Project Evaluation Report

Report title:	Somali Girls' Education Promotion Project–Transition Midline Evaluation Round 2
Evaluator:	Consilient
GEC Project:	Somali Girls' Education Promotion Project–Transition (SOMGEP-T)
Country:	Somalia
GEC window:	GEC-Transition
Evaluation point:	Midline 2
Report date:	August 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.

Girls' Education Challenge

UKaid from the British people

consilient

SOMALI GIRLS EDUCATION PROMOTION PROJECT - TRANSITION

August 28, 2020
Version 5

MIDLINE EVALUATION - ROUND 2

Prepared By:

JUUSO MIETTUNEN, PH.D.
BRENTON D. PETERSON
SEAN ROBERT

care

Girls' Education Challenge

UKaid
from the British people

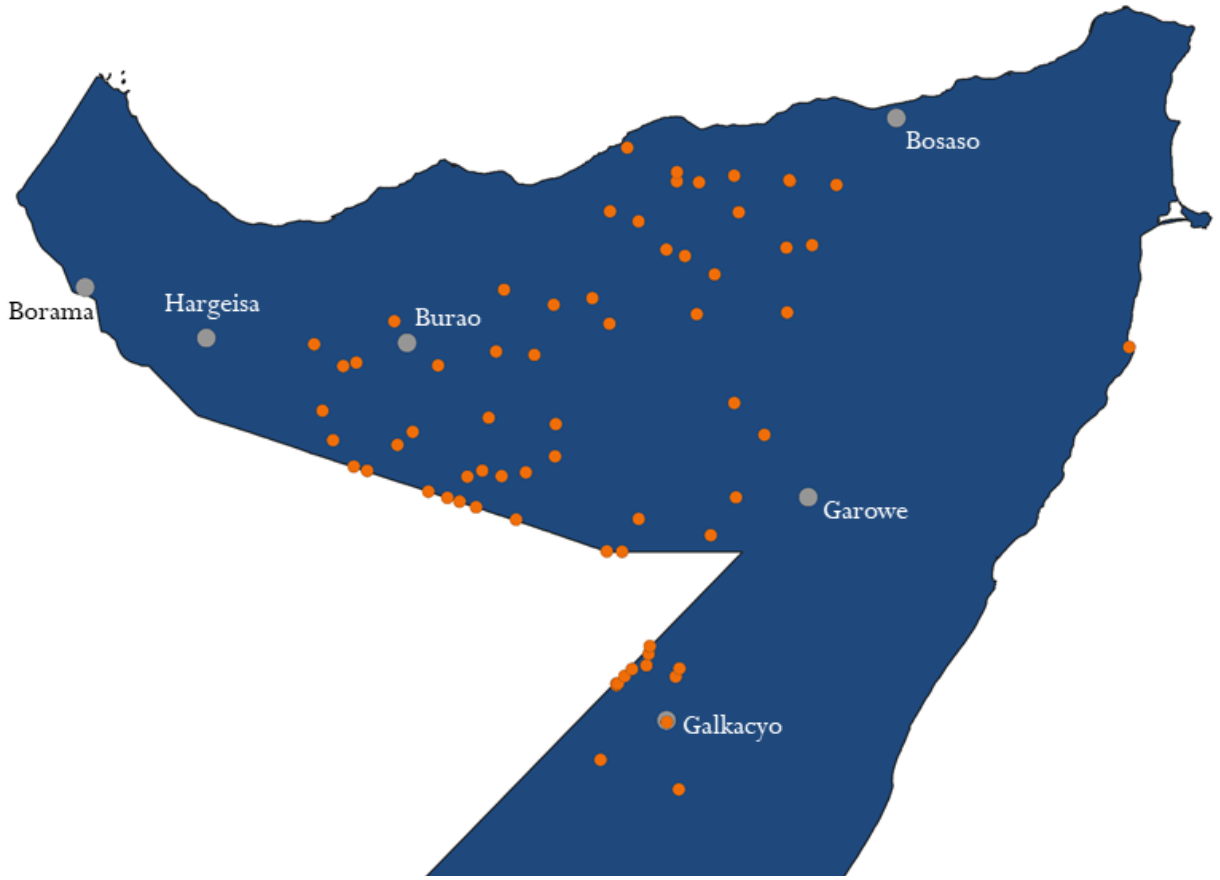
TABLE OF CONTENTS

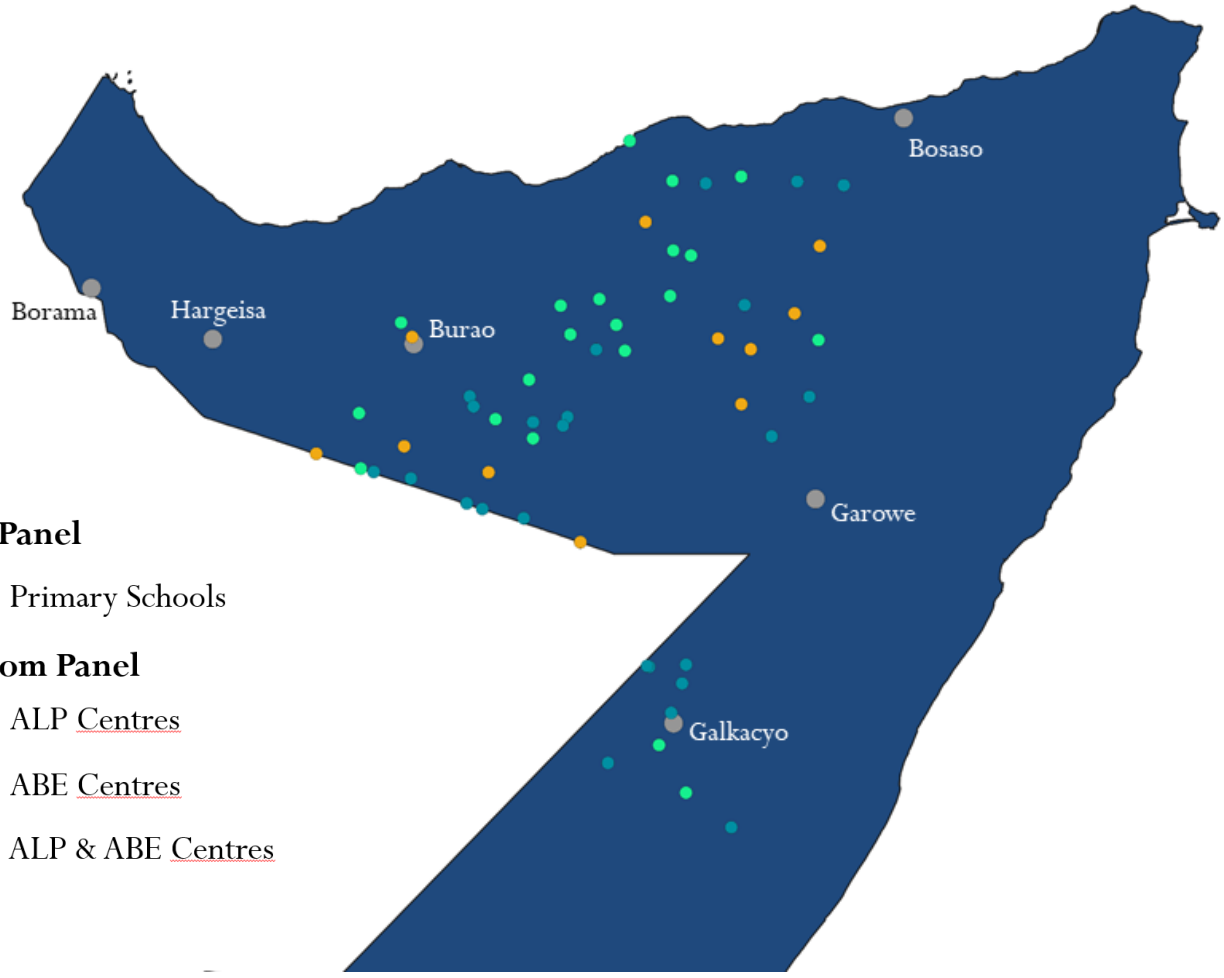
LIST OF TABLES

List of Figures

ABBREVIATIONS

ABE	Accelerated Basic Education
ALP	Alternative Learning Program
CEC	Community Education Committee
EGRA	Early Grade Reading Assessment
EGMA	Early Grade Math assessment
FGD	Focus Groups Discussion/s
GEC	Girls Education Challenge
GEC-T	Girls Education Challenge - Transition
GEF	Girls Empowerment Forum
GwD	Girls with Disabilities
PSUs	Primary Sampling Units
PPS	Probability Proportionate to Size
SOMGEP-T	Somali Girls' Education Promotion Project -Transition
DFID	UK's Department for International Development
YLI	Youth Leadership Index
WHO	World Health Organisation





Top Panel

Primary Schools

Bottom Panel

ALP Centres

ABE Centres

ALP & ABE Centres

EXECUTIVE SUMMARY

Background

The long-term goal of SOMGEP-T is to bring about sustainable improvements to the learning and transition outcomes of marginalised Somali girls. Marginalised girls who are targeted under SOMGEP-T are expected to exhibit meaningful improvements in learning outcomes (literacy, numeracy, and financial literacy) and transition outcomes (transition rate) as compared to a comparison group; targeted schools, communities and government institutions are expected to demonstrate indications of sustainability. The project targets the underlying causes of marginalization, specifically through influencing stakeholder attitudes and promoting social change at the household, school, community and policy/governance levels.

In addressing the barriers to girls' education in Somalia, SOMGEP-T will focus on four key domains of change, or direct outputs: (1) improving access to post-primary options; (2) fostering supportive school practices and conditions for marginalised girls; (3) promoting positive shifts on gender and social norms; (4) enhancing the capacity of Ministries of Education to deliver quality education. According to SOMGEP-T's Theory of Change (ToC), positive changes in these outputs will lead to improvements in four intermediate outcomes – school attendance; school governance; teaching quality; and girls' life skills. The combination of the four intermediate outcomes will in turn contribute to positive changes in the number of girls acquiring literacy, numeracy, and financial literacy skills through accessing and completing a quality primary and secondary education, either through formal or non-formal education. The combination of changes in these four intermediate outcomes is also expected to contribute to the sustainability of the expected results through shifts in practices at community, school, and system levels.

The SOMGEP-T evaluation employs a quasi-experimental design, tracking a longitudinal panel of girls (and their schools) from a baseline that was conducted in late 2017 and through a first-round midline, completed in late 2018. This report documents the state of affairs at the second-round midline, compared primarily to the baseline. This round's sample includes 69 primary schools, split between 37 intervention and 32 comparison schools. In total, 1,290 cohort girls from the baseline were re-contacted and interviewed successfully. The report also presents baseline findings for a new sample of girls – girls enrolled in Accelerated Basic Education (ABE), with a sample size of 482 girls, selected from 35 distinct ABE centres. Finally, the evaluation also reports findings for Alternative Learning Program (ALP) girls recruited at last year, of which there are 336 in this round, split among 32 distinct ALP centres.

Learning

Three general observations emerge from the aggregate learning analysis. The first is the program's apparent impact on financial literacy. When using the pure longitudinal panel of all the individuals who overlap between Baseline and Midline Round 2, the impact is a substantive 8.4 percentage in favour of intervention schools. Secondly, increases in numeracy outcomes are systematically higher among intervention girls. The panel consisting of girls who have been enrolled since baseline has improved their results on average by 4.6 points more than the comparison group since the baseline. This divergence has almost entirely occurred between the two midline evaluation rounds, likely because impacts of the program in this regard are not immediate. Thirdly, despite indications of program impact in financial literacy and numeracy, literacy outcomes in comparison schools have often shown more marked improvement than intervention schools since the baseline.

Transition

Transition rates have increased sharply since the baseline in both intervention and comparison communities. Successful transition increased 3.2 points in intervention communities, over and above, comparison communities. The impact is not limited to one particular type of girl or one particular transition pathway, either; the program has increased within-school grade progression, reduced dropout rates, and increased enrolment into non-formal education, when judged relative to comparison communities.

ABE Girls Baseline

ABE girls face many challenges in terms of attending or staying in school. These include displacement, health conditions, poor economic conditions, and caregivers with little or no formal education. The gaps between ABE girls and their ISG or ALP counterparts are large in Somali literacy (-49 percent compared to ISG girls and -35 percent compared to ALP girls), numeracy (-30 and -23 respectively), and financial literacy (-17 and -15 respectively). The existence of a large literacy gap between ABE girls and their peers was expected, given the unique challenges that ABE girls face and that the ABE program was their first opportunity to attend formal school. However, these learning outcomes begin splintering at the most basic levels of literacy, emphasizing that ABE girls face a steep learning curve if they are to close the gap with their peers.

Attendance

The program appears to be having a positive impact on attendance rates, in general, though these effects are noisy in a statistical sense. The attendance analysis used three different quantitative data sources—derived from entirely different data collection tools—to triangulate attendance rates at ML2 and in previous rounds. The findings across these sources were consistent: attendance rates in intervention schools had increased vis-à-vis comparison schools in small but meaningful ways. Our best estimate of the program's impact on attendance is that girls' attendance has increased 1.4 percentage points from baseline to midline – a substantively meaningful impact, though smaller than might be desired after two years of program implementation. There is no evidence that the program is having particularly large or minimal effects among specific subgroups of girls or schools, aside from the fact that the program's impact is significantly higher in Somaliland than in Puntland.

Teaching Quality

When looking at the indicators of teaching quality, there is some reason to be concerned about the overall effect of the intervention on teaching quality. While perception of teaching quality remains high, indicators of student-centred pedagogy do not support this optimism. When all indicators were aggregated into an index of classroom activity and participation, schools scored higher in the baseline than they did during the second midline. Self-reported use of formative assessments increased dramatically from baseline to the second midline. However, the proportion of teachers who reported having documentation of their use remained unchanged – founding reason to be cautious about widespread use of formative assessments among intervention classrooms.

Similarly, observed use of physical punishment in the classroom has gone from 'common' at the baseline to 'unobserved' during ML2. However, intervention girls reported that the use of corporal punishment by teachers in the classroom rose from 23.2 percent at baseline to 30.9 percent at ML2, whereas, students at interventions schools who reported that their teacher would punish them for getting things wrong in a lesson increased from 40.8 percent to 54.7 percent. These results suggest that without persistent reinforcement of teaching best-practices, long-term sustainability of program interventions may be diminished.

School Management and Governance

For school management, several important variables have improved since the baseline. According to the head teacher surveys, all intervention schools now have an active CEC, and nearly 72 percent of intervention schools have developed their management plan. The same increase is not witnessed in comparison schools where the schools that have a management plan remains nearly unchanged.

However, the findings are mixed on progress towards program goals as per the logical framework. One of the goals was to increase CEC monitoring of student attendance to 80 percent by ML2. According to both caregivers and headteachers, fewer than 80 percent of CEC's are fulfilling this responsibility (68.2 percent and 65.6 percent, respectively). Similarly, in terms of monitoring retention, only 15.9 percent of parents in intervention areas say that the CEC is engaged in this and 37.5 percent of head teachers share the same opinion. Meanwhile, the program has met its objective of 30 percent of CECs taking action against corporal punishment and other child protection issues.

Finally, while the CECs' consideration of the needs of marginalised sub-groups is evident, most committee members viewed addressing barriers to their education as beyond their means due to the very severely limited financial resources at their disposal. Although findings from the head teacher survey indicate at least partial improvements – as 78.3 percent of school management plans in the intervention group now include plans to follow-up with dropouts – this, nonetheless, calls into question the feasibility of some of the factors required for the long-term sustainability of the project. In particular, targeting the most marginalised is largely seen as beyond the means (in terms of financial or other resources) of the CECs. As such, it is most likely that members of these groups will continue to fall through the cracks.

Life Skills

As with several other outcomes (e.g., attendance, learning) studied in this evaluation, the program's impact on life skills appears to be positive, meaningful in a substantive sense, but too small to distinguish it, statistically, from a null result. Using CARE's Youth Leadership Index (YLI) as a metric for problem-solving abilities, self-confidence, organisation, and ability to motivate their peers, the program has increased girls' scores by around 1.9 points on a 100-point scale. While this effect is small, it is fairly clear – cohort girls in intervention and comparison schools had nearly identical scores at baseline, and girls in intervention schools have improved markedly since that time, while scores for girls in comparison schools have remained flat. Notably, we interpret these results unequivocally as the program's causal impact on YLI scores: thanks to the quasi-experimental design, the findings are intended to represent the program's true impact; moreover, our analysis suggests that girls with greater exposure to programming—those who remained enrolled across rounds—saw larger improvements in self-confidence, consistent with the program itself being the mechanism of change. As with attendance rates, we find that participation in the GEF was associated with even larger gains in self-confidence and leadership skills.

Our results are less clear-cut when it comes to the life skills index, which differs from the YLI in the specific questions asked, but often has conceptual overlap with the YLI. On this second index, the program does not show any meaningful impact over time. However, when we split the life skills index into its constituent parts, the evidence suggests that the program improved the self-confidence aspects of the life skills index. The conclusion that emerges is that the program—and the GEF intervention in particular—has had a positive impact on girls' self-confidence and confidence-adjacent outcomes, such as willingness to speak up at home and in school. But the program has had less, or no, impact on other aspects of "life skills", such as feelings of loneliness, agency over life decisions, and desire to stay in school.

Community Attitudes

In the context of the SOMGEP-T evaluation, community attitudes are expected to have effects on transition rates by encouraging re-enrolment of OOS girls and continued enrolment for girls already enrolled. The value that caregivers place on girls' education has improved as a result of the program, though the improvements have not been dramatic. In general, improvements have been largest from the perspective of girls: the program has increased the share of girls who feel they receive support from their family to stay in school and perform well by about 7.2 percentage points since baseline, over and above the improvements seen in comparison schools. Caregivers also report an increased role of girls in making decisions that influence their education, and an increased belief that a girl's education is a worthwhile investment, even if funds are limited. However, the impacts among caregivers are of small or moderate size—none are statistically significant, though some approach that standard—and head teachers' impressions of attitudes in their community have not improved as a result of the program.

School-Related Gender-Based Violence

Girls in SOMGEP-T schools report a wide range of experiences related to gender-based violence and, more generally, safety in their communities and schools. While overall perceptions of safety en route to, and at, school have improved over time, girls still face a number of threats to their physical safety, and endure verbal and sexual harassment in both contexts. Some girls describe going to school in the shadow of outright violence, in the form of clan-based conflicts; at least one school is partially occupied by military personnel keeping the peace locally. Outside of these extreme cases, the most prominent threat to girls' safety seems to be gendered harassment and assault by boys and men in their community and by their peers at school.

1. BACKGROUND

1.1 PROJECT CONTEXT

The Somali Girls Education Promotion Project – Transition (SOMGEP-T) is a Girls' Education Challenge (GEC-T) project focused on improving educational outcomes among some of the most marginalized girls in Somalia and Somaliland. In this section, we provide a brief overview of the historical, economic, and educational context in which the program is being implemented. This overview is concise, and we refer readers to the baseline and previous midline report for a more detailed discussion. Instead, we provide a brief overview and then describe contextual factors that have changed since the baseline, as these changes may influence program implementation and our analysis of program impacts since the baseline.

While traditional overviews of recent Somali history tend to focus on the overthrow of President Siad Barre in 1991 and the subsequent decline into a long-running civil war, the historical background of relevance to SOMGEP-T does not neatly fit this narrative. As the maps of the sample locations – provided on previous pages – show, SOMGEP-T is primarily being implemented in a wide swath of central-to-eastern Somaliland, running from the area around Burao to the disputed border regions between Somaliland and Puntland. The primary exception is a set of schools in the western portion of Galmudug, near the Ethiopian border.

The history of these areas is affected by the civil war that began in earnest in 1991, but a prior history of conflict is also relevant. Beginning in 1987, the Barre regime and a rebel movement operating in Somaliland – the Somali National Movement – engaged in conflict throughout Somaliland, which prompted hundreds of thousands of residents to flee into neighboring Ethiopia. The Barre regime engaged in indiscriminate bombing; one consequence was the destruction of state authority in Somaliland, and the destruction of most buildings – including schools – in most urban areas. Following the ouster of Siad Barre in 1991, Somaliland declared its independence, and has been a de facto independence – but universally non-recognized internationally – since that time.

In contrast, Puntland and Galmudug were more directly impacted by the long-running Somali civil war that began in 1991 and has continued – depending on one's definition – to the present. Puntland declared its status as an autonomous state in 1998, but it is also a Federal Member State (FMS) in the Federal Government of Somalia (FGS). Galmudug is, likewise, a Federal Member State.

The different historical trajectories of governance in the three regions shape many aspects of life therein, influencing conflict dynamics, government administration and other factors that have the potential to influence SOMGEP-T programming and performance. In terms of conflict, there are two vectors of conflict that are particularly relevant: the territorial dispute between Somaliland and Puntland, the breakaway state of Khatumo, and the related clan conflicts that occur in the area; and, in Galmudug, the conflict and destabilization by al-Shabaab, the conflict between Ahlu Sunna Waljama'a – a moderate Sufi militia and other political forces especially supporting the FGS efforts in State making in Galmudug.

As noted above, many SOMGEP-T schools are located in the eastern portion of Somaliland, parts of which are disputed with the government of Puntland. Much of this region has limited state presence from either side in the dispute, a fact which underscores and reinforces the nature of the border disagreement. In addition, in 2012 a secessionist movement arose in Sool, Sanaag, and Cayn, in southeastern Somaliland, along the

Ethiopian border, which sought to gain independence from both Somaliland and Puntland. This movement has receded, and Somaliland exercises partial control over the area at this time.

SOMGEP-T schools operate in the midst of many of these conflicts. For instance, following the 2017 Somaliland elections, long-simmering tensions between two clans in Ceel-Afweyn morphed into open fighting, which has continued to the present. As recently as March, the town of Ceel-Afweyn was effectively divided between the two groups, and travel from one side of town to the other was risky. Conflict has prompted migration out of the town and into neighboring towns and rural villages. Rural and peri-urban areas in the vicinity of Ceel-Afweyn have also been impacted, and many members of the professional class – including teachers – have left the area.

In another instance of conflict in this region, the area around Yubbe has recently been affected by intermittent violence between different clans. While the proximate cause appears to be control over natural resource wealth (especially in the form of gold mines) in the area, the dispute is oriented around deeper divisions that tie into politics, support for or opposition to secession, and clan loyalties. Again, this has consequences for SOMGEP-T schools, which are not present in Yubbe, but which are in the rural areas in its vicinity.

In Galmudug, an entirely different set of conflict dynamics potentially impact SOMGEP-T programming. Ahlu Sunna Waljama'a (ASWJ), a Sufi militia that has fought al-Shabaab for many years, has recently come into conflict with the FGS, which has sought to disarm the militia and displace it as the sole government and armed force in Galmudug. Most recently, conflict arose in Dhusamareb, with the Somali National Army taking control of the town from ASWJ. While this conflict has not directly impacted Galkayo and its environs – where most of the project's evaluation schools in Galmudug are located – the potential for spillover exists.

