have never completed

¹⁷⁹ Formulation of the answer options: Yes (acceptable to not attend) versus no (not acceptable to not attend).

¹⁸⁰ Due to the low variation of the distribution for the caregivers' aspirations for the primary and secondary education, the sub-groups were rather coded as the whether caregivers also aspire for the university education or not.

primary/secondary education¹⁸¹. However, the coefficients remain low (ranging between 0 to 0.04 percentage points) with a small effect on the caregivers' likelihood to support the girls' education.

TABLE 77: CAREGIVER ASPIRATIONS FOR GIRL, BY COHORT GROUP

Caregiver aspires to university education for a girl		Coef.	Standard error
Overall	87.39%	-	-
FE Girls	88.60%	0.02	0.02
NFE Girls	84.66%	-0.04*	0.02
ABE Girls	89.26%	0.03	0.02

*** significant at 99% confidence interval, ** significant at 95% confidence interval, * significant at 90% confidence interval

Overall, most of the variables had strong and significant coefficients. The strongest and statistically significant negative impact was observed for pastoralist households, indicating that the caregivers in the pastoralist households are significantly less likely to aspire for the university education for their girls, with 0.24 percentage points difference in the means. This is in line with the previous findings indicating that the pastoral households are less likely to invest in the girls' education if the funds are limited and it is also consistent with general opinions expressed during the FGDs, that girls from the pastoral communities are more likely to marry earlier. On the other hand, the fact that the HoH belongs to one of the pastoralist clans is significant but has also a small positive effect (0.04 percentage points difference in means). These results suggest that when HoH belongs to one of the pastoralist clans, which are traditionally dominant, caregivers are significantly more likely to support girls' education.

The fact that the HoH belongs to one of the agricultural clans¹⁸² has a significant and negative impact on the caregivers' aspirations for the girls' education, with 0.13 percentage points difference in means, while the fact of belonging to the marginalized clans and occupational minorities has no impact on the caregiver's aspirations. While a large part of the agricultural clans lives in the South West State, after controlling for the geographic location, the results remain unchanged.

The second strongest and negative impact was recorded for the lack of no-formal education for the caregiver and the HoH. On the other hand, the fact that the HoH/caregiver has no education at all seems to have a positive and significant impact on the caregiver's aspiration. These results suggest that the fact

¹⁸¹ In addition, teachers were asked complimentary questions about the community and their own attitudes to girls pursuing further education. For both sets of questions teachers responded in the similar way, equally supporting the girls and boys' education.

¹⁸² Namely the Digil-Mirifle.

of having attended Madrasa or another type of informal education has a negative impact on caregivers' aspirations¹⁸³. The overall effect of the education subgroups varies between 0.07 to 0.09 percentage points mean difference, which is a relatively small effect compared to some of the more significant sub-groups.

All the variables with the significant and positive impact on the caregiver's aspirations were related to the lower economic status of the household, such as limited access to water, medication, and cash income. While the effect on parental attitudes is small for all of these variables, the results seem to be aligned with the previous analysis of the attendance rates, suggesting that the poorer and more vulnerable households are more likely to support their children's education. Also, unlike the previous indicators, the question is more hypothetical, and respondents may not necessarily incorporate the financial constraints while responding.

Despite the fact that the journey to school is unsafe, the caregiver is more likely to have significantly higher aspirations for the girls' education, with 0.13 percentage points mean difference. The caregivers may send girls to school even if it is unsafe because they have higher aspirations for their girls.

TABLE 78: CAREGIVER ASPIRATIONS FOR GIRL, BY SUBGROUP

Sub-group	Obs.	Coef.	St. err.
Female HoH	530	-0.02	0.02
HoH has no education	302	0.08***	0.02
HoH has no formal education	1162	-0.07***	0.02
Caregiver has no education	383	0.08***	0.02
Caregiver has no formal education	1209	-0.09***	0.02
HoH doesn't have wage/occupation	558	0.01	0.02
Pastoral household	25	-0.24**	0.09
Poor roof	284	-0.01	0.03
Went to sleep hungry many (more than 10) nights, last 12 months	385	0.02	0.03
Went without water for home use most/all days, last 12 months	74	0.08**	0.03
Went without medicines most/all days, last 12 months	238	0.07***	0.02
Went without cash income most/all days, last 12 months	210	0.09***	0.02
HH owns land	462	0.02	0.02

¹⁸³ When considering if the impact of being exposed to informal education (different from Madrasa) or qur'anic education is driving the differences between the two sub-groups, we found that the effect on caregivers' aspirations when caregiver had informal education or went to madrasa had no effect on the caregiver's aspirations. Thus we cannot confirm that exposure to informal education is driving the differences between the two subgroups. However, the fact that the HoH received no education at all may also indicate that the household is facing higher economic insecurity or marginalization than households where HoH or caregiver received some form of informal education.

HH has savings	27	0.05	0.05
HH owns phone	1229	0.01	0.03
HoH belongs to pastoralist clan	671	0.04**	0.02
HoH belongs to agricultural clan	598	-0.06***	0.02
HoH belongs to marginalized group or occupational minority	143	0.03	0.03
Girl was mistreated by her teacher	28	-0.01	0.08
Caregiver considers that journey to school is unsafe	23	0.13***	0.01
Caregiver considers that school is unsafe	36	-0.11	0.09
Girl is not enrolled in school because there is not enough money to pay the costs of her education	367	0.04	0.04

*** significant at 99% confidence interval, ** significant at 95% confidence interval, * significant at 90% confidence interval

During the household survey, caregivers were also asked whether they agree with the statement that the work or HH chores are an acceptable reason to not attend the school. **Overall, 40% of the respondents agree with the statement.**

While looking at the distribution of the mean scores across different cohort groups, the highest percentage of the caregivers who agreed with the statement was observed among the caregivers of the girls in the formal schools (44%), followed by the caregivers of the girls participating in the NFE classes (39%).

TABLE 79: CAREGIVER ATTITUDES TOWARD DOMESTIC WORK, BY COHORT GROUP

Caregiver believes work or HH chores are an acceptable reason to not attend school		Coef.	St. Err.
Overall	40.07%	-	-
FE Girls	43.47%	0.05	0.03
NFE Girls	39.22%	-0.01	0.03
ABE Girls	38.02%	-0.03	0.03

*** significant at 99% confidence interval, ** significant at 95% confidence interval, * significant at 90% confidence interval

Subsequently, the regression analysis results of the additional variables, presented in the table below, suggest that in the female-headed households, households with some form of the land ownership, households with some savings, caregivers are more likely to prioritize household chores and work over the girls' education. While these results seem to be contradictory to the previous analysis, where female-headed households were more likely to financially support girls' educational costs, it is possible that with female HoH, mother or other female caregiver bears additional responsibilities to provide sufficient financial resources for the household, which results in relegating additional responsibilities to the girl, such as taking care of younger sibling or household. Besides, one of the outcomes from the FGDs is that household chores were one of the biggest challenges for girls' attendance and that it was caregivers, who decided on how much of the household chores girls will be responsible for.

“The obstacle is for a mother to do the house chores that the girl was doing. It won’t be possible to tell her to go to school and learn and still do the chores. The mother should do the chores, and the girl should go and study.” (Mother FGD)

In this case, the fact that the household has some savings seems to be a relatively bigger predictor of the caregiver’s negative attitudes towards the girls’ education opportunities, representing 0.35 percentage points increase. However, such a strong effect could be also explained by a very low number of observations.

Also, in households where girls reported to have been bullied by their peers, mistreated by their teachers or when the caregiver considers that the school environment is not safe enough for their daughters, the caregivers are significantly more likely to agree with the statement. The especially strong effect was found in the case when girls reported they were bullied by their peers, representing 0.70 percentage points change in the caregivers’ attitudes. These findings are also in line with the general opinion dominating across FGDs that girls are more likely to be bullied or harassed by boys and that there should be a strong gender division between the girls and boys in the schools. In addition, it was often mentioned that parent sees the risk of pregnancy as one of the reasons why it is riskier for girls to go to the schools. Equally strong negative effect on the caregivers’ attitudes was observed when girls reported they were mistreated by their teachers, representing 0.38 percentage points mean difference and when the caregiver considered that the school was unsafe for girls, representing 0.36 percentage points mean difference.

*“Families mostly love that girls should stay at home and boys should go out. Because families are scared that the girls should go somewhere else. Often, many girls are seen pregnant at a mall, and this is why many families are very scared to send their daughters to schools.”
(Mothers FGD)*

“I also feel unsafe in the schoolyard. When boys come out of the classes, they sit there and they sometimes harass and insult girls. So, yes, this area is unsafe.” (Risk mapping)

Moreover, in households where the girls are currently not enrolled because of the lack of funds to pay the educational costs, the caregivers are more likely to agree with the statement and the results are also statistically significant, with 0.22 percentage points change in the mean difference. This finding is in line with the previous conclusions when the more vulnerable households, such as food insecure households or with limited access to water, medication and cash income, were less likely to invest in the girls' education and as well delegate additional household chores to the girls.

In addition, the fact that the HoH or caregiver has no education at all has a negative impact on the caregivers’ attitudes, while some form of informal education has a positive impact on the caregivers’ attitudes. The effect of these subgroups is ranging between 0.07 to 0.16 percentage points and is statistically significant when the HoH and the caregiver have no formal education. However, when we have specifically compared groups when HoH / caregiver received no education, informal education

(different from Madrasa) and went to Madrasa, the effect on the caregiver’s attitudes was close to zero and statistically insignificant¹⁸⁴.

On the other hand, a handful of variables seem to act as predictors of caregiver’s more positive attitudes towards the girls’ education. The biggest and significant difference in the mean scores was observed in the pastoral households, suggesting that the caregivers in these households are 0.29 percentage points less inclined to prioritize work or HH chores over girls’ education. While considering whether the HoH belongs to the pastoralist, agricultural or marginalize clans, only the fact that the caregiver belongs to the marginalized group or occupational minority has a statistically significant effect on the caregiver’s attitudes. In this case, the caregiver is 0.10 percentage points more likely to disagree with the statement. This finding is in line with the findings from the attendance section where schools with a higher percentage of girls from pastoral households were significantly more likely to have higher school attendance.

Interestingly, all the variables indicating households’ lower economic status, such as food insecurity, limited access to water, medication and cash income had negative and significant coefficients, suggesting caregivers in the economically more vulnerable households are more likely to hold positive attitudes towards the girls’ education. This was especially the case for households with limited access to medication and cash income, followed by limited access to water. In addition, caregivers were significantly more likely to disagree with the statement in the households, where the HoH does not have a stable occupation or does not earn income.

TABLE 80: CAREGIVER ATTITUDES TOWARD DOMESTIC WORK, BY SUBGROUP

Sub-group	Obs.	Coef.	St. err.
Female HoH	530	0.09**	0.03
HoH has no education	302	0.11***	0.03
HoH has no formal education	1162	-0.12***	0.03
Caregiver has no education	383	0.07	0.03
Caregiver has no formal education	1209	-0.16***	0.04
HoH doesn't have wage/occupation	558	-0.08***	0.03
Pastoral household	25	-0.29***	0.10
Poor roof	284	-0.01	0.03
Went to sleep hungry many (more than 10) nights, last 12 months	385	-0.04	0.04
Went without water for home use most/all days, last 12 months	74	-0.11*	0.06
Went without medicines most/all days, last 12 months	238	-0.09*	0.03
Went without cash income most/all days, last 12 months	210	-0.08*	0.04
HH owns land	462	0.10***	0.03
HH has savings	27	0.35***	0.09

¹⁸⁴ For possible explanation of the results refer to the previous case (Ibid 9).

HoH belongs to pastoralist clan	671	0.04	0.04
HoH belongs to agricultural clan	598	0.00	0.03
HoH belongs to marginalized group or occupational minority	143	-0.10**	0.05
Girl was bullied in school by her peers	20	0.70***	0.06
Girl was mistreated by her teacher	28	0.38***	0.13
Caregiver considers that journey to school is unsafe	23	-0.14	0.09
Caregiver considers that school is unsafe	36	0.36***	0.11
Girl is not enrolled in school because there is not enough money to pay the costs of her education	367	0.22***	0.07

*** significant at 99% confidence interval, ** significant at 95% confidence interval, * significant at 90% confidence interval

Moreover, the analysis of additional qualitative data suggests that mothers held overall positive views of the value of the girls' education. In some cases, mothers emphasized the importance of educating girls especially because of their future roles as mothers, to build the base of the family, better support her children's education and the community. Another reason was it was mostly the mother's role to support girls' education. During the FGDs with the CEC committees, teachers expressed more mixed opinions, often stating that girls had more limited access to education in past, however nowadays, due to the sensitization efforts, the situation is changing, and girls are getting enrolled at schools.

Lack of parents' sensitization and education was also commonly mentioned by both teachers, CEC, and mothers while asking about the reasons why girls would be less likely to attend schools. This is in line with the previous results when the HoH or caregiver was not exposed to any form of education were more likely to undervalue the importance of educating girls.

"If a parent is educated, they send their child to school whether they are boy or girl but if they are ignorant, they send only boys to school while they said girls stay at home to cook food."
(FGD Mothers)

When the household faced financial constraints, the preference for the education of boys rather than girls and the higher risk of early marriage were commonly mentioned challenges to the girls' education. However, both were often perceived as a function of the financial situation of the households, which was also reflected in the quantitative analysis.

"So when they cannot afford to send them all to school they will send the boys and tell the girls that they will end up in the kitchen later on, whilst the boy will benefit from it." (Mother FGD)

However, the highly gendered expectations of girls' behaviors seem to persist among mothers and teachers. Mothers frequently mentioned fears of pregnancy if girls would attend the schools, as previously discussed. Mothers often mentioned that girls and boys should be physically separated in schools. Also, it

seemed that girls are in general expected to be shy, quieter and fear to ask questions. However, in some cases, teachers reported girls were less shy or equally participating in the classes. Some respondents hold the same opinions, implying that girls should be treated differently in schools, compared to boys.

“Some of the others hold the same opinions, implying that girls should be treated differently in schools.” (Mother FGD)

Girls often expressed fears of being harassed in the toilets, fears that anyone, even boys could access girls’ toilets and lack of proper separation or doors. Moreover, girls seem to be more likely stigmatized in schools because of their menstrual needs. On a few occasions, teachers mentioned that girls are ashamed of asking teachers to go to toilets during the class or being seen to use the toilets in the school.

“Yes, of course, there are special toilets for the girls. It is difficult for girls to use the toilets since the teachers and boys are present in the school. Sometimes they ask the teacher to go outside as they can’t use the school toilets because they are ashamed (Teachers FGD)

Lastly, on a few occasions, CEC members mentioned that involvement of the religious leaders has a positive impact on the perceptions of the value of girls’ education, especially through directly talking to the parents or holding seminars for the girls.

“The religious scholars play a very significant role in our community. They inspire girls to go to school and they work with us. For instance, there is one religious scholar and he always sits down with the parents and he inspires them to bring their daughter into schools.” (CEC FGD)

More concrete and in debt analysis on how households and caregivers make their financial decisions and how they affect the attitudes towards girls’ education and their schooling opportunities would be necessary, especially to measure the change in the households’ sending on girls’ education and better understand economic vulnerabilities of these households.

The main findings of this section suggest that overall, caregivers are more likely to invest in girls’ education even if their funds are limited, compared to the number of caregivers who agreed that the cost of education is an acceptable reason to not to attend the school. Overall, caregivers in the female-headed households seemed to be more supportive while more vulnerable households, with limited access t water, medication, food or cash income, mere less likely to support girls’ education, if faced with financial constraints. In some instances, households with some exposure to informal education (measured as the minimum level of education achieved) were less likely to agree that financial costs are an acceptable reason to not to attend the school.

While pastoral households seemed to disagree with both statements. These results may indicate a degree of self-selection effect. As the pastoral households belong to nomadic or semi-nomadic communities, the barriers for girls from these communities attending schools are higher compared to other girls, which would explain the results for the first indicator. On the other hand, once these girls attend the schools, the parents are more determined to ensure that girls continue their education, despite the cost of education. This also suggests that girls from pastoral households in the schools in urbanized areas may be less marginalized compared to the girls from pastoral communities in rural areas.

Additionally, while the majority of respondents stated that they aspired for girls' university education, pastoralist, and HoH or caregivers with no education at all were less likely to aspire for higher education. In both instances, households are significantly more likely to be more disadvantaged, vulnerable and likely to come from the more marginalized environment, which would explain very low aspirations for girls' education.

Last but not the least, a higher proportion of the caregivers were more likely to prioritize housework over girls' education. This was especially the case for female-headed households when HoH or caregiver had no education at all and then households have no savings or when they owned land. On the other hand, pastoral households were significantly less likely to agree with the statement.

8. AUXILIARY ANALYSIS

8.1 MENSTRUAL HEALTH AND HYGIENE

A goal of the AGES intervention is to promote better menstrual hygiene management among girls, an outcome which falls outside the standard primary and intermediate outcomes incorporated into the broader Girls' Education Challenge programme. As mentioned in our discussion of attendance rates, menstruation is seen as a common challenge for girls' attendance, as girls often stay out of school during their period, either due to shame/embarrassment, a lack of appropriate menstrual hygiene products, for religious reasons, or due to the pain associated with periods.

The evaluation incorporated a survey module focused on menstrual health and hygiene that was directed at girls who had begun menstruating. Girls being interviewed by a female enumerator – or participating in interviews where a female team member was readily available to take over the interview at this stage – were asked whether they had reached the age of menstruation; for those who had, they were asked a series of follow-up questions regarding menstrual health practices (reuse of pads/rags, washing of pads/rags, and so forth), information sources regarding menstruation, and their experience of menstruation-related illness. Of the 1,420 girls in the sample, 54.7 percent (n = 777) were asked about their experience of menstruation, and 53.4 percent of those girls indicated they had experienced their first menstrual cycle (n = 415).¹⁸⁵ Our discussion below focuses on this smaller subset of girls, who answered the menstruation module questions.

According to this group of girls, menstruation may present less significant of a barrier to attendance than our other analysis suggests. Overall, 9.6 percent of girls feel they cannot attend school during their periods; while this represents a substantial share of girls, it is less dramatic a result than the qualitative data around attendance rates might suggest, while still a substantive barrier for this subgroup. In addition, many of the girls surveyed cited pain/discomfort as the reason they cannot attend school while menstruating – of the girls who felt they could not attend school, 45.0 percent cited a lack of sanitary pads, while 42.5 percent cited pain/discomfort. Failing to attend school while menstruating is a problem, regardless of the reason, but girls who cite discomfort are – in some sense – exercising choice over their attendance decisions, rather than being kept out of school due to embarrassment or a lack of sanitary products. At the same time, it is possible this question understates the extent to which girls miss school due to their period, as asking whether girls *can* attend school during their monthly period is not the same as asking them *whether* they attend school and how consistently they do so.

The focus of this section is on practices related to menstruation. However, to put these practices in context, it is necessary to understand where girls learn about menstruation and proper menstrual hygiene. The overwhelming majority (86.0 percent) of girls learn about menstrual hygiene from their mothers, and most of the remainder (another 8.9 percent) receive information on the topic from another female relative. Just 2.7 percent of girls receive their information from any other source, including those who

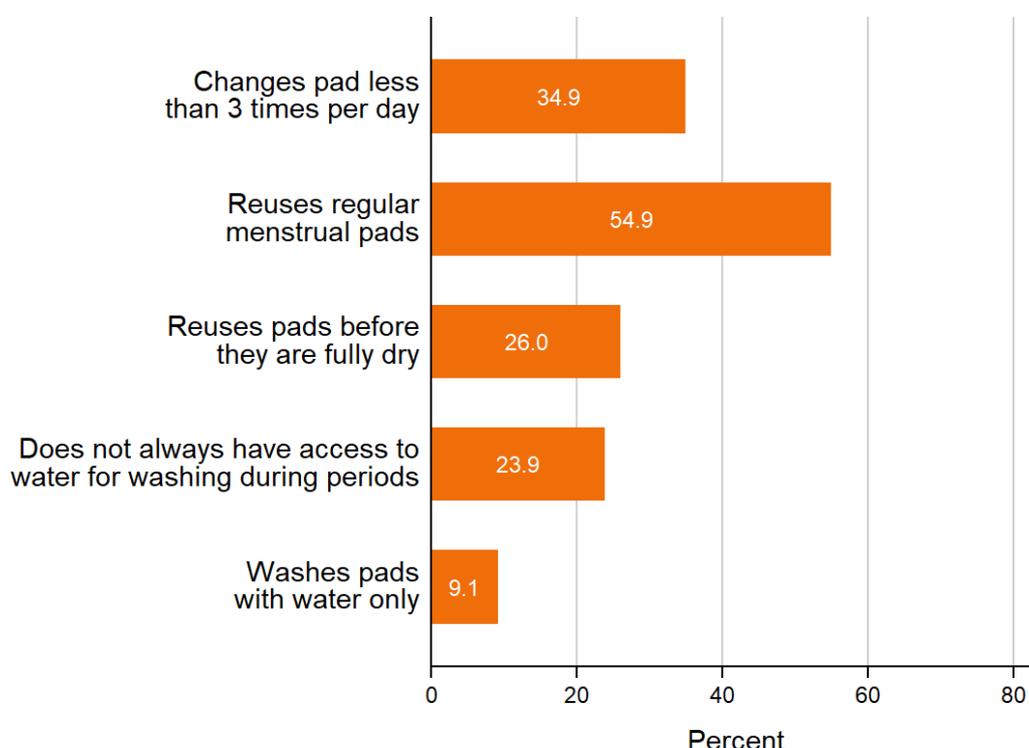
¹⁸⁵ The share of girls who indicated they had experienced their first menstrual cycle is relatively low, but not surprisingly so, given the relatively youth of the sample. Participation in the menstruation module is highly correlated with age, as one would expect.

learned about menstruation from a friend, a teacher, or a radio or media campaign. The remainder claimed they have not received any information on the topic.

We do not make firm judgments regarding the quality of information available from these disparate sources. For instance, many mothers may provide effective, accurate information to their daughters regarding menstrual hygiene management. However, this is far from guaranteed, especially given the relatively low education levels of most caregivers in our sample.

It is telling that many girls in the sample use unhygienic practices, because it suggests they may be receiving poor information on the topic. As shown in the figure below, a large share of girls reuse regular menstrual pads that are not designed for multiple uses, and many do not change their menstrual pads sufficiently often (fewer than 3 times per day, implying use of the same pad for 8 or more hours). Perhaps most problematic is the fact that 26.0 percent of girls are, at least occasionally, forced to reuse menstrual pads before they are fully dry, a practice which can cause vaginal infections. Along the same lines, 9.1 percent of girls who reuse pads or rags wash them with water alone, rather than soap or another cleansing agent.

FIGURE 20: MENSTRUAL HYGIENE PRACTICES



Age is somewhat correlated to the problematic practices described above. Younger girls are more likely to reuse pads or rags, and are less likely to have reliable access to water during menstruation. However, the sample of girls under the age of 15 who responded to these questions is quite limited (n = 53), so conclusions of this kind should be considered tentative.

Poor menstrual hygiene practices are important because they can have serious, deleterious health effects. Infections derived from poor hygiene can be a nuisance and source of embarrassment, but can also become life-threatening, cause long-term health problems and infertility. Among the girls interviewed, 74.9 percent reported experiencing a menstruation-related illness at some point since they began menstruating. The wording leaves the

nature of these illnesses open to interpretation, which may explain the high share of girls who report experiencing such illnesses. On the other hand, given the poor hygiene practices and lack of alternative information sources of the typical girl in the sample, the rate of menstruation-related illness should not be surprising.¹⁸⁶ In fact, poor menstrual hygiene practices are significant predictors of this class of illnesses.

To analyse the relationship between menstrual hygiene and the experience of menstruation-related illness, we estimated a series of linear (OLS) regression models predicting self-reported menstruation-related illness, as a function of individual binary variables that indicate whether a girl uses a given poor hygiene practice. In all cases, we control for age of the girl, because girls who have a longer history of menstruation have had more opportunity to experience a related illness.¹⁸⁷ To illustrate, consider the practice of reusing non-reusable sanitary pads/rags: we estimate a regression predicting experience of illness as a function of reusing non-reusable sanitary pads/rags and the respondent's age.

The results confirm our prior expectations regarding the importance of menstrual hygiene. In separate regressions, reuse of non-reusable sanitary pads/rags, reusing a pad/rag before it was fully dry, and having unreliable access to water during one's period were all significantly and positively associated with higher incidence of menstruation-related illness. Occasional reuse of pads when they were still wet – even pads intended for reuse -- was associated as 12.3 point increase in illness incidence, and unreliable access to water was associated with a 15.5 point increase in incidence.¹⁸⁸ Surprisingly, girls who change their pads less often, in general, and less often than conventional standards dictate, specifically, did not experience higher rates of menstruation-related illness. As noted above, however, it is possible that this population may experience high rates of type III FGM, which may account for severe pain during menstruation.

Due to the possibility that poor hygiene practices are correlated, we took the analysis further by estimating a model that incorporated multiple hygiene practices into a single model predicting illness incidence. In practice, hygiene practices are less correlated than we expected – most girls who reuse non-reusable pads,

¹⁸⁶ It is also important to note the impact of practices, such as female genital cutting (FGM/C), which is widespread in many Somali communities and has been shown to increase the rate of sexual and reproductive health problems (Farage, Miranda A., et al. 2015. "Female Genital Cutting: Confronting Cultural Challenges and Health Complications Across the Lifespan." *Women's Health* 11 (1): 79-94).

¹⁸⁷ In practice, age is not strongly correlated with reports of illness in our sample, but we assume this may be due to mis-reports. Note that age is controlled for in the regression model with a series of binary variables for each age group, rather than a single continuous age variable, as the former approach allows the age to have a fully non-linear effect on menstrual illness.

¹⁸⁸ All three variables noted were statistically significant at the 5 percent level, at minimum. For girls who reuse pads/rags, using water alone to wash them was also positively correlated with illness incidence, but this effect (a 2.8 point increase in likelihood) was not statistically distinguishable from a null result, likely due to the relatively small sample size (n = 328) available for analysis. The effect size, if accurate, is still substantively meaningful, but should not be relied upon, given the uncertainty associated with it.

for instance, do not report ever reusing pads before they are fully dry. Regardless, in a more fully-specified regression model, which also controlled for the age of the respondent, we found that the practices noted above – reuse of pads, reuse of pads before they fully dried, and inconsistent access to water during menstruation – were still associated with higher rates of illness incidence. Only the first and third practice were statistically significant at conventional levels ($p = .04$ and $p = .07$, respectively); however, the substantive magnitude of the relationship was not dramatically different from the simpler regressions, suggesting that the relationship is likely still valid, but would require a larger sample and/or more extensive analysis for confirmation.

In regression models that incorporate multiple independent variables that capture aspects of the same phenomenon – in this case, menstrual hygiene practices – it is common to calculate a combined index for use as a regressor, to study the impact of practices, writ large, on illness incidence. Rather than construct a menstrual hygiene index, we calculated the combined effect of all five practices captured in our model, and found that a test of their joint impact was statistically significant at the 1 percent level.

The share of girls in our sample impacted by menstruation-related illness is extremely high, and appears to be related – at least in part – to widespread poor menstrual hygiene practices. These circumstances are made worse by the fact that few girls feel they need to see a doctor or healthcare professional when they experience illnesses of this kind. As shown in the table below, which echoes the sources of menstruation-related information described earlier in this section, mothers and female relatives are the overwhelming destination for girls who experience menstruation-related illnesses. While mothers and female relatives are an acceptable avenue in some cases – for instance, if the issue in question is a simple infection and the mother has the requisite knowledge to understand the problem and procure the necessary antibiotics or antifungal medications – this is typically not the case, especially among the target population. Moreover, failure to adequately treat bacterial vaginosis or urinary tract infections can cause significant, life-threatening health complications. Relying exclusively on female family members for proper diagnosis and treatment of menstruation-related illnesses puts girls at risk insofar as illnesses with the potential for serious complications can go untreated.

TABLE 81: SOURCES OF MENSTRUATION RELATED INFORMATION

Type of source	
Mother	86.9%
Female relative	8.8%
Friends	2.0%
Does not get help	1.2%
Nurse/Doctor	0.5%
Other	0.5%
Health Volunteer	0.2%

8.2 INTERMEDIATE OUTCOMES AS PREDICTORS OF LEARNING

The previous section of the report looked at learning outcomes or different learner’s sub-groups and for various potential barriers for learning. This section of the report turns to the intermediate outcomes of the project, namely attendance, teacher quality, school management and institutional governance, girls’ self-esteem and empowerment, as well as community attitudes and behaviour. The specific focus here is to assess the theory of change assumptions linking the intermediate outcomes to learning and transition outcomes by studying whether improved intermediate outcomes are associated with increased learning

and higher rates of transition. This, however, can only be done to an extent at baseline. The reason is that the cohort girls recruited for the intervention have been selected specifically due to the fact that they were not enrolled before. As such, many of the factors have not had time to influence the wider outcomes of learning and transition. For example, the factors relating to school management and teaching quality naturally will not affect the learning outcomes of girls (enrolled in formal education) who have not in fact attended school yet at baseline. The analysis thus seeks to establish the veracity of the assumptions underpinning the theory of change in a more qualitative and theoretical fashion.

As in the previous section, we investigate the relationship between the intermediate outcomes and learning by looking at each variable in turn first in a bivariate regression and then in a multivariate regression where the control of respondent type or program track is used.¹⁸⁹ In addition, we study aggregate learning outcomes by combining numeracy and literacy scores into a single dependent variable which has the range from 0 to 200.¹⁹⁰ In this section we regress the intermediate outcome variables with the aggregate learning score as the dependent variable while controlling for the respondent type and calculating the standard errors accounting for the effect of a clustered sample.

Community attitudes and learning

One of the biggest challenges that a girl faces is society that often does not believe in girls' education. And because the girl can do more housework than the boy does, I can say that there is a lot of pressure on girls in the house. So, society still doesn't understand girls' education. – FGD Teachers

If her aunt takes care of the chores, she can go [to school].

And if she doesn't?

She can't go. Vignette qualitative exercise

Similarly to the various barriers and sub-groups analysed above, we looked at some of the intermediate outcomes that are assumed to affect learning. Of course, as above, we do not expect these variables to have much effect learning outcomes at baseline as the intervention is only beginning.

TABLE 82: ANALYSIS OF LEARNING SCORES AND COMMUNITY ATTITUDES

	Literacy			Numeracy			Coefficient after control	
	Mean	P value	n	Mean	P value	N	Lit.	Num.
Overall	44.3		1420	67.1		1420		

¹⁸⁹ Please refer to the discussion in the beginning of the previous section for the reasoning.

¹⁹⁰ More precisely stated, the dependent variable in this analysis is the aggregate of literacy and numeracy outcomes.

Girl spends a few hours or more per day doing HH chores	42.8	0.120	919	67.2	0.837	919	- 6.6** 0.009	
Caregiver aspires to university education for girl	45.2	0.023	1241	67.2	0.614	1241	9.0** 0.003	
Caregiver believes girls' education worthwhile, even if funds are limited	47.4	0.312	174	69.3	0.392	174		
Caregiver believes work or HH chores are acceptable reason to not attend school	41.1	0.075	569	66.7	0.706	569		
Caregiver believes cost of education is acceptable reason to not attend school	45.6	0.091	884	68.6	0.015	884		3.7** 0.025

We first looked at community attitudes and support for girl’s education by looking at caregiver’s responses to various questions. The results are presented in the table formatted similarly to the results in the previous section. Two variables are significant predictors of learning outcomes when controlling for respondent type. If girl has to spend much time with household chores, it is, as predicted, correlated with negative learning outcomes, in literacy. This is well supported by the qualitative interview data:

Well, girls have different treatments then boys. For example, if a girl is at home and has five brothers, they don’t count. When talking about housework and taking care of the parents, it’s all on the girl. She’s the one that cooks, cleans and shops. – FGD Teachers

Meanwhile, when caregiver aspires university education for her girl, the girl has overall a better reading baseline than others in the sample.

As for many of the sub-groups and barriers discussed before, the findings for numeracy outcomes are not always consistent. This time, those whose caregivers believe that the cost of education is an acceptable reason for a girl not to attend school in fact score slightly higher than others when controlling for program track. It is difficult to hypothesise what explains this contradictory finding. However, one possible explanation is that there are slightly more salespersons among those who support the statement. Being involved in a sales profession might make the caregiver both become more aware of cost in general as well as pass numeracy skills onto her child. Yet, when looking at the learning scores between those who have at least one parent working in sales compared to the rest, no systematic difference emerges.

Yet, all cohort girls are being selected and enrolled into the program due to the fact that they were not already going to school. As such, it can be assumed that either the attitudes or the material reality did not permit the girl to attend until now. The fact that a girl who spends a lot of time doing chores in the house has a lower literacy starting point is intuitive as it should indicate that the girl has had less idle time, to be used for learning or anything at all. Meanwhile, it can be assumed that the attitude-related factors captured here through the caregivers responses to question relating to the importance of girl’s education would

really only start to affect the learning outcomes as the program progresses where these attitudes may impact such intermediate factors as girl's ability to attend school and remain enrolled as well as time available to do homework and other learning outside of school. The ultimate test for these intermediate outcome indicators will thus happen at midline.

Yet, at least in logical terms, the focus on community attitudes seems very merited. Community and family support, or lack thereof was a key theme discussed in the qualitative interviews. The logic was that there are many more barriers for girls attending school including those that have to do with attitudes and concerns around safety and possible inappropriate behaviour outside of the home. Community members also discussed the fact that if the girl's household does not support her education, she will not have enough time to do homework and other learning that takes place outside of the immediate school hours:

Boys are better at school than girls because girls do a lot of housework that boys don't do. The boy has a lot of time to read or time to go and read with someone else. – FGD Teachers

The approach engaging religious scholars in the intervention is also seen by some community members as relevant as some linked the shift in overall community support for girls' education to religious scholars promoting girl's education.

Community attitudes and practices are also fundamental to ensure support for the enrolment, attendance and retention of girls belonging to marginalised groups. As noted in the above analysis, belonging to a marginalised group is a predictor of poorer learning outcomes; qualitative data highlights the negative beliefs of community members regarding the ability of marginalised children to enrol in school and attend classes. Results from another recent study¹⁹¹ carried out in south central Somalia indicate that children from occupational minorities may face discrimination and insults when attending school, again highlighting the need for shifts in community attitudes and practices to ensure regular attendance and a safe learning environment for girls from those groups. As previously noted in this report, negative attitudes and practices towards girls' education may also be more prevalent in marginalised girls' own households, particularly in relation to early marriage – potentially as a result of a combination of (i) historic lack of access to education affecting perceptions and (ii) economic need. Indeed, when looking at rates of marriage among cohort girls, 12.7 percent of girls from marginalized groups in the sample are mothers – in contrast with 5.3 percent of pastoral clan respondents and 3.7 percent of agricultural clan members. This implies that there is an intersectionality as it pertains to the challenges faced by girls from marginalised clans. They are on the one hand, more likely to face economic challenges. These girls can also encounter harassment and bullying at school. The fact that they are more likely to be married off at an early age further compounds to the challenges they have to overcome to enrol, attend and learn.

Teaching quality and learning

I think one of the common problems these schools face is having unqualified teachers. There should be expert and qualified teachers that produce excellent students in the future. If teachers

¹⁹¹ Consilient (2020) Education Sector Program Implementation Grant – Baseline Evaluation, p.36

are unqualified, then students would undoubtedly be unqualified with their subjects and this causes many to fail exams. – FGD CEC

The AGES ToC hypothesises that improved skill-specific teaching quality – i.e. addressing specific teaching skill gaps that are reflected in girls’ learning – will translate into improved learning outcomes for girls. For formal school girls, it is not possible at this stage to actually test the hypothesis through the learning outcomes as the intervention is in its early stages and the formal education cohort girls are all recently enrolled and as such the teaching quality related factors cannot have had any systematic effect on the learning outcomes for formal education students. However, for the ABE and NFE girls the intervention had attended class for some 4-5 months before the baseline. As such, the analysis in this section will be split in two, where we first show the results for the FE cohort and then for the ABE and NFE cohort separately.