Beyond conflict itself, the complex history and territorial disputes in the program's operating area affect the manner in which education is administered. SOMGEP-T is attempting to work actively with the Ministries of Education in each respective area, but the proliferation of ministries is likely to make coordination more difficult and increase the logistical and financial burden of interventions targeting them. As one example, consider the nature of curricula development: Somaliland and Puntland have previously established grade-level standards for their students, and a curriculum tied to those standards. Within the last year[1], the FGS has also completed a curriculum development exercise, and has begun shipping textbooks and other materials to schools under FGS jurisdiction. SOMGEP-T schools, therefore, operate under at least four partially-overlapping sources of administration.

Before we discuss the extent of marginalization experienced in targeted communities *relative* to the rest of Somalia and Somaliland, it is important to understand the challenges facing education in the region more broadly. While geographically-disaggregated data is extremely limited, the Somali context as a whole presents a challenging environment for girls' education. To illustrate: just 36.1 percent of girls 15-24 years old in Puntland were capable of reading a short, common sentence, in 2011.[2] The Multiple Indicator Cluster Survey from 2011 in Puntland also documented high rates of child labour, and showed that the majority of children – 75.2 percent in the sample – had experienced "psychological aggression or physical punishment during the last month". The state of education and child development in Somaliland – described in a separate Multiple Indicator Cluster Survey (MICS) report from 2011 – was not dramatically better: just 44.1 percent

[1] Note from the project: The new FGS National Curriculum Framework and syllabi was completed in 2017. The new textbooks were initially developed in 2018 started being distributed in 2019.
[2] UNICEF and Ministry of Planning and International Cooperation. 2011. "Multiple Indicator Cluster Survey – Northeast Zone, Somalia, 2011."

of girls 15-24 years old were literate at a very basic level, and just 18.3 percent of girls were entering primary school at the appropriate age level.

Although the most recent rainy season (Deyr, occurring in mid-fall) was productive, the drylands of Somalia and Somaliland are still in the general grip of a widespread drought. The drought has prompted significant displacement, with pastoralist families who lose their livestock to drought migrating to urban centers, often enrolling their children in school for the first time. This in-migration to towns has occasionally overwhelmed local schools in terms of capacity, while schools in rural areas are nearly empty. Internal displacement, by some estimates, affects 40,000-80,000 people in Somaliland, 130,000 people in Puntland, and 870,000 people in South and Central Somalia.[3] In the areas studied here, the region with the highest concentration of Internally Displaced Persons (IDPs) is Mudug (Galmudug), and portions of Puntland, such as Bari.

Even relative to these general challenges, the context in which SOMGEP-T is being implemented is especially daunting. Not only is the evaluation sample exclusively rural, the particular rural areas tend to be remote. As mentioned previously, eastern Somaliland is not only rural, but also remote, and often beyond the scope of state control. These issues come through in the findings in this report, insofar as Ministry of Education (MoE) officials are often unable to visit SOMGEP-T more often than once per year, because they lack access to their own transportation[4]. Other schools are in the Ethiopian borderlands, where state presence is minimal.

The economic characteristics of SOMGEP-T communities underscore the difficulties. Sool and Sanaag – which together comprise 48.0 percent of the sample of SOMGEP-T girls – are the two regions of Somalia and Somaliland with the highest concentration of nomads as a share of their population. In Sool, the Population Estimation Survey of Somalia estimated that 57.3 percent of the population was nomadic, and this same rate was 64.8 percent in Sanaag.[5] Even less nomadic regions represented in the sample – such as Mudug, in Galmudug, and Toghdeer, in Somaliland – have nomadic populations comprising more than 20 percent of their populace. Moreover, the rural areas of these regions have much *higher* rates of nomadism, and these are precisely the areas where SOMGEP-T schools are located.[6]

Compared to the already-low levels of educational attainment and primary enrolment reported in Somaliland and Puntland, as a whole, the specific regions targeted by SOMGEP-T fare worse. For instance, while Somaliland's female (15-24 years) literacy rate was 44.1 percent in 2011, literacy was just 36.5 percent in Sanaag and 37.7 percent in Sool, two of the most heavily-represented areas in the evaluation sample. In Sool, just 11.1 percent of girls entered primary school on-time, and attendance rates were the lowest of any region in Somaliland.

In essence, SOMGEP-T schools operate in an environment of extreme resource limitations, population mobility, and economic and environmental fragility, even by Somali standards. While geographically-disaggregated data is extremely limited, the Somali context as a whole presents a challenging environment for girls' education.

Given these aggregate figures, we expect the situation in the areas targeted by SOMGEP-T to start from even lower baselines. SOMGEP-T communities face a cluster of correlated challenges, from reliance on rainfall for most livelihoods, to nomadism, to lack of access to state officials, that all inhibit aspects of educational and economic development. Even more problematically, they interact in ways that reinforce the challenges:

[3] Drumtra, J. (2014). Internal Displacement in Somalia. Washington DC.: Brookings Institution. Retrieved from Brookings Institution: https://www.brookings.edu/wp-content/uploads/2016/06/Brookings-IDP-Study-Somalia-December-2014.pdf.
[4] Note from the project: SOMGEP-T provides transportation and logistical support for MoE visits.
[5] United Nations Population Fund. 2014. *Population Estimation Survey 2014: For the 18 Pre-War Regions of Somalia*.
[6] For example, SOMGEP-T schools in Mudug are concentrated in the dry pasturelands along the Ethiopian border.

remote communities do not receive assistance from the government; but these communities are too poor to self-fund schools; they are less connected to international diaspora networks that might help with fundraising; they are more dependent on rainfall, which leaves them susceptible to wide swings in economic outcomes from year to year; but they are too isolated to engage effectively with informal social safety nets that operate in many towns in the region. In short, these communities face an assortment of challenges that are mutually reinforcing and particularly difficult to overcome.

1.2 Theory of Change and Assumptions

Theory of Change Overview

The long-term goal of SOMGEP-T is to bring about sustainable improvements to the learning and transition outcomes of marginalised Somali girls. Marginalised girls who are targeted under SOMGEP-T are expected to exhibit statistically significant improvements in learning outcomes (literacy, numeracy, and financial literacy) and transition outcomes (transition rate) as compared to a comparison group; targeted schools, communities and government institutions are expected to demonstrate indications of sustainability. To achieve its long-term outcomes and create a more supportive environment for girls, the project will focus on addressing the underlying causes of marginalisation through influencing stakeholder attitudes and promoting social change at the household, school, community and policy/governance levels.

SOMGEP-T defines marginalised girls as those who face the intersection of multiple barriers to access education and once enrolled, to remain in school after Grade 3.[7] SOMGEP's studies have identified that the barriers marginalised girls face include extreme poverty, pastoralism, displacement, being over age for their grade, a high degree of exposure to violence/ conflict, orphan status, disability, belonging to a minority clan, and having an illiterate mother (who is often experiencing financial hardship as a female head of household).

More generally, barriers to girls' education in Somalia can be categorized as demand-side barriers and supply-side barriers. Demand-side barriers include traditional gender and social norms (early marriage, chores, girls' low agency, gender-based violence (GBV)), poverty and high vulnerability to the negative effects of climate change, high absenteeism (seasonal migration, chores), perceptions of disconnect between education and the local market, and armed conflict. Supply-side barriers include limited provision of secondary education and poor infrastructure, limited number of qualified teachers, low teacher capacity to teach higher numeracy skills, Somali literacy and English as a second language, lack of catch-up opportunities/ remedial education for pastoralist children, and limited capacity of school leadership and education officials to address absenteeism, dropout and poor learning outcomes.

In addressing the barriers to girls' education in Somalia, SOMGEP-T will focus on four key domains of change, or direct outputs: (1) improving access to post-primary options; (2) fostering supportive school practices and conditions for marginalised girls; (3) promoting positive shifts on gender and social norms; (4) enhancing the capacity of MoEs to deliver quality education. According to SOMGEP-T's Theory of Change (ToC), positive change in these outputs will lead to improvements in four intermediate outcomes – school attendance; school governance; teaching quality; and girls' life skills. The combination of the four intermediate outcomes will in turn contribute to positive changes in the number of girls acquiring literacy, numeracy and financial literacy skills through accessing and completing a quality primary and secondary education, either through formal or non-formal education. The development of financial literacy skills in

[7] CARE SOMGEP-T MELF Final

particular is hypothesised to depend on the combination of (i) teachers' improved capacity to teach numeracy; (ii) increased agency and voice enhancing participation in class; and (iii) opportunities to practice financial literacy and numeracy through savings groups as well as daily transactions. The combination of changes in the four intermediate outcomes is also expected to contribute to the sustainability of the expected results through shifts in practices at community, school and system levels.

EXPECTED OUTCOMES

Long-term outcomes:

(1) Learning: The number of marginalised girls supported by GEC with improved learning outcomes (literacy, numeracy, financial literacy).
(2) Transition: The number of marginalised girls who have transitioned through key stages of education, training, or employment.
(3) Sustainability: The changes brought about through the project which increase learning and transition through education cycles are sustainable at the community, school, and system levels.

Intermediate Outcomes:

(1) Attendance
(2) School governance
(3) Improved quality of teaching
(4) Life skills development

Outputs and Key Activities

Through its key activities, SOMGEP-T will deliver four key outputs to improve the learning and transition outcomes of marginalised Somali girls and empower them to engage in the local economy and decision-making processes in the future. SOMGEP-T's outputs and activities are outlined below.

Output 1: Improved access to post-primary options

Girls will be supported to transition into either formal secondary school through grants (bursaries)[8] for poor families, or into accelerated education through the provision of an accelerated learning program (ALP) developed in partnership with the MoEs and communities through the development and strengthening of community education committees (CECs).

Key Activities:

- Work with MoEs to develop and implement ALP
- Develop girls' life skills in upper primary through ALP, including leadership skills, financial literacy and training on business selection and management of income generation activities
- Develop CECs skills and activities to improve retention and transition
- Provide partial grants to girls from poor families
- Equip two boarding schools for girls with furniture / learning materials and promote girls' enrolment

Gender Equity and Social Inclusion (GESI) Transformative Adaptations, Gender:

- Prioritise the recruitment of female teachers for ALP.

[8] School fees are paid directly to schools; in addition, girls receiving grants also receive uniforms and basic support items from the projects

- Advocate for teachers training colleges to have a quota for female graduates.
- Lobby for inclusion of women in CECs structure and encourage effective participation
- Work with CECs to promote participatory and inclusive review of SIP's and ensure that the SIP clearly outline feasible and relevant initiatives for promoting girls' education

GESI Transformative Adaptations, Social Inclusion:

- Provision of Alternative Basic Education (ABE) to support overage girls who have never attended school to transition into primary school (response to findings showing the influx of a large number of out-of-school girls who never attended school in targeted areas).
- Mapping the spread of nomadic and pastoralists households and develop appropriate strategies for reaching these socially excluded sub-populations
- CEC and host communities intensify efforts to enrol girls from nomadic/pastoralist households

Output 2: Supportive school practices and conditions for marginalised girls

The project will boost numeracy outcomes and English skills among primary and secondary students, providing remedial support to struggling students as well as those with high absenteeism rates, particularly pastoralist girls, and supporting the school leadership to track attendance, learning, retention and transitions, therefore increasing the chances of marginalised girls building foundational skills, completing primary school and succeeding in secondary education.

Key Activities:

- Train teachers on improved delivery of literacy and English language, supported by digital content in all 148 primary and 51 secondary schools
- Train teachers on improved delivery of numeracy in all 148 primary and 51 secondary schools, strengthening their capacity to teach students to apply basic numeracy concepts in daily life, including in basic transactions and business plans (financial literacy)
- Train teachers to provide structured remedial support to students at primary and secondary level
- Train and coach teachers to deliver the ALP curriculum, including financial literacy, leadership skills development, adolescent savings and basic principles of business selection and management
- Construct additional classrooms in remote primary schools; build water facilities in new secondary schools; and provide solar chargers for mobile devices/ tablets and sanitary pads to schools
- Incorporate life skills and financial literacy training into Girl's Empowerment Forums (GEFs) and Boys' Empowerment Forums (BEFs)
- Provide career guidance in schools

GESI Transformative Adaptations, Gender:

- Review of inactive empowerment forums and revitalises non-functional empowerment forums
- Identify women role models in private sector to participate in school-based career guidance sessions
- Security risks as a result of lack of perimeter wall - the project need to explore options with CECs, especially use of durable, less costly, locally available and environmentally friendly materials

GESI Transformative Adaptations, Social Inclusion:

- Training of teachers on ABE delivery.
- Remedial sessions to be rolled out at the same time in a number of schools, preferably starting with remote schools to maximise exposure.

- Schools with few and less qualified teachers will be prioritised in all trainings.

Output 3: Positive shifts on gender and social norms at community and individual girl level

Through promoting positive shifts on gender and social norms, the project will create an environment where girls and boys are equally supported to attend school, their skills are valued, there are higher expectations for their achievement, and where girls and boys are safe from harmful practices. It is expected that these activities will not only encourage parents to send their girls to school but will also encourage girls to stay in school by creating a safe environment for them and emphasizing the importance of education in relation to other social pressures that typically cause girls to drop out, such as marriage.

Key Activities:

- Engage community-level stakeholders including religious leaders, women's groups, men and boys
- Expand and strengthen GEFs and create BEFs to develop leadership and mentorship skills
- Provide adult literacy and financial literacy classes for mothers
- Support the financial empowerment of mothers through savings groups (VSLA), business selection, and business coaching and mentoring

GESI Transformative Adaptations, Gender:

- Greater inclusion of men as agents of change to spearhead the mobilisation of the community to promote girls' education

GESI Transformative Adaptations, Social Inclusion:

- Community awareness to reach various sub-groups including nomadic and pastoralists, rather than focusing only on host communities

Output 4: Enhanced MoEs' capacity to deliver quality and relevant formal and informal education

MoEs' staff, local education officers will be supported to develop robust governance and support structures, taking an active role in improving girls' retention and transition rates, overseeing the implementation of quality standards and data management systems, and identifying and addressing barriers to learning using a gendered lens. MoEs are uniquely positioned to send a strong, positive message about the importance of girls' education to the FGS and the governments of Somaliland and Puntland.

Key Activities:

- Strengthen Gender Units capacity to improve girls' education outcomes through trainings, development of action planning and provision of incentives to retain the gender focal points especially in rural areas
- Support quality assurance and standards (QAS) functions at all MoE levels
- Provide support to Regional Education Officers (REOs) and District Education Officers (DEOs) to mainstream improved teaching practices and address retention/ transition issues
- Work closely with MoE on non-formal education (NFE) for mothers and entrepreneurship skills for girls
- Development of project Information, Education and Communication (IEC) materials in conjunction with MoE for use at stakeholder advocacy and promotion events

GESI Transformative Adaptations, Gender:

- Include the recruitment of female, teachers, head teachers, Regional Education Officers (REOs) and District Education Officers (DEOs) in the advocacy activities.

GESI Transformative Adaptations, Social Inclusion:

- Support Regional Education Officers (REOs) and District Education Officers (DEOs) to access remote schools, which are often neglected because of lack of transport by incorporating school visits by the REOs during project implementation

In addition to these initially planned activities and GESI-relevant adaptations, the following general adaptations were made by the project in response to baseline findings:

Adaptation 1, to improve learning especially in numeracy:

- Extensive coaching sessions for teachers in struggling schools (from April to May) before schools close using the numeracy Level 1 module developed in Phase One for basic number operations and coaching guidelines developed recently. This will help build teachers' delivery of the content and students' foundational skills in maths.
- Subject to availability of funds, the project is exploring how to achieve maximum exposure by running remedial learning sessions during the June to August school break. Funds will be particularly required to commit teachers on monthly performance-based contracts on number of lessons delivered during the three months. The remedial support will be guided by grade specific tasks identified through the baseline process.
- The project will also consider inclusion of participatory and interactive basic maths games in GEF and BEF manuals. The original plan was to only focus on life skills and include aspects of financial literacy in the GEF manuals. This will be borrowed from the Numeracy and ALP modules.

Adaptation 2, in response to low baseline achievement levels in literacy:

- Trimming the English literacy test (removing the upper reading comprehension task and the written tasks),
- Addressing limited reading fluency and vocabulary, matching the electronic platform with a clear messaging to the teachers on its use (when, which dosage, with whom), as many of the students are not learning it in school at all.
- Exploring positive deviants (the five comparison schools that performed very well, particularly Ilays), to learn what are the successful strategies there.

Adaptation 3, involves the project adopting a stronger (and more specific) monitoring of CECs which encompasses the periodic assessment of CEC functionality and fidelity of implementation. In order to provide additional support to CECs, adaptations will be tailored to respond to the specific issues arising from the assessments. The identified issues will be addressed during the coaching of CECs. Additionally, the project intended to:

- Further query data to see where the problem is in relation to the functionality of CECs.
- Develop a coaching guide for CECs to be reviewed and include more details for staff/MOE.

Adaptation 4, in order to improve GEF impact, includes:

- Tracking GEF activities and verifying their functionality, as there is a likelihood that many aren't meeting/ functional after the pioneer group of older girl's transition to higher levels of education.

- For functional groups, use self-monitoring/ reflection/storytelling tools whenever possible, as opposed to more traditional checklists.
- Non-functional groups will require mobilization and refresher trainings, identifying and addressing reasons for disbanding/ lack of functionality.
- Focus the Gender Units Annual Work Plan on supporting the GEFs as a matter of priority.

Adaptation 5, in order to address high proportion of teachers observed using corporal punishment at the baseline, adaptations include:

- Including coaching that models alternative, non-violent ways of disciplining children.
- Mainstreaming alternative, non-corporal discipline strategies throughout project activities: i.e. teacher trainings, teacher coaching, GEFs, training of mentors, CECs coaching, MOE consultations/supervisors/TOTs, introducing score cards for schools in the use of positive discipline, digital platform, ESL, ALP modules, etc.
- Print and distribute the code of conduct for teachers in all schools.

The following adaptations were adopted in light of project monitoring:

- Adaptation 1: Address the exclusion of OOS girls who have never been to school and those who dropped out in Grades 1, 2 or 3
    - Explore the possibility of running ABE classes for OOS girls who reached at most Grade 3.
    - Ensure ALP students of schooling age and who have recently dropped out (less than two years) to enrol back to formal schools.
    - Engage CECs to mobilise resources for an additional teacher salary/incentive.
    - Expand the ALP to other villages where there is need.
- Adaptation 2: Improving the capacity of ALP teachers to teach all four ALP subjects with special emphasis placed on the delivery of English lessons
    - Engage additional facilitator with complementary skills
    - Ongoing tailored coaching with emphasis on gaps to improve classroom practice
- Adaptation 3: Reach out to nomadic and pastoralist households
    - Mapping the spread of nomadic and pastoralists households and develop appropriate strategies for reaching to sub populations to ensure inclusion. Engaging CEC and host communities to intensify efforts to enrol girls from nomadic households as well those from pastoralist families.
- Adaptation 4: Promote the use of local and culturally appropriate visual aids
    - Encourage teachers to use locally available teaching and learning materials e.g pebbles, animal droppings, sticks, bottle caps etc to promote participatory learning.
    - Support teachers and students both to draw pictures related to math subjects and post to class walls (provision of manila papers, drawing colour pencils, posters etc)

- Adaptation 5: Increasing student contact time/exposure
    - Address multiple grade/ split classes - Due to learners being at different levels [especially in NFE, and Mathematics] the exposure time to learning in multi-grade classes is limited. Each class is expected to take one hour; therefore, level one students should be provided with one hour of instruction, subsequently starting a level 2 class, instead of running one-hour multi-grade sessions.
    - The project will discuss with MoE, CECs and school administration about the possibility of having multiple shifts (morning and afternoon sessions for separate groups of students) in schools with insufficient teachers to cover all grades.
- Adaptation 6: Promote the use of VSLA loans and social funds to primarily focus on supporting girls' education
    - Set off plan to increase the awareness of VSLA groups about the importance of girls' education through village events and gatherings and using VSLA as a platform for discussing issues of education equity and socio-cultural norms affecting learning and transition.
    - Create visibility posters for all villages with VSLA groups on social funding borrowing from best practice in livelihood projects.
    - Linkage between VSLA groups and CECs/ school administration
    - Encourage participants to support their children's education by utilizing VSLA loans.
    - Document successful groups and the evidence of how the VSLA mothers used their loan, and the proportion of participants using loans/ savings to support costs related to girls' education.
- And to address baseline findings related to marginalisation:
    - Expand ALP to enrol further 2,345 out-of-school girls within an additional 34 villages, increasing ALP coverage from 76 villages to 110 villages.
    - Provision of two-year Alternative Basic Education (ABE) classes for 2,029 marginalised girls and link them with existing schools to join formal education depending on learning achievement
    - Training CECs across 199 villages in identification of different types of disabilities and support to girls and boys with disabilities
    - Work with CECs to liaise with parents of displaced out-of-school girls and girls with disabilities, provide targeted social support and track their attendance.
    - Assess girls with disabilities for placement in regular schools or referrals to special needs facilities.
    - Provide specialised equipment and learning materials for 300 girls with disabilities. Assistive devices will be provided on the basis of need and may include mobility aids, hearing aids, wheelchairs, glasses and walking equipment. Teaching-learning materials will include large-print textbooks, pen grips/holders, and equipment for cognitive and perception development.

- Train MoE staff and teachers in 199 schools to identify and support girls with disabilities, and train at least 110 teachers and MoE staff on inclusive and special needs education. The training will include basic special education; identification and basic assessment of girls with disabilities; building inclusive classroom environments; guidance and counselling. The training will include residential training and follow up on-site sessions.
- Encourage girls and boys from pastoralist families to participate in empowerment forums to enhance their confidence and address negative stereotypes associated with their way of living.
- Lead annual social mobilisation campaigns in 70 villages to encourage pastoralists to bring their children to school and actively participate in their education.
- Provide psycho-social counselling for development of self-esteem among girls with disabilities, and treatment of anxiety and depression
- Work with CECs and teachers to address corporal punishment, particularly against over-age and displaced adolescents and those who are struggling to learn, and promote community-managed self-monitoring of community efforts in addressing corporal punishment. Encourage teachers to employ positive disciplinary measure to deter corporal punishment.
- Incorporate sessions on identification and support for girls with disabilities in stakeholder, as well as NFE and VSL groups.
- Provide support to VSLA groups to start business upon completion of the VSLA cycle through competitive selection of most viable business ideas.
- Reinforce and encourage CECs to continue supporting need-based "tuition waiver".
- Increase reading time by establishing and supporting community managed reading clubs associated to GEFs/BEFs and promote the use of culturally appropriate local learning materials.