When looking at the formal school students, only one variable has a statistically significant correlation with learning levels. Namely, girls identifying that teachers’ lesson pace was too fast for them in fact scored higher on both assessments. The explanation here could be that, like with perceptions of safety, those who are capable of admitting this are in fact on average more outspoken and critical than their counterparts. The quality that makes one admit to this, or identify this as an issue, could be positively correlated with learning capacities.

TABLE 83: ANALYSIS OF LEARNING OUTCOMES AND TEACHING QUALITY – FORMAL SCHOOL GIRLS

	Literacy			Numeracy		
	Mean	P value	n	Mean	P value	n
Overall mean	33.7		421	58.4		421
Agree a little/Disagree a little/lot: My teacher makes me feel welcome	34.8	0.818	57	51.9	0.105	57
Agree a little/lot: my teachers are often absent	33.0	0.802	125	56.9	0.479	125
My teacher rarely/never encourages participation	34.8	0.882	35	49.9	0.076	35
Teacher rarely/never explains how things learning are useful in our lives	46.4	0.139	9	70.9	0.156	9
My teachers’ lessons move too fast for me	40.1	0.008* *	177	64.0	0.008* *	177
Teacher punishes students who get things wrong in a lesson	34.3	0.520	331	58.5	0.815	331
Teacher used corporal punishment in last week	34.3	0.829	152	55.6	0.259	152
Agree little/lot: My teacher treats boys/girls differently in classroom	30.7	0.226	155	56.6	0.384	155

Meanwhile, for the ABE and NFE girls who have been exposed to teaching activities for some months, more variables relating to teaching quality are significant. Teacher absenteeism is correlated with significantly lower literacy levels. The same applies to those who said their teacher used corporal punishment. As with formal school students, the ones who viewed teaching pace as too quick were likely

to score higher in both assessments. This, again, is most a spurious measure that in fact captures critical thinking or some other omitted variable that contributes positively to learning.

TABLE 84: ANALYSIS OF LEARNING OUTCOMES AND TEACHING QUALITY – ABE AND NFE GIRLS

	Literacy			Numeracy		
	Mean	P value	n	Mean	P value	n
Overall mean	48.7		999	70.8		999
Agree a little/Disagree a little/lot: My teacher makes me feel welcome	43.3	0.198	119	69.4	0.583	119
Agree a little/lot: my teachers are often absent	41.1	0.007 **	281	67.6	0.105	281
My teacher rarely/never encourages participation	43.7	0.219	109	70.9	0.951	109
Teacher rarely/never explains how things learning are useful in our lives	46.1	0.724	20	75.6	0.365	20
My teachers' lessons move too fast for me	54.4	0.005 **	476	73.6	0.012 *	476
Teacher punishes students who get things wrong in a lesson	48.2	0.565	637	70.8	0.997	637
Teacher used corporal punishment in last week	40.4	0.020 *	147	67.7	0.168	147
Agree little/lot: My teacher treats boys/girls differently in classroom	46.8	0.380	375	69.2	0.276	375

School governance and learning

Yes, the monitoring is always good for the school. When there is a monitoring, the monitor will know the weakness of the school and this will ultimately help the school to improve. So monitoring is always good for the school. – FGD Teachers

The next set of results focuses on school management and institutional governance. In this regard we look at three factors, whether the school has a CEC; if the CEC has monitored attendance; and whether they have monitored teaching quality. The data presented comes from the school survey and as such it relates to formal schools only. As has been discussed elsewhere, these variables are not expected to be significant at baseline, given the limited time the program and school in general has had to affect the learning process of the girls. However, one of the variables is already significant, namely whether the CEC has monitored student attendance in the past year at the school or not. If the CEC has affected the early attendance of newly enrolled cohort girls, it is possible that learning outcomes will have been positively affected already.

TABLE 85: ANALYSIS OF LEARNING SCORES AND SCHOOL GOVERNANCE

	Literacy			Numeracy			Coefficient after control	
	Mean	P value	n	Mean	P value	n	Lit.	Num.
Overall mean	44.3		1420	67.1		1420		
School has a CEC	34.1	0.495	397	58.4	0.734	397		
CEC has monitored student attendance in last year	40.5	0.005	189	63.0	0.051	189	12.3* * 0.005	8.3* 0.051
CEC has monitored teaching quality in last year	37.2	0.111	217	60.0	0.434	217		

The qualitative interviews firmly link effective school management and oversight to learning and overall performance of the school. Teachers, when discussing what improves teaching quality, often mentioned having a clear curriculum and guidance from the ministry as one of the key pillars. They also welcomed the oversight by the CEC:

Because sometimes teachers can come and leave the class early, but when there is a monitoring, then teachers will try to be in the class till the time is finished and this will help the students to understand their lessons. – FGD Teachers

Monitoring is most useful for the teacher, because he says to himself: “you are coming to do your work the best way.” – FGD Teachers

Life skills and girls’ learning

One day a teacher was sick, and no one was teaching the class in that time. So, a girl came to me and said: “Sir, the teacher is not here today, and the students are chasing each other around class. So, I want to take the opportunity and teach the class”. I was really happy. I went to her class and saw how she was teaching, and I was very impressed. - FGD Teachers

The final intermediate outcome we assess is girls' leadership skills (defined as voice, vision, self-confidence, decision-making and organisation¹⁹²) and empowerment. Self-perceptions of leadership was captured through a series of over 20 questions. We utilize the Youth Leadership Index, created by aggregating results from all related questions.¹⁹³ This scale ranges from 0 to 84, with higher values representing stronger self-perceptions of leadership skills. Empowerment was captured with a single survey question, posed to female students: "who makes decisions about whether you will go to school"? Girls who indicate that they decide solely or that they decide jointly with their family are coded as exercising (partial or full) agency over schooling decisions. The results presented here use a standardised version of the YLI where the scale is further converted to one ranging from 0 to 100 in order to better illustrate the relationship between the index and the learning outcomes that range between 0 and 100.

The project ToC hypothesises that if girls improve their life skills in terms of leadership skills, self-confidence and self-efficacy, that they may perform better in school as a result of being more confident in their abilities and being able to participate more actively in the classroom. A straightforward test of this hypothesis is whether girls' learning outcomes are strongly correlated with their YLI scores. The panel of graphs below presents girls' literacy and numeracy scores as a function of their YLI scores, with the orange line on each graph indicating the slope of the fitted regression line. In the case of numeracy, the correlation with YLI score is positive and statistically significant.¹⁹⁴ In the case of literacy, the correlation is positive but is not statistically significant. Thus, the hypothesis about more leadership skills leading to higher academic performance finds a moderate level of support in the baseline data.

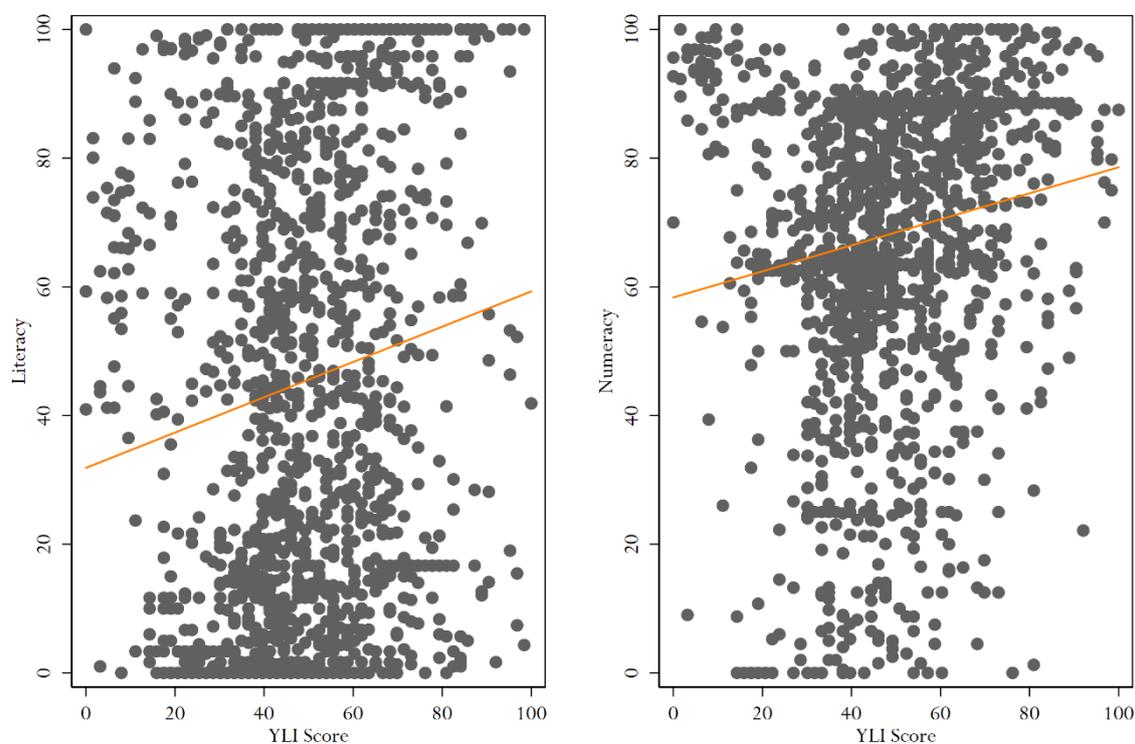
¹⁹² CARE (2014) Youth Leadership Index Toolkit, p.4

¹⁹³ See Section 5.2 for a more detailed description of how this index was created.

¹⁹⁴ YLI score and numeracy score are correlated at $p=0.009$ in a regression with cluster-robust standard errors. YLI score and literacy score are correlated at $p=0.393$ in a regression with cluster-robust standard errors.

FIGURE 21: YLI SCORES AND LEARNING OUTCOMES

YLI Scores and learning outcomes

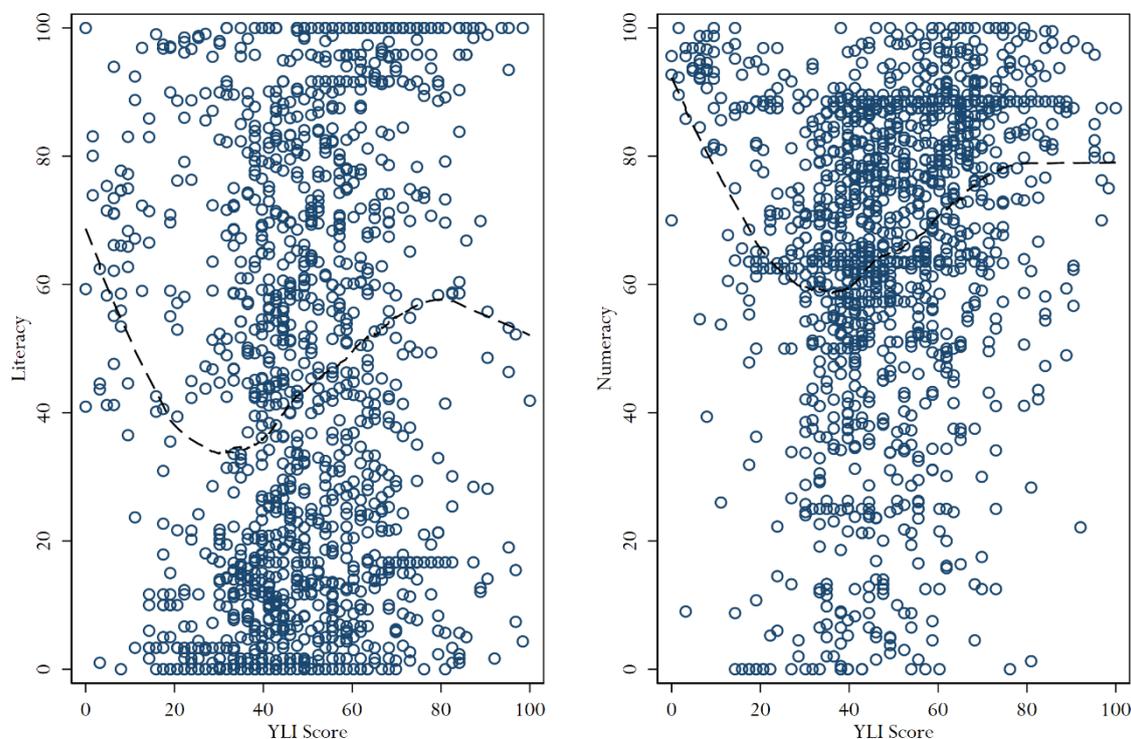


YLI turns out to be a good predictor of learning outcomes at baseline. Every one-point increase in the YLI score (from 0 to 100) is associated with a 0.2 percent increase in the numeracy score. The relationship is significant at the 99% confidence level. ($p=0.003$). Similarly, for literacy, every one-point increase in the YLI is associated with a 0.27 percent increase in the literacy score ($p=0.002$). Both relationships survive the control with respondent type without losing much of their effect or significance. As such, we can conclude that the theory of change as it pertains to life skills is appropriate. Yet, while the relationship between YLI and both learning measures seems to be positive and significant, it does not appear to be linear. The initial scatterplot below represents the relationship between YLI and literacy on the left and YLI and numeracy on the right while assuming a linear relationship.

When we abandon the assumption of linearity and plot the line using the locally weighted smoothing (LOWESS) curve, which plots a local regression of the relationship between numeracy and literacy at each level of literacy, the relationship looks quite different. The below graphs demonstrate the relationship. For both literacy and numeracy the line looks quite similar in that high learning scores in the low extreme of the YLI cause the line plotting the relationship to start high only to start receding as YLI scores increase. For the range where most of the cohort girls are, the score co-increase, which is to say that the YLI covaries well with the learning scores for most learners, but struggles to predict learning outcomes in its extremes, whereby extremely low or extremely high scores in the YLI do not automatically translate to corresponding levels in learning.

FIGURE 22: YLI SCORES AND LEARNING OUTCOMES - LOWESS

YLI Scores and learning outcomes - Lowess



Qualitative data also links questions of leadership skills (self-confidence, vision) and learning. Consider these girls in vignette exercises discussing the potential of one of the characters in the stories to continue learning:

If she is very confident, nothing can stop her from going to school.

Many things could happen to her. She will start thinking positively. She will start understanding the topics she didn't know before. She will learn new things. Then she will ultimately help community and herself too. – Vignette

Other community members also link factors relating leadership skills to education. Like the teacher who stated: “Boys usually show themselves because girls are shy. Shyness gets in their way of everything.” As such, it seems that CARE’s focus on life skills as measured through the YLI is well justified.

9. MEASUREMENT AND INDICATOR DEVELOPMENT

As part of the baseline evaluation, we undertook a review of the project’s stated approach to measuring impact over time. For instance, the program has selected indicators for each of the five intermediate outcomes, but these indicators may or may not be methodologically sound or capture critical outcomes from the program’s chosen interventions. Baseline evaluations also often reveal problems in the measurement of indicators, as they are applied to real-world data for the first time (at least in the context of a given project). In this section, we discuss the indicators themselves, their validity, their connection to the activities AGES is undertaking, and whether they could be fruitfully adjusted going forward.

Learning

The learning assessments, as currently deployed, are well-designed, likely as a result of CARE’s extensive experience implementing learning assessments in previous GEC and GEC-T projects. We would recommend making the assessments more difficult in the next evaluation round, while maintaining a subset of subtasks that are as precisely comparable as possible with the baseline version. There are two reasons for this recommendation. First, some girls have already achieved the upper bound of the existing assessment, and we expect ceiling effects to be more prominent in later evaluations. Second, while the assessment captures the most critical aspects of functional literacy and numeracy (e.g., simple addition and subtraction), girls enrolled in formal school for multiple years can be expected to reach a higher level of literacy and numeracy than these tests assess. Inclusion of multiplication and division, for instance, would make the assessments more in line with the skills we expect girls to be taught in formal school.¹⁹⁵

Transition

The transition indicator itself does not require revision. The current approach captures the varied transition pathways that are possible for girls in the Somali context.

Sustainability

Perhaps the largest gap in this baseline evaluation was our inability to measure most of the project’s indicators of sustainability. The project has selected 12 indicators – divided across the implementation and impact levels – with which to assess sustainability. As a result, we do not describe how we would suggest measuring each indicator, in part because the project may plan to measure these indicators from internal monitoring data or other sources. However, we strongly believe that the external evaluation can, in principle, shed light on sustainability, in addition to whatever internal measures are used. To illustrate, consider two sustainability metrics and how we would suggest measuring them:

- **Proportion of GEFs and BEFs implementing community actions to support attendance and retention** – this indicator could be measured by asking girls who identify as participating in GEF activities to select which activities their school’s GEF has undertaken.
- **Proportion of umbrella schools providing ABE** – this indicator likely cannot be credibly measured without expanding the data sources used in the evaluation. While the simplest

¹⁹⁵ This point may be less relevant for girls enrolled in non-formal education, in which case the harder subtasks could be applied only to girls enrolled in formal schools or ABE.

approach would be to assess this indicator through the use of qualitative interviews, an alternative approach would be randomly sample schools that are part of an umbrella association and call their head teachers to complete a very brief telephone-based survey. This would provide a more credible measure, at relatively minimal cost.

Attendance

This report relied very heavily on physical headcounts for assessing attendance rates in schools, ABE centres, and NFE centres. Headcounts are preferable in most ways to alternative metrics, including measures based on school attendance records, which are often inaccurate.

At the same time, we strongly recommend expanding the evaluation's repertoire of attendance measures, for the sake of triangulating information across multiple sources. Headcounts alone are limiting, insofar as they capture attendance on a single day out of an entire school year.

We recommend incorporating a measure based on school records. In the past, we have measured attendance over the entire school year thus far, but this is time-consuming during fieldwork and often subject to significant inaccuracy. Instead, we suggest measuring attendance rates over the previous one month, using school records. At the same time, AGES should encourage better record-keeping, both to facilitate this type of measure and to encourage good school management practices.

We also recommend incorporating survey-based measures of attendance, based on responses from both caregivers and girls. Our preferred approach would ask both caregivers and girls how many days of school they have attended in the previous two weeks. This time period is a necessary compromise to ensure adequate recall, as it is difficult for girls to recall how many school days they missed several weeks prior.

Teaching Quality

Our approach to measuring teaching quality sought to mirror the project's logframe indicator, which was somewhat vaguely stated as the "percentage of teachers/ ABE, life skills facilitators who demonstrate the use of appropriate methodologies for literacy/ numeracy teaching." Our approach captured aspects of gender-equity, use of corporal punishment, use of formative assessments, and specific pedagogical practices, the latter of which were primarily focused on encouraging a participatory classroom environment.

However, given the program's actual teacher training activities, these measures could be targeted more effectively. For instance, training of ABE facilitators related to child protection could be assessed by asking ABE teachers to describe the proper mechanisms for handling child protection reports (e.g., where should they be reported). Similarly, the project's training will seek to improve pedagogical practices around reading and numeracy teaching, but our current measurement strategy does not assess reading-specific pedagogy; rather, it measures general teaching practices only. We recommend re-designing the classroom observation tool to capture the use of specific teaching practices that should be commonly employed when teaching basic literacy.

Life Skills

The measure of life skills used in this evaluation is the Youth Leadership Index (YLI) developed by CARE for use across many education and girls' empowerment projects across multiple contexts. While the index includes 21 distinct questions, it primarily measures leadership and self-confidence of girls. There are two ways the measure of life skills could be improved.

First, it may be useful to disaggregate the YLI or decompose it into specific thematic areas. For instance, if the project's activities are designed to increase self-confidence, it may be possible to separate the YLI's constituent parts and identify a subset of questions that generally identify self-confidence in girls. If this is pursued, we would recommend that it

be undertaken as part of a broader effort to assess and improve the YLI.¹⁹⁶ The YLI has been employed widely across many contexts, and a large-scale meta-analysis of the index could reveal ways to improve it by removing specific questions that do not accurately separate high- and low-achieving girls, or by identifying multiple axes which are measured by the index.

Second, we would suggest incorporating a secondary indicator of life skills, designed to capture additional aspects of the life skills training AGES will employ. Per Section 1.3, above, life skills training will seek to develop girls' business skills and financial literacy. Rather than employing an entire additional assessment around financial literacy, it may be possible to assess girls' business skills in a simpler way. For example, asking girls whether they would feel confident in their ability to start a small business, or their confidence when engaging in market transactions, could provide information on both self-confidence and basic financial literacy. Alternatively, girls could be asked two or three simple questions asking them to calculate, as one example, profit after expenses. Our goal in this approach would be to provide a second metric of life skills without dramatically lengthening the existing household survey.

School Management

As noted in Section 7.4, our analysis of school management and governance only touched on the focus of the program tangentially. The program's goal is for CECs to promote enrolment by ultra-marginalised girls, and for the relevant educational authorities to use quality assurance procedures that prioritise inclusion. However, our measures of CEC activity levels do not specifically address inclusivity in a substantively satisfying way. We are able to determine how active CECs are in terms of promoting enrolment and encouraging retention, but neither measure is specific to girls who the project considers the most marginalised.

We suggest redesigning the head teacher and teacher surveys to better capture these aspects of CEC activity. Alternatively, surveying CEC members directly may be more productive. One option would be to ask head teachers targeted questions regarding CEC member outreach to individual marginalized groups, such as pastoralist families, minority groups, and so forth. If a survey with CEC members were conducted, it could be very brief and emphasise capturing what is necessary for this indicator by asking CEC members how many outreach visits they have completed to different *types* of girls, how much financial support has been provided to girls in marginalized groups versus other girls, etc. Assessing the use of inclusive quality assurance procedures by MOECHE officials is slightly more difficult, but could be achieved through a telephone survey of local MOECHE officials, as one option. Our suggestion is that – if the FM and CARE's Monitoring & Evaluation team wish to revise the specific operationalization of this indicator – a variety of creative approaches to measurement be considered.¹⁹⁷

Community Attitudes

The measurement of community attitudes is broadly sound. Currently, community attitudes are assessed by asking girls' caregivers what they consider to be acceptable reasons for a girl to miss school, and about

¹⁹⁶ We note that the YLI has been widely used and we expect CARE has, in the past, conducted extensive methodological analysis of the YLI. Therefore, our recommendation may be something that has already been completed.

¹⁹⁷ A telephone survey could also be implemented with CEC members, with phone numbers collected from head teachers during fieldwork.

their aspirations for their girls' education. These are appropriate metrics of caregiver attitudes, but this indicator would benefit from more extensive triangulation across data sources.

With regard to triangulation across data sources, we would encourage asking girls to assess attitudes toward girls' education in their communities. While girls' views may not be objective measures of community attitudes, their perceptions of how community members view them and the importance of their education is, in itself, of interest. Moreover, asking girls to assess community attitudes provides a lens on wider attitudes, rather than exclusively those of their caregiver. Finally, girls' responses are less likely to be subject to social desirability bias – whereas caregivers may wish to appear pro-education in front of an enumerator, when girls are asked about attitudes in their community, they have little incentive to misrepresent those attitudes.¹⁹⁸

¹⁹⁸ Triangulation using responses from surveys with head teachers or teachers may also be useful, for many of the same reasons. Responses from head teachers and teachers likely come with a bias – many teachers, who have dedicated a portion of their lives to education, may feel that their communities do not prioritize education adequately. At the same time, teachers and head teachers are well-informed observers when it comes to education, and are likely to be more attuned to community attitudes than other members of the same communities.

10. CONCLUSIONS AND RECOMMENDATIONS

10.1 CONCLUSIONS

Improved beneficiary targeting

Several beneficiary characteristics and barriers were identified as a possible indication of educational vulnerability, such as the fact of belonging to agricultural clans and simultaneously living in economically more insecure households or being a mother and facing a higher chores burden. These barriers – and the geographic distribution of these barriers across states – was recognised early on by the project, and incorporated into their Theory of Change and the design of their interventions. Although the cohort studied in this evaluation includes many girls who are not members of minority groups or marginalised clans, the project is effectively targeting marginalised girls in general. It is worth noting that the project has limited access to the countryside surrounding its project locations, where many members of the agricultural clans and minority groups actually live. Considering the rate of affect disabilities, agricultural clan members, minority group members, and economically vulnerable households in the sample, the project’s approach to marginalisation appears to be both well-targeted and context-sensitive.

Learning

Learning scores in the cohort are strongly correlated with age and -- by extension -- the learning institution in which girls are enrolled. NFE girls, who are the oldest, achieved the highest average scores, despite the fact that they have generally not been enrolled in any formal education in the past. Girls' different starting points should be factored into curriculum design for ABE and NFE centres, in particular, and the project should consider designing learning assessments that can capture a higher upper bound of learning scores, to accommodate NFE girls.

The learning assessments reveal breakpoints which seem to sharply distinguish low-achieving and higher-achieving girls. For instance, almost all girls performed well on single-digit addition tasks, but double-digit addition saw a significant drop in performance, implying that the procedures necessary to add 2-digit numbers (carrying values) is a skill that a girl either has or does not have. Similar patterns were observed for 1- and 2-digit subtraction, suggesting that learning gains could be catalysed by focus on arithmetic rules that are widely applicable. Curriculum development efforts should focus on identifying additional breakpoints, especially in the area of literacy.

The intermediate outcomes that underpin the AGES Theory of Change generally appear to be correlated with learning outcomes, suggesting that project interventions have strong potential to influence learning going forward. Of particular relevance at baseline is the existing association between leadership skills (YLI scores) and learning outcomes -- girls with a 10-point improvement in YLI score 2 points higher on numeracy and 2.7 points higher on Somali literacy, on average. While it is important to temper causal conclusions based on this finding, given the potential for reverse causation, the finding is suggestive, and in line with reasonable theoretical expectations.

Transition

Our analysis of transition outcomes relied heavily on the sample of benchmark girls and on qualitative data. Unfortunately, both have shortcomings: in the former, the sample is limited to women aged 20-24 years, which is less applicable to the younger girls in the cohort; in the latter, the qualitative data provided relatively rich insights into subgroup-specific issues around transition, but less concrete findings overall.

Transition rates among the sample of benchmark girls was extremely low. In fact, the vast majority of women (85.5 percent) were not enrolled in school, employed, or engaged in other productive activity. In terms of individual-level factors that influence enrolment and transition, household characteristics – especially the educational background of one’s parents – impinged somewhat on the likelihood of being enrolled among this older cohort. Beyond that, the most consistent themes regarding transition were the impacts of marriage and motherhood, the barriers encountered by girls in pastoralist households, the financial strain education places on households and the impact that financial strain has on enrolment decisions, and the reliance of households on girl-children for domestic work. These dynamics highlight the extent to which the project’s emphasis on inclusivity is well-placed, but may require directing greater financial resources toward CECs and communities than is currently planned.

Attendance

The overall attendance based on the headcounts accounted for 86% present of children present in classes. The attendance was high across all cohort groups, with the highest rates observed in the formal schools for both, girls and boys, while the lowest values of 79.6% were recorded for the ABE groups.

Higher attendance rates were observed when girls were more likely to come from pastoral households, more girls lived in households with phones, facilitating communication with teachers, schools had reliable electricity and separate toilets for girls and boys and a functioning CEC. On the other hand, the main predictors of negative attendance rates were observed when girls were more likely to be single orphans and or to come from households with land ownership.

Community attitudes

We have first analysed parental willingness to invest in girls’ education, despite the financial constraints of the households. Overall, 62% of caregivers agreed that the cost of education is an acceptable reason for girls to not to attend schools, while nearly 87% agreed that it is worthwhile to invest in girls’ education, even if the funds are limited. We have identified several main predictors of parental willingness to support the educational costs of girls, such as female-headed households, pastoral households and when the caregiver or the HoH had some exposure to informal education. On the other hand, caregivers in less financially secure households and food-insecure households were more likely to prioritize other types of expenses over girls’ education.

During the analysis of caregivers’ aspirations for girls’ education and their attitudes towards the girls’ roles within the households, we found that while 87% of caregivers aspired for girls’ university education, 40% of caregivers agreed that housework is an acceptable reason to not attend school. Caregivers were more likely to prioritize housework in economically more insecure households, when HoH did not receive any education and in female-headed households. In pastoral households and when HoH or caregiver received at least informal education, caregivers were significantly less likely to prioritize housework over girls’ education.

The main qualitative findings suggest that poorer household are more likely to prioritize boys’ education over girls’ education, especially in combination with higher housework burden. Moreover, the prospects of being married and the inability to attend school during the menstrual cycle were seen as other major barriers to girls’ education. Girls from nomadic communities were said to face higher barriers to learning, compared to other students.

School Management and Governance

School management in the AGES context is focused on two main actors: the government, the federal-level MOECHE and the state-level educational ministries, on one hand, and Community Education Committees (CECs) on the other. At all levels, the goal of the program is to promote inclusivity of educational opportunities. CECs overseeing the schools in the sample are almost universally functional.

However, the extent to which they actively encourage enrolment and retention is more limited, with just 60.5 percent of CECs promoting the enrolment of out-of-school children, and even fewer following up on cases where students dropped out.

More problematically, these metrics are not specific to marginalised children, and do not indicate whether special effort is made to reach out to the families of marginalised children. In qualitative interviews, CEC members report taking steps to specifically assist minority groups, pastoralist families and – in some cases – girls with disabilities. However, given the relatively low level of activity of many CECs, and the fact that assisting marginalised girls is often more resource-intensive than promoting enrolment more generally, our interpretation is that very few CECs are actively and systematically engaged in the kind of outreach envisioned by the program.

Self-perceptions of leadership skills

The main findings indicate that the girls' self-perceptions of leadership skills are low, with only 12% of girls, who scored 70% or higher in the YLI. Girls were more likely to have stronger self-perceptions of leadership skills when their caregivers received informal education, in households under higher financial stress and then caregivers held positive attitudes towards girls' education. Lower self-esteem was negatively correlated with the inadequate learning environment, especially when teachers did not encourage participation in class when girls could not access learning materials, such as textbooks or then girls reported they witnessed physical punishment during the past week.

The main findings from the qualitative data suggested that the mother and daughter relationship was especially important for girls' self-perceptions of leadership skills. Respondents often suggested that mothers played an important role in helping girls to express their opinions. Teachers were also identified as important stakeholders in the development of girls' self-perceptions of leadership skills, especially through encouraging girls to express their opinion and pursue higher education.

Teaching Quality

Students enrolled in project schools and centres face a number of barriers to learning stemming from the practices of their teachers. As a general rule, participation is actively encouraged in only about 62 percent of classes, and teachers do not use group work or games as effectively as they could. Beyond participation, many students feel lessons move too fast, and few teachers use the formative assessments that would allow them to detect that they are moving too fast for their students' current levels of understanding. While there are significant regional differences in specific teaching practices, such as the case of physical punishment, there are not widespread, systematic differences across regions in overall teaching quality.

Regressive attitudes toward both girls' education and corporal punishment are widespread among the teachers corps. Some of the data sources that identify gender bias among teachers are not dispositive – for instance, teachers may direct questions specifically to one gender for a number of reasons aside from bias or gender exclusivity. However, other indicators paint a stark picture: 28.1 percent of teachers believe that it is more culturally acceptable for boys to attend school than girls, and 21.6 percent believe it is more important for boys to complete secondary school than girls. Findings regarding corporal punishment are even more compelling, in 38.6 percent of classrooms observed during fieldwork, female students were physically punished (39.2 percent involved physical punishment applied to male students). And, when teachers were asked directly about use of physical punishment, 63.3 percent indicated that students – whether boys, girls, or both – should be punished physically.

10.2 EVALUATION RECOMMENDATIONS

- The learning assessment scores calculated in this baseline suggest that ceiling effects are already affecting girls' scores, and this problem will become more pronounced in the midline and endline evaluation rounds. As part of the midline evaluation, a more difficult set of subtasks should be added to both the Somali literacy and numeracy assessments, to reduce the impact of ceiling effects going forward. Two considerations are important to note: first, the midline and endline assessments should be carefully designed to include a set of subtasks that are *comparable* to the full baseline assessment, and the comparability of the baseline assessment to the comparable portion of future assessments should be rigorously and empirically evaluated.¹⁹⁹ Given the pre-post design of the assessment, changes in the difficulty of the assessment from one round to another have much more severe consequences for the rigor of the evaluation than in evaluations with a comparison group. Second, the more difficult midline subtasks need to be sufficiently difficult that ceiling effects at the endline will be avoided without making further adjustments in difficulty between the midline and endline. If the midline assessment is still subject to moderate ceiling effects, correcting the problem at endline will not be possible – even if the endline assessment is made more difficult, the analysis will rely exclusively on subtasks of comparable difficulty between midline and endline, eliminating entirely the benefit of making the endline assessment more difficult. In short, extreme care should be taken at midline to ensure comparability to the baseline and forward-looking levels of difficulty for midline-to-endline comparisons.
- Comparability between learning assessments implemented in different rounds is of paramount importance, given the pre-post nature of the evaluation design. Put frankly, even minor differences (e.g., 1-2 percentage points) in difficulty from one round to the next present problems of bias that cannot be resolved in the pre-post design. Even more problematically, the direction of this bias is not known and cannot be accounted for with the comparison testing typically implemented as part of GEC and GEC-T evaluations.²⁰⁰ Given the potential implications, we recommend designing the midline and endline learning assessments proactively prior to the midline evaluation round. With both assessments prepared for the midline, respondents can be randomly assigned to one of the two assessments at midline, and assigned to

¹⁹⁹ We suggest formally testing comparability in advance of the midline evaluation, on a fairly large sample of students.

²⁰⁰ The standard method employed in GEC evaluations involves administering both assessments to a set of 20-100 students and comparing their mean scores on the two assessments. If the two means are statistically similar, the two assessments are considered comparable. However, this procedure is not valid as a test of comparability, because a difference-of-means (or other available tests, such as a non-parametric rank-sum test) that fails to reject the null hypothesis of a difference between the two tests is emphatically *not* confirmation that the two tests are of equivalent difficulty. Such tests might have confidence intervals around the estimate of difference of anywhere from 2 to 10 percentage points, meaning that large true differences in difficulty cannot be detected or ruled out in such a small sample size. Even in a comparison procedure with a very large sample, the small degree of uncertainty that would remain will introduce potential bias into analysis of changes in learning scores over time.

take the other assessment at endline. Random assignment of this kind will ensure that, on average, any difference in difficulty between the two assessments is equally distributed across the midline and endline evaluation rounds, removing one critical source of potential bias.

- In general, additional effort should be dedicated to tool design prior to the midline evaluation, in an effort to better capture all of the indicators of interest, as outlined previously in Section 9. This recommendation applies to both intermediate outcomes and sustainability indicators; in addition, effort should be made to ensure indicators can be properly and extensively triangulated across multiple survey questions, data sources, and data types. For instance, attendance rates in this baseline were assessed exclusively on the basis of physical headcounts. Headcounts are generally preferable to the use of school attendance records – which are often incomplete and inaccurate – but are subject to their own limitations, such as the fact that they are focused on attendance during a single day. Data on attendance can and should be triangulated through questions directed at girls or their caregivers about their attendance and, potentially, via the teacher survey, in addition to qualitative interviews. The midline evaluation should include a focus on improving tool design, to better assess baseline-to-midline impact and to prepare for a more rigorous assessment of midline-to-endline impact going forward.
- As part of the midline and endline evaluations, the training curriculum and schedule should be revamped to provide additional time and focus on proper administration of the learning assessments. Most or all of the field team members utilized during the baseline have extensive experience collecting quantitative survey data, including with ODK Collect specifically. However, learning assessments like EGRA and EGMA are less commonly used than typical quantitative questionnaires, and many enumerators may not have prior experience to fall back upon when administering them. Moreover, some aspects of assessment administration are admittedly difficult, especially timed exercises that involve counting the number of words a student read correctly. Training time should be divided such that team leaders can be given targeted training on the data collection tools they alone will complete, and their team members can be given broader instruction and more practice time for the learning assessments. In addition, the evaluation team should ensure that team leaders accompany their team members for multiple interviews during the first week of fieldwork; while this was done during the baseline, it is an essential quality assurance practice, and especially important in the context of complex survey instruments.

10.3 PROGRAMMING RECOMMENDATIONS

Our recommendations below are organized in terms of intermediate outcome and sustainability outcomes. We do not include programmatic recommendations focused specifically on learning or transition outcomes, because the intermediate outcomes below are – according to the project's Theory of Change – the link between program outputs or activities, and primary outcomes like learning and transition. For instance, where our recommendations concern teaching quality or community attitudes, we assume, in line with the Theory of Change, that these will have downstream impacts on learning and transition, but we focus on their more proximate effects.