Assumptions

The success of SOMGEP-T is predicated on a number of assumptions which have affected, and will continue to affect, the ability of project staff to carry out, monitor, evaluate, and effect change through project activities. The project's major assumptions include:

- Schools remain open during most of the year; absence of major disruptions (widespread conflict, famine, political disturbances, economic shocks)
- Most schools adhere strongly to the intervention procedures and protocols, ensuring fidelity of implementation
- Project partners adhering to implementation guidelines/protocols
- MoE efficiency
- ABE and ALP acceptance
- High retention of out-of-school girls
- Complementary emergency support in case of severe drought
- Timely deployment of facilitators for ALP and ABE/ teachers are available
- ALP curriculum includes skills considered as relevant in the local job market/ businesses
- Local authorities and religious leaders are supportive

- No major disruptions to government functionality post-elections, allowing for timely implementation
- Absence of major economic shocks
- Absence of major disasters and widespread conflict
- Parents are supportive of girls' participation in GEFs and BEF's

It should be noted that at least two of these assumptions articulated as part of the project MEL-F were violated during the initial phase of implementation, with significant implications for the efficacy of project interventions. Most notably:

- Schools in conflict-affected areas of Somaliland closed for significant periods of time during the past year, resulting in potentially attenuated learning outcomes for all students attending those schools.
- The aftermath of the drought has resulted in ongoing economic hardship for vulnerable households, particularly pastoralists who in many cases lost their main source of livelihood as a result of the drought and are now struggling at the household level with basic issues of food insecurity.

The table below links each intervention to specific intermediate outcomes and provides a comprehensive explanation of how these will in turn contribute to achieving SOMGEP-T's long-term outcomes of learning, transition, and sustainability.

TABLE 1: PROJECT DESIGN AND INTERVENTION

Intervention types	What is the intervention?	What Intermediate Outcome will the intervention will contribute to and how?	How will the intervention contribute to achieving the learning, transition and sustainability outcomes?
Improving access to post-primary options	Work with MoE to develop and implement ALP and ABE	Attendance and life skills development. ALP offers out of school girls and students who are unable to/ do not wish to attend formal secondary school with an alternative option, thereby encouraging them to remain in school. The program will focus in part on developing life skills that will be relevant to the job market. ABE provides a pathway for girls who have never attended school or dropped out in early grades to enrol and complete lower primary	By offering an alternative pathway for girls who may have otherwise dropped out, transition rates will improve. Girls will have increased exposure to higher learning, which will boost learning outcomes. ALP's particular focus on developing life skills will ensure this intervention produces sustainable outcomes, or outcomes that are relevant to the individual and

		education through an accelerated curriculum.	community.[9] ABE allows for transitions into upper primary or entrepreneurship after completion of the accelerated course, thus contributing to learning and transition.
	Develop girls' life skills through ALP and ABE, including leadership skills, financial literacy and business selection and management of income generation activities; participation in Girls' Empowerment Fora	Life skills development. Girls will learn relevant life skills that will not only boost their learning outcomes and attendance, but will also enable them to contribute to the local economy once they leave school.	The project's learning outcomes are focused on literacy, numeracy, and financial literacy. This intervention is designed to boost these specific learning outcomes, as well as increasing the likelihood of transition into formal primary (for ABE graduates), ALP or secondary education. Additionally, the focus on leadership skills and other skills relevant to the job market contributes to the sustainability of SOMGEP-T.
	Develop CECs to improve retention and transition	Attendance and retention. The enhanced capacity of CECs will enable them to develop context-appropriate strategies for improving retention and transition, which will in turn have a positive effect on attendance rates.	A focus on retention and transition is expected to have a direct impact on transition rates and learning outcomes, as girls will have better access to higher education levels. The focus on the community level will ensure buy-in and contribute to the project's sustainability at the community level.

[9] The project has worked closely with the MoEs to develop the ALP model and policies related to non-formal education, thus building the foundation for the future replication of the model through government and partner-led efforts. The ALP is directly aligned with key objectives of the ESSPs to increase enrolment and provide alternative learning opportunities for marginalized groups of girls, particularly those who dropped out after early primary.

	Provide partial grants to girls from poor families	Attendance and retention. Poverty is one of the leading reasons parents are unable to send their children, and girls in particular, to school. Providing partial grants to girls from poor families will alleviate some of the financial burden impoverished families face in sending their children to school.	Increased attendance and retention is expected to improve transition rates and learning outcomes, as girls who are in school and are properly equipped are more likely to succeed. Girls from poor families who may not have otherwise had access to education will be better equipped to participate in decision-making and economic activities.
	Equip and enrol girls in 2 boarding schools	Attendance and retention for girls living in remote areas where upper grades are not available within a short distance from villages. Many families are unable to afford the fees associated with sending their children to school, including fees associated with school enrolment, textbooks, uniforms, and other supplies. By equipping and enrolling girls in boarding schools, the burden families face will be alleviated, and girls will have the equipment they need to remain in school and succeed.	Increased attendance and retention is expected to improve transition rates and learning outcomes, as girls who are in school and are properly equipped are more likely to succeed. Girls from poor families who may not have otherwise had access to education will be better equipped to participate in decision-making and economic activities.
Supportive school practices and conditions for marginalised girls	Train teachers on improved delivery of literacy and English language, supported by digital content in all 148 primary and 51 secondary schools	Improved quality of teaching. Qualified teachers are in low supply in all project areas. Teacher trainings will develop the skills of teachers, thereby improving their teaching quality; increased student performance and	Improved teaching quality contributes to enhance learning and transition outcomes, as children are equipped with the literacy skills in Somali and basic English skills necessary to progress to higher levels of education. Interventions focused on

		motivation is likely to have a positive effect on attendance.	improving teaching quality are expected to boost transition rates and learning outcomes in a sustainable way, by equipping children with the skills they need to succeed not only in school, but outside school as well.
	Train teachers on improved delivery of numeracy in all 148 primary and 51 secondary schools	Improved quality of teaching, addressing specific gaps. Qualified teachers are in low supply in all project areas. Teacher trainings will develop the skills of teachers, thereby improving their teaching quality.	Poor teaching quality contributes to poor learning and transition outcomes, as children are not equipped with the basic numeracy skills necessary to progress to higher levels of education and to develop financial literacy, particularly when focusing on the application of numeracy to basic tasks. Interventions focused on improving teaching quality are expected to boost transition rates and learning outcomes in a sustainable way, by equipping children with the skills they need to succeed not only in school, but outside school as well.
	Train teachers to provide structured remedial support to students at primary and secondary level	Improved quality of teaching. Qualified teachers are in low supply in all project areas. Teacher trainings will develop the skills of teachers, thereby improving their teaching quality.	Poor teaching quality contributes to poor learning and transition outcomes, as children are not equipped with the literacy, numeracy, and English skills necessary to progress to higher levels of education. In relation to this intervention in particular, students are

			more likely to drop out if they do not have proper support. Interventions focused on improving teaching quality are expected to boost transition rates and learning outcomes in a sustainable way, by equipping children with the skills they need to succeed not only in school, but outside school as well.
	Train and coach teachers to deliver the ALP and ABE curricula	Improved quality of teaching and life skills development. Qualified teachers are in low supply in all project areas. Teacher trainings will develop the skills of teachers, thereby improving their teaching quality. Additionally, the ALP and ABE curricula offer life skills development.	Offering alternative pathways to lower and upper primary completion will increase transition rates and boost learning outcomes by keeping girls in school. Ensuring the proper delivery of the ABE and ALP curricula, which include a focus on life skills development, will make the intervention relevant to students and the community and contribute to its sustainability.
	Construct additional classrooms in remote primary schools; build water facilities in new secondary schools; and provide solar chargers for mobile devices/tablets and sanitary pads to schools	Attendance and retention. Lack of infrastructure is a major issue facing all areas of the FRS, Somaliland, and Puntland, but marginalised communities in particular. Lack of proper facilities makes it difficult for students to attend and learn well in school, particularly when schools face an increase in enrolment. Additionally, girls who do not have access to sanitary pads are	Boosts to attendance and retention are expected to contribute to improvements in transition and learning outcomes. Infrastructure development will benefit not just the current cohort of students with which SOMGEP-T is engaged, but will also benefit future students.

		more likely to stay home, or drop out of school entirely. Therefore, this intervention is expected to boost attendance and retention.	
	Incorporate life skills and financial literacy training into GEFs and BEFs	Life skills development. This intervention is focused on providing relevant life skills training through community-based forums, enhancing attendance and learning (through increased participation in class and enhanced financial literacy skills).	Financial literacy training is one of the specific learning outcomes SOMGEP-T is expecting to influence. Financial literacy and life skills training will increase the likelihood of girls succeeding in higher levels of education, and will also equip them to contribute to the local economy through income-generating activities. These skills are expected to increase the relevance of education for students and families. Life skills – specifically leadership skills – are expected to boost students' voice and self-confidence, enhancing classroom participation among girls.
	Provide career guidance in schools	Life skills development. Providing career guidance will help develop an appropriate support system for girls and will encourage them to seek out ways in which to achieve their future career goals.	Encouraging girls to think about their futures and how to achieve their aspirations will impress on them the importance of knowledge and education. It will also give them a clear pathway to achieving their goals.
Positive shifts on gender and social norms at community and	Engage community-level stakeholders including religious leaders, women's groups, men and boys	Attendance and retention. Gender and social norms are a major barrier to girls' education. Gender norms such as those that keep girls at home helping	Boosts to attendance and retention are expected to contribute to improvements in transition and learning outcomes. Shifts in gender

individual girl level		their mothers with chores negatively affect attendance and retention rates. Through engaging with community-level stakeholders, the project will contribute to community-level understanding of the importance of girls' education.	and social norms are expected to have a long-term, sustainable impact on the communities in which SOMGEP-T will operate.
	Expand and strengthen GEFs and create BEFs to develop leadership and mentorship skills	Life skills development, attendance and retention. In addition to providing life skills development, GEFs and BEFs will be engaged in participatory tracking of graduates during project implementation, which will help them assess their own progress in increasing transition rates.	Girls who receive leadership and mentorship skills through life skills development will be better equipped to participate in class, breaking traditional norms that restrict girls' voice; to engage in the local economy; and to contribute to their communities in the future. Additionally, the capacity of GEFs and BEFs to track attendance and retention rates will contribute to improvements in learning and transition outcomes, and will encourage community-based organizations to think about how their actions have a direct effect on important student outcomes.
	Provide adult literacy and financial literacy classes for mothers	Attendance and retention. Evidence from SOMGEP indicates that literate mothers are supportive of their daughters spending time with their schoolwork at home, and are also more likely to	Boosts to attendance and retention are expected to contribute to improvements in transition and learning outcomes. Shifts in gender and social norms are expected to have a long-

		appreciate the importance of girls receiving an education. Mothers who place a higher value on education are expected to understand the importance of enrolling their girls in school and encouraging them to remain in school.	term, sustainable impact on the communities in which SOMGEP-T will operate.
	Support the financial empowerment of mothers through savings groups (VSLA), business selection, and business coaching and mentoring	Attendance and retention. Female heads of household are often struggling to meet the financial and opportunity costs of education, affecting girls' attendance. Mothers who participate in VSLA are able to access funds to build small businesses and support their children's education, and are also more likely to appreciate the importance of girls receiving an education. Mothers who place a higher value on education are expected to understand the importance of enrolling their girls in school and encouraging them to remain in school.	Boosts to attendance and retention, linked to increased financial capacity of vulnerable households, are expected to contribute to improvements in transition and learning outcomes. Shifts in gender and social norms are expected to have a long-term, sustainable impact on the communities in which SOMGEP-T operates.
Enhanced MoEs' capacity to deliver quality and relevant formal and informal education	1) Strengthen Gender Departments' capacity to improve girls' education outcomes through trainings, development of action planning and provision of incentives to	Improved school governance, quality of teaching, retention, attendance, and life skills development. Enhancing the capacity of MoEs to develop plans, administer trainings, and provide incentives will contribute to all four intermediate outcomes by sending a strong, positive message	Enhancing the capacity of MoEs to take action on girls' education will have long-term effects on the communities in which SOMGEP-T operates. It will encourage positive shifts in gender and social norms, and will give MoEs actionable ways to contribute to improving

	retain the gender focal points especially in rural areas 2) Provide support to Regional Education Officers (REOs) and District Education Officers (DEOs) to mainstream improved teaching practices and address retention/ transition	about the importance of girls' education from the government, and by giving the government clear and actionable ways to contribute to positive changes in girls' education outcomes.	learning and transition outcomes.
	Support quality assurance and standards (QAS) functions at all MoE levels	Quality of teaching, school governance, attendance, retention, and life skills development. Enhancing the ability of MoEs to monitor and evaluate their actions will enable them to understand the current educational situation and develop effective plans for addressing any gaps that exist.	Enhancing the capacity of MoEs to take action on girls' education will have long-term effects on the communities in which SOMGEP-T operates. It will encourage positive shifts in gender and social norms, and will give MoEs actionable ways to contribute to improving learning and transition outcomes.
	Provide support to Regional Education Officers (REOs) and District Education Officers (DEOs) to mainstream improved teaching practices and address retention/ transition	Improved quality of teaching, school governance, attendance, retention. This intervention is focused specifically on increasing the capacity of officers who have more direct oversight over the education system in their areas to address issues related to attendance and retention and mainstream improved teaching practices.	Enhancing the capacity of MoEs to take action on girls' education will have long-term effects on the communities in which SOMGEP-T operates. It will encourage positive shifts in gender and social norms, and will give MoEs actionable ways to contribute to improving learning and transition outcomes.

	Work closely with MoE on NFE for mothers and entrepreneurships skills for girls	Life skills development, attendance and retention. Encouraging and equipping MoEs to engage with mothers and girls will have a positive influence on social and gender norms, which will increase attendance and retention rates, and will contribute directly to the life skills development of girls.	Enhancing the capacity of MoEs to take action on girls' education will have long-term effects on the communities in which SOMGEP-T operates. It will encourage positive shifts in gender and social norms, and will give MoEs actionable ways to contribute to improving learning and transition outcomes.
	Development of project IEC materials in conjunction with MoE for use at stakeholder advocacy and promotion events	Life skills development, attendance, and retention. IEC materials are specific knowledge products that will be shares with Parent Teacher Associations (PTA) forums, GEFs, and BEFs. These forums contribute directly to life skills development, attendance, and retention.	Enhancing the capacity of MoEs to take action on girls' education will have long-term effects on the communities in which SOMGEP-T operates. It will encourage positive shifts in gender and social norms, and will give MoEs actionable ways to contribute to improving learning and transition outcomes.

BENEFICIARIES

SOMGEP-T was designed to target girls who are most marginalised in terms of educational outcomes. In the context of the program, this means that girls are targeted who are thought to be the least likely to enrol in school in the first place, and the most likely to drop out, fail to complete primary school, or otherwise fail to reach their educational potential. The program's definition of marginalization focuses primarily on individual- and household-level characteristics to define marginalisation, though broader community-level considerations both influence relative marginalisation and can act as independent considerations themselves. SOMGEP-T defines marginalisation according to their household's relative poverty, participation in pastoralism as their primary livelihood, internal displacement, and minority clan membership. At an individual level, girls are considered marginalised if they are orphans, older than the typical student in their grade (e.g., girls who are 15 years old, but enrolled in grade 2), or have an illiterate mother. Exposure to conflict or violence is also seen as a vector of marginalisation.

In practice, there are a number of aspects of marginalisation which are not specifically listed here, but which can operate as barriers to educational attainment in many cases. In other cases, our metrics of marginalisation are actually imperfect indicators of those characteristics listed above. To illustrate additional aspects of marginalisation not captured above, we occasionally make reference to girls living in households in which

neither parent has obtained any education. To illustrate a case where we use a metric that is an imperfect indicator of a marginalisation characteristic, consider food security – food security and poverty are not precisely synonymous. However, food insecurity is a very useful marker of poverty, especially in a context where most households own livestock (i.e. livestock ownership is a particularly useful way to differentiate household wealth). Throughout this report, we occasionally reference indicators of this kind, such as measures of food insecurity, dietary diversity, and so forth, which we consider markers of marginalisation.

In this section we describe the main sample employed in the ML2 evaluation, including the primary cohort girls, ALP girls, and boys. This round constitutes an evaluation of the SOMGEP-T program two years into implementation; however, it also serves as the baseline for girls enrolled in Accelerated Basic Education (hereafter, "ABE girls"). Note that we do not describe the sample of ABE girls extensively in this section; instead, we relegate discussion of ABE girls to their own section (Section 8), where we provide more detailed analysis of this new cohort sample.

The table below describes the geographic breakdown of the main samples that we analyse in this evaluation. The top panel describes the sample of cohort girls, separated by baseline and ML2, and further disaggregated by intervention and comparison groups. Astute readers will note that the sample is identical from baseline to ML2. Indeed, this is intentional, as we report results primarily using the "true panel" of cohort girls who were successfully re-contacted at ML2. In short, these are girls who were originally recruited at baseline and were re-contacted at ML2, regardless of what occurred at ML1.[10] For the learning analysis, this is our core sample; whenever we study trends over time in girls' outcomes, caregiver attitudes, and so forth, the true panel is or preferred sample. As the table shows, the sample is heavily oriented toward Somaliland, while Puntland makes up approximately one-third of the total sample. Schools in Puntland are over-represented in the intervention group, while Somaliland makes up a greater share of the comparison sample. Note that the Galmudug sample is very small, in general – it is important to bear this in mind, as we often report zone-level analyses but exclude Galmudug from this approach, due to its small sample size.

TABLE 2: GEOGRAPHIC BREAKDOWN OF THE COHORT GIRLS' AND BOYS' SAMPLE AT BASELINE AND ML2

	Baseline				Midline #2			
	Intervention		Comparison		Intervention		Comparison	
Sample Breakdown (Cohort Girls)								
Zone	**N**	**Pct.**	**N**	**Pct.**	**N**	**Pct.**	**N**	**Pct.**
Somaliland	345	60.0%	353	69.4%	345	60.0%	353	69.4%
Puntland	191	33.2%	142	27.9%	191	33.2%	142	27.9%
Galmudug	39	6.8%	14	2.8%	39	6.8%	14	2.8%
Total Girls	575	100.0%	509	100.0%	575	100.0%	509	100.0 %
Sample Breakdown (Boys)								
Zone	**N**	**Pct.**	**N**	**Pct.**	**N**	**Pct.**	**N**	**Pct.**
Somaliland	159	64.4%	155	70.8%	92	57.5%	70	64.8%

[10] We do not constrain the sample to require re-contact at ML1, because the ML1 evaluation only sought to re-contact in-school girls, and this would unnecessarily limit the sample we employ for analysis. In addition, our core interest is in changes from baseline to ML2, rather than in the shorter-term changes that could be observed from ML1 to ML2.

Puntland	85	34.4%	60	27.4%	60	37.5%	33	30.6%
Galmudug	3	1.2%	4	1.8%	8	5.0%	5	4.6%
Total Boys	247	100.0%	219	100.0%	160	100.0%	108	100.0 %

The bottom panel of the table reports the geographic breakdown of the boys' sample. As we describe in the methodology section below, boys were recruited through household surveys conducted with girls. At baseline, 510 boys were recruited into the sample; however, our effective sample of boys is smaller, due to school-level attrition (the removal of schools from the sample following baseline). Unlike girls, who were re-contacted in each round, a new cross-section of boys was recruited in each round – this explains the fact that the girls' sample does not change from baseline to ML2 (because it is a panel), while the boys' sample composition changes (because it is a repeated cross-section in the same communities).

The table below extends this discussion to ALP girls, who were recruited into the evaluation at ML1. These girls constitute an entirely separate sample, drawn from 32 ALP centres. The table is much simpler than that presented above, because the ALP girls' sample does not include a comparison group of any kind. ALP girls were re-contacted at ML2; the table reports the full sample of ALP girls contacted at each round, though our analysis in the report often focuses exclusively on the "true panel" of ALP girls (n = 257).

TABLE 3: GEOGRAPHIC BREAKDOWN OF THE ALP GIRLS' SAMPLE AT BASELINE AND ML2

	Midline #1		Midline #2	
Sample Breakdown (ALP Girls)				
Zone	**N**	**Pct.**	**N**	**Pct.**
Somaliland	253	69.3%	238	70.8%
Puntland	97	26.6%	83	24.7%
Galmudug	15	4.1%	15	4.5%
Total Girls	365	100.0%	336	100.0%

Beyond geography, the table below reports the breakdown of cohort girls and boys by their grade level (and enrolment status). As before, the top panel reports information for cohort girls, and the grade-level breakdown is identical across baseline and ML2 specifically because we disaggregate the samples by *baseline grade*.[11] Again, the top panel of girls is the "true panel", which we prefer to use for methodological reasons. The actual sample of girls contacted at baseline and ML2 was much larger, and we occasionally analyse this larger sample, but it is the exception, rather than the rule.

TABLE 4: GRADE-LEVEL BREAKDOWN OF COHORT GIRLS' AND BOYS' SAMPLES

	Baseline		Midline #2	
BL Grade	**Intervention**	**Comparison**	**Intervention**	**Comparison**

[11] This approach is in line with guidance from the Fund Manager, who advises reporting results by referring to girls according to their grade at baseline. We follow this approach and, additionally, we refer to girls throughout this report by the girls' enrolment status or cohort grouping (in-school girl, OOS girl, ALP girl, ABE girl) at the time of her initial recruitment at baseline (where baseline refers to the baseline for her specific cohort type, i.e. baseline for in-school and OOS girls, but ML1 for ALP girls, who entered the sample later).

Sample breakdown (Cohort Girls)								
OOS	172	29.9%	196	38.5%	172	29.9%	196	38.5%
Primary 1	1	0.2%	1	0.2%	1	0.2%	1	0.2%
Primary 2	1	0.2%	0	0.0%	1	0.2%	0	0.0%
Primary 3	127	22.1%	106	20.8%	127	22.1%	106	20.8%
Primary 4	90	15.7%	62	12.2%	90	15.7%	62	12.2%
Primary 5	99	17.2%	83	16.3%	99	17.2%	83	16.3%
Primary 6	84	14.6%	61	12.0%	84	14.6%	61	12.0%
Primary 7	1	0.2%	0	0.0%	1	0.2%	0	0.0%
Primary 8	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Form 1 and above	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Grade not known/ALP	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Total girls	575	100.0%	509	100.0%	575	100.0%	509	100.0%
Sample breakdown (Boys)								
OOS	83	33.6%	74	33.8%	26	16.3%	25	23.1%
Primary 1	0	0.0%	0	0.0%	2	1.3%	2	1.9%
Primary 2	0	0.0%	0	0.0%	6	3.8%	3	2.8%
Primary 3	56	22.7%	35	16.0%	5	3.1%	3	2.8%
Primary 4	30	12.1%	36	16.4%	9	5.6%	4	3.7%
Primary 5	44	17.8%	41	18.7%	23	14.4%	14	13.0%
Primary 6	34	13.8%	33	15.1%	36	22.5%	8	7.4%
Primary 7	0	0.0%	0	0.0%	24	15.0%	30	27.8%
Primary 8	0	0.0%	0	0.0%	25	15.6%	15	13.9%
Form 1 and above	0	0.0%	0	0.0%	0	0.0%	3	2.8%
Grade not known/ALP	0	0.0%	0	0.0%	4	2.5%	1	0.9%
Total boys	247	100.0%	219	100.0%	160	100.0%	108	100.0%

As the bottom panel of the table shows, boys at baseline were recruited exclusively from grades 3-6 and from boys who were not enrolled in school at all. It is also worth pointing out that the share of boys who are out-of-school has declined from baseline to ML2. As before, the boys' sample is *not* a longitudinal panel, and the composition of the sample has changed over time, not least in the number of boys interviewed.