Improved attendance

- The analysis of the attendance rates points at the fact that single orphaned girls are significantly more likely to have lower attendance compared to other cohort girls. Project activities should include these girls as a target group and provide them with additional support, such as scholarships. Also, better insight into the household decision-making processes of single orphaned girls would allow future programming to identify and address barriers preventing girls from attending schools.
- Toilet facilities for girls in the schools should be separated from boy's toilets and hygiene and cleanliness should be better monitored. Moreover, school management should be consulted on possible solutions on how to ensure the security of girls' toilets and make toilets easier to access for girls. Our findings suggest that girls are more likely to miss classes as they prefer to use toilets outside their schools. Girls often felt less secure and comfortable using toilets at schools, especially during their menstrual cycle.
- The household burdens were identified as a significant factor decreasing girls attendance rates. Our findings also suggest that mothers are the main decision-makers concerning the housework responsibilities delegated to girls. Further engagement of mothers to identify the best approaches to reduce housework burden for girls, would improve their attendance. In addition, working with male household members and further involvement of school going boys should be a priority group for project interventions addressing change in the social norms, gender expectations and division of household roles to reduce housework burden for school going girls
- Our findings suggest that girls living in households with land ownerships are significantly less likely to attend schools, especially during the rainy seasons. Thus, seasonality and agricultural practices should be further considered when designing interventions for girls engaged in labour, such as scholarships. Moreover, CEC should be involved while designing interventions, especially due to their knowledge of local context and social interactions.
- Parental involvement in the monitoring of girls' attendance and direct contact with CEC seems to be particularly important for sustainable and positive change in girls' attendance. Further support of the CEC activities and more participatory design of approaches to engage parents would contribute to the development of better-targeted strategies to improve attendance.
- Project activities should also focus on working with teachers and school personnel to improve the quality of school and class attendance records and to further sensitize them on implications of poor record keeping for school attendance and project activities.

- Support for girls with disabilities should additionally target barriers to attending schools, such as inadequate transport options, lack of equipment addressing specific needs of girls and lack of financial resources of their households to secure girls' access to schools and adequate care.
- Lower quality of attendance records for ABE centres should be addressed by the future project activities, especially by improved monitoring systems, sensitization of teachers or relevant personnel and parents' or caretakers' involvement in the monitoring of girls' attendance.

Improved life skills

- Our findings suggest that girls with disabilities are more likely to have significantly lower YLI scores, compared to other girls. To ensure that girls with disabilities have equal chance to participate in Girls' Empowerment Forums (GEFs), GEFs should have sufficient capacity to accommodate the specific needs of these girls.
- Mothers' support for girls expressing their opinions and gaining self-confidence was found to be especially important for girls' self-esteem. Thus, future programming should explore ways to further involve mothers or caregivers in activities aimed at improving girls' self-esteem.
- Our findings suggest that girls feel especially uncomfortable in schools during their menstrual cycles. Improved agency and capacity to advocate for own needs in the school environment should be incorporated in the SHR sessions of the GEFs. These sessions should aim to create a safe space for girls to discuss their experiences and strategies on how to deal with their menstrual cycle in a school environment.

Positive change in community attitudes and practices

- Our finding suggests that low perceptions of the value of girls' education and preferences for educating girls over boys are mostly a function of economic vulnerability and HoH and caregivers' exposure formal or informal education. Project activities should apply a more nuanced approach to the provision of financial support to households of marginalized girls and prioritize among the most disadvantaged girls, especially, as caregivers in these households are significantly more likely to invest financial resources to household needs rather than girls' education.
- Community attitudes, as discussed in this report and in most GEC and GEC-T evaluations, focus on the preferences and attitudes of caregivers or community members toward education, in general. But an additional facet of community attitudes will be especially pertinent to encouraging enrolment and retention of the marginalized girls targeted by AGES: attitudes toward marginalized groups, specifically, and their education. Girls from minority groups and girls with disabilities face barriers in the form of discrimination from teachers and bullying from other students. These factors risk undermining efforts to encourage enrolment or improve the financial viability of schooling, and need to be addressed alongside more tangible efforts such as these.

Teaching Quality and Practices

- While instilling better teaching approaches is essential to the project's Theory of Change and its success in producing positive learning outcomes, learning new pedagogical practices is a time- and energy-intensive task for both teachers and project staff. At the same time, most teachers in project learning centers would benefit from smaller interventions focused on attitudinal and behavioral change, especially around gender equity and corporal punishment. For teachers not included in broader training programs, the project could hold 2-hour seminars promoting gender-equitable attitudes among teachers by highlighting the role of prominent female Somali public figures, emphasizing the value for child development of educated mothers, and describing religious teachings that are

consistent with girls' education. Similar efforts could change, at the margins, attitudes toward corporal punishment. More frequent efforts could be led by the GEFs and BEFs, sponsoring school-wide assemblies that focus on gender equity and the value of educating girls. While these efforts are unlikely to be transformational, especially in the short- and medium-term, they are relatively low-cost and can improve teaching outcomes in a context where the baseline of teacher attitudes toward both topics are so poor to start.

- Teacher absenteeism is rampant, at least according to student reports, though the length of exposure of most girls to schooling should be taken into consideration. Improving data collection around teacher attendance during the midline evaluation will be useful – especially if it allows triangulation of girls' opinions with more objective and systematic data – but teacher attendance should also be tracked as part of the project's own monitoring efforts. Moreover, emphasizing teacher attendance during training sessions and tying any teacher incentives – including participation in the training program – to observed levels of teacher attendance during school-days would be a fairly simple method for reducing absenteeism among the teaching staff.
- The program intends to promote teachers' use of formative assessments. These efforts should be coupled with practical illustrations of their value, built around findings from this report and from additional research in individual classrooms. For instance, lessons that move too fast are cited as a common problem by girls, and this is something that even intermittent use of formative assessments would reveal to a teacher. Additional, class-specific research could build upon findings in this report that show skills gaps at particular points in addition and subtraction (e.g., 1-digit proficiency versus 2-digit difficulties). If teachers can be shown how to identify these skill gaps in their own classrooms in a way that doesn't require a significant investment of time, it could illustrate the true power of formative assessments, which is that they allow teachers to identify issues in their own classrooms, rather than at some less practical aggregate level, and that they allow teachers to identify students with particular knowledge or skill shortcomings.

School Management and Governance

- Many of the CEC members interviewed cited tangible actions their committees take to encourage enrolment by marginalized girls. However, these interviews were conducted among the most visibly active CECs, and the actual rate of CEC engagement in school management, enrolment promotion, and monitoring, is much lower than the qualitative data suggests. Many CECs engage in only superficial and infrequent monitoring, and relatively few provide specific financial or in-kind support to their schools. Frequent and deeper engagement is needed by CECs in general and, especially, in the realm of marginalized girls. CECs should be given practical ideas – such as hosting community forums, or engaging with local religious leaders – to promote enrolment in ways that do not require scarce committee resources. Further, committees should be assisted in fundraising from their communities to provide partial scholarships or learning materials for marginalized girls. Given the urban areas in which AGES is being implemented, there may be greater capacity for community fundraising than in more rural, more economically vulnerable areas.
- Efforts to encourage enrolment and retention for girls with disabilities should focus on actions that provide the greatest benefit at manageable cost levels. Retrofitting all schools to be accessible to all girls with disabilities is not financially feasible. But some barriers to attendance or enrolment are more manageable, such as facilitating or supporting transport to school. GEFs could be mobilized in support of girls with disabilities by, for instance, assigning a different girl to help her walk to school each day during the week, such that the time burden does not fall exclusively on a single friend.
- The most prevalent form of disability among cohort girls appears to be mental health difficulties. While this may stem partially from measurement issues, it is also partially consistent with the conflict-affected

environment in which cohort girls have been raised. According to the project's documentation, GEFs will be a venue that provides peer support for girls with disabilities. These efforts should also target mental health difficulties directly, as this may have the broadest application in the population. Mental health support via small peer groups, perhaps with older girls mentoring and informally counselling younger girls, could make a moderate impact at minimal opportunity cost.

ANNEXES

ANNEX 1 - PROJECT LOGFRAME

The project's logframe, with values entered for the baseline evaluation round, is attached as a separate annex.

ANNEX 2 - COHORT EVALUATION APPROACH

Project to complete

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

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

All cohorts will be evaluated through an external evaluation.

ANNEX 3 - BENEFICIARIES TABLE FROM SAMPLE DATA

The table below indicates summary statistics for the overall sample of cohort girls and for distinct beneficiary or respondent types. Information for the subgroups 'Difficult to afford for a girl to go to school' and 'Didn't get support to stay in education and do well' were not collected. We classified households as owning land if the household owned their own land, owned land jointly with others, or owned both types of land. In practice, joint land ownership was relatively uncommon, with just 5.5 percent of households owning land exclusively with other households; we focus on the distinction between any land ownership and entire landlessness, because complete landlessness is a stronger marker of poverty or occupational differences than sole versus joint land ownership.

Beyond land ownership, it is useful to clarify how other variables below were coded. When the roof of the house where the beneficiary girls lived was made from mud, thatch, cardboard, tarp or plastic, we categorized the household as having a poor-quality roof. We considered households unable to meet their needs when the caregiver described the household's situation as "unable to meet basic needs without charity". The subgroup "gone to sleep hungry for many days in the past year" was the case then the household reported its members went to sleep hungry many or most days during the past 12 months. We added a second category, which captures a more liberal coding of household food insecurity or hunger: if a caregiver reported that members of their household had gone to sleep hungry "just one or two days" or more in the last 12 months.

Finally, we also included subgroups defined by barriers that specific girls face to educational achievement. We identified girls who speak a language other than language of instruction used in their school by asking caregivers which language a girl spoke at home, and asking head teachers to identify the language used in their schools. Unfortunately, the survey with head teachers did not distinguish between schools that used

the two different dialects of Somali, af-Maay and af-Mahatiri.²⁰¹ Therefore, we reviewed the school list with fieldwork team leaders who had visited different locations, identifying the areas where instruction was primarily in af-Maay versus af-Mahatiri (among those schools using Somali, instead of English). The subgroup identified in the table below is the set of formal school girls who speak a language at home that differs from the language of instruction. Finally, we identified girls who have a high chore burden as those whose caregivers reported they typically spend a whole day or half a day or several hours doing chores on a normal day.

**TABLE 86: CHARACTERISTIC SUBGROUPS AND BARRIERS OF SAMPLE FOR PORTFOLIO LEVEL
AGGREGATION AND ANALYSIS**

Characteristic/Barrier	FE Girls	ABE Girls	NFE Girls	Full Sample
Single orphans	10.7%	10.1%	17.5%	13.0%
Double orphans	0.7%	1.2%	1.0%	1.0%
Living without both parents	3.6%	5.0%	8.5%	5.8%
Living in female headed household	34.9%	33.7%	42.7%	37.3%
Married	0.2%	0.8%	9.3%	3.7%
Mother under 18	0.0%	0.4%	1.6%	0.7%
Mother under 16	0.0%	0.0%	0.0%	0.0%
Difficult to afford for girl to go to school	N/A	N/A	N/A	N/A
Household doesn't own land for themselves	38.0%	31.4%	29.1%	32.5%
Material of the roof (mud, thatch, cardboard and plastic/tarp)	18.1%	22.1%	19.6%	20.0%
Household unable to meet basic needs	39.0%	48.1%	47.6%	45.2%
Gone to sleep hungry for many days in past year	25.9%	30.6%	24.9%	7.0%
Gone to sleep hungry at least 1-2 days in past year	78.4%	81.6%	78.6%	79.6%
LOI different from mother tongue	22.6%	N/A	N/A	22.6%
HoH has no education at all	22.1%	22.1%	19.8%	21.3%
HoH has no formal education	80.8%	82.9%	81.7%	81.8%
Primary caregiver has no education at all	29.0%	27.9%	24.5%	27.0%
Primary caregiver has no formal education	85.7%	86.0%	83.9%	85.1%
Didn't get support to stay in education and do well (%)	N/A	N/A	N/A	N/A

²⁰¹ The same questions posed to girls and their caregivers made a clear distinction between the two dialects.

Sufficient time to study: High chore burden (whole day, half day or quarter day/few hours)	57.7%	63.4%	71.7%	64.7%
---	-------	-------	-------	-------

ANNEX 4 - BENEFICIARIES TABLE FROM PROJECT MAPPING DATA

TABLE 87: DIRECT BENEFICIARIES, BY AGE

Age	Proportion of cohort 1 direct beneficiaries	Data source
Aged 10	11.80%	Project data collected during the identification and enrolment process of the beneficiaries
Aged 11	8.20%	Project data collected during the identification and enrolment process of the beneficiaries
Aged 12	12.36%	Project data collected during the identification and enrolment process of the beneficiaries
Aged 13	15.40%	Data collected at the start of the project during the identification and enrolment process of the beneficiaries
Aged 14	8.97%	Data collected at the start of the project during the identification and enrolment process of the beneficiaries
Aged 15	5.98%	Data collected at the start of the project during the identification and enrolment process of the beneficiaries
Aged 16	5.03%	Data collected at the start of the project during the identification and enrolment process of the beneficiaries
Aged 17	10.56%	Data collected at the start of the project during the identification and enrolment process of the beneficiaries
Aged 18	9.43%	Data collected at the start of the project during the identification and enrolment process of the beneficiaries
Aged 19	12.27%	Data collected at the start of the project during the identification and enrolment process of the beneficiaries
N = 20468		

TABLE 88: TARGET GROUPS BY OUT-OF-SCHOOL STATUS

Status	Proportion of cohort 1 direct beneficiaries	Data source
E.g. Never been to formal school	96.12%	Data collected at the start of the project during the identification and enrolment process of the beneficiaries

E.g. Been to formal school, but dropped out	3.88%	Data collected at the start of the project during the identification and enrolment process of the beneficiaries
E.g. Enrolled in formal school	32.37%	Data collected at the start of the project during the identification and enrolment process of the beneficiaries
N = 20468		

TABLE 89: DIRECT BENEFICIARIES BY DROPOUT GRADE

Level of schooling before dropping out	Proportion of cohort 1 direct beneficiaries	Data source
Grade 1	31.86%	Data collected at the start of the project during the identification and enrolment process of the beneficiaries
Grade 2	68.14%	Data collected at the start of the project during the identification and enrolment process of the beneficiaries
N = 794		

TABLE 90: OTHER SELECTION CRITERIA

Selection criteria	Proportion of cohort 1 direct beneficiaries (%)	Data source – Project monitoring data, data from sample used in external evaluation or assumption?
Married	5.48%	Data collected at the start of the project during the identification and enrolment process of the beneficiaries
Divorced	5.29%	Data collected at the start of the project during the identification and enrolment process of the beneficiaries
Have children	7.42%	Data collected at the start of the project during the identification and enrolment process of the beneficiaries
From minority clans	19.03%	Data collected at the start of the project during the identification and enrolment process of the beneficiaries
Engaged in some casual labour	3.88%	Data collected at the start of the project during the identification and enrolment process of the beneficiaries

Orphaned (have lost either one or both parents)	13.74%	Data collected at the start of the project during the identification and enrolment process of the beneficiaries
Double orphaned	1.01%	Data collected at the start of the project during the identification and enrolment process of the beneficiaries
N = 20468		

TABLE 91: OTHER BENEFICIARIES

Beneficiary type	Total project number for cohort 1	Total number by end of project	Comments	Data source
Learning beneficiaries (boys) – as above, but specifically counting boys who will get the same exposure and therefore be expected to also achieve learning gains, if applicable.	E.g. 1000	E.g. 3000	E.g. 3 cohorts – aiming for 1000 per cohort.	Learning beneficiaries (boys) – as above, but specifically counting boys who will get the same exposure and therefore be expected to also achieve learning gains, if applicable.
Broader student beneficiaries (boys) – boys who will benefit from the interventions in a less direct way, and therefore may benefit from aspects such as attitudinal change, etc. but not necessarily achieve improvements in learning outcomes.	14,076	14,076. This will remain same as cohort 1 as this is the total population of boys enrolled in the target schools	These are the total number of boys enrolled in schools where the project has enrolled girls for formal education. These boys will benefit from the activities of the Boys Empowerment Forums that will be established in those schools. Such benefits include improved leadership skills and positive attitude towards girls' education.	Broader student beneficiaries (boys) – boys who will benefit from the interventions in a less direct way, and therefore may benefit from aspects such as attitudinal change, etc. but not necessarily achieve improvements in learning outcomes.
Broader student beneficiaries (girls) – girls who will benefit from the interventions	11,701	11,701 (This is estimated to remain the same during cohort 2)	These are the total number of girls enrolled in the Formal Primary	Broader student beneficiaries (girls) – girls

<p>in a less direct way, and therefore may benefit from aspects such as attitudinal change, etc. but not necessarily achieve improvements in learning outcomes.</p>		<p>since it is the school's entire enrolment of girls. This is however less the project's girls as these have been already counted above</p>	<p>Schools that host the Project's Alternative Basic Education (ABE) Centres. These girls will benefit from activities (girls' agency/leadership skills) of the Girls' Empowerment Forums that are being established in those schools</p>	<p>who will benefit from the interventions in a less direct way, and therefore may benefit from aspects such as attitudinal change, etc. but not necessarily achieve improvements in learning outcomes.</p>
<p>Teacher / tutors beneficiaries – number of teachers/tutors who benefit from training or related interventions. If possible /applicable, please disaggregate by gender and type of training, with the comments box used to describe the type of training provided.</p>	<p>582 Teachers, ABE/NFE Facilitators and Quranic Teachers</p>	<p>582 (The number will remain the same as the same teachers will teach cohort 2</p>	<p>168 (75F:93M) ABE Facilitators trained on delivery of ABE curriculum, 168 (88F : 80M) NFE facilitators trained on delivery of NFE curriculum, 106 Teachers in Formal schools who will be trained on improved methodologies, positive classroom management techniques, gender transformative practices and individual support to learners and 140 (134 : 6 F) Quranic teachers trained on inclusive education especially identification of learners with special needs.</p>	<p>Teacher / tutors beneficiaries – number of teachers/tutors who benefit from training or related interventions. If possible /applicable, please disaggregate by gender and type of training, with the comments box used to describe the type of training provided.</p>
<p>Broader community beneficiaries (adults) – adults who benefit from broader interventions, such as community messaging /dialogues, community advocacy, economic empowerment interventions, etc.</p>	<p>7,000 parents</p>	<p>14,000. This includes another 7,000 targeted for cohort 2</p>	<p>These are targeted for economic empowerment – Village Savings and Loans interventions</p>	<p>Broader community beneficiaries (adults) – adults who benefit from broader interventions, such as community messaging /dialogues,</p>

				community advocacy, economic empowerment interventions, etc.
--	--	--	--	--

[Review of methodology to be completed by external evaluation team]

ANNEX 5 - MEL FRAMEWORK

The project's Monitoring, Evaluation and Learning (MEL) Framework is attached as a separate annex.

ANNEX 6 - EVALUATION INCEPTION REPORT

The inception report for the baseline evaluation is attached as a separate annex.

ANNEX 7 - BASELINE DATA COLLECTION TOOLS

The data collection tools – both qualitative and quantitative – used during the baseline evaluation are attached as a separate annex.

ANNEX 8 - DATASETS, CODEBOOKS, AND PROGRAMS

The datasets collected for the AGES baseline evaluation are attached as a separate annex, as are the programs (Stata .do files) that produce the primary learning and transition results. A brief codebook documenting the main variables necessary for replicating results is provided below.

TABLE 92: KEY VARIABLES FOR LEARNING AND TRANSITION ANALYSIS, HOUSEHOLD SURVEY

Variable	Variable name in dataset	Comments
Girl ID	uniqueid	Unique ID for every girl in the dataset (FE, ABE, and NFE girls). Girls who are tracked over time will use the same unique ID for the purpose of linking baseline data to future rounds of data collection.
Zone	zone	Identifies the state in which the interview took place
District	district	
Sampling point code (or school code)	scode	Uniquely identifies sampling locations. Some sampling locations contain more than one learning institution, as formal schools, ABE centers, and NFE centers overlap in some cases. In essence, this variable captures the different communities or neighborhoods where data was collected.
Learning institution or center	center	Uniquely identifies learning institutions. Formal schools, ABE centers, and NFE centers occupying the same sampling location are assigned different center codes, so as to identify them. In total, there are 58 unique sampling locations and 108 unique center codes. The variable <i>center</i> is the clustering variable – the variable by which we cluster standard errors.

Sampling weight	weight	The weight used to adjust for unequal cluster sizes in the household survey and learning assessment. See the methodological discussion in Section 2 for more details.
Age	age	Girls' age, as reported by the girl at the start of the household survey/girls survey.
Grade	grade	Identifies the grade level or learning center level of the girl at baseline. Grades range from 1 to 2; values of 3 indicate the girl is enrolled in NFE; values of 4 indicate the girl is enrolled in ABE.
Learning institution type or track	resptype	Identifies girls as being part of the following learning tracks: 1. FE (formal education) = 1 2. NFE = 2 3. ABE = 3
Learning Outcomes		
Somali Literacy Score	som	Calculated from all Somali literacy subtasks of learning assessment (equally weighted)
Numeracy Score	num	Calculated from all numeracy subtasks of learning assessment (equally weighted)
Literacy subtasks 1 through 6	som_sb1 through som_sb6	Subtask scores for Somali literacy
Numeracy subtasks 1 through 8	num_sb1 through num_sb8	Subtask scores for numeracy
Transition Outcomes		
Current "transition" status	transition	Calculated as part of the transition analysis, for the transition benchmark sample. It is derived from enrolment status in formal school (trans_pcg_1tc), enrolment status in non-formal education (trans_pcg_2tc), and employment status (trans_pcg_12tc).
Binary transition indicator	trans_bin	Derived from the "transition" variable, it identifies girls who are in a productive activity that is age-appropriate, among the transition benchmark sample only. Girls enrolled in education, NFE, or employed are all coded as "1".

TABLE 93: KEY VARIABLES IN OTHER EVALUATION DATASETS

Variable	Variable name in dataset	Comments
Sampling point code (or school code)	scode	Uniquely identifies sampling locations. Some sampling locations contain more than one learning institution, as formal schools, ABE centers, and NFE centers overlap in some cases. In essence, this variable captures the different communities or neighborhoods where data was collected.

ANNEX 9 - LEARNING TEST PILOT AND CALIBRATION

1. Design

- The EGRA and EGMA versions developed for this study used a simplified version of the structure of more advanced learning assessments previously used with primary school students in Somaliland, Puntland and Galmudug by SOMGEP-T.
- The EGRA includes the core tasks outlined in the 2016 EGRA Toolkit, second edition²⁰², namely letter sound identification; timed familiar word reading; timed passage reading; and reading comprehension. Given previous results in Somalia showing more advanced literacy skills (likely through the acquisition of basic decoding skills during Quranic education), the project has opted for including multiple levels of reading comprehension tasks. The design included a multilingual analysis, where staff reviewed the vocabulary used in the tasks to ensure an adequate progression in difficulty levels by allocating simpler words familiar to both Af-Maay and Af-Mahatiri speakers at the beginning of tasks, and then progressing towards words with less familiar phonemes for minority language speakers/ more complex words (longer, complex phonemes). All texts were checked to avoid using passages/ tales children might be familiar with, thus preventing bias. Multiple changes were made to comprehension questions and text passages based on feedback on clarity, language and level of difficulty. Tasks were originally developed in Somali (Af-Mahatiri) and applied in this language²⁰³.
- The EGMA test included standard tasks – number identification, quantity discrimination, missing numbers, two levels of addition and subtraction (single and double-digit) and problem-solving. Instructions were developed in Af-Mahatiri²⁰⁴.

2. Pilot, calibration and implications

The learning assessments were piloted with a total of 120 students in five formal schools between February 11-26, 2019. Students in grades 1, 3 and 5 were assessed to identify the level of difficulty for children with limited exposure to education (expected to mirror the results of the AGES students) and also to determine the potential for ceiling effects once children reach mid-primary (expected end level for AGES participants). The pilot sample size was limited due to restrictions on staff mobility during this

²⁰² USAID/ RTI (2016) Early Grade Reading Assessment (EGRA) Toolkit Second Edition

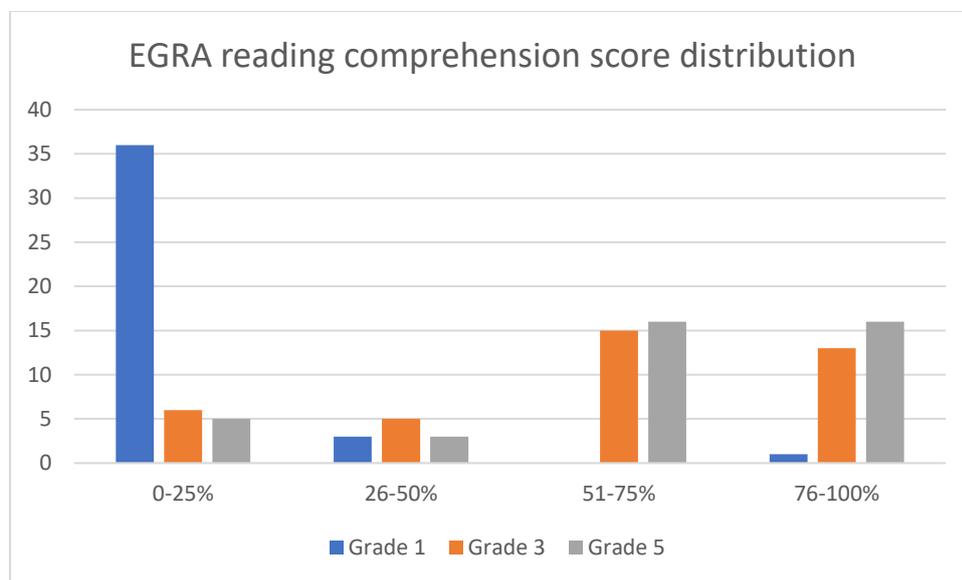
• ²⁰³ During application, instructions were provided in Af-Mahatiri or in Af-Maay, depending on the student's background.

²⁰⁴ Instructions were translated into Af-Maay depending on the students' background.

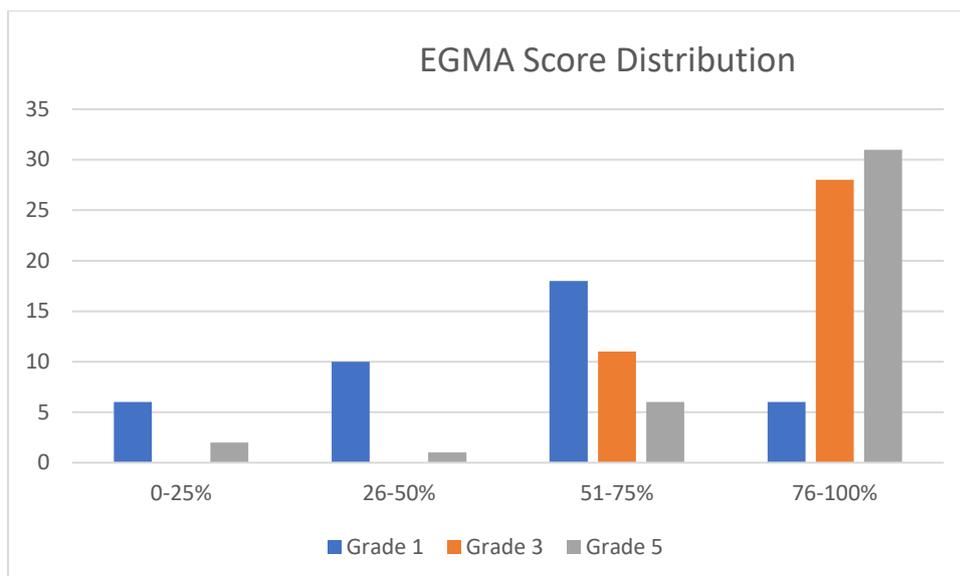
period (during which a sequence of major attacks took place) and the inability to engage an external party to carry out assessments due to cost implications²⁰⁵.

The results indicated that the literacy test considerably difficult for grade 1 students, as 90% of the students had scored zero on reading comprehension, but the distribution of results for grades 3-5 excluded a potential floor effect. On numeracy, students in grade 1 reached high scores in basic tasks (number identification, quantity discrimination), potentially due to previous exposure to these skills in class; nonetheless, a floor effect was observed in other tasks (missing numbers, subtraction, double-digit addition). As 96% of the identified girls had never attended school, it was decided that the inclusion of all tasks would be advisable to avoid the risk of a floor effect. It was also considered that in case a similar situation (high scores in initial tasks) was observed, these tasks could be discarded and an aggregated EGMA score could be calculated based on the subsequent, more advanced tasks.

The distribution of pilot scores can be seen in the graphs below:



²⁰⁵ The cost of the baseline study exceeded considerably the original budget, resulting in a need to re-allocate resources from other lines and limiting other investments in M&E activities.



The results of the pilot were shared with the GEC Fund Manager on March 5 for approval. The learning assessments were also submitted to the MOECHE for approval. The final version of the EGMA incorporated a larger number of responses for letter identification (allowing for a better assessment of difficult sounds for Af-Maay and Af-Mahatiri speakers) and slight modifications in reading comprehension questions to ensure clarity. The final version of the EGMA included slight modifications on the numbers used in different tasks (both for calibration in terms of difficulty levels and to prevent bias since the test had already been shared and might have been disclosed to participants).

3. Methodology for marking

Tasks were individually marked and in the case of EGMA, two tasks were timed (familiar word reading and passage reading). For timed tasks, the number of words read per minute was used for measurement of reading fluency, while for other tasks, marking was based on the percentage score achieved. CARE considers that the individual marking of tasks is important to allow for identification of particularly challenging skills which might not be taught, or correctly taught, in classrooms. In the case of EGMA, an overall score was calculated (the combined scale has a Cronbach-alpha 0.877) as per regular practice.

From Consilient: At baseline, no calibration across learning assessments was completed, and only one set of learning assessments was developed. In future evaluation rounds, it will be important – if new assessments are to be utilised – for the development to focus on ensuring the assessments are of equivalent difficulty to the baseline. This is especially important, because the pre-post design employed by AGES is not robust to changes in assessment difficulty over time. For instance, if a future assessment is of significantly greater difficulty – resulting in lower scores at endline than at baseline – it will not be possible to determine with a high degree of confidence whether the decline in scores is the result of changes in the true learning scores of girls or the result of changes in assessment difficulty. Even in less extreme scenarios, in which girls’ learning scores have improved, very small changes in assessment difficulty produce bias in estimation of program impact in terms of learning. Therefore, prior to future evaluation rounds, considerable effort should be made to pilot the learning assessment and investigate their difficulty relative to the baseline.

ANNEX 10 - SAMPLING FRAMEWORK

The distinct sampling frames used for drawing the baseline sample are attached as a separate annex, along with an explanatory note describing each.

ANNEX 13 – SAMPLING REVISIONS AND STATISTICAL POWER

In the methodology section of the main report, we report on differences in the target and achieved sample of cohort girls, which were fairly substantial for a baseline evaluation. As we discuss in more detail in that section, the number of clusters or Primary Sampling Units (PSUs) was reduced during fieldwork, and a number of interviews were cleaned from the data due to misadministration of the learning assessments. Specifically, the target sample size was 1,530 girls across 45 formal schools, 44 NFE centres, and 43 ABE centres. In practice, the achieved sample size was 1,420 girls across 38 formal schools, 36 NFE centres, and 34 ABE centres.

The reduction in overall sample size and number of clusters could be consequential in terms of statistical power. Depending on how one defines clusters, the number of clusters may be considered either fairly small or quite large: if each sampling point – including overlapping formal schools, ABE centres, and NFE centres – is a single cluster, the achieved sample includes just 58 clusters; if each sampling point – learning institution combination is a distinct cluster – e.g., NFE and ABE centres at the same sampling point are considered separate clusters – the number of clusters rises to 108. During our analysis, we use the latter definition of clusters to adjust for correlation in the error term among girls in the same clusters, as we discuss in more detail in the report. A reduction in the number of clusters from 132 to 108 is a considerable reduction, implying that statistical power may be dramatically reduced from the power analysis performed during the design stage.

To assess the severity of the issue, we performed supplementary power analysis, the results of which are reported below. In standard power analysis, researchers must make a number of assumptions regarding the structure of data, the likely effect size of the intervention, and so forth. While we must make many of these same assumptions, our analysis is also tied to the empirical data available from the baseline, which ensures the simulations are based on plausible assumptions. Nonetheless, it is important to note that the analysis below is intended to highlight the potential reduction in power due to changes in the sample design, and should not be interpreted as valid measures of statistical power in general. In other words, the results are not meant to suggest a change in sample design is needed, or that the sample is properly or insufficiently powered. Rather, the calculations are illustrative: if the target sample design was expected to yield, for instance, 80 percent power for a given effect size, the results below suggest power may have declined by anywhere from 1 to 10 percentage points, depending on a number of assumptions. Our goal in this section is to illustrate the possible reduction in power from design changes, rather than specify the precise quantity of those reductions.

The focus of our power analysis is on Somali literacy scores, though the results should apply equally to numeracy and to transition.²⁰⁶ Our simulation derives a number of assumptions from the baseline data, including the approximate distribution of Somali literacy scores (i.e. the sample mean and standard deviation). To simulate the clustered nature of the data, we allow incorporate cluster-specific coefficients, analogous to cluster fixed effects or mean scores by cluster that differ systematically from the aggregate mean. This yields an intra-cluster correlation – the relative variance of learning scores within, versus between, clusters – very similar to the observed value in the baseline data, at $\rho = 0.27$. To complete one

²⁰⁶ In general, power to detect differences in a binary outcome, such as transition, is lower. However, there is no reason to expect the *relative reduction* in power arising from the sample changes discussed here would be smaller or greater for a binary outcome, compared to a continuous measure, such as learning scores.

round of power analysis, we randomly simulate 1,000 samples of cohort girls with the structure listed above. We then simulate an endline learning score based on an expected improvement in scores of 2.5 percentage points, where the effect size is also correlated within clusters. We apply hypothetical attrition rates of 25 percent from baseline to midline and a further 20 percent from midline to endline, yielding final sample sizes of 918 (target) and 852 (achieved) girls for baseline-to-endline comparisons, which represents the sample that would theoretically be available for panel analysis.

Under this general setup, we consider the scenario in which the target sample of 1,530 girls was achieved, with 132 unique clusters, and then consider the true outcome, in which 1,420 girls were contacted, across 108 unique clusters. By comparing statistical power achieved between the two samples, under otherwise identical conditions – identical expected effect size, correlation structure across clusters, and distribution of learning scores – we can quantify the decline in power from the target to the achieved sample.²⁰⁷

TABLE 94: STATISTICAL POWER COMPARISONS BETWEEN TARGET AND ACHIEVED SAMPLE SIZES

Effect Size	Target Sample	Achieved Sample
2.5 points	94.2%	91.6%
1.5 points	63.9%	55.4%

The table reports statistical power, at the 95 percent confidence level (i.e. with alpha set to .05) for effect sizes from baseline to endline of 2.5 and 1.5 percentage points, respectively. Each simulation was repeated over 1,000 iterations, such that the results represent the share of iterations in which a statistically significant effect was found, as is standard in power analysis simulations. Assuming an effect size of 2.5 points, power declines by an estimated 2.6 points when moving from the target sample design to the achieved sample. The results are starker when the effect size becomes smaller, with an 8.5 point reduction from the target sample design to the achieved sample.

To reiterate our earlier caution regarding these results, it is essential to note that these results do not represent the "true" statistical power of the evaluation. The simulations were based on a series of assumptions that may or may not apply in practice and almost certainly differ from those utilized by CARE's Monitoring & Evaluation team and the FM when agreeing to sample sizes in the MEL Framework. The results are provided to highlight the consequences we should expect from the reduction in sample size and number of clusters – if the MEL Framework estimated statistical power under the target sample design at 80 percent, for instance, we might expect a reduction of approximately 2 to 10 points as a result of the sampling changes that occurred during the baseline. While the loss of power is not dramatic, it is a meaningful reduction, especially if the evaluation was properly powered, rather than over-powered, to start.

²⁰⁷ Again, it is important to note that the analysis is still based on a number of assumptions regarding attrition rates, and the expected effect size. Power analysis performed prior to the project is unlikely to use such a small expected effect size, but this value is useful for illustrative purposes.