Finally, the table below disaggregates the girls' and boys' samples by age. As with grade, we refer to girls at ML2 by their baseline age. For instance, a girl who was in grade 4 and age 10 at baseline will be reported in this table as grade 4 and aged 10 years even two years later, during ML2. The girls' sample is mostly composed of girls age 10-14 years – these girls comprise 74.6 percent of the combined intervention and comparison group sample. In contrast, we do not refer to boys by their age at baseline, because the boys' sample is not a panel, and the set of boys interviewed changed from baseline to ML2, for a number of reasons. There is no guarantee that a boy interviewed at ML2 was also interviewed at baseline, even if his sister is in the true panel sample and was interviewed in both rounds.[12]

[12] This is because boys are randomly selected within households, and some households may have multiple eligible boys. Further, boys are only interviewed if they are eligible and available for interviewing at the time of the household survey.

TABLE 5: AGE BREAKDOWN OF GIRLS' COHORT AND BOYS' SAMPLES, BASELINE AND ML2

	Baseline				Midline #2			
	Intervention		Comparison		Intervention		Comparison	
Sample Breakdown (Cohort Girls)								
Aged 9-11	169	29.4%	153	30.1%	169	29.4%	153	30.1%
Aged 12-13	195	33.9%	147	28.9%	195	33.9%	147	28.9%
Aged 14-15	123	21.4%	118	23.2%	123	21.4%	118	23.2%
Aged 16-17	60	10.4%	59	11.6%	60	10.4%	59	11.6%
Aged 18-19	28	4.9%	32	6.3%	28	4.9%	32	6.3%
Aged 20+	0	0.0%	0	0.0%	0	0.0%	0	0.0%
Total girls	575	100.0%	509	100.0%	575	100.0%	509	100.0%
Sample Breakdown (Boys)								
Aged 9-11	72	29.2%	69	31.5%	0	0.0%	0	0.0%
Aged 12-13	85	34.4%	83	37.9%	53	33.1%	29	26.9%
Aged 14-15	58	23.5%	36	16.4%	58	36.3%	36	33.3%
Aged 16-17	21	8.5%	23	10.5%	35	21.9%	32	29.6%
Aged 18-19	11	8.5%	8	3.7%	9	5.6%	8	7.4%
Aged 20+	0	0.0%	0	0.0%	5	3.1%	3	2.8%
Total boys	247	100.0%	219	100.0%	160	100.0%	108	100.0%

2. EVALUATION METHODOLOGY

2.1 Evaluation Objectives and Design

Overall Evaluation Questions

SOMGEP-T is being implemented as part of the broader GEC-T program, which shapes the outcomes of interest to the evaluation and the metrics which are used for assessing impact. The research questions addressed in this evaluation are tightly focused on two goals. The first set of questions focus on whether SOMGEP-T has had a discernible impact on intervention girls, schools, and communities, across many different core and primary outcomes. The second set of questions are focused on understanding the nature of impact, how it is mediated, through which activities and mechanisms it operates, whether some girls are being left behind or could be better targeted, and whether gains made are likely to be sustained past the end of active SOMGEP-T implementation.

The first set of questions are broader and generally employ similar methodologies for obtaining answers. These questions include:

- What impact did the GEC Funding have on the transition of marginalised girls through education stages and their learning?
- To what extent did the intervention result in additional gains in learning (literacy and numeracy) among the intervention group, in relation to the comparison group?
- What impact has the program had on intermediate outcomes in the intervention group, relative to the comparison group.

The second set of questions are slightly more varied:

- What works to facilitate transition of marginalised girls through education stages and increase their learning?
- How sustainable were the activities funded by the GEC and was the program successful in leveraging additional interest and investment?
- Is there a significant difference between the acquisition of literacy/ numeracy/ English skills among ASLP participants and marginalised girls enrolled in formal secondary school?
- Are the intermediate outcomes identified by the project contributing to the accelerated acquisition of literacy/ numeracy skills and improved transition rates? Are there intermediate outcomes that do not seem to be influencing outcomes at all? Do the findings support the ToC or challenge its assumptions?
- What are the key factors influencing the acquisition of literacy, numeracy and English language skills?
- What are the specific literacy/ numeracy/ English competencies that marginalised girls are lagging behind on?
- Is there a difference in the rate of acquisition of literacy / numeracy subtasks that girls are able to practice in their daily lives, vis-à-vis the acquisition rate of subtasks that are not used on a regular basis by the girls targeted by the project?
- To what extent are extremely marginalised sub-groups, such as pastoralist girls and disabled girls, able to attain basic competencies in literacy, numeracy and English? Are there other sub-groups who

are demonstrating a pattern of lagging behind in performance? What are the potential reasons for this pattern?

- Is there a relationship between the acquisition of leadership skills and learning outcomes? If yes, how?
- To what extent the participation of mothers in VSL may influence the acquisition of numeracy and financial literacy skills by girls?
- What are the key factors influencing transitions to more advanced levels of education?
- To what extent are extremely marginalised sub-groups, such as pastoralist girls and disabled girls, able to transition into upper primary/ post-primary education? Are there other groups who are lagging behind in transition rates? What are the potential reasons for this pattern?
- To what extent is the acquisition of leadership skills influencing transition outcomes?
- Did the intervention contribute to a shift in traditional gender norms and power relationships at the household and community levels? If yes, what types of changes have occurred? How are these changes affecting adolescent girls and boys?
- How did the intervention affect boys' learning and retention?
- What are the key changes identified by the girls themselves in terms of their capacity to engage in non-traditional roles at the household, school and community? To what extent are those claims supported by quantitative evidence?
- Is there a difference in the learning outcomes for students targeted in areas heavily affected by drought, compared to those that were less affected? Likewise, is there any difference for transition?

The overall evaluation design, described in the next section, was designed with these questions in mind.

Overall Evaluation Design

SOMGEP-T employs a quasi-experimental evaluation design, which relies on pre-selected comparison and intervention groups of schools to draw inferences about program impact. Prior to the baseline, CARE identified a set of schools where their interventions would be implemented, based on their experience implementing the first round (SOMGEP) in the same areas. These schools became the intervention group. At the same time, CARE identified a set of schools where the program would not be implemented, and which would be suitable as comparison schools, as they operated in similarly rural areas in the same districts as intervention schools.

From the set of intervention and comparison schools, the evaluation team first matched intervention and comparison schools into pairs based on their similarity across a number of characteristics, particularly geographic zone, school size, and whether the school was receiving support from an NGO, and existence of a Community Education Committee. Schools were not matched on urbanicity, because all urban schools were dropped from the sample (in other words, all schools in the sample were implicitly matched as being rural). Matched-pairs were generated using Coarsened Exact Matching, which matches schools within a set of bounds, rather than enforcing exact matches. This approach is especially useful when matching schools on integer variables, such as school size, because few schools will have precisely the same number of students; matching schools that are in a similar school size range is sufficient to ensure similarity between the two schools. Because the matching method and variables employed yielded just 32 matched pairs, a second and third round of matching were undertaken, loosening the requirements slightly, until 36 matched pairs were achieved. As noted in the baseline report, the resulting sample was "perfectly balanced (between intervention and comparison) in terms of zone, and is nearly balanced in terms of school size and known involvement of other NGOs (39% of intervention schools with NGO involvement, vis-à-vis 29% of comparison schools)."

In practice, the value of the matched-pair design has been diminished slightly since the baseline, due to differential school-level attrition from the sample. Following the baseline evaluation, five comparison schools were identified as particularly high-performing, and a joint decision was made between CARE and the evaluation team to remove the schools from the sample. In addition, two schools from the baseline do not appear in the ML2 sample due to security and accessibility issues, as we discuss in the sampling methodology section. In total, six comparison schools and one intervention school have been dropped from the original baseline sample, resulting in less balance between the groups – in terms of sample size, as well as pre-existing school characteristics – than originally envisioned. Importantly, this lack of balance does not undermine the validity of the research design, a fact which we highlight below.

The quasi-experimental evaluation employs a difference-in-differences design, which makes comparisons between the intervention and comparison groups over time, looking for differences in their trends from baseline to endline. The analysis does not rely on differences between comparison and intervention groups at endline or at any single point in time. Rather, we investigate whether – for instance – girls in intervention schools improved more than girls in comparison schools, regardless of their relative starting point at baseline or relative ending point at endline.

The difference-in-differences design is robust to a number of sources of bias that plague other research designs, especially pre-post analyses without a comparison group. Specifically:

- **Pre-intervention differences between intervention and comparison groups** – if the intervention and comparison group have different baseline performance in terms of learning or transition, this gap is controlled for explicitly in the design. If project schools already had higher learning achievement prior to the program's start, a simple comparison between intervention and comparison schools at endline would not be valid, because intervention schools were *already* higher-achieving than their comparison counterparts. By accounting for baseline differences in learning outcomes, this source of bias is eliminated.
- **Systematic changes in outcomes over time that are not attributable to the project** – if the areas where SOMGEP-T is being implemented experience a broad-based change, such as multiple productive rainy seasons in a row, this would bias simple pre-post analyses of program impact. In the example given, we might expect enrolment and attendance rates to rise broadly in all areas where the rains have been reliably good. If the evaluation used a pre-post design without a comparison group, and the analysis revealed improving transition rates from baseline to endline, this increase would be incorrectly attributed to the program, when it may have arisen from the improved rains. By incorporating a comparison group, the SOMGEP-T design allows us to identify general trends in the area and isolate those from program-specific impacts.

These two primary advantages motivate most uses of the difference-in-differences design, and they are critical to our application here, especially given that there are significant secular trends in the outcomes that need to be accounted for through the use of a comparison group. Beyond these motivations, the design also reduces or eliminates bias that could otherwise result for changes in the learning assessment across rounds – a point we discuss in more detail when we compare the difficulty of learning assessments in a later section. In short, the difference-in-differences design eliminates many of the methodological concerns that reduce the validity of inferences drawn from other designs.

The overall evaluation design outlined here has not changed appreciably since the baseline. In fact, the only substantive change made to the primary design has been in the sample itself, which was adjusted at the school level in the manner noted above. Where the evaluation design has changed, it has been supplemented in the

form of bringing in additional cohorts of girls, such as ALP and ABE girls. However, the recruitment of ALP and ABE girls into the evaluation does not influence the primary design used for studying learning, transition, or intermediate outcomes, as these additional samples are treated separately from the primary cohort recruited at baseline.

Role of the Midline Evaluation

This evaluation constitutes the second midline evaluation of the SOMGEP-T program. The baseline was conducted in late 2017, and a slightly limited round 1 midline evaluation (ML1) was conducted in late 2018. In contrast to ML1, this round assesses both in-school and OOS girls for impacts since baseline; it also constitutes the second round of contact with ALP girls, who were initially recruited into the evaluation during ML1. The primary goal of this evaluation, which is reflected in the amount of space dedicated to this task, is to assess the program's impact in formal primary schools and among the main cohort recruited at baseline. A secondary goal is to assess the impact of the program on ALP girls, primarily in the context of learning outcomes.

Beyond these efforts at assessing program impact, the ML2 serves as a baseline for girls enrolled in Accelerated Basic Education (ABE) centres. ABE centres were established by CARE in response to findings at the baseline, which showed a higher-than-expected number of relatively young (pre-teen and early teenage) girls were not enrolled in school. ABE girls are incorporated into this evaluation, which serves as a baseline against which their improvements in learning and their performance in terms of transition will be tracked.

In some ways, the SOMGEP-T evaluation seems to comprise three different evaluations: one of the primary cohort, which included both in-school and OOS girls, selected to be representative of their communities; one of ALP girls; and one of ABE girls. However, these groups of girls are interrelated and overlap with one another in important ways, despite the fact that the *samples* of each group are mutually exclusive.[13] Specifically, in-school and OOS girls selected at baseline may, in practice, transition into ALP or ABE programs. Again, they cannot become ALP or ABE girls in terms of the cohorts evaluated, but they can and do enrol into ALP or ABE programs. For instance, among girls in the main cohort re-contacted at ML2, 15 are enrolled in either ALP or ABE programs currently.[14] In short, the existence of ABE and ALP programs has an effect on cohort girls, providing them additional avenues of education, but the cohorts of girls recruited for the evaluation are considered in isolation from one another.

Data Collection Tools

The design of the ML2 evaluation mirrors, in broad ways, the baseline and ML1 evaluations. As we highlighted in the previous section on evaluation design, a new sample of beneficiaries – Accelerated Basic Education (ABE) girls – was added to the evaluation, and girls comprising the out-of-school girl cohort, originally recruited at the baseline, was re-contacted during ML2. This latter fact stands in contrast to the ML1 evaluation, when OOS girls were not re-contacted for budgetary reasons (additionally, activities specifically targeting OOS girls had been limited at the time, as the ABE component had not yet been implemented).

[13] That is, if a girl is included in the original cohort selected at baseline, she was ineligible to be selected into either the ALP or ABE samples. The three samples consist entirely of different girls.

[14] Again, these girls are still classified as in-school and OOS girls in cohort terms, based on their status at their initial recruitment.

The ML2 evaluation mirrors the BL and ML1 evaluations in terms of the data collection tools employed as well. The evaluation uses both quantitative and qualitative tools, which have not changed dramatically over time. The quantitative tools included:

- Household survey
    - Modules with girls, the head of their household, their caregiver, and – where available – a boy in the household aged 12-22 years
    - Topics covered:
        - Household characteristics
        - Girl's characteristics, including disability status
        - Girl's relative sense of empowerment, life skills (Youth Leadership Index module)
        - Enrolment, employment, and transition status of girl
        - Teaching quality, according to girls and caregivers
        - School management, according to caregivers
        - Attitudes toward girls' education
- Learning assessment
    - English literacy based on SeGRA
    - Somali literacy based on SeGRA
    - Numeracy based on SeGMA
    - Financial literacy
    - Assessment of working memory
- Classroom observation tool
    - Captures aspects of classroom environment, teaching quality, gender equity, and pedagogical practices
- Attendance headcount tool
- Head teacher survey
    - Topics covered:
        - School record-keeping and policies
        - School infrastructure and availability of materials
        - Information on teaching corps and teacher absenteeism
        - Attendance of cohort girls, drawn from school records

Relative to the two prior rounds, there are only four significant changes to the quantitative tools. First, an extended battery of questions related to food security was added to the household survey for the first time. Respondents were asked to list all of the foods they had eaten in the previous 24 hours, to allow measurement of dietary intake and diversity.

Second, an assessment of girls' and boys' working memory was added to the evaluation for the first time. Children were shown a series of 19 images of everyday items with which they would be familiar (a tire, a ball, etc.). After reviewing the images, the paper was removed and children were asked to list all of the images they could recall. This brief assessment tests the quality of children's working memory, defined as the ability to retain information for a specific task during a limited period of time[15].It is widely accepted that working

[15] Baddeley, A. (2010) Working Memory. *Current Biology*, Volume 20, Issue 4, 23 February 2010, p.R136-140

memory is a predictor of reading comprehension, as well as having a direct impact on the ability to perform other complex cognitive tasks.[16]

Third, the ML2 round brought assessments of English literacy and financial intelligence back into the evaluation. Both assessments had been included in the baseline evaluation, but were not completed during ML1 as the related components had not yet been implemented. The ML1 evaluation targeted in-school girls from the baseline and newly-recruited ALP girls. Neither group had been exposed to significant teaching on English literacy – therefore, it was agreed that the English literacy assessment could be skipped until ML2. The financial intelligence assessment was similarly removed from ML1 for budgetary and practical reasons. At ML2, these assessments are included, which will allow valid comparisons from baseline to ML2, using the difference-in-differences approach.

Fourth, the Somali literacy assessment includes an added subtask that was not included in previous rounds. Somali literacy scores during ML1 showed very mild evidence of ceiling effects, with 9.0 percent of in-school girls scoring above 95 percent in Somali literacy. At the same time, in-school girls were the most likely to experience ceiling effects, and the possibility that ceiling effects will influence OOS girls, ALP girls, and ABE girls appears to be even more remote. However, to guard against the possibility of ceiling effects going forward, a subtask was added to the assessment; while inclusion of this subtask in the analysis would render the assessment incomparable to baseline and ML1, this is not a major methodological concern. Precise comparability of the assessments from round to round is only necessary in the absence of a difference-in-differences design, as we discuss in more detail in Section 2.3, below, where we empirically analyse the comparability of the assessments. To the extent that precise comparability is important, valid comparisons from baseline and ML1 to ML2 are possible by analysing the ML2 Somali literacy assessment without subtask 9; valid comparisons from ML2 to the endline can be made using the full assessment, including subtask 9.

Beyond these changes, it is important to note that revisions made to the data collection tools between baseline and ML1 were maintained during ML2. During the inception stage for ML1, the evaluation team reviewed the baseline evaluation report, looking for indicators or thematic areas of interest that were not addressed well by quantitative measures, and where triangulation across data sources was limited. At that time, a number of questions were added to the ML1 data collection tools, with the goal of improving data collection and reporting going forward. This process mirrored CARE's own revision and improvements to the qualitative data collection tools from baseline to ML1, which we address below. Where improvements to the quantitative tools were made from baseline to ML1, these changes were maintained into ML2. In some cases, this means our comparisons are limited to ML1-to-ML2; where this is necessary due to the limitations of the BL data, we note it explicitly.

The qualitative tools were largely unchanged from the ML1 round. The set of interviews and the groups they targeted were identical to the ML1 round, with the following FGDs and KIIS:

- FGD with Community Education Committee (CEC) members
- FGD with teachers
- FGD with female community members
- KII with Ministry of Education (MOE) officials
- KII with girls with disabilities
- Participatory risk mapping with girls who are members of the Girls Empowerment Forum

[16] Cowan, N. Working Memory Underpins Cognitive Development, Learning and Education. *Educ Psychol Rev.* 2014 Jun 1; 26(2): 197–223.

- Participatory story-telling exercises (vignettes) with girls who are members of the Girls Empowerment Forum

As noted in the ML1 evaluation report, these tools represented a dramatic departure – in some ways – from the baseline. CARE's monitoring and evaluation team reconsidered the qualitative tools between baseline and ML1, developing valuable new participatory exercises to be completed with girls. In other cases, the existing interview guides were revised significantly, to better capture program indicators. At ML2, these updates have been maintained, and relatively minimal changes were made between ML1 and ML2.

2.2 Sampling Methodology

The sample employed for this evaluation is a panel of schools and girls that have been re-contacted over either two or three rounds, depending on the specific school and girl in question. The design of the evaluation relies on the fact that the sample is mostly stable over time, as this ensures enough overlap of girls from one round to another to facilitate difference-in-differences analysis. In the sections below, we detail the sample's design in stages: the first stage of sample selection is the selection of schools or learning centers, which constitute the primary sampling unit (PSU); the second stage is the selection of individual girls, classrooms for headcounts, and teachers for classroom observations. Because the sample has changed minimally since baseline, we review the sample design very briefly and focus our attention on the changes that are relevant for the analysis or for the evaluation going forward, toward endline. The one exception is the sample of ABE girls, which is new in this round – in that case, we discuss the sample design more extensively.

School Sample

The baseline school sample was drawn with the evaluation's quasi-experimental design explicitly in mind. At baseline, a sample of 76 schools, evenly divided between intervention and comparison schools, was selected. Intervention and comparison schools were drawn from a broader sample frame using matched pairs. First, matched pairs of intervention and comparison schools were created, matching schools on the basis of observed pre-intervention characteristics.[17] Once matched pairs were created, these pairs were sampled, including only those pairs that were met a specific level of similarity. By selecting matched pairs in this way, balance on pre-intervention characteristics is ensured. To the extent that pre-intervention characteristics are correlated with later developments, pre-baseline matching makes the parallel trends assumption underlying the difference-in-differences design more plausible.

During ML1, the sample of schools visited was truncated for budgetary and logistical reasons. In total, 13 schools were removed from the ML1 sample; however, it is important to note that the majority of these removals were intended to apply to ML1 only. Five schools were dropped from the sample because they were outliers in terms of learning scores at baseline.[18] All five schools were in the comparison group, and it was decided to exclude them going forward. Two schools were excluded due to intra-clan conflict. One school was excluded by mistake, by the evaluation team, a mistake which was not discovered until data analysis had already started. Finally, five schools were removed from the ML1 sample because they had very few in-school girls at baseline. Because the ML1 evaluation only sought to interview in-school girls – OOS girls were not

[17] The evaluation team used Coarsened Exact Matching (CEM) for this purpose.

[18] We discuss this issue in greater detail in the baseline and ML1 reports. In short, these schools exhibited especially high baseline scores, particularly in terms of English literacy. Together, the evaluation team and CARE's monitoring and evaluation staff followed up in detail and found that English literacy scores in these schools were likely driven by the impact of specific teachers and the fact that instruction was largely in English, both of which set them apart from their matched schools.

re-contacted at this stage for budgetary reasons – visiting schools in which only 4-8 girls would be re-contacted was not justifiable from a budgetary or logistical perspective. The intention at the time of ML1 was that all schools in this list, with the exception of the five outlier schools, would be re-incorporated into the sample at ML2. This list is reproduced, grouped by the reason they were excluded at ML1, in the table below.

The table documents what occurred in this round of data collection for each of the schools excluded at ML1. For instance, among the schools that were excluded as outliers, none were added back to the sample, an outcome which is in line with the intentions of both CARE and the evaluation team. Among the two schools that were impacted by localized clan conflicts, one school was added back to the sample; Dhumay Primary School had been entirely destroyed since it was last visited during baseline. Four of the five schools that had been excluded from ML1 because they had too few in-school girls to justify a visit were added back to the sample; the single exception was Xingod Primary School, which was inaccessible due to conflict at the time of fieldwork. In total, the sample included 76 schools at baseline, 63 schools at ML1 and 69 schools at ML2. No schools that were included in ML1 were removed at ML2; all schools excluded from ML2 are documented in the table below.

TABLE 6: SCHOOLS EXCLUDED AT ML1 AND INCLUSION STATUS AT ML2

School Name	Zone	Intervention Status	Included at ML2	Reason for exclusion at ML2
Reason for ML1 Exclusion: Outlier Schools at Baseline				
Ilays Primary	Puntland	Comparison	No	Outlier
Al-Rahma Primary	Puntland	Comparison	No	Outlier
Rako Primary School	Puntland	Comparison	No	Outlier
Umada	Galmudug	Comparison	No	Outlier
Salaxudii Primary	Puntland	Comparison	No	Outlier
Reason for ML1 Exclusion: Intra-clan Conflict				
Dharkeyn Primary	Puntland	Intervention	Yes	
Dhumay Primary School	Puntland	Intervention	No	School was destroyed
Reason for ML1 Exclusion: Very few in-school girls at baseline				
Imam-nawawi Primary School	Somaliland	Intervention	Yes	
Xingod	Galmudug	Comparison	No	Lack of accessibliity
Baran Primary School	Somaliland	Intervention	Yes	
Al-Safa	Galmudug	Intervention	Yes	
Ceel-xume Primary School	Somaliland	Comparison	Yes	
Reason for ML1 Exclusion: Dropped accidentally				
Rajo Primary School	Puntland	Intervention	Yes	

The discussion above concerned formal primary schools selected into the evaluation sample at baseline. At ML1, ALP girls were brought into a supplementary sample, and this process necessitated visiting 32 ALP centers, only about half of which overlapped with primary schools that were already in the sample. During ML2, all 32 ALP centers were revisited successfully, with no replacements needed. At ML2, an additional

cohort of girls participating in ABE programs, was brought into the evaluation; we discuss sampling for this group in a separate section below.

RE-CONTACTING COHORT GIRLS

Moving below the first sampling stage, the discussion in this section briefly outlines the method used to re-contact cohort girls and replace those who could not be re-contacted. As aforementioned, the design of the SOMGEP-T evaluation relies on a panel of cohort girls that are tracked over time. Therefore, new girls were not selected into the sample until it became necessary to replace an existing cohort girl.[19]

Re-contact procedures were specified precisely during training, with specific requirements needing to be met before a girl could be replaced. The specific steps are provided in the analysis of re-contact rates in a later section, but included two phone calls to every contact number available for the girl's household, two visits to the girl's household, and other steps to facilitate re-contact. If a girl could not be located during a typical two-day visit to their community, enumerators attempted to find information about the girl's current location, enrolment status, and other details that would allow her to be incorporated into the transition analysis.[20]

If a girl could not be located or was unavailable for an interview after completing the required steps, the team leader authorized either a replacement or documenting the girl as not able to be re-contacted. The replacement process that followed varied by the *type* of girl in question.

- In-school girls - replaced
- Out-of-school girls – not replaced
- ALP girls – replaced
- In-school girls that had fallen out of the sample at ML1 – not replaced

At the most basic level, in-school girls and ALP girls were replaced when they could not be re-contacted. In both cases, girls were selected from the grade or ALP level (if applicable) that the original girl *would* have been in, if she had advanced one grade per year and had been located at ML2. For instance, a girl in grade 5 at ML1 who could not be located at ML2 would be replaced with a girl in grade 6. A girl who was last contacted at baseline – because her school was excluded from ML1 data collection – who was in grade 5 at baseline would be replaced with a grade 7 girl at ML2.

To replace a girl, team leaders listed all girls in the eligible grade range and used a random number generator to select the replacement girl. If no girls in the correct grade were available, a list was made of girls in the adjoining grades (e.g., if replacing a girl who would have been in grade 5 at ML2, girls in grades 4 and 6 would be listed).[21] In total, of 154 in-school girls who were replaced at ML2, 66 were from a grade that did not correctly correspond to the replaced girl's expected grade – typically because no girls were available, or all girls in the grade were already members of the cohort.

[19] Here we use the term cohort to refer to the in-school and OOS girls that were selected into the sample at baseline, as well as the ALP girls who were selected into a new ALP-specific sample during ML1.
[20] A girl who has migrated temporarily or permanently out of the area can still be included in topline transition analysis, if her enrolment status is known with confidence.
[21] A lack of available girls for replacement occasionally occurred in smaller schools and higher grades, where the number of girls available is small. In addition, girls were only eligible for selection as a replacement if they were not already in the cohort, which further limited the available pool of girls.

Two groups of girls were not replaced if they could not be re-contacted, though the methods used to re-contact them were identical to in-school and ALP girls. First, OOS girls were not replaced in the sample, because no logical and logistically feasible method for identifying eligible OOS girls was available. OOS girls were originally recruited into the sample, alongside in-school girls, as part of a random household sample. However, finding a replacement OOS girl would require visiting many households; finding a replacement who matched the approximate age of the girl being replaced would have been too burdensome logistically. Therefore, girls who were originally recruited into the sample as OOS girls were not replaced if they could not be re-contacted. The evaluation team plans – and strongly recommends – that the endline evaluation will include efforts to re-contact the OOS girls who fell out of the sample at ML2.

Second, in-school girls who had fallen out of the sample at ML1 were not replaced. These were girls who we attempted to contact during ML1 but who ultimately fell out of the sample because they could not be re-contacted. The evaluation team made the decision to seek these girls out, as we expected to find some of the girls during ML2. For example, some girls had refused participation or were simply unavailable during the fieldwork timeline of ML1. Other girls had travelled temporarily to a new location. In both cases, we expected to locate some of these girls during ML2, and we recruited them back into the sample. We discuss the positive benefits of this decision on aggregate attrition rates in our analysis of re-contact rates. However, if these girls still could not be located at ML2, they were not replaced by a new girl, as they had already been replaced when they were not located during ML1.

ACHIEVED COHORT SAMPLE

Building on the discussion in the previous section, we briefly report the achieved sample of cohort girls re-contacted from previous rounds. These results exclude ABE girls, who were sampled for the first time in this evaluation wave – we discuss sampling of ABE girls in more detail in the next section. Further, we discuss re-contact rates among cohort girls – comprising the main cohort recruited at baseline, in addition to ALP girls recruited during ML1 and ABE girls recruited as part of ML2 – in great detail in Section 2.4, below. Here we simply report the total number of girls in each cohort that we sought to re-contact in this round, and how many girls were re-contacted, replaced, or fell out of the sample entirely.

The total column, on the right, represents a type of sample target, insofar as it was the maximum number of girls of a particular cohort that would be sought out during data collection. Note that in-school girls who were not contacted at Ml1 constitutes a set of girls who were recruited into the sample at baseline but who could not be located at ML1 or who lived in communities not visited during the ML1 evaluation. Also of note is the fact that OOS girls were not replaced if they could not be located, accounting for their high attrition rate and replacement rate of zero.

TABLE 7: ACHIEVED SAMPLE AMONG COHORT GIRLS AT ML2

Cohort of Girl	Re-Contacted	Replaced	Fell out of Sample	Total
In-School Girls	661	137	9	807
In-School Girls Re-Contacted from BL	116	8	116	240
Out-of-School Girls	368	0 (none attempted)	325	693
ALP Girls	257	79	29	365

Finally, although ABE girls are not included in the table above – because this round constituted their first contact as part of the evaluation – it is important to note that their sample target was not met. The original target was set at 507 ABE girls. However, just 482 ABE girls were recruited into the ABE sample, due primarily to the fact that a number of ABE centres selected into the sample did not have a sufficient number (15) of eligible girls to comprise a full cluster.

To make the total sample size in each cohort clear, the table below provides the per-round sample size for each cohort. In the table, "N/A" refers to cases where the cohort was not included in the evaluation sample (e.g., out-of-school girls were not contacted at ML1, and ALP girls had not entered the sample at the time of the baseline). Note that the sample sizes reported are the total sample, without taking into account adjustments we make prior to analysis to ensure the samples are comparable across rounds.

TABLE 8: TOTAL SAMPLE SIZE, BY COHORT AND ROUND

Sample	Baseline	Midline #1	Midline #2
In-School Girls	987	807	922
Out-of-School Girls	754	N/A	368
ALP Girls	N/A	365	336
ABE Girls	N/A	N/A	482

Throughout the report, we draw on different subsamples for different analytical purposes. The decision regarding which sample to analyse depends on three primary considerations:

- Maximizing comparability across rounds
- Maximizing the sample available for analysis
- Maximizing the representativeness of the sample

The second and third considerations generally run in parallel – maximizing the breadth of the sample typically also makes it more representative of the underlying population, though this is not always the case. Unfortunately, maximizing comparability, as a goal, nearly always conflicts with the goal of maximizing the sample size for analysis. The decisions we make regarding the sample to analyse must balance these goals, in light of the goal of the particular analysis.

To shed light on the different samples we analyse in this evaluation, the table below breaks down the number of observations that fall into different "slices" of the sample, by cohort. To illustrate, consider the top panel, focused on in-school girls: the first row simply documents the total number of in-school girls interviewed in each round, without considerations of comparability – all girls, whether replaced, replacement, or re-contacted are included in the "total sample." The "full panel" represents the set of girls who were contacted at baseline and re-contacted in each subsequent round – we use the term "panel" specifically to refer to cases where *the same girl* appears in both or all rounds. The "baseline panel" shows the set of girls contacted at baseline and re-contacted at ML2, regardless of their inclusion, exclusion, or replacement at ML1. Finally, we also report samples that we deem "cross-sectional" – these samples include all girls in the equivalent panel sample *and* all girls who were replaced or who serve as their replacements.

Walking through an example with the in-school girl sample may be useful to fix ideas regarding the construction of these samples. The relevant distinction between a panel and cross-sectional sample over the same time period (e.g., "baseline panel" versus "baseline cross-section") is in the treatment of replaced and replacement girls. A cross-section includes girls that were replaced – but not those who fell out of the sample and were *not* replaced – whereas the panel would exclude such girls. This distinction explains the difference in sample sizes – e.g., 716 girls in the baseline panel versus 849 girls in the baseline cross-section, with the difference arising from the 133 girls who were replaced between baseline and ML2. The relevant distinction between the cross-sectional sample (n = 849) and the total sample (n = 987 at baseline, 922 at ML2) stems from the girls who fell out of the sample entirely and *were not* replaced. Girls fell out of the sample either because their school was removed from the sample after baseline, or because a girl could not be re-contacted and no replacement was available (attrition without replacement).

TABLE 9: SAMPLE SIZE FOR VARIOUS SAMPLE CONSTRUCTIONS, BY COHORT

Sample	Baseline	Midline #1	Midline #2
In-School Girls			
Total Sample (All Interviews)	987	807	922
Full Panel (BL-ML1-ML2)	529	529	529
Full Cross-Section (BL-ML1-ML2)	798	798	798
Baseline Panel (BL-ML2)	716	0	716
Baseline Cross-Section (BL-ML2)	849	0	849
Midline #1 Panel (ML1-ML2)	0	661	661
Midline #1 Cross-Section (ML1-ML2)	0	798	798
Out-of-School Girls			
Total Sample (All Interviews)	754	N/A	368
Baseline Panel (BL-ML2)	368	N/A	368
Baseline Cross-Section (BL-ML2)	368	N/A	368
ALP Girls			
Total Sample (All Interviews)	N/A	365	336
Midline #1 Panel (ML1-ML2)	N/A	257	257
Midline #1 Cross-Section (ML1-ML2)	N/A	336	336
ABE Girls			
Total Sample (All Interviews)	N/A	N/A	482

The table illustrates several core realities about the sample. First, there is a clear trade-off between sample size and comparability of the sample across rounds. Within a given cohort and over a given time period, the panel sample always allows the most rigorous comparisons, but comes with sometimes significant reductions in sample size. Second, there is a clear trade-off between sample size and the number of rounds that can be included in analysis. Within a given cohort and sample construction – panel or cross-sectional – per-round sample sizes are larger when we limit our attention to BL-to-ML2 comparisons or ML1-to-ML2 comparisons, rather than drawing comparisons across all three rounds (BL-ML1-ML2) simultaneously.

We provide a fuller discussion of the representativeness of various samples in Annex 2, discussing the trade-offs between different approaches to sample construction and their relative representativeness of the underlying population of girls in SOMGEP-T schools. Because we employ a number of different samples across the report, it can become confusing for readers trying to understand which sample is being used in which section. The table below lists the main samples employed in the learning and transition sections of the report. Occasionally, we draw on additional subsamples not explicitly listed in this table, but primarily report these results in Annex 18.

TABLE 10: SAMPLES EMPLOYED FOR VARIOUS ANALYSES

Outcome and Cohort	Sample Employed and Rationale
Learning – Somali literacy and Numeracy	Primary: Full panel (BL-ML1-ML2) Secondary: Baseline panel (BL-ML2) Considerations: Targets and program impact use the full panel; main results in report drawn from both full panel and baseline panel; additional analysis in Annex 18 drawn from other samples (cross-sectional, ML1 panel, etc.)
Learning – English literacy and financial literacy	Primary: Baseline panel (BL-ML2) Considerations: Full panel not available for these outcomes, as English literacy and financial literacy were not tested at ML1. Supplemental analysis uses baseline cross-sectional sample.
Transition – In-school girls	Primary: Full panel (BL-ML1-ML2) Secondary: Baseline panel (BL-ML2) Considerations: Targets and program impact use the full panel; supplemental analysis uses baseline panel to maximize available sample size.
Transition – Out-of-school girls	Primary: Baseline panel (BL-ML2) Considerations: Targets and program impact among this cohort use the baseline panel, because out-of-school girls were not tracked at ML1 (i.e. they are not part of the "full panel").

The table above focuses on learning and transition analyses. We do not list the samples employed for the intermediate outcomes, because the data collection tools used vary so much from outcome to outcome. However, as a general rule, we employ either the full panel or baseline panel for these analyses. The full panel is used most in cases where we are analysing household survey data (life skills and attendance). The baseline panel is used most in cases where the sample size available for analysis is already small (e.g., school management, where much of the data is drawn from the head teacher survey, with fewer than 75 observations

in each round) and we wish to preserve as many observations in the data as possible.[22] Throughout the report, where we refer to BL-to-ML2 comparisons – unless otherwise specified – we are referring to the "baseline panel," i.e. the set of schools, girls, or households who are precisely comparable from baseline to ML2. Similarly, when we refer to ML1-to-ML2 comparisons, we are referring to the "midline panel," the set of schools, girls, or households who are precisely comparable between ML1 and ML2. In those rare instances where we use a non-panel dataset for analysis, we specify this fact explicitly.

SAMPLING ABE GIRLS

In contrast to in-school, OOS, and ALP girls, for whom this report acts as a midline evaluation, ABE girls were recruited for the first time during this round of data collection. For ABE girls, this evaluation constitutes a baseline against which their progress will be measured during the SOMGEP-T endline evaluation in late 2020. Owing to this difference, we analyse ABE girls as a separate group in Section 8 of this report. Here we discuss the sampling procedures used to select ABE girls.

Prior to the ML2 round, CARE's monitoring and evaluation team produced a list of ABE centers targeted for intervention, with a count of ABE girls enrolled in each center. The sample frame covered 66 ABE centers, with an estimated 1,605 girls enrolled in the centers at the time the frame was constructed.

In line with guidance from the Fund Manager, a target sample size of 509 ABE girls was established. The evaluation team opted for a sample of 35 ABE centers, with a target cluster size of 15 girls per center, which would yield a sample size of 525 girls. The oversample was intentional and strategic, as we expected some centers to have too few girls to constitute a full cluster, especially as some ABE girls likely were already members of the baseline cohort (as OOS girls) and would not be eligible for recruitment as ABE girls.

Among the 66 ABE centers, 21 were in locations that overlapped with formal primary schools or ALP centers that were already in the evaluation sample. Due to the timeframe available for data collection and budget limitations for such a large sample (with over 2,500 girls sought to be contacted or re-contacted during ML2), it was decided that these 21 ABE centers would be selected into the ABE sample. A further 14 ABE centers were selected randomly, with stratification by zone. Importantly, the remaining 14 centers were selected explicitly to recover the zone-by-zone population distribution of ABE girls. The overlapping 21 centers were disproportionately in Somaliland; therefore, the remaining 14 centers were drawn heavily from Puntland, in order to recover the overall zone-level population distribution. The table below describes the population of ABE girls in the sample frame by zone, the allocated (targeted) interviews by zone, and the achieved interviews. As the table shows, the sample matches the underlying distribution of ABE girls across zones, within an expected level of approximation, given the fact that a clustered sample can rarely match population distributions precisely.[23]

TABLE 11: GEOGRAPHIC DISTRIBUTION OF ABE POPULATION AND SAMPLE

Zone	ABE Population	ABE Targeted Sample	ABE Achieved Sample

[22] In these cases, focusing on the baseline and ML2 data exclusively by using the "baseline panel" maximizes the number of unique schools included in the sample, because the ML1 sample had the fewest schools overall. The "full panel" covers fewer schools as a result, because it requires schools to appear in all three rounds in order to be included.

[23] This is due to the fact that clustered samples are allocated by cluster, comprising 15 interviews in this case, instead of by individual interviews. If 1, 2, or 3 clusters were allocated to Galmudug, it would comprise 2.9, 5.7, or 8.6 percent of the target sample; following standard practice, we allocated the number of clusters that would most closely recover the true population distribution.

Galmudug	6.9%	5.7%	6.2%
Puntland	38.9%	40.0%	40.5%
Somaliland	54.2%	54.3%	53.3%
Total	100.0%	100.0%	100.0%

Within ABE centers, girls were selected randomly. Team leaders composed a list of all ABE girls in the center who were eligible for inclusion, where inclusion was defined by their age range (12-22 years), their non-inclusion in another cohort sample as part of the SOMGEP-T evaluation, and their attendance at the ABE center during the two days of data collection.[24] Once the list was constructed, the team leader selected 15 girls randomly, using a random number generator. In centers with fewer than 15 eligible girls, all eligible girls were recruited into the sample. The teams were able to complete a full cluster of 15 interviews in 27 centers, with as few as 5 and 7 girls eligible in some centers. As a result, the sample achieved for ABE girls was just 482 girls.

Only one ABE center selected as part of the original sample draw required replacement, because no eligible ABE girls were present at the time of fieldwork.[25] This center was located near Fadhigab Primary School, in Ceel-Afweyn, Sanaag; it was replaced with an ABE center located in Aynabo, in neighboring Sool.[26]

2.3 Methodological Analysis

In this section, we consider critical methodological issues related to the third evaluation round and investigate their potential impact on the results presented throughout this report. We do not analyse every potential methodological pitfall of the evaluation; rather, we focus on those that present particularly thorny problems for drawing causal inferences regarding the program's impact. The SOMGEP-T evaluation design is quite rigorous, with a well-powered intervention and comparison group that has been tracked systematically over two – or, for about half the cohort group, three – rounds.

The discussion below is not intended to imply that the methodology of the evaluation is systematically flawed or invalid. Instead, it is intended to serve two goals: to systematically consider threats to inference that may occur to readers, but which can be discounted through supplementary analysis; to briefly discuss common issues that are not actually true threats to inference, given the overall design of the evaluation; and to highlight genuine threats to inference so that, in discussing our substantive results below, we can refer readers to those concerns and make clear the extent to which methodological challenges are actually problematic. In short,

[24] On one hand, the last criterion produces a sample that is not truly representative of ABE girls, some of whom were not present during fieldwork. To the extent that this occurred, it biases the sample toward girls who attend the center more consistently. On the other hand, if the evaluation's goal is to understand the impact of the ABE intervention vis-à-vis non-enrolment, it is reasonable to select girls who we can verify are being exposed to the intervention.

[25] Specifically, girls were present at the ABE center, but they were all reportedly too young (under 12 years of age) for the eligibility criteria.

[26] The decision to replace the Fadhigab ABE center did not impact the status of Fadhigab Primary School in the sample, nor did it impact the inclusion of the local ALP center (both were included in the sample, as in previous rounds).

the goal of this section is to make clear under what circumstances and the extent to which methodological issues are of true concern, so that readers can interpret our findings with the appropriate degree of caution.[27]

RE-CONTACT RATES

The design employed by the SOMGEP-T evaluation is quite rigorous overall. Much of the inferential power of the design relies on the difference-in-differences design, which can be employed with either a panel or a repeated cross-section of respondents.[28] The primary advantage of the panel design is that it reduces or eliminates the potential for random sampling variation to drive our results. In a repeated cross-sectional design, a particularly unusual sample of intervention girls at endline could result in positive or negative estimates of impact that are driven by the unusual nature of the sample. On average, a repeated cross-sectional design is still unbiased, but there is no guarantee that individual iterations of the design will produce unbiased results, due to the potential for sampling variation. A panel design, on the other hand, eliminates this specific threat to inference.

At the same time, a panel design is only inferentially valid if the rate of panel attrition is minimized. If many girls fall out of the sample and are not replaced, the design is weakened in terms of statistical power. If many girls fall out of the sample and are replaced, the design begins to resemble a repeated cross-section, rather than a panel study. Minimizing attrition, maximizing re-contact rates, and ensuring replacement girls are selected in a non-biased fashion – to mirror the girls they are replacing, to the extent possible – is of the utmost importance.

In this section, we analyse re-contact rates for the second-round midline evaluation. We use the term re-contact to refer to a cohort girl (ALP, in-school girl, OOS girl) who was successfully located and interviewed in the current round. Under this definition, re-contact rates are the inverse of attrition rates. We report re-contact rates for each cohort group, and then discuss a complicating factor in the calculation of re-contact rates, which actually improves baseline-to-midline re-contact rates. Next, we analyse predictors of successful re-contact, to determine whether panel attrition is "as if random" or correlated with project outcomes in a way that might produce bias in our analysis. Finally, we discuss the process by which replacement girls were selected, and investigate whether replacement girls are statistically similar to the girls they are replacing.

Before turning to a discussion of re-contact rates, note that our procedures for re-contact were in line with the first-round midline evaluation. For girls who were recruited into the panel previously, data collections completed these steps to locate them:

- Called every phone number listed for the household a minimum of two times, allowing at least six hours between phone call attempts. In each round of data collection, households were asked to provide two contact phone numbers; in practice, this means that each household had between one and four numbers listed.
- Visited her household a minimum of two times, allowing at least six hours between visit attempts.
- Asked the head teacher and other teachers in the school for contact information for the girl or her family, and how they could be reached.

[27] The analyses in this section focus on re-contact and attrition exclusively. For the sake of space, we have relegated to Annex 13 analysis of whether the learning assessments used in this round are of equivalent difficulty to those used in prior rounds.
[28] A repeated cross-section would target the same intervention and comparison groups in multiple rounds of data collection, but would not seek to track the same specific individuals within those groups over time. In contrast, in a panel design, we seek to contact the exact same set of respondents in multiple rounds of data collection.

- Asked other girls in the same grade or age group if they know the girl and how she could be reached.

These procedures generally mirror standard practice in panel surveys in underdeveloped areas. One point on which our approach differs is in the number of telephone contact and visitation attempts made and the wait time required before making a second attempt. Many large household panel surveys will require contact attempts to occur on separate days, or even specify that a full day must pass between days with unique contact attempts. This is not possible in the context of SOMGEP-T, because fieldwork logistics dictate that teams typically spend two days at each sample location. A typical site visit would include a contact attempt in the early afternoon of the first day, and a second attempt the following morning. In general, it is not possible to *require* longer wait times between attempts, given the logistics of fieldwork on this project. While the short timeframe and challenging environment likely exacerbated panel attrition, other characteristics of the SOMGEP-T sample – especially its largely rural nature – improved re-contact rates in spite of these difficulties.[29]

Re-contact rates across four distinct cohort groups are reported in the table below. The first three categories are relatively straightforward to define. The first cohort consists of in-school girls, who were either recruited into the cohort at baseline and successfully re-contacted during ML1, or were recruited as replacements during ML1. In all cases, girls in this cohort were contacted at ML1, which means that one year has passed since they were last interviewed. Among this group, re-contact rates were 81.9 percent, which is marginally higher than re-contact rates during ML1 among the same *type* of girl.