ANNEX 14 – CEILING EFFECTS SIMULATION

In our discussion of learning results in the main report, we documented significant ceiling effects in Somali literacy scores and less severe, but still meaningful, ceiling effects in numeracy scores. We do not replicate the analysis here, but refer readers to our discussion in Section 4 of the report. The primary risk of floor and ceiling effects are that they obscure true changes in learning outcomes over time.

A useful way to think about measuring learning outcomes is to consider true learning a latent, unobserved variable, which exists on a wide spectrum. Observable learning scores, on the other hand, capture only a portion of this spectrum, from 0 to 100, levels which are defined by the difficulty of the test. No test captures the full spectrum of learning, so some potential test-takers always have true learning values outside the bounds of 0 to 100. If a student's true learning score changes from 105 to 110 over time, but the test only captures learning outcomes from 0 to 100, their improvements are not captured by the test. Students of this kind actually drag down the true change in learning scores from one round to the next, producing bias in estimates of project impact. In some contexts, similar effects can produce a bias toward a positive result (i.e. bias estimates upward), but this is extremely unlikely in the research design used by AGES.

Note that the potential for ceiling effects in AGES is exacerbated by the fact that there is no comparison group, for two reasons. The first reason, as we discuss in the limitations section of the report, is that making the assessment more difficult at midline is not a simple solution in a pre-post design without a comparison group. Therefore, there is no easy solution to ceiling effects. Second, and more relevant to the analysis in this annex, comparing girls at midline to benchmarks established at baseline accounts for growth or natural maturation effects as girls age; this approach implies that ceiling effects will worsen over time as a result of natural growth in scores. The simulation below takes this into account in an attempt to quantify the level of bias that we can expect as a result of ceiling effects.

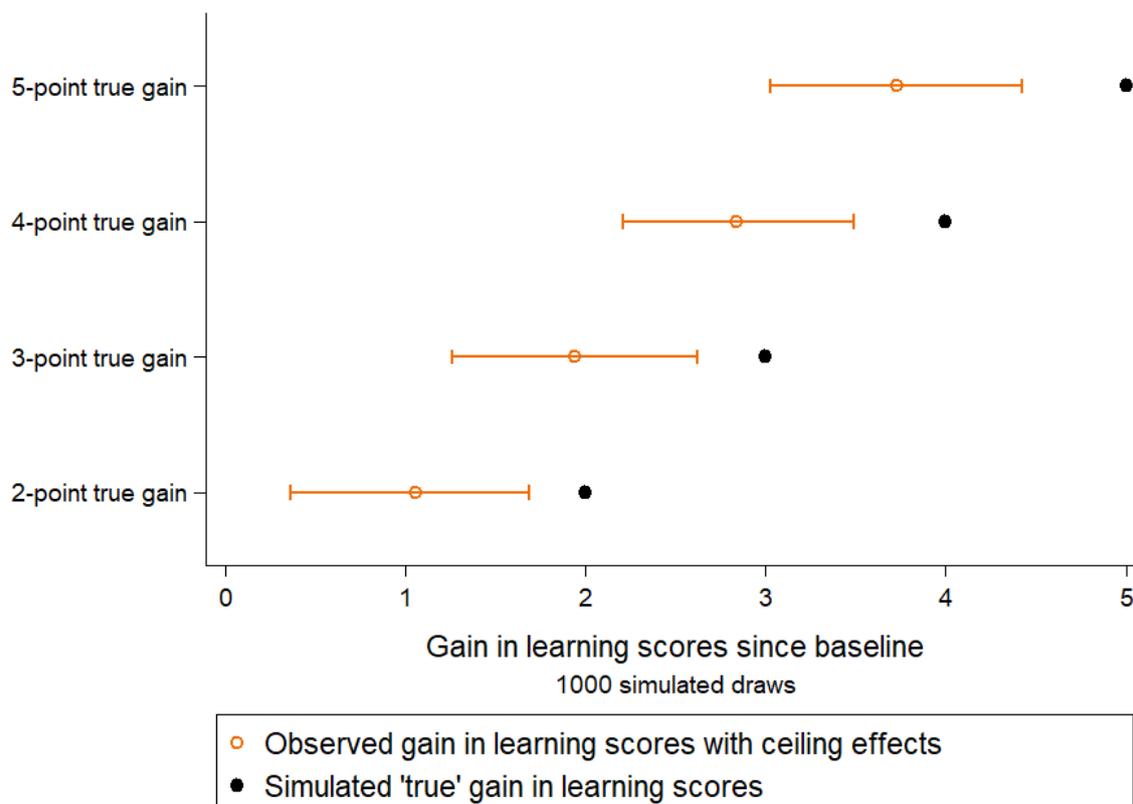
Our simulation starts from the baseline learning data, using the empirical distribution of Somali literacy scores. We calculate the differences in learning scores between age cohorts, which become the benchmarks against which girls will be judged in the midline. These benchmarks highlight the fact that ceiling effects will get worse over time, as they are – on average – positive, implying gains in score and a movement toward 100 percent. Note that we limit our simulations to formal school (FE) girls, because these girls will be tracked for the full life of the project. This decision reduces the magnitude of the ceiling effects we observe, as FE girls achieved lower mean scores on both numeracy and literacy than their ABE and NFE counterparts. Just 2.1 percent of FE girls achieved perfect scores in Somali literacy, compared to 15.9 percent of NFE girls, for instance.

Next, we make an assumption about the true impact of the project, in terms of latent learning scores. Our first simulation assumes a 5 percentage point gain in scores, which is a substantively meaningful effect size by any standard. We simulate improvements in learning scores by randomly drawing values from a normal distribution with a mean of 5 and a standard deviation of 9, which is roughly consistent with the distribution of learning score improvements we have observed in other educational interventions, at least in those cases that have had discernible positive impacts. Random draws imply that some girls in the sample regressed, while others progressed, but that – on average – the gain in latent learning was 5 points. To determine *observed* learning scores, we truncate the latent learning levels to a 0-100 scale; to determine the project's observed impact on learning scores, we calculate the gains in learning scores, truncated to 0-100, vis-à-vis the benchmarks established from the baseline data. Because our calculations are based on a random draw of gains in latent learning, we repeat this simulation for 1,000 iterations, calculating the mean improvement in observed learning scores as a result of the project, in our simulated environment.

The results of the simulations are presented in the figure below. We repeated the simulations with gradually decreasing true impact levels, from 5 percentage points down to 2 percentage points. The figure

shows the "true" impact as a black dot on the right-hand side of the graph (5 points, 4 points, etc.) and the impact we would expect to observe, in practice, as an orange dot to the left. The horizontal bars represent the 95 percent confidence intervals around our mean estimate. The take-away from this analysis is that a 5-point true gain in learning scores would manifest as 3.7 point gain in practice, under plausible assumptions about the distribution of impact, growth effects from year to year, and based on the distribution of scores and severity of ceiling effects at the baseline. If the true impact were smaller, the potential consequences actually become larger, as a share of the overall effect size: at a 3-point true gain, the observed effect would be around 1.9 points, a 36.7 percent reduction, and at a 2-point true gain, the observed effect would be 1.0 points, a 47.5 percent reduction.

FIGURE 23: SIMULATED LEARNING EFFECT SIZES, WITH CEILING EFFECTS



Setting up the simulation to include only FE girls reflects the reality of the program's evaluation design; nonetheless, even among FE girls exclusively, the potential for ceiling effects to reduce the magnitude of observable program impact is significant, as our simulations show. If we expand the simulations to include ABE and NFE girls, a 5-point gain in latent learning scores would actually be observed as a 3.4 point increase, while a 3-point gain would be observed as a 1.7 point increase (as opposed to a 1.9 point increase among FE girls alone). These are more drastic reductions – they do not reflect the most likely outcome, as AGES will only track FE girls over the life of the program, but they reflect the greater potential for ceiling effects to bias analysis of learning outcomes toward a null result if ABE and NFE girls are included in shorter-term tracking and evaluation (in, for instance, baseline-to-midline comparisons).

Two additional factors should be kept in mind while considering these results. First, we have assumed a one-year gap from baseline to midline, with one year of realized maturation during that time. In practice, if the midline occurs two years after the baseline, the biasing impact of ceiling effects will be amplified.

This fact holds true for the endline as well, because maturation effects will continue to undermine our ability to capture latent true learning outcomes as girls age and progress. Second, we did not perform power analysis to understand the effect this bias would have on statistical power to discern project impact over time. In general, a cursory analysis suggested that reductions in power would be very minor at the effect sizes we simulated. However, if the true effect of the project is 2 points or less, reductions in power could be substantial – sufficiently large that they could turn a properly-powered study into one with entirely insufficient power. In fact, this problem may be worse than we estimate, because our analysis did not take into account reductions in sample size due to attrition over time, which would simultaneously reduce statistical power. Depending on the available sample size margin, statistical power could be affected by ceiling effects at midline and endline.

ANNEX 15 – MEASUREMENT OF DISABILITY STATUS

The discussion in the main report surrounding the prevalence of disability was fairly extensive. However, we made a number of choices around measurement – including testing alternative metrics for disability status – that were not fully explained in the text. This annex describes the main coding scheme used for AGES, which mirrors that of the Washington Group guidance. In the text of the report, we employ two alternative coding schemes, which are detailed elsewhere in the report,

Our main approach to measuring disability status follows the Washington Group's Child Functioning module and coding guidelines. Where the Washington Group does not provide explicit guidance is in the aggregation of linked questions or domains. We opt to aggregate some of the domains identified by the Washington Group, because they are either intrinsically or thematically linked. The categorizations we make include the following broader areas; we also refer to the domains and questions that feed into each definition from the Washington Group questions.

- Seeing
 - Seeing domain – based on difficulty seeing with or without glasses.
- Hearing
 - Hearing domain – based on difficulty hearing with or without a hearing aid.
- Walking/mobility
 - Walking domain – based on difficulty walking 100 or 500 meters with or without equipment.
- Self-care
 - Self-care domain – based on difficulties with dressing or feeding oneself.
- Communicating
 - Communicating domain – based on difficulties communicating with people either inside or outside the household.
- Cognitive
 - Learning domain – based on difficulties learning things relative to girls their own age.
 - Remembering domain – based on difficulties remembering things relative to girls their own age.
 - Concentrating domain – based on difficulties concentrating on topics relative to girls their own age.
- Behavioral
 - Accepting change domain – based on difficulties accepting change in their routine.
 - Behavior domain – based on difficulties controlling their behaviour.
 - Making friends domain - based on difficulties making friends.
- Affect
 - Anxiety domain – based on the frequency a girl feels very anxious, nervous, or worried
 - Depression domain – based on the frequency a girl feels very sad or depressed

In all domains except those nested under affect, the survey questions assess the difficulty girls face when trying to complete a functional task, asking them or their caregiver whether they can complete the task without any difficulty, with some difficulty, only with a lot of difficulty, or whether they cannot do the task at all. In line with the Washington Group guidance, we coded girls as having a disability in a particular area if she cannot do a task at all or can only do it with a lot of difficulty. For instance, if a girl cannot remember things at all, or can only accept changes to her routine with a lot of difficulty, she is classified as having a cognitive or behavioural disability, respectively.

The single exception to this coding scheme applies to the "affect" domains of anxiety and depression.²⁰⁸ Girls and their caregivers were asked the frequency with which a girl feels, for instance, nervous, anxious or worried. The Washington Group guidance suggests that girls who feel anxious or depressed *daily* should be classified as having an "affect disability." In contrast, and in consultation with CARE's Monitoring and Evaluation team, we have classified girls as experiencing anxiety or depression if they report episodes of either outcome daily or weekly. This dramatically alters the share of girls classified as having an "affect disability", increasing the share of girls reporting an affect disability from 2.5 percent (daily only) to 11.4 percent (daily or weekly). There is little research on mental health disabilities in developing countries, but we view the higher rate as a more plausible representation of mental health issues in the sample, given the conflict-affected regions in which AGES is being implemented.

²⁰⁸ In earlier drafts of this report, we referred to issues of anxiety and depression as "mental health", but we have since adjusted the wording to be in line with the Washington Group.

ANNEX 16 – EVALUATION METHODOLOGY

This annex reviews a number of methodological details of the AGES baseline evaluation. Some of the discussion in this section is relevant for understanding the analysis in this report; other details will be necessary for the design of future evaluation rounds, to ensure comparable approaches are taken or to understand where improvements to the design could be made.

Quantitative Sampling

As noted in Section 2.3 of the main report, a number of changes were made to the quantitative sampling approach immediately prior to, and during, fieldwork. Specifically:

- Schools were replaced
- Schools were excluded, and other sampling points were assigned a larger cluster size to compensate for this fact
- Some pre-selected girls could not be located or were not eligible for recruitment into the cohort, and they were replaced
- CARE and the evaluation team identified two enumerators who were conducting the learning assessments incorrectly, and respondents whose scores were affected by their administration were removed from the sample.

Replacement of Specific Girls

The sampling methodology for formal schools differed from that of ABE and NFE centres in one important way: the selection of girls within the institution. For ABE and NFE centres, team leaders selected girls randomly from the set of eligible girls upon arrival at the centre, because no pre-defined sample frame of girls in these centres was available prior to the start of fieldwork. For formal schools, on the other hand, CARE provided a full, individual-level sample frame of girls. When drawing the sample of formal schools, we selected schools using probability proportionate to size (PPS) sampling, and then drew specific girls into the sample using random sampling within each selected school (in other words, the sample is a classic clustered sample).

As we would expect, some girls listed in the sample frame were unavailable or ineligible for interviewing during fieldwork, and required replacement. For each formal school, 12 cohort girls and 4 replacement girls were randomly sampled. In total, 160 girls from the original cohort could not be located, 60 replacement girls were located, and 100 additional replacement girls were randomly sampled and

located²⁰⁹. The replacements were made in 13 locations in Jubaland, 10 locations in Banadir and 6 locations in South West.

In some instances, we had to replace the identified cohort girls who were not eligible, due to the following reasons²¹⁰:

TABLE 95: NUMBER OF REPLACEMENTS BY CASE

Reason for replacement	Number of cases	% of replacements made
Girls were too young	3	2%
Girls were too old	3	2%
Girls we in a grade 3 or higher	12	8%
Girls dropped out of the school	42	26%
Girls never attended the school	14	9%
Girls could not be located in school, unreachable via phone and/or no records of the girl in the school	45	28%
Girls were attending ABE or NFE classes	3	2%
Girls relocated to another school / moved to another place	19	12%
Girls were now not enrolled in the school	10	6%
Girls were enrolled in the school before being registered for the project	4	3%
Other reasons (such as sickness, could not access school because of the flooding at the time of the visit, could not afford to pay school uniform etc.)	4	3%

When the teams could not identify 12 eligible cohort girls, team leaders were required to first inform the fieldwork manager and research officer on how many girls are missing to achieve the required sample and the reasons why the pre-selected cohort girls were not eligible. Once the replacements were confirmed, team leaders proceeded to locate girls from the list of 4 pre-selected replacement girls, which was provided to the team leader before the beginning of the fieldwork. If the teams were not able to identify

²⁰⁹ While a number of replacements represent a substantial share compared to the size of the final sample and could be indicative of the beneficiary selection and record-keeping practices, the higher share of replacement is to be expected in especially low-resource settings.

²¹⁰ The number of cases presented in the table is an approximation. While all the replacement cases were reported to Consilient’s technical research team, not all the cases were reported in the data collection reports. The information presented in the table is based on the data extracted from the data collection report and fieldwork notes indicated in the cohort tracker managed by the evaluation team. Note that replacement of girls does not represent a significant methodological issue, with respect to inferences drawn in this report or future evaluation rounds, because the girls were removed from the sample and their replacements were selected randomly from the existing sample frame.

any eligible girls among the pre-selected replacement girls, team leaders informed the fieldwork manager and research officer, who randomly drew an additional list of replacement girls from the list of all the enrolled girls, provided by CARE.

If the number of girls who were not eligible was significant, the fieldwork manager and research officer would notify the research manager or the director of research for further guidance on how to proceed. If even after several rounds of replacements, teams could not identify enough girls to complete the sample for the sample point, we have contacted CARE for further guidance. During the fieldwork, in all such cases, we were advised to exclude the schools from the sample.

Enumerator-Specific Issues and Exclusion from the Sample

The final major sampling-related issue encountered during fieldwork actually relates to survey administration. While the removal of sampling pointed reduced the achieved sample size somewhat, data for two enumerators also had to be excluded from the final sample due to incorrect administration of the learning assessments. Concerning patterns in the learning assessments were flagged early in fieldwork by both CARE's Monitoring & Evaluation team and Consilient's technical evaluation staff. These issues were centered on two enumerators:

- Enumerator 1 reported particularly high scores for Somali literacy, which prompted a review of their scores, additional accompaniment by their team leader, and a discussion of how they were administering the learning assessments. Based on that discussion, Consilient's technical team decided to re-administer the learning assessments to all girls this enumerator had interviewed (n = 29 at that stage of data collection).²¹¹ The team re-administered learning assessments to all 29 girls, with the problematic enumerator participating in the re-administration (n = 10) and completing an additional 33 interviews later in fieldwork. Unfortunately, a second review of his performance by CARE and Consilient revealed that the enumerator's scores for reading comprehension subtasks continued to be much higher than other enumerators. Given the pattern of problems, it was decided that all interviews by this enumerator (n = 43) would be removed from the data. The interviews re-administered by other team members (n = 19) remain in the data.
- Enumerator 2 was not identified as a potential problem until the end of fieldwork, when reading comprehension scores for their respondents were compared to the full sample and to the other enumerators on their team. Girls' performance on reading comprehension, when paired with this enumerator, were abnormally high, and often inconsistent with their performance on other reading subtasks. Both CARE and Consilient agreed that the enumerators interviews (n = 48) should be removed from the data.

²¹¹ Although there were concerns regarding the possibility that re-administering the assessment would result in higher scores, this was ruled out, because the enumerator in question had generally skipped later subtasks of the assessment and assigned the girls scores based on their performance on the first two subtasks. Thus, the girls in question had not been exposed to the full set of test questions, and would not have an undue advantage vis-à-vis other girls.

In total, poor administration of the learning assessments resulted in the loss of 91 cohort girls from the sample.²¹²

Qualitative Sampling

Section 2.3 provides the most important information regarding the qualitative sample, including a description of the tools employed, the populations sought out, and the achieved sample in terms of specific respondent groups and its geographic composition. Below, we break down the qualitative tools and sample in greater detail.