The second cohort consists of out-of-school girls, who were recruited during the baseline and were not enrolled in school at that time. Girls in this cohort were not re-contacted during ML1, meaning that two years have passed since they were last interviewed. Re-contact rates among this group were much lower, at 53.8 percent. Lower re-contact rates fit three possible explanations, all of which are likely operative to varying degrees: the first is that out-of-school girls are less connected to the local school, such that teachers and other girls in the school are less useful as informants regarding their location; the second is that out-of-school girls are, arguably, at greater risk of early marriage, migration, or other outcomes that would take them out of their previous location; the third is that two years had passed since their last interview, providing an additional year of potential out-migration, marriage, and so forth.

TABLE 12: COHORT-SPECIFIC RE-CONTACT RATES

Cohort group	Cohort size	Re-contact Rate
In-school girls	807	81.9%
Out-of-school girls	693	53.8%
ALP girls	365	70.7%
In-school girls who were not re-contacted at ML1	240	48.3%
Total	2,105	66.9%

[29] Rural areas and small villages are particularly valuable for facilitating re-contact, because teachers and students are familiar with most or all members of their community, and can provide guidance to the data collection teams on locating the targeted girls. Even if the girl is no longer living in the village, they often know why and the location to which she migrated.

The third cohort group re-contacted during this round are ALP girls, who were recruited into the sample for the first time during ML1. Among these girls, re-contact rates fell somewhere in the middle, at 70.7 percent. This result is somewhat surprising, given that ALP girls are engaged in a local education initiative and only one year had passed since their last contact. On the other hand, ALP girls are older than their in-school and out-of-school counterparts in the cohort, because they followed a different recruitment logic – at the time of their recruitment, their average age was 17.0 years, compared to 13.6 years among in-school and out-of-school girls, combined, in the same data collection round. Their older age means that ALP girls are more susceptible to life outcomes that would remove them from the sample, such as early marriage and out-migration.

Finally, we report re-contact rates for in-school girls that were not contacted during ML1. These girls fall into two categories: those whose schools were not visited, for budgetary and logistical reasons, during ML1, and those who could not be located during ML1. In the former case, eight schools were excluded from the ML1 sample with the intention of revisiting them during ML2 and the endline. Girls in this category are different from regular in-school girls in that two years passed between contact attempts. In the latter case, we attempted to contact in-school girls who we did not successfully re-contact – despite visiting their schools and attempting to contact them – during ML1. These girls had fallen out of the sample, but there were indications that we might find some of these girls if we made a second attempt to contact them, as some were only *temporarily* unavailable (due to illness or short-term travel) during ML1 data collection. Our idea was to seek these girls out, in the hope that we could reduce total panel attrition by recruiting them back into the cohort sample. Among this group, unsurprisingly, re-contact rates are very low: in the case of schools that fell out of the sample, two years had passed since we contacted the girls; in the case of girls who could not be located at ML1, they had already fallen out of the sample, so finding them now is necessarily a low-likelihood event. Given these challenges, the 48.3 percent re-contact rate among these girls is surprising and a genuinely positive result.

The decision to re-contact girls who fell out of the sample at ML1 does complicate the nature of the sample, as girls fell out of the sample and reappear at ML2. Moreover, girls who fell out of the sample at ML1 were generally replaced; in cases where they were brought back into the sample at ML2, their replacement and the original girl both appear in the sample. It is for this reason that we tend to prefer the "true panel" sample for most of our analysis below, because this limits the sample to the set of girls who are the same in both periods under consideration. For instance, the true panel for BL-to-ML2 comparisons includes all girls who appear in both the baseline and ML2, regardless of whether they were re-contacted successfully during ML1.

The picture is further complicated by the fact that some schools were permanently dropped from the sample after the baseline, while other schools were temporarily excluded during ML1 for logistical or other reasons described above. The figure below attempts to make clear the points at which sample attrition occurred.

Between baseline and ML1, two waves of school-level attrition occurred. The first was removal of outlier schools from the baseline, who showed particularly high English literacy scores for idiosyncratic reasons explored in more detail during the baseline evaluation. The second wave was the set of schools that were excluded from ML1 but still constituted part of the sample in the long-term (i.e. the plan was to re-visit them at ML2). A further bout of individual-level attrition occurred at ML1 when girls from the baseline could not be located or re-contacted. This set of girls is represented in the figure by the line connecting baseline and ML1.

Between ML1 and ML2, the figure distinguishes between schools that were successfully brought back into the sample and schools that continued to be excluded. The five high-performing schools were excluded

permanently; the schools excluded for logistical and other reasons at ML1 were brought back into the sample in most cases, with two exceptions that were specific to security and accessibility concerns.

Like school-level attrition, individual-level attrition also occurred in each round, but was two-way in nature (i.e. girls did not exclusively *leave* the sample; they also re-entered the sample, in later rounds). During the ML1 evaluation, individual-level attrition occurred when girls could not be located. Such girls were replaced, where possible, or fell entirely out of the sample if no suitable replacement girl was available.

Between ML1 and ML2, individual-level attrition was two-way. Cohort girls of several different kinds fell out of the sample: those who were OOS at baseline and had not been contacted since that time; girls whose schools were excluded from the ML1 sample and had not been contacted since baseline; girls who had been successfully re-contacted at ML1; and replacement girls selected at ML1. Each of these groups experienced non-zero attrition, as we would expect. At the same time, girls re-entered the sample at ML2, especially those who we attempted to contact at ML1 but who could not be located.

The goal of the figure is to illustrate the complexity of the re-contact and replacement dynamics at work in the SOMGEP-T evaluation. It also illustrates the relative magnitude of different sources of attrition and re-entry into the sample. For instance, 116 in-school girls who had fallen out of the sample at ML1 due to individual-level attrition were brought back into the sample at ML2, while 146 in-school girls were replaced (137) or fell out of the sample entirely (9) at ML2. This means that, while ML1-to-ML2 attrition among in-school girls was 18.1 percent, *net* attrition – taking into account the girls who were brought back into the sample at ML2 – was just 3.7 percent.

The table below documents the net re-contact rates from baseline to ML2, where net enrolment takes into account the re-entry of girls over time. Our interest here is in the overall share of girls from baseline who were successfully re-contacted at ML2, whether they were re-contacted at ML1 or not.[30] The first column reports re-contact rates for girls whose schools remained in the sample between baseline and ML2. Re-contact here means the ability to re-contact the same girl, assuming we visited her school at ML2. The right column reports the aggregate re-contact rate, including schools that were dropped from the sample wholesale. Naturally, re-contact rates here are lower, due to the fact that some schools were dropped from the sample entirely following baseline.

TABLE 13: NET RE-CONTACT RATES FROM BASELINE TO MIDLINE ROUND 2

Cohort group	Individual-Level Re-Contact Rate	Total Re-Contact Rate
In-school girls	72.6%	65.4%
Out-of-school girls	53.1%	48.8%
Total	64.1%	58.2%

As the results make clear, re-contact rates are much lower among OOS girls than among girls who were enrolled in school at the time of the baseline. Re-contact rates are low, despite efforts to bring girls back into the panel sample at ML2. At the endline, all girls – OOS and in-school girls alike – who have fallen out of the sample since baseline should be sought out, and brought back into the sample if they can be located. This will

[30] This focus fits with our tendency to rely on BL-to-ML2 comparisons throughout this report in most cases.

reduce the overall attrition rate from baseline to endline, resulting in a stronger evidence base for understanding program impact.

PREDICTORS OF SUCCESSFUL RE-CONTACT

The re-contact rates described above are relatively low. However, given the context in which SOMGEP-T is being implemented, CARE's Monitoring and Evaluation team anticipated a high attrition rate, and designed the sample with relatively high attrition rates in mind. If net attrition from ML2 to endline follows the trend from previous evaluation rounds, we anticipate additional attrition of approximately 8 percentage points, resulting in a total "true panel" sample size of around 932 girls.

Beyond aggregate re-contact rates, an additional methodological concern is the extent to which attrition or re-contact is non-randomly distributed across the sample. The analysis in the previous section showed – in line with our existing expectations and prior experience – that attrition is much higher among OOS girls than in-school girls. Other factors can also shape attrition rates. Attrition that is as-if random – i.e. if attrition is not correlated with the outcomes of interest for the evaluation, such as learning – does not produce bias in conclusions drawn regarding change in program outcomes over time. The only methodological concern in that case is a reduction in sample size that may reduce the statistical power of the analysis.

In contrast, non-random attrition can produce bias in estimates of program impact. If girls who are predisposed to lower learning outcomes tend to fall out of the sample over time, we will overestimate the improvement in learning scores that have occurred over time. Importantly, if attrition occurs equivalently in intervention and comparison schools, bias will not result, even in the case of non-random attrition. That is, if girls predisposed to lower learning outcomes fall out of the sample at higher rates, but this occurs equally in both intervention and comparison schools, no aggregate bias plagues our estimates of program impact. A cross-sectional analysis – one in which replacement girls are included in the sample – will be biased in the context of differential attrition between intervention and comparison schools, if replacement girls are not equivalent to the girls they replaced. In this section, we seek to understand the nature of attrition overall, and whether it is likely to result in biased inferences, by analysing the nature of attrition and the similarity of replacement girls and those they replaced.

We begin by reporting aggregate re-contact rates across geographic zone and age group, in the table below. To be clear, our analysis in this section is concerned with individual-level attrition, ignoring the removal of entire schools. The overall re-contact rate from baseline to ML2 is 64.1 percent, across both in-school and out-of-school cohorts, where cohort is defined by the girl's enrolment status at baseline.

Geographic zone does not strongly predict re-contact rates, in the case of Somaliland and Puntland, where re-contact rates are very similar to one another. By comparison, the re-contact rate is much lower in Galmudug. Despite the small sample size in Galmudug (n = 93 girls at baseline), the difference between Galmudug and the other two zones is statistically significant at the 5 percent level, even after accounting for clustering at the school level.

The results also show that successful re-contact is highly correlated with age. We report results based on the girl's age at baseline. Re-contact rates are highest among 10- and 11-year old girls, who make up a combined 28.0 percent of the sample. While rates do not decline monotonically with age, the trend is unmistakeable: among girls 10-13 years old, re-contact rates were 70.8 percent; among girls 14-15 years old, re-contact rates were 59.0 percent; and among girls 16 years or older, re-contact rates fell to just 51.0 percent.

TABLE 14: RE-CONTACT RATE BY SUBGROUP

Subgroup	Re-Contact Rate
Overall	64.1
Zone	
Somaliland	64.7
Puntland	65.7
Galmudug	48.4
Age	
10	72.7
11	71.8
12	68.1
13	70.5
14	58.1
15	60.3
16	51.5
17	50.7
18	51.4
19	48.7

The results above are based on a bivariate analysis, calculating re-contact rates among a specific subgroup, but without taking into account the correlation between demographic characteristics of the girls. For instance, girls in Galmudug are slightly older than their counterparts in the other zones, which may partially explain the lower re-contact rates in Galmudug. Similarly, older girls were more likely to be out-of-school at the baseline, and this may also explain the relationship between age and lower re-contact rates.

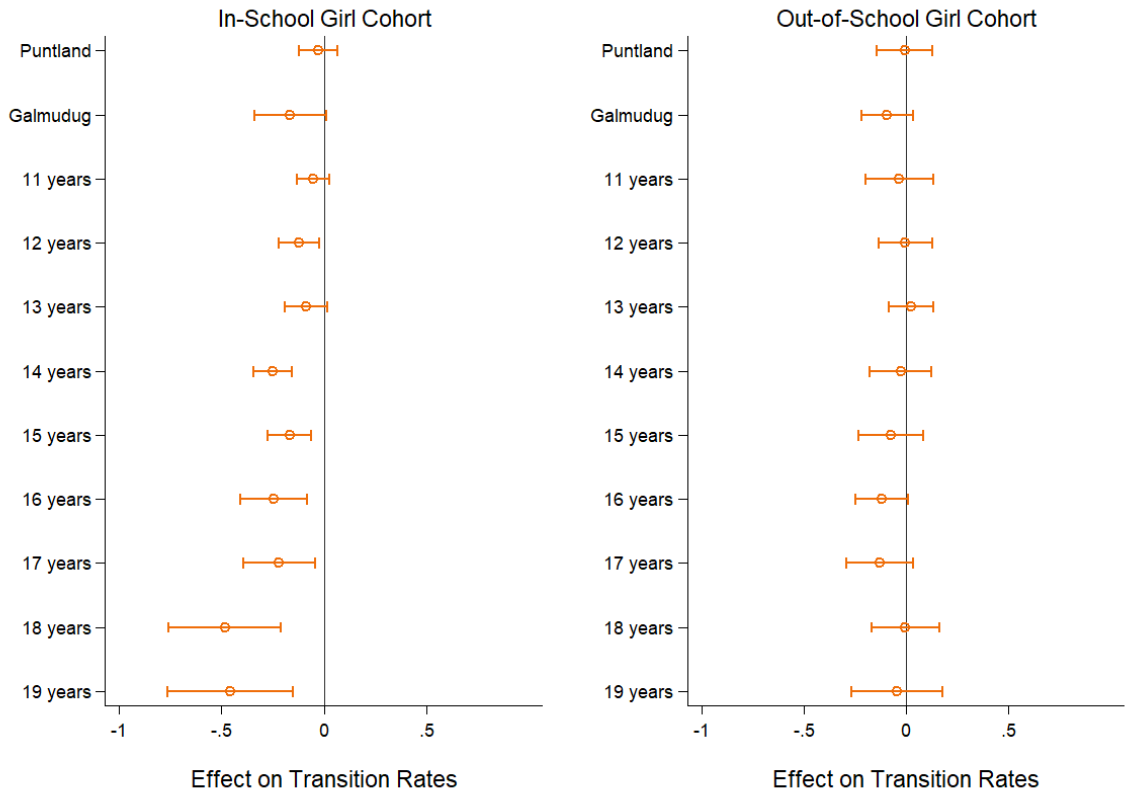
The figure below illustrates the relationship between age and geographic zone, on one hand, and re-contact rates, on the other. Importantly, we analyse separate subsamples for in-school and out-of-school girls, to ensure that differences between these two groups in terms of age or geography are not driving the results. The left panel reports findings for in-school girls, while the right panel limits our analysis to out-of-school girls. Each panel is derived from a linear regression predicting successful re-contact, where the dot represents the regression coefficient for a given characteristic (i.e. age 10) and the horizontal bar represents the 95 percent confidence interval around that coefficient. Where the 95 percent confidence interval does not cross the vertical line at 0, it means the correlation between that variable and successful re-contact is statistically significant at the 5 percent level. Each age group's relationship to re-contact is calculated as a binary or dummy variable for that age group, where girls aged 10 are the omitted category. Therefore, the impact of being 11 years old on re-contact rates is measured as the difference in re-contact rates between 11-year old girls and 10-year old girls.

Several findings emerge from the figure. First, even after controlling for age and accounting for differences between in-school and out-of-school girls, Galmudug appears to have systematically lower re-contact rates

than the other zones. The result is not statistically significant in either subsample – denoted by the fact that the 95 percent confidence interval crosses the vertical line in both cases – but the negative correlation is consistent across the two samples.

The second finding that emerges is the fact that age is a better predictor of successful re-contact among in-school girls than OOS girls. Re-contact rates decline with age in both cases, but re-contact rates recover among older OOS girls, and any negative trend among OOS girls is less dramatic than that observed among in-school girls, where successful re-contact becomes rarer with age across the board. One interpretation of this finding is that being out of school is strongly associated with lower re-contact success, and older in-school girls are more likely to drop out of school. In other words, older girls in the in-school girl cohort are more likely to drop out of school, and it is the act of dropping out that reduces the likelihood of being successfully re-contacted. In contrast, for girls who are already out-of-school are less, re-contact rates are low across all ages, and age does not have as strong an impact on re-contact rates.

FIGURE 1: PREDICTORS OF SUCCESSFUL RE-CONTACT AT ML2, BY COHORT



The findings above are based on a linear regression predicting successful re-contact, as a function of age and geographic zone. We expanded on this analysis by incorporating additional demographic characteristics and program outcomes into the model of re-contact. Loosely speaking, the characteristics we incorporate capture aspects of household structure (female-headed households and girls who are married), economic status (pastoralist households, roof quality, ownership of assets, and experiences of hunger), community or school characteristics (safety of traveling in the area, availability of food via school attendance), and program outcomes (learning and YLI scores). The set of characteristics we analyse is generally limited to those that

apply to a significant minority of girls and which were captured for both in-school and out-of-school girls – it is for this reason that we do not incorporate aspects of teaching or school quality, as these factors do not apply to – and were not captured for – out-of-school girls.[31]

In the figure below, we report the results of three linear regression models predicting successful re-contact. The variables on the y-axis of each graph are included as predictors in the model; each model also controls for geographic zone and age, as in the results above. The results presented document the regression coefficient (the relationship between belonging to a subgroup, such as female-headed households, and successful re-contact) and the 95 percent confidence interval, which indicates our confidence in the result.

Among in-school girls, there are few strong predictors of successful re-contact, beyond age and geographic zone, which are not shown in the graph. Girls in female-household households have re-contact rates 8.0 percentage points lower than girls in male-headed households, after accounting for the effect of age, zone, and the other characteristics in the graph. Near the bottom of the graph for in-school girls, the results indicate that girls who achieved higher numeracy, Somali literacy, and YLI scores at baseline are slightly more likely to be re-contacted successfully at ML2, though none of these results approach statistical significance.[32]

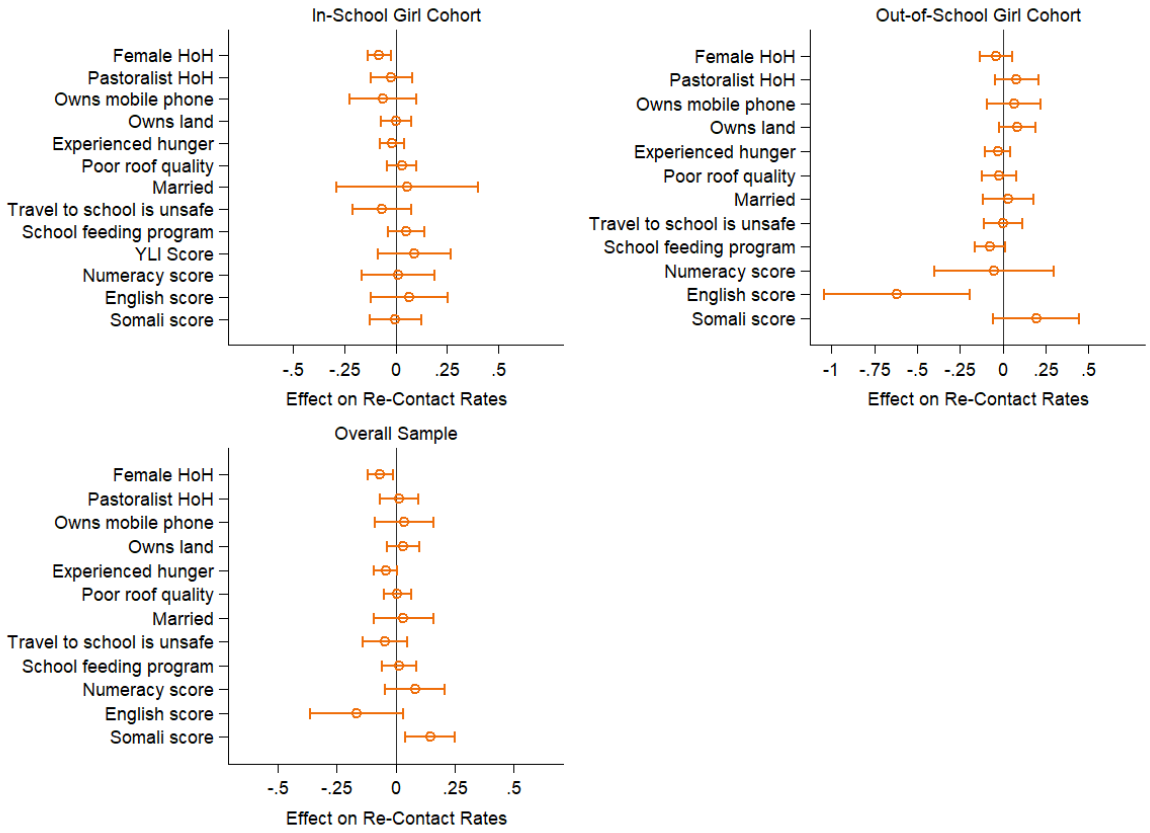
Among OOS girls, there are few consistent predictors of successful re-contact. For instance, girls in households that own a mobile phone or land are slightly more likely to be re-contacted successfully. At the same time, though, girls in households that have experienced hunger in the previous year are also more likely to be re-contacted successfully. It does not appear that household economic status is a strong predictor of re-contact rates among either in-school or OOS girls. The starkest finding among OOS girls concerns English literacy scores: for every 10-point increase in a girl's English literacy score at baseline, the likelihood of re-contacting her declined by 6.2 percentage points. There is not a clear explanation for this finding, as we would not expect any specific correlation between English performance and re-contact rates among OOS girls, unless girls who achieved higher scores are more likely to migrate away from the rural areas where SOMGEP-T is being implemented and into urban areas.[33]

[31] The single exception is the YLI score, which was not captured for OOS girls at baseline. Nonetheless, we include this variable because it is of particular interest, as we might expect girls with lower self-esteem or less empowerment within their households and communities to be more likely to migrate, get married, or otherwise drop out if the sample.

[32] Although each learning score does not, in isolation, approach statistical significance, correlated variables may – in combination – predict successful re-contact. This argument does not apply in this case, however; even when tested jointly, the impact of the three learning scores and YLI scores on successful re-contact is indistinguishable from a null effect.

[33] For instance, it is possible that OOS girls who achieved higher English literacy scores were temporarily out-of-school and have since enrolled in school elsewhere, perhaps while living with relatives.

FIGURE 2: DEMOGRAPHIC CHARACTERISTICS AND LEARNING OUTCOMES AS PREDICTORS OF SUCCESSFUL RE-CONTACT



Across the different analyses, there are few patterns with respect to successful re-contact. The most robust finding is that older girls are less likely to be re-contacted, a finding that applies primarily to in-school girls. The second robust finding centres on Galmudug, where re-contact rates were systematically lower. After controlling for a number of demographic and other characteristics, re-contact rates were still lower in Galmudug, at least among in-school girls – re-contact rates among in-school girls were 17.0 points lower in Galmudug than among equivalent girls in Somaliland, and similarly low compared to Puntland.

The biggest inferential concern surrounding attrition is that attrition may operate in a systematically different way between intervention and comparison communities. Unfortunately, there does appear to be evidence to support this concern. First, re-contact rates in Galmudug are strongly associated with intervention status: in intervention communities, re-contact rates were 54.3 percent, compared to 30.4 percent in comparison communities, though this finding utilizes a very small sample of comparison schools in Galmudug (just 1). For comparison, re-contact rates in the overall sample were 63.4 percent among intervention schools and 64.9 percent among comparison schools.

In addition, re-contact rates among married girls are related to intervention status. Among non-married girls, re-contact rates are very similar; however, among married girls, re-contact rates were 44.8 percent and 70.0 percent for intervention and comparison communities, respectively.

It is possible that these findings do not represent a systematically different mechanism by which girls self-sort into re-contact or attrition between intervention and comparison communities. We have tested a number of relationships between girls' characteristics and re-contact success, and occasional false positive results are consistent with basic statistical theory. The next section tests whether girls who fall out of the sample are replaced by functionally equivalent – in terms of learning outcomes and demographic characteristics – girls. Insofar as replacements mirror the girls they replace, our concern about differential attrition will be reduced, but not entirely eliminated.

COMPARISON OF SAMPLE COMPOSITION

Replacement procedures employed during data collection were designed to ensure comparability of replacement girls and the girls who they were selected to replace. Replacement girls were selected from the grade the original girl *would* have been enrolled in, if she had progressed grades as expected (i.e. 1 grade per year). However, in many cases, no girls who met this criterion were available – out of 145 replacement girls selected at ML2, only 54.5 percent were selected from the correct grade. The remainder were selected from a different grade, because all girls in the grade were already part of the existing cohort. Team leaders were instructed to replace girls using the closest available grade level to the guidelines (e.g., move up or down one grade from the target grade, until an eligible girl is located).

In total, 145 girls were replaced during ML2, and 159 girls were replaced previously during ML1.[34] In this discussion, we continue our focus on cohort girls selected at baseline, and relegate discussion of ALP girls – who entered the evaluation via a different sampling method – to a separate section. After accounting for re-recruitment into the sample, we analyse 277 pairs of replaced and replacement girls.

To assess the similarity between replaced girls and their replacements, we performed t-tests that accounted for clustering by school level. For each of the variables in the table below, we compare the share of replacement girls with that characteristic to the share of replaced girls with the same characteristic; we report the share for each group and identify those gaps that are statistically significant in the right-most column.[35]

The first row reports age differences between replaced (original) and replacement girls. Replacement girls are older on average, but this finding is expected, as we sought to draw replacements from the appropriate grade, and girls in the round-appropriate grade – such as grade 6 at ML2 for a girl who was originally recruited in grade 4 at baseline – are necessarily older, on average.

Beyond age, we find a number of small, but statistically insignificant differences between the two groups of girls. For instance, replacement girls are somewhat more likely to live in female-headed households. However, only asset ownership is systematically different between the two groups, with replacement girls drawn from households that are less likely to own a mobile phone and less likely to own land.

We also find that caregivers of replacement girls are much less likely to view girls traveling in the area as unsafe. The share of caregivers who believe traveling to local schools is unsafe for girls in their community was 6.9 percent among the original girls, and 1.9 percent among their replacements. It is important to note, however, that a similar decline occurred among girls who were successfully re-contacted at ML2, meaning that – even among a set of caregivers who remained stable from baseline to ML2 – safety rates seemed to

[34] A portion of these 159 girls were re-contacted successfully at ML2 and recruited back into the sample, as discussed previously.
[35] A single asterisk or star represents differences that are significant at the 10 percent level; two stars indicate results that are significant at the 5 percent level. We report findings based on a 10-percent threshold partially due to the small sample size of replacement girls we have available for analysis.

improve. Caregivers of replaced girls answered this question at baseline, while caregivers of replacement girls answered it at ML2; therefore, any structural shift in community safety would manifest as a difference between replaced and replacement girls, but may not reflect a true difference between these two groups. Our interpretation of this finding is that safety broadly improved over time, and this explains the gap between baseline girls and their replacements.

TABLE 15: DIFFERENCES BETWEEN REPLACEMENT AND REPLACED GIRLS, BL TO ML2

Girl/Household Characteristic	Original Girl	Replacement Girl	Significant Difference
Age	13.4	14.2	**
Female head of household	50.2	56.6	
Living without either parent	8.7	10.6	
Disability of any kind	5.1	5.8	
Pastoralist head of household	9.7	9.9	
Household owns a mobile phone	96.4	91.6	*
Household owns land	81.5	72.8	**
Household experienced hunger in last year	30.7	27.8	
Household has a poor-quality roof	30.4	30.9	
Girl is married	2.2	1.1	
Girl has a heavy chore burden	11.9	9.2	
Traveling to schools in the area is unsafe	6.9	1.9	*
School has a feeding program	61.7	66.1	
Numeracy score	43	58.8	**
English literacy score	11.5	22.4	**
Somali literacy score	45.9	65.3	**

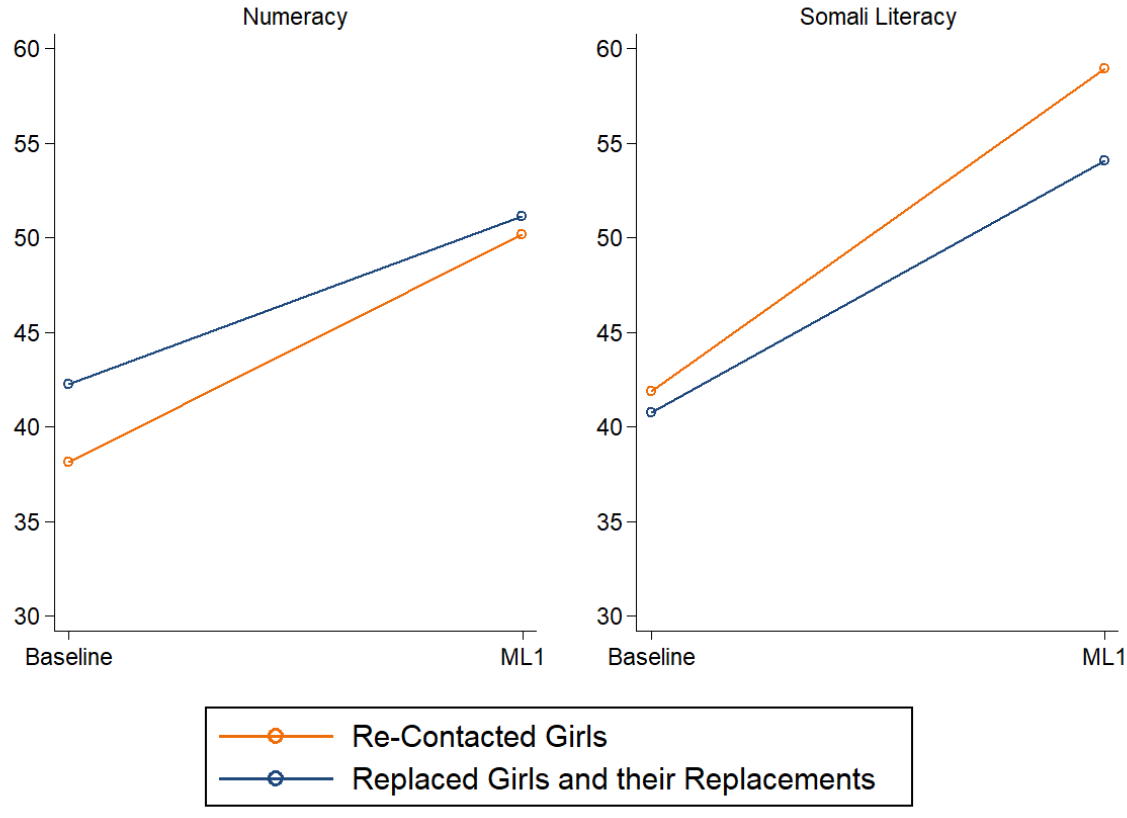
The most problematic findings are those in the final three rows of the table, which show that replacement girls performed significantly better on all three learning assessments than the girls they replaced. Somali literacy scores, for instance, were almost 20 points higher among replacement girls than among the girls they replaced.

Findings around learning scores need to be considered with a significant caveat, though, due to the relationship between replacement status, age, and grade. Replacement girls are approximately one year older – at the time they took the learning assessment (ML2) – than were the original girls when they took the learning assessment (at baseline). Moreover, the median replacement girl was in grade 6, while the median girl being replaced was in grade 5 when she completed the learning assessment. Naturally, we expect girls who are older and in a higher grade to score more highly on learning tests.

We investigated this issue at length because it is of such importance to the overall evaluation. If replacement girls achieve higher scores than the girls they replaced, it would invite significant scepticism regarding any reported changes in learning scores over time; the sample is becoming higher-achieving over time as a result of replacing girls, not exclusively as a result of maturation effects and program impacts.

To understand the nature of learning scores among replacement girls, we plot learning scores for two groups of girls over time. The first group, represented in orange, are in-school girls who were successfully re-contacted, and appear at both baseline and ML1. The second group are replaced girls and their replacements.[36] As noted above, we expect girls who are older and in a higher grade to score more highly on the assessments. But we would not expect their gains in performance to outpace those of "true panel" girls who are also advancing in age and grade level.

FIGURE 3: LEARNING SCORES AMONG RE-CONTACTED AND REPLACED IN-SCHOOL GIRLS, BL TO ML1



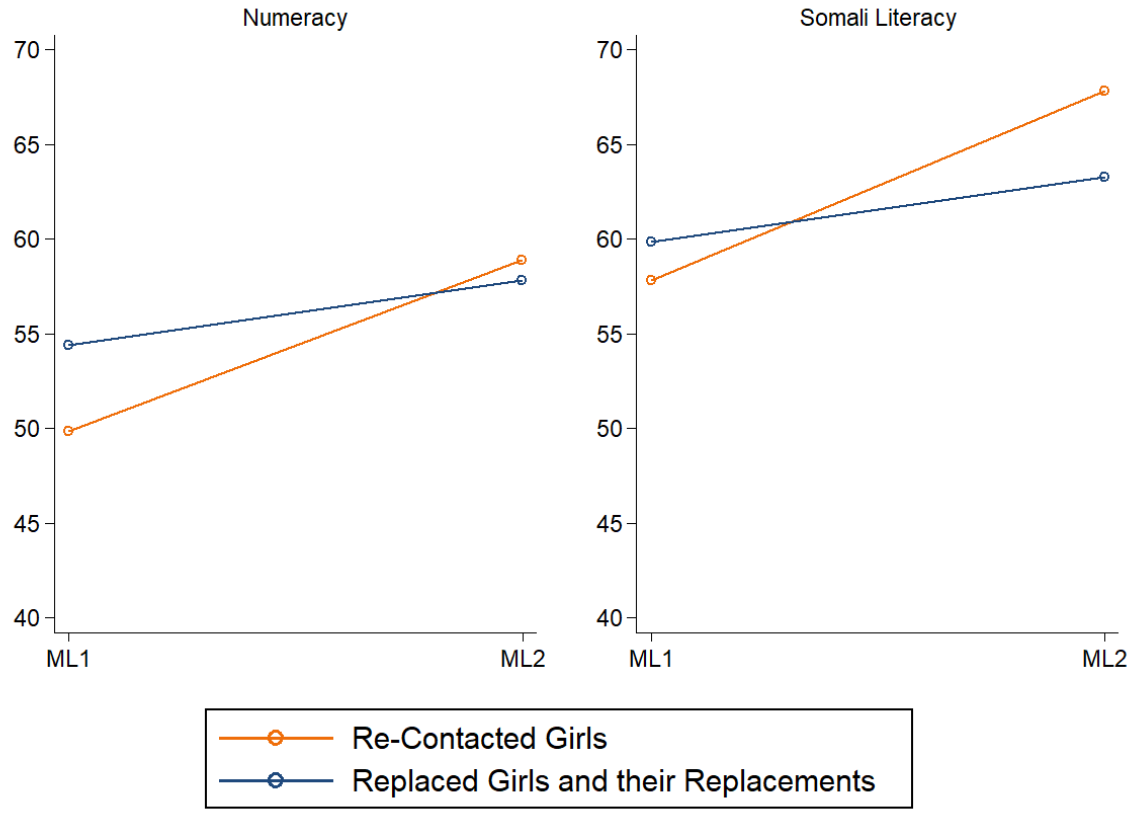
In the left panel, panel girls underperform replaced and replacement girls in numeracy in both rounds, while they outperform replaced and replacement girls in Somali literacy (right panel). However, our interest is in the evolution of scores over time. In both cases, panel girls gained – relative to replaced/replacement girls – from baseline to ML1. Contrary to our discussion above, replacement girls do not appear to be systematically pulling up learning scores, via a mechanism in which lower-scoring girls and replaced by higher-scoring girls.

[36] We limit the sample of re-contacted girls to in-school girls, because replaced/replacement girls are exclusively in-school at the time of their initial recruitment, ensuring a like-for-like comparison.

Rather, the gap between replaced and replacement girls is *smaller* than the gains in learning that panel girls achieve over the same time period. That is, panel girls improved their numeracy performance by 12 points from baseline to ML1. This represents a true improvement, as the set of girls assessed stayed constant across the two rounds. If replacement girls were being selected in a way that would drive up average scores, we would expect their ML1 performance to outpace the girls they replaced by *more* than 12 points; in reality, their performance outpaced the girls they replaced by just 8.9 points.

We repeated this analysis for replacements that occurred between ML1 and ML2 as well, and report the results in the figure below. Again, the gap between replacement girls at ML2 and the girls they replaced from ML1 are *smaller* than the gains achieved by true panel girls over this time period.

FIGURE 4: LEARNING SCORES AMONG RE-CONTACTED AND REPLACED IN-SCHOOL GIRLS, ML1 TO ML2



These results should not be interpreted as a blanket indication that replacement girls are identical or even comparable to the girls they replaced. In practice, even if replaced girls mirror their replacements on a number of characteristics that we measured as part of the evaluation, there is no guarantee that the two sets of girls are comparable across all characteristics.[37]

[37] This distinction is often framed as the difference between observed and unobserved variables, either in the context of regression analysis (where controlling for observed factors does not necessarily account for unobserved factors), in quasi-

In reality, the nexus of our concern is about differences between intervention and comparison girls in terms of *who falls out of the sample* and, subsequently, *who is selected as a replacement*. If a different *type* of girl falls out of the sample in intervention communities, vis-à-vis comparison communities, this would produce bias in difference-in-differences analysis. Likewise, if a different *type* of girl is selected as a replacement in intervention communities, relative to comparison communities, this difference also produces bias.

The table below compares replaced and replacement girls' learning scores in intervention and comparison communities. The results are broken down into two rounds of replacement, for ease of interpretation – for instance, the first panel reports baseline and ML1 numeracy scores for girls who were replaced (baseline) and the girls who replaced them (ML1). The second panel repeats this analysis for girls who were replaced at ML2, by reporting their ML1 learning scores to the ML2 scores of those girls who replaced them. The comparison of interest is reported in the final column – if the difference between replaced and replacement girls in intervention and comparison communities are similar, there is no reason to be concerned about bias.

In the context of replacements taken during ML1, intervention and comparison schools seem to be very similar. Numeracy scores are much higher (about nine points) among replacement girls, but this is true for both intervention and comparison schools; a similar pattern emerges in Somali literacy, where replacement girls in intervention and comparison communities scored 12.8 and 14.1 points higher than the girls they replaced, respectively.[38]

Where problematic results arise in the table below is in the context of replacements made at ML2. In these cases, replacement girls outperform the girls they replaced only in comparison communities, but not in intervention communities. For instance, in intervention communities, replacement girls drawn at ML2 scored just 0.6 points higher than the girls they replaced had scored at ML1. In comparison communities, replacement girls outperformed the girls they replaced by 6.4 points. This gap produces a comparison group whose scores have increased – relative to the intervention group – due to the replacement process, rather than due to learning achieved through the school.[39]

TABLE 16: LEARNING SCORES AMONG REPLACEMENT AND REPLACED GIRLS, BY INTERVENTION STATUS

	Baseline	ML1	ML2	Difference
Numeracy – Baseline to ML1				
Intervention	41.4%	50.6%		9.2%

experimental settings such as that studied here, or even in experimental settings, where researchers often want to verify the similarity of their treatment and control groups across a range of observed characteristics, but are left wondering whether the two groups are truly balanced on unobserved factors.

[38] We do not report similar findings for English literacy, because English literacy was not assessed at ML1. Therefore, the contrasts we are able to draw are less direct. In practice, we expect similar findings to emerge across subject areas, due to the broad correlation between English literacy, Somali literacy, and numeracy scores we observe in the data at large.

[39] We attribute this gap to the replacement process and not to improvements in learning, more broadly, in comparison schools, because we do not observe a similar gap between intervention and comparison communities among true panel girls. If the difference between replaced and replacement girls were reflective of dramatic improvements in learning in comparison communities – which is certainly possible – we would expect this to also be reflected in dramatic learning gains by *all girls* in comparison communities, not only girls who are selected as replacements. The data contradict this idea: improvements in comparison communities within the true panel do not outpace those in intervention communities. Because the gap between intervention and comparison groups manifests exclusively in the set of girls drawn as replacements, it suggests that the replacement process itself is responsible for the gap.

Comparison	43.4%	51.8%		8.4%
Numeracy – ML1 to ML2				
Intervention		54.5%	55.1%	0.6%
Comparison		54.3%	60.7%	6.4%
Somali Literacy- Baseline to ML1				
Intervention	40.4%	53.2%		12.8%
Comparison	41.1%	55.2%		14.1%
Somali Literacy – ML1 to ML2				
Intervention		61.2%	61.9%	0.7%
Comparison		58.5%	64.7%	6.2%

These results suggest we should be very cautious when using the "cross-sectional sample", which we define as the sample of girls who appear in both rounds of an analysis, including girls who were replaced, and their replacements.[40] The cross-sectional sample is likely biased due to the fact that replacement girls are pulling the mean learning score in comparison communities upward at a faster rate than in intervention communities.

Beyond the replacement procedure, we may also be concerned about differential attrition-without-replacement. Some girls who were not re-contacted successfully cannot be replaced, either because no eligible girls are available for replacement, or because the original girl was an OOS girl, who cannot be replaced.[41] In addition to the replacement procedure, a second potential source of bias comes from differential attrition and non-replacement, i.e. differences between intervention and comparison communities in terms of the type of girls who fall out of the sample entirely, without being replaced. In practice, this more extreme form of attrition is not a concern for our analysis for two reasons: first, the girls who fall out of the sample and are not replaced appear to be broadly similar between intervention and comparison communities, such that no bias should result.[42] Second, and more fundamentally, we do not include girls who fell out of the sample entirely and could not be replaced in either the panel or cross-sectional samples for analysis. That is, even in the repeated cross-sectional analysis – which uses the "cross-section sample" – we do not include girls who fell out of the sample and were not replaced. The distinction between the panel and cross-section samples focuses on whether they include girls who were successfully replaced, and their replacements; neither sample includes girls who could not be replaced. Therefore, this potential source of bias will not influence our inferences.

[40] We refer to this sample as a cross-sectional sample because it is equivalent to a repeated cross-sections design. The difference-in-differences approach can be implemented with either a panel dataset (the same units of analysis, in this case cohort girls, are observed repeatedly over time) or with a set of repeated cross-sectional samples (repeated random samples drawn from the same populations over time). Both data structures produce valid inferences using difference-in-differences. As noted elsewhere in this report, the panel structure has advantages due to a reduction in sampling variation across time periods or evaluation rounds.

[41] OOS girls were not replaced because there is not a clear and logistically-feasible mechanism for identifying and sampling OOS girls. A random sample of the community would enable location of OOS girls, but

[42] Between baseline and ML2, 202 comparison girls and 248 intervention girls fell entirely out of the sample. Their learning scores were similar across the two types of communities: in the subject area with the largest gap – English literacy – the difference between such girls in intervention and comparison communities was just 1.4 points.

In this section, we have considered the re-contact and replacement process from a variety of angles, probing myriad potential sources of bias. To summarize our findings and how they should be applied to the analysis and results presented in the body of the report, we consider the following findings to be of particular note:

- Attrition (or, inversely, successful re-contact) is not as-if random. Attrition is most common in Galmudug, is correlated with age, and may be correlated with unobserved characteristics of girls, or characteristics that we did not explicitly analyse.
- Replacement procedures were designed to ensure replacement by an equivalent girl in terms of grade level, but there are a number of systematic differences between replacement girls and the girls they are replacing, including in household asset ownership. The extent of difference between replacement and replaced girls is not entirely clear, but could be meaningful.
- Replacement girls outperformed the girls they replaced in terms of learning performance, in general, which can be attributed to the replacement girls' older average age and higher grade level at the time they were assessed. However, replacement girls in comparison communities outperformed the girls they are replacing by a much higher margin than occurred in intervention communities, meaning that replacement itself is driving comparison community learning scores higher, vis-à-vis intervention communities.

In combination, these findings suggest we should be very cautious employing replaced and replacement girls in our analysis. Doing so is likely to bias our findings regarding learning, and may produce bias in analysis of other project outcomes, including intermediate outcomes. This argument does not apply to subsamples analysed between specific rounds: for instance, if we study changes in learning scores from ML1 to ML2, including girls who were selected as replacements at ML1 is not inherently problematic, because many such girls were successfully re-contacted at ML2. In that case, they form part of the "true panel" from ML1 to ML2, in that the same girl appears in both rounds. But these same girls – alongside the baseline girls they replaced – should generally be excluded from baseline-to-ML1 comparisons, because they do not form part of the true panel over this time period. Throughout this report, we rely heavily on the "true panel" of girls or households, where the true panel is explicitly defined by the time period included in the analysis; we often also report results using the cross-sectional sample, but generally rely less on these findings, in line with the discussion here.

PREDICTORS OF SUCCESSFUL RE-CONTACT - ALP GIRLS

In our final analysis of re-contact and replacement outcomes, we report results for ALP girls. This cohort of girls was recruited during the first midline (ML1) evaluation round, and they were recruited exclusively from areas where SOMGEP-T is being implemented. No "comparison" ALP centres were selected. For these reasons, ALP girls are fundamentally different from the main cohort, recruited at baseline, and concerns regarding re-contact rates are somewhat different. In the context of ALP girls, maintaining a high re-contact rate is important, because assessment of the impact of ALP programming is based on a pre-post comparison of learning scores at ML1 to learning scores at the endline. If most ALP girls are replaced between ML1 and the endline, the validity of the pre-post comparison is adversely affected.

In the aggregate, 70.4 percent of ALP girls were successfully re-contacted at ML2. Re-contact rates are lower than in-school girls over the same time period, but this is not surprising, given that ALP girls are older, on average, than in-school cohort girls. The average age of ALP girls when first recruited was 17.0 years. As the bottom panel of the table below shows, re-contact rates among younger ALP girls are on-par with those of in-school girls at equivalent age levels. To illustrate: in-school girls aged 10-13 at the baseline had one-year

re-contact rates – from baseline to ML1 – of 85.2 percent, while ALP girls in the same age range recorded a re-contact rate of 87.0 percent.[43]

The other finding that emerges from the table below concerns geographic location. Unlike in-school girls, re-contact rates among ALP girls are actually highest in Galmudug, though this finding is based on a sample size of just 15 Galmudug-area ALP girls. On the other hand, the gap in re-contact rates between Puntland and Somaliland are also substantively large – ALP girls in Somaliland are significantly less likely to be re-contacted successfully.[44]

TABLE 17: RE-CONTACT RATES AMONG ALP GIRLS

	Re-Contact Rate	Sample Size
Overall	70.4%	365
Zone		
Somaliland	66.4%	253
Puntland	78.4%	97
Galmudug	86.7%	15
Age		
10-13	87.0%	23
14-15	63.5%	74
16-17	70.0%	100
18-19	71.3%	108
20-22	71.7%	60

[43] The latter result, for ALP girls, is derived from a one-year re-contact period, ML1 to ML2. While there may be differences between BL-to-ML1 and ML1-to-ML2 re-contact periods that militate against drawing firm conclusions from their comparison, this finding, at minimum, illustrates that re-contact rates for ALP girls are not fundamentally different from their primary school peers of equivalent ages.