The specific list of tools, containing relevant indicators and information related to their piloting is summarized in the table below.

TABLE 96: QUALITATIVE TOOL DEVELOPMENT

Tool name	Relevant indicator(s)	Who developed the tool?	Was the tool piloted?	How were piloting findings acted upon (if applicable)	Was FM feedback provided?
FGD with CEC	<ul style="list-style-type: none"> Barriers to learning for girls; girls from minority clans and ethnic groups; girls with disabilities; pastoralist children Effectiveness of ABE courses in supporting needs of marginalized girls Effectiveness of life skills courses in supporting needs of marginalized girls Community support for girls education CEC management 		Tool piloted by CARE prior to the training		
FGD with mothers	<ul style="list-style-type: none"> Access to education Barriers to school attendance Learning and teaching quality Barriers to learning for girls with disabilities and girls from marginalized groups 		Tool piloted by CARE prior to the training		

²¹² Overall, 35 formal school girls, 27 NFE girls, and 29 ABE girls were removed from the sample for this reason.

	<ul style="list-style-type: none"> • School governance • Girls' leadership skills • Security risks • Transition 				
FGD with teachers	<ul style="list-style-type: none"> • Learning and teaching quality • Barriers to school attendance • Barriers to enrolment and attendance for girls from minority clans and ethnic groups; girls with disabilities; pastoralist children • Security risks • Barriers to learning for girls and boys • Boys' and girls' participation in class • Access to girls' toilets 		Tool piloted by CARE prior to the training		
Risk mapping	<ul style="list-style-type: none"> • Girls perceptions of the most/least secure sites on school groups • Girls perceptions of the most/least secure sites in their communities • Girls perceptions of the risks faces by boys versus girls 		Tool piloted by CARE prior to the training		
Vignette exercise	<ul style="list-style-type: none"> • Girls perceptions of barriers to learning • Girls perceptions of barriers to attending schools 		Tool piloted by CARE prior to the training		

The planned sample was achieved as planned, with 10 interviews of each type conducted, across 10 “qualitative sample points.” No changes were made to the geographic locations of the qualitative interviews, except for Howlwadaag, where, despite several attempts, the team was not able to organize the FGD with the CEC due to their unavailability. Thus, while FGDs with mothers and teachers and risk mapping/vignette exercises were conducted in Howlwadaag, the FGD with the CEC was conducted in another sample point, located in Wadajir District.

While selecting the participants for the sessions, team leaders were instructed to select the girls from the same age category corresponding to the cohort groups (10-13 or 13-17 or 17-20), 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.

For the FGDs with the CEC committee, the only selection criteria were that the CEC was active and functioning (also used as the selection requirement for the sampling point). The team leaders would

receive a list of the CEC members with their phone contacts and contact each member individually until a minimum or a maximum number of participants were confirmed. However, in the few cases, only 4 CEC members would attend the FGDs.

To identify the teachers for the FGD sessions, the team leaders were asked to select the teachers teaching in grades 1 and 2 and the classes with the girls participating in the LNGB AGES program. The minimum number of participants was 6 and the list of all the eligible teachers was provided to the team leaders by the school principal. Team leaders subsequently selected the teachers and contacted the teachers for the session. Teachers teaching in the morning and afternoon classes were included.

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. The minimum number of participants was also 6.

The qualitative data collection took place at the same time as the quantitative data collections mostly because of budgetary and logistical constraints and because of the timeline pressures, as schools in some of the locations were closed during the December/January holidays.

During the qualitative data collection, we have encountered a few methodological issues, such as one case during the risk mapping when a female researcher omitted the questions related to the school groups and asked girls only about their community. In this case, the researcher conducted an additional session with the same girls to complete the missing information. Also, in one case, the team leader conducted the first round of the FGDs (CEC, teachers, mothers, and girls) and did not wait until all the participants responded to the question, before asking the next question. In this case, the team leader was asked to conduct all 4 sessions again.

Pre-Data Collection and Training

Enumerator Selection, Training, and Pilot Testing

To collect the baseline data, 7 teams, each with one team leader and 3 enumerators were deployed. Enumerators and team leaders were selected based on their experiences, gender and language requirements for the fieldwork locations. Altogether, 13 female and 13 male team leaders and enumerators were deployed to conduct the data collection.

All teams participated in six days of training, including one day dedicated to a pilot test conducted in non-sampled schools. 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 the fieldwork management (i.e. filling and keeping the tracking sheets) and on the random selection using the Random Number

Generator application. 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), minimizing 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.

While assembling the fieldwork teams, we made sure an overall 50-50 gender split was respected. In total 4 teams had two female and two male members, while 2 more teams had one female and three male members, and 1 team had three female members and one male member. Two team leaders were women. However, 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 were 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 stigmatization, girls may feel especially uncomfortable or even refuse to respond to these questions, in the presence of men.

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.

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 non-sampled 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, most 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.²¹³ Second, the data collection tools team leaders are tasked with can be fairly complicated, and require strong organizational skills. They also, at least in the case of the classroom observation, require some degree of subjective judgment. 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.

Quality Assurance and Data Cleaning

Quality Assurance

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.

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

²¹³ Team leaders and enumerators engaged in other GEC evaluations in Somaliland, Puntland and Galmudug are often unable to travel safely to the regions where AGES is being implemented.

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.

During this process, we have noted that the risk mapping exercises were the most challenging. During three sessions, the team members conducting the sessions forgot to ask specific questions and needed to return to the school and complete the missing information.

In the case of all 4 types of qualitative interviews conducted in one district, the team had to re-do the sessions as the team leader did not dedicate equal time to all the participants to express their opinions. Moreover, the team conducting the qualitative interviews in another district had to re-do an FGD with mothers due to the technical issues encountered while recording the session.

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 benchmark learning assessment data were reviewed daily, the review and cleaning of the data from other surveys were done bi-weekly. In addition, during the fieldwork, the anonymized data were shared with CARE Somalia on a weekly basis.

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. The data from the household survey, benchmark learning assessment, headcount, and disability survey were checked and cleaned more regularly.

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. To facilitate data management and analysis, all the files were ordered and coded using the RQDA analysis software.

Fieldwork-Specific Challenges

Learning assessments: as noted previously, systematic problems were noted with the manner in which two enumerators were administering the learning assessments. In the case of one enumerator, many of the assessments were re-administered by a member of the same team, and – after thorough review by the evaluation team – the re-administered assessments were included in the final sample. The remainder of the assessments, from both problematic enumerators, were excluded.

Missing grades: as some formal schools did not have all the grades from 1 to 8, not all the headcount surveys, benchmark assessments and classroom observations could be completed.

Data Analysis

Survey Weights

The sampling approach taken for AGES, which employed PPS sampling in the first stage selection of formal schools and ABE centres or Primary Sampling Units (PSUs), does not generally necessitate the use of survey weights to correct for unequal probabilities of selection. With equal cluster sizes (i.e. number of girls selected during second-stage sampling), PPS ensures each girl in the sample frame has an equal overall probability of selection.²¹⁴ Given that the evaluation team did not stratify the sample by state, the sample is theoretically self-weighting, at least insofar as equal cluster sizes were employed.

It is important to note that survey weights can be used to achieve ends other than ensuring equal probability of selection for all potential respondents in the sample frame. For instance, it is common to use post-stratification weights to adjust the sample to mirror known population parameters, such as the population's known or expected sex ratio, age distribution, geographic distribution across states, and so forth. Population surveys also often use survey weights to adjust for non-response, especially differential non-response across population groups, or partial non-coverage of key groups in the sample frame.

In this report, we employ weights solely to adjust for unequal cluster sizes. For reference, cluster sizes vary between 8 and 12 girls in formal schools, and 8 and 17 girls in both ABE and NFE centres, though most formal schools are represented by 12 girls in the sample and most ABE and NFE centres are represented by 17 girls. Oversampling from the standard cluster size of 12 girls was the result of a need to preserve the largest possible sample size, despite the dwindling number of clusters/sampling sites as fieldwork progressed and sites were dropped from the sample. The data also include clusters with a smaller-than-expected number of girls, either because the site had too few girls meeting the eligibility criteria or because interviews were cleaned from the data after the fact.²¹⁵

We do not use weights to recover population parameters, as in post-stratification, because the sample frame from which the sample was drawn is of unknown accuracy. Several of the schools and centres visited by field teams did not have the requisite number of girls to constitute a full cluster, despite being listed in the sample frame as having 20 or more girls enrolled. While weighting to ensure each state comprises a share of the sample proportional to their share of girls in the sample frame is possible with the available data, this would entail using weights to recover a state population distribution that may not be reliable. Given that the use of weights reduces the efficiency or precision of statistical estimates, reducing statistical power, their use should be carefully justified. In a context in which the sample frame is of unknown reliability, this standard does not appear to be met.

Similarly, we do not use weights to adjust the sample in line with post-hoc corrections to the sample frame. In other words, it is theoretically possible to – during fieldwork – obtain more accurate population

²¹⁴ In PPS, schools with a larger number of girls have a higher probability of first-stage selection, but each girl has a proportionally lower probability of selection during the second stage, as a given cluster size (e.g., $n = 12$) represents 12 opportunities for selection into the sample at the school level, within a larger school population. Assuming accurate school-level population information and equal cluster sizes, the PPS sample is self-weighting.

²¹⁵ The most prominent case of this post-fieldwork cleaning involved the two enumerators described earlier in this section, whose poor administration of the learning assessment prompted removal of their interviews.

figures for each PSU sampled and then use weights to account for inaccurate PSU populations in the sample frame. However, doing so relies on collecting extremely reliable and complete enrolment data during fieldwork and, nonetheless, will not account for errors population counts for PSUs present in the sample frame but not selected into the sample. This latter issue subsequently generates incorrect probabilities of selection that cannot be corrected post-hoc, and renders the sample non-representative in an entirely new way.

By comparison, our approach to survey weighting is conservative, from an inferential standpoint, and simple. We adjust cluster sizes *within* types of PSUs to be equal, where types of PSUs are formal schools, NFE centres, and ABE centres. Hence, in our aggregate analysis, every formal school is given equal overall weight to every other formal school; however, each formal school is *not* given equal weight to every other NFE centre, because the weights only account for unequal cluster sizes within a given type of PSU. It is important to establish this equilibrium between clusters to ensure that when capturing learning improvements all learning centres would be treated fairly. The decision is pertinent as in the absence of weighting, the outcomes in larger educational centres (or ones with more girls in the cohort, for whatever reason), would disproportionately drive the project outcomes in learning and transition. The following table shows how the weight factors depending on the number of cohort girls in the sample for the sample cluster.

Our approach corrects for the fact that unequal cluster sizes yield unequal probabilities of selection for girls in different PSUs. This adjustment is strictly limited to the set of sampled PSUs, however, and we make no claim that the sample is more or less representative of the AGES beneficiary population with weights applied than without. Our focus is analogous to protecting internal validity – the validity of conclusions drawn within the sample – as opposed to external validity. Specifically, we wish to ensure, to the extent possible, girls in the sampled schools were selected with equal overall probability to one another, and that each PSU is weighted equally, to prevent a few large clusters from driving conclusions about project impact or other outcomes.²¹⁶

As noted, our approach is conservative: we do not make tenuous claims that we can validly recover population parameters or make post-hoc adjustments to the sample frame. The weights we apply range in value from 1 to 2.125, well below extreme weights that might impact efficiency and statistical power, and well within the bounds of methodological rules of thumb.²¹⁷ Nor is the approach here applied with finality.²¹⁸ In addition, note that we apply survey weights exclusively to the household survey and learning assessments, and use weights only in the context of aggregate analysis. Where we study subgroups of girls, such as girls with specific physical impairments, or girls who come from specific ideological backgrounds, we do not apply weights, as adjusting for unequal cluster size is no longer relevant when

²¹⁶ In practice, cluster sizes were sufficiently similar that it is unlikely any small subset of schools could unduly influence our conclusions.

²¹⁷ Potter, Frank, and Yuhong Zheng. 2015. "Methods and Issues in Trimming Extreme Weights in Sample Surveys." *Proceedings of the American Statistical Association, Section on Survey Research Methods*.

²¹⁸ If desired, the weighting approach can always be adjusted during the midline or endline evaluation rounds.

only a portion of girls in each cluster are included in the subgroup. Finally, in a subsequent section that investigates the representativeness of the sample vis-à-vis the target population, we check the extent to which our survey weights alter the composition of the sample, whether they could change key findings, and whether they push the sample closer to, or further away from, the population parameters from the sample frame.²¹⁹

Clustering

Throughout the analysis, we cluster standard errors wherever appropriate. Specifically, we utilize clustered errors wherever multiple observations or data points were collected from the same learning institution. Therefore, we cluster errors when analyzing 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.²²⁰

ANNEX 17 - WEIGHTING AND SAMPLE CHARACTERISTICS

The survey weights that we employ do not specifically seek to correct discrepancies between the composition of the sample and the composition of the population, nor do they adjust the sample to match known population parameters, as we discuss elsewhere in this report. The survey weights we apply

²¹⁹ Of course, as the earlier discussion alluded to, assessments of representativeness comparing the sample to the sample frame should be interpreted with caution, given the tentative nature of the sample frame. The same caveat applies equally to judging the weighting scheme based on its ability to recover population parameters from the sample frame, as these population parameters are not known with certainty.

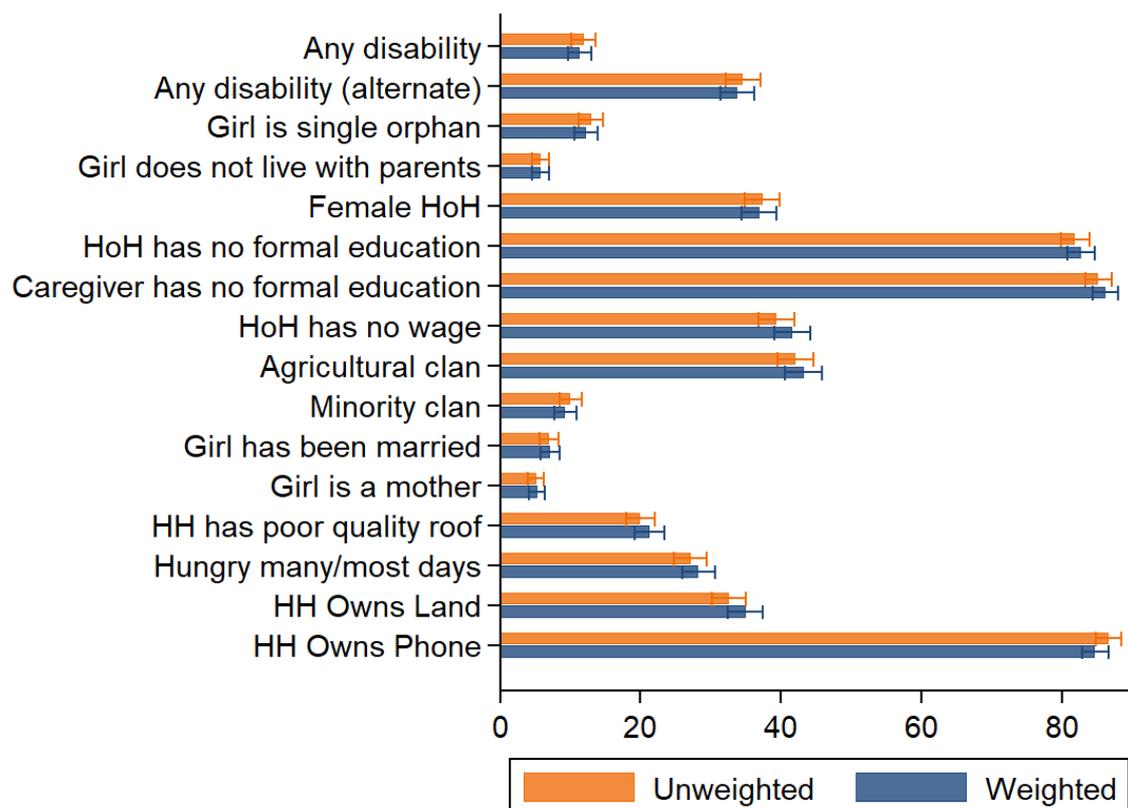
²²⁰ Empirically, this decision is validated by the baseline data, which shows high degrees of clustering within learning institutions (e.g., separating formal schools, ABE centres, and NFE centres). Institution type is a very strong predictor of learning scores, even after controlling for their shared sampling point or geographic location. Moreover, intra-cluster correlation is just 0.14 when calculated across geographic locations ($n = 58$), but 0.27 when calculated across the 108 location-institution combinations. The higher intra-cluster correlation observed when considering institution type makes clear the importance of accounting for clustering within institutions, rather than sampling points/locations alone.

throughout this report are constructed to adjust for unequal cluster sizes, to ensure schools are given equal weight in the analysis – a necessity, given that institutions targeted for supplemental data collection were not selected randomly.

Survey weights force changes in sample composition. Given the discussion in the previous two sections, it is worth considering the extent to which weighting alters sample composition. Note that our descriptions of the sample above did not include weights.

The net effect of the weights on the composition of the sample is generally very small, as shown in the figure below. Weighted estimates of sample composition are never statistically distinguishable, at the 5 percent level, from unweighted estimates, illustrated in the figure by the fact that the confidence intervals around each estimate overlap one another in each instance. For example, our survey weights shift the share of the sample composed of minority group members from 10.1 percent down to 9.3 percent, but this difference does not approach conventional levels of statistical significance.

FIGURE 24: WEIGHTING AND SAMPLE CHARACTERISTICS



Survey weights can, of course, shift outcomes of interest, in addition to shift demographic characteristics. As an example, Somali literacy scores are approximately 1 point higher when we use survey weights than their unweighted values.

The uniformly small changes in sample characteristics illustrated above are not surprising, given the structure of the sampling weights. Our weights range from a value of 1 to 2.125, a narrow range for survey weights in almost any context. The majority of respondents (69.2 percent) are given a weight of 1, with only small shares of the sample (6.2 percent) receiving weights greater than 1.5. Weights are arithmetically applied, meaning

that individuals with weight of 2.125 are "counted" approximately double in the sample compared to respondents with a weight of 1.0. However, given the small number of respondents to which these weights are applied, the net effect on sample characteristics and outcomes is typically quite small.