[44] We limit our analysis of ALP re-contact status, relative to our analysis of the broader cohort, because of the smaller sample size available for analysis. It is also worth noting that many ALP girls were no longer enrolled in ALP centres when they were located during ML2 data collection. Of those re-contacted successfully (n = 336), 75.9 percent were enrolled in ALP, 2.1 percent were enrolled in formal school, and 0.9 percent were enrolled in ABE centres. The remainder – 21.1 percent – were not enrolled in any educational program when they were located at ML2.

3. EDUCATIONAL MARGINALISATION

This section discusses the characteristics and barriers related to education marginalization and their distribution among sample girls across the baseline and ML2 evaluations. SOMGEP-T defines educational marginalization characteristics to include poverty, family conditions, language obstacles, and illiteracy among caretakers. Student characteristics also include disabilities, including physical, cognitive, and mental health impairments. The figures on characteristics provide a snapshot of the entire sample; the calculations are exclusively based on ML2 data but capture data from all sample girls – including both intervention and comparison groups (both in school and out of school girls) as well as ABE and ALP girls. Change in characteristics between the baseline and ML2 is not assessed, as this is a panel study[45] and many of the educational marginalisation characteristics possess one or more of the following attributes which undermine the value of intertemporal analysis:

- **The characteristic is either static or mutable in only one direction.** For example, being an orphan is a characteristic associated with educational marginalization. However, once a girl is an orphan, this status cannot change. Given this, the number of orphan girls cannot 'improve' and therefore measuring its change overtime is not informative or
- **The characteristic is a dimension of the girl's demographic that, while being associated with greater educational marginalization, does not innately require 'improvement'.** Girls where the head of household is female, for example, are more likely to face educational marginalization. However, transitioning female headed households to male headed households is not necessarily desirable. As such, measuring to find a reduction in female headed households is both unlikely and unrelated to intervention activities.

Given the nature of these characteristics, changes over time are not particularly meaningful to programme evaluation (within existing circumstances, where no educational marginalization characteristics have degraded so significantly during the project cycle that strategic or operational changes to implementation should be considered). Rather, these are discussed to describe the general circumstances in which sampled girls are functioning from within.

Educational marginalization barriers include household dynamics that may handicap a girl's ability to access quality education, including the extent to which family members encourage and help facilitate school enrolment and attendance. Barriers also include school quality variables which may affect a girl's likelihood to access a quality education. Distinguishing boundaries between 'characteristics' and 'barriers' in this analysis are not clear-cut. However, barriers differ in that their change over time are indicative of progress in the broad ecosystem of girls' education – where the ecosystem encapsulates girls' social and household environments as well as in school educational experiences. Barriers are dynamic within these systems, and are likely affected by project activities as well as exogenous factors. As such, tables which present figures on educational marginalisation barriers look at change over time. These tables examine the differences between baseline and ML2 values as well as the differences-in-differences between intervention and comparison groups

[45] As a panel evaluation, data is collected with the same girls and households over time. This means that changes in demographics and other relatively static characteristics do not (and are not expected to) change throughout the project cycle.

between the baseline and ML2 evaluations.[46] The difference-in-differences (abbreviated DiD in the tables below) calculation does not intend to attribute change to the intervention, but rather serves to stage a backdrop with which analyses in subsequent sections can be understood. In short, the educational barriers faced by girls, and the extent to which they have ebbed and flowed throughout the project cycle, illustrate the ways in which girls interact with their home and school environments. Educational characteristics can be understood as the constraints under which girls navigate those barriers.

3.1 Sample Characteristics and Barriers

The table below presents the number and proportion of girls with characteristics associated with educational marginalization. A few characteristics appear to be applicable to many of the girls, with no substantial variation between groups. Most girls, for example, have a head of household with no formal education and half of the girls in the main treatment groups (intervention and comparison) have a female head of household. The difference in the prevalence of girls who are orphans, who are married, and who have ever been married is marginal between intervention and comparison groups, though each is slightly (<3%) more prevalent within the comparison group. The comparison cohort also has a notably higher proportion of mothers than the treatment groups.

Table 18: Characteristics of education marginalization

Girl Characteristics	Intervention		Comparison		ALP		ABE	
	N	%	N	%	N	%	N	%
Living without both parents	69	10.1 %	79	13.3 %	150	21.4 %	50	10.4 %
Orphan	2	0.3%	6	1.0%	1	0.1%	0	0.0%
HoH is female	355	51.9 %	318	53.5 %	317	45.3 %	208	43.3 %
HoH has no formal education	467	72.1 %	373	70.5 %	430	64.0 %	362	80.3 %
Primary caregiver has no formal education	545	79.8 %	480	80.8 %	504	72.0 %	415	86.5 %
Currently married	30	4.4%	39	6.6%	160	22.9 %	4	0.8%
Has ever been married	37	5.4%	46	7.7%	201	28.7 %	7	1.5%
Mother, under 16	1	0.2%	0	0.0%	2	1.3%	0	0.0%
Mother, over 16	18	8.4%	32	15.6 %	152	27.8 %	5	6.7%
HH Characteristics	Intervention		Comparison		ALP		ABE	
	N	%	N	%	N	%	N	%

[46] The figures in these table only include the panel of girls interviewed both at the baseline and midline evaluation. Thus, the comparison between rounds presents information regarding the changes in the distribution of characteristics and barriers for the same individuals and are not reflective of changes in sample composition.

HH is pastoralist	55	8.0%	49	8.2%	64	9.1%	27	5.6%
HoH has no wage-earning occupation	311	45.5 %	283	47.6 %	306	43.7 %	272	56.7 %
Moved in the past 12 months	19	2.8%	9	1.5%	6	0.9%	6	1.3%
Seasonal migration	31	4.5%	20	3.4%	24	3.4%	5	1.0%
Owns camels	73	10.7 %	125	21.0 %	107	15.3 %	52	10.8 %
Owns medium-sized livestock	437	64.0 %	420	70.7 %	471	67.3 %	286	59.7 %
Owns land	462	68.8 %	404	68.7 %	476	68.7 %	312	67.4 %
House is informal / temporary structure	23	4.3%	31	6.6%	44	6.5%	24	5.2%
Poor quality roof	139	24.6 %	170	34.3 %	233	33.8 %	205	43.0 %
Household Deprivation and Food Security	Intervention		Comparison		ALP		ABE	
	N	**%**	**N**	**%**	**N**	**%**	**N**	**%**
Goes to sleep hungry many/most days	79	14.0 %	57	11.4 %	27	8.1%	58	12.1 %
Goes to sleep hungry some days or more often	198	35.1 %	160	31.9 %	107	31.9 %	213	44.6 %
Gone without clean water many/most days	175	31.0 %	153	30.5 %	117	35.0 %	158	32.9 %
Gone without clean water some days or more often	350	62.0 %	320	63.8 %	215	64.4 %	344	71.7 %
Gone without medicines many/most days	239	42.8 %	215	43.6 %	144	43.2 %	198	41.5 %
Gone without medicines some days or more often	373	66.7 %	348	70.6 %	224	67.3 %	340	71.3 %
No protein source, last 24 hours	132	23.0 %	86	16.9 %	75	22.3 %	186	38.6 %
Reduced food expenditure, last 3 months	216	38.0 %	176	34.9 %	110	32.8 %	186	38.8 %
Went entire day without food at least once, last 30 days	146	25.7 %	80	15.9 %	65	19.4 %	109	22.7 %
Reduced number of meals at least once, last 30 days	185	32.5 %	150	29.8 %	104	31.0 %	168	35.0 %
Disabilities and Impairments	Intervention		Comparison		ALP		ABE	
	N	**%**	**N**	**%**	**N**	**%**	**N**	**%**
Girls with any disability	66	9.6%	48	8.0%	62	8.8%	31	6.4%

Girls with any disability except mental health	25	3.6%	18	3.0%	17	2.4%	11	2.3%
Vision impairment	5	0.7%	4	0.7%	3	0.4%	5	1.0%
Hearing impairment	3	0.4%	1	0.2%	1	0.1%	0	0.0%
Mobility impairment	3	0.4%	1	0.2%	1	0.1%	2	0.4%
Cognitive impairment	1	0.1%	1	0.2%	1	0.1%	0	0.0%
Self-care impairment	0	0.0%	0	0.0%	0	0.0%	1	0.2%
Communication impairment	2	0.3%	1	0.2%	1	0.1%	0	0.0%
Mental health impairment	48	7.0%	36	6.0%	47	6.7%	23	4.8%
Disability of arms/hands	13	1.9%	11	1.8%	11	1.6%	5	1.0%
Anxiety	45	6.5%	31	5.2%	37	5.3%	18	3.7%
Depression	32	4.6%	24	4.0%	30	4.3%	17	3.5%
	Intervention		Comparison		ALP		ABE	
School characteristics	**N**	**%**	**N**	**%**	**N**	**%**	**N**	**%**
Language of instruction not Somali	66	11.5 %	19	3.7%	0	0.0%	0	0.0%
School was affected by conflict	114	20.0 %	84	16.7 %	20	6.0%	37	7.7%

Household economic structure, defined by a number of economic and livelihood indicators, is also classified as a characteristic of educational marginalization by care. The table illustrates the portion (and number) of households within each cohort who depend on income sources such as pastoralism, seasonal migration, and productive assets (livestock or cows). As with the girls' characteristics discussed above, there is little variation between treatment and control groups for most household economic structure characteristics. Some exceptions to this trend include a higher tendency of comparison households to own camels or to own medium-sized livestock. Both of these would indicate a higher socio-economic status.

In contrast to general household characteristics, there are more consistent differences between intervention and comparison groups in terms of household food security, which is a better metric of short-term economic deprivation. For instance, 35.1 percent of intervention households have experienced hunger – going to sleep hungry – at least once in the past year, compared to 31.9 percent of comparison households. This small gap is replicated in other measures of food security and dietary diversity: girls in intervention households are less likely to have consumed protein in the previous 24 hours, more likely to have gone an entire day without food in the previous 30 days, and more likely to have reduced their food expenditure in the previous three months. All of these results are consistent with the program's targeting of more marginalized, more rural, and more pastoralist communities. However, given the evaluation employs a difference in difference methodology, such variations are of little consequence to analysis.

Few girls throughout the sample report any disability; less than 10% in each cohort (intervention, comparison, ALP, and ABE) have any disability at all. Most disabilities reported are mental health impairments. The questions to the caregivers used to evaluate whether a girl had a mental health impairment were "How often does GIRL seem very anxious, nervous or worried?" and "How often does GIRL seem very sad or depressed?" With CARE, it was decided that if the caregiver said that the girl experienced such feelings

daily, weekly, or monthly, then the girl was evaluated to have a mental health impairment. Given the evaluation operates within a low-information context with regards to mental health, interpretation of terms such as anxiety or depression may vary substantially between caretakers and do not necessarily correlate with accepted clinical definitions. Alternatively, other disabilities, including hearing, using hands or arms, and communication, were determined by asking girls "Do you have difficulty [using your hands or arms]?" If girls responded that they have no difficulty or some difficulty, they were not considered to be disabled or impaired. Only girls who reported a lot of difficulty or complete inability to complete the task were characterized with the disability. This threshold may exclude girls who, among a variety of other drivers, a) are compelled by cultural factors or social desirability bases to minimize the degree of their disability, or b) are only peripherally aware of their disability, which they have been living and coping with (potentially without notice of adults or caretakers) for a majority of their lives. In general, our expectation is that disability is under-reported within the sample and does not represent the true prevalence of disability within the cohort, for a variety of methodological and socio-cultural reasons.[47]

The table below examines changes in educational barriers over time. We report these trends for three distinct groups: ALP girls, who constitute a separate cohort of girls, the main cohort girls in intervention schools, and the main cohort girls in comparison schools. For ALP girls, we make comparisons from ML1 to ML2, as this cohort was originally recruited during the ML1 evaluation. For the intervention and comparison cohorts, the comparisons are from baseline to ML2. In both cases, the figures in the table only include the panel of girls interviewed in both rounds and enrolled in the same type of learning institution, to ensure comparability over time. In other words, the ALP sample includes only ALP girls who were contacted at both ML1 and ML2, who were enrolled in an ALP program in both rounds. The intervention and comparison girls comprise only those who were successfully re-contacted at baseline and ML2 and remained enrolled in both rounds, as the questions focus on girls' experiences at school. Girls who were not enrolled in school were not asked these questions.

In addition, the right-most column reports the difference-in-differences (DiD) estimate of change between intervention and comparison groups from the baseline and ML2 evaluations. The figures in this column represent the *relative* change between intervention and comparison girls (limited to enrolled girls, and excluding ALP girls) from baseline to ML2.

Each of the barriers listed in the table is phrased as a negative statement and as such, a decrease (or negative value) represents reduction in that barrier. For example, in the intervention group, 3% fewer caregivers believe that travel to school is unsafe for girls in the ML2 than did so in the baseline. During the same period, the percent of caregivers in the comparison group who believe travel to school is unsafe for girls dropped by 8.2%. This indicates that, over the same period, this barrier was reduced more successfully among non-intervention girls than with intervention girls – with a DiD of 5%. It is important to note, however, that intervention schools have been systematically more affected by conflict than comparison schools, due to the project's own targeting criteria (which includes disputed border areas). On the other hand, the proportion

[47] We discuss these points in more detail in Section 8.4, which deals specifically with the new sample of ABE girls recruited during this round of the evaluation. However, the same concerns apply to both samples of girls – indeed, any survey attempting to assess disability rates in Somalia. Anecdotal experience suggests that disability is highly stigmatized among Somalis, which may decrease reporting; moreover, self-reporting disability may be seen as admitting "weakness", something which is generally avoided in Somali culture. Combined with potential misunderstandings regarding the nuanced meaning of concepts like anxiety and depression, we believe the prevalence rate reported here is biased downward and does not represent the full extent of disability in the sample.

of girls who feel unsafe at the school fell more in the intervention group – potentially reflecting the contribution of project interventions.

TABLE 19: EDUCATIONAL MARGINALIZATION BARRIERS, SCHOOL-ORIENTED BARRIERS

School Barriers	ALP (ML1-ML2)	Intervention (BL-ML2)	Comparison (BL-ML2)	DiD (BL-ML2)
Caregiver believes travel to school is unsafe for girls	-5.5%	-3.0%	**-8.2%**	5%
Girl feels unsafe travelling to school	-3.7%	-3.5%	**-7.8%**	4%
Doesn't feel safe at school	**-1.8%**	**-6.0%**	-4.4%	2%
Difficult to move around school	+0.3%	-4.0%	**-6.1%**	2%
Doesn't use drinking water facilities	-1.0%	-9.5%	**-17.9%**	8%
Doesn't use toilet at school	+6.0%	+1.1%	**-9.2%**	10%
Doesn't use areas where children play/socialize	+4.5%	**-13.1%**	-10.1%	3%
Disagrees teachers make them feel welcome	+2.4%	-2.8%	**-3.8%**	1%
Agrees teachers treat boys and girls differently in the classroom	-0.9%	**-7.3%**	-5.0%	2%
Agrees teachers often absent from class	**+15.2%**	-9.8%	**-15.4%**	6%
Afraid of teacher	+5.5%	**+11.2%**	+17.5%	6%
Does not feel comfortable asking teacher questions	-2.0%	-2.4%	**-3.3%**	1%
Teacher punishes students who get things wrong in a lesson	-9.6%	-8.3%	**-10.9%**	3%
Teacher uses physical punishment*	+0.5%	-17.7%	**-21.2%**	4%
No computers at school	-1.2%	+1.3%	**-1.8%**	3%
Cannot use books or other learning materials at school	-2.6%	-2.8%	**-7.0%**	4%
Not enough seats for all students	**-11.2%**	-21.3%	**-24.7%**	3%
Caregiver says principal is of poor quality	N/A	-3.0%	**-4.4%**	1%
Caregiver says teaching is of poor quality	N/A	-1.5%	**-4.8%**	3%
Girls says teacher does not asks girls and boys an equal amount of questions	**-18.7%**	-2.1%	**-5.2%**	3%

Girl says teacher does not ask girls and boys questions of equal difficulty	-17.7%	**-2.0%**	-1.8%	0%

* The number of girls who report teachers' use of physical punishment compares ML1 and ML2, but not baseline. This is a result of changes to the survey sequencing between the baseline and ML1. The change, made to improve the clarity of the question and ultimately the validity of the data, maintained consistency between the ML1 and ML2 surveys, making it preferable to compare these two rounds.

Overall, the most substantial improvements in school barriers have been the availability of seats for all students, the use of areas where children play and socialize, and a reduction in teacher absenteeism. Both intervention and comparison student cohorts also report more widespread use of drinking water facilities and a decline in teachers' use of punishment when students get something incorrect in a lesson. For each of these, however, greater strides appear to have been made among comparison girls than intervention girls. In fact, many of the barriers appear to have larger reductions among comparison girls than intervention girls. Only seven of the twenty-one barriers in the table above indicate greater improvements were made among intervention girls than comparison girls in reducing the educational barrier. Among those where barriers were reduced to a greater degree among comparison girls than intervention girls, the DiD averages at just under 4%. The largest DiD within this group is in the use of drinking water facilities and the use of school toilets.

For most barriers listed above, both intervention and comparison groups demonstrate improvements in reducing educational barriers. However, positive values – present for four of the barriers – suggest the prevalence of the barrier has increased among surveyed girls. Most notably, both intervention girls and comparison girls report fear of their teacher at substantially higher rates in the midline than at the baseline. This variable, however, should be interpreted with caution, as other indicators of student-teacher relationships do not follow this trend. Both groups of girls report improvements in perceptions of school safety as well as the welcoming of their teachers. Likewise, fewer students in both intervention and comparison groups report a discomfort in asking the teacher questions and there is a sharp decline in the number of girls who say that their teacher punishes students who get things wrong in a lesson.

As the ALP and ABE cohorts differ so fundamentally from the intervention's main cohorts (intervention and comparison samples of in school and out of school girls), difference in differences are not calculated for household barriers. However, to provide an overview of how the different groups compare overall, the portion of girls – in all cohorts – with each barrier type at the time of the ML2 evaluation is reported. (These are displayed in rows denoted "ML2" for main cohorts and are the only values presented for ALP and ABE girls.) Figures are comparable among cohorts for several household barriers, including family support, autonomy in education-related decision making, distance to school, and caregiver literacy.

Similar to the above analysis of school-quality barriers, differences in household barriers between baseline and ML2 for both intervention and comparison groups are calculated (these are in the rows denoted with the "Δ" symbol. Likewise, a difference and difference between the two main cohorts is computed. Of the household barriers evaluated, prohibitively intense chore burdens (whole day spent on chores) for girls decreased the most between baseline and ML2, for both intervention and comparison groups. Intervention girls in the ML2 evaluation were more likely to report support from their family to stay in school than in the baseline, while the portion of comparison girls who feel the family does not support their staying in school increased over the same period. A greater portion of intervention girls feel that they have a choice in whether she will attend or stay in school than in the baseline, however comparison girls still outperform intervention girls, despite this progress. Both sample groups report higher rates of feelings that the family make schooling decisions for her, with the change more notable among intervention girls. Overall, DiD between intervention

and comparison groups between the baseline and endline for household barriers remain low, indicating slightly more progress in mitigating barriers among intervention girls.

TABLE 20: EDUCATIONAL MARGINALIZATION BARRIERS, HOUSEHOLD BARRIERS

HH Barriers		Main Cohorts: change (Δ) BL – ML2			Other girls: ML2 only	
		Intervention	Comparison	DiD	ALP	ABE
High chore burden	Δ	**-5.9%**	-5.1%			
	ML2	21.2%	20.6%	0.8%	26.6%	9.4%
Doesn't get support from family to stay in school	Δ	**-1.2%**	+3.8%			
	ML2	5.1%	6.0%	5.1%	4.7%	6.2%
Girl feels that she has no choice whether to attend or stay in school	Δ	**-3.2%**	-0.9%			
	ML2	88.8%	83.9%	2.3%	71.6%	81.7%
Girl feels family makes schooling decisions for her	Δ	+7.5%	**+1.3%**			
	ML2	24.5%	32.2%	6.2%	16.4%	33.2%
Distance to school is greater than 30 minutes	Δ	**+0.8%**	+2.2%			
	ML2	1.0%	2.2%	1.5%	1.4%	3.5%
PCD and family members are not involved in CEC	Δ	**+4.3%**	+9.7%			
	ML2	16.7%	28.2%	5.4%	20.3%	16.3%

Project response

The findings reflect the major challenges experienced by rural and remote populations in Somalia, particularly those dependent on pastoralism for a livelihood. In this evaluation round, we have included measures of dietary diversity for the first time in the SOMGEP-T study; the findings show that despite the pastoralist setting, 23% of the girls in formal education and 39% of the girls in ABE did not consume protein-rich foods, a surprising result in a context where milk is a stapleThe results shed additional light on the needs of students and highlight the vulnerability of ABE participants -particularly since the results clearly indicate that low levels of protein consumption are affecting learning outcomes. . The results also further illustrate the differences between the comparison and treatment group, as only 17% of the girls in comparison sites did not consume

protein. In this evaluation round, the project has also included specific questions on the effect of conflict on schools, which highlighted the higher proportion of intervention schools (20%) impacted by the clashes between clans/ states in 2019.

Positive findings include improvements in safety and use of WASH facilities at school, as well as more equitable teaching practices, in particular for formal school students, reflecting efforts from the project and CECs to address barriers to girls' attendance. The results show a mixed scenario, where some barriers have decreased while others (teacher absenteeism, for instance) have increased. The proportion of students with mental health issues has also decreased considerably in relation to previous evaluation rounds.

3.2 Intersection of Barriers and Key Characteristics

In our discussion of sample composition above, we distinguished between the characteristics of girls and the barriers to education that girls face. Loosely speaking, we argued that the "characteristics" highlighted are immutable or nearly-immutable, in the sense that they *generally* do not change over time. For instance, girls' disability status can change over time, but typically does not. Assessing change over time on such characteristics is unlikely to reflect changes in sample composition that are meaningful from a programmatic or evaluation standpoint.

On the other end of the spectrum are barriers to education that girls face, which emphatically do change over time – a fact reflected in the previous section. At the same time, we do not believe it serves an evaluation or program design purpose to analyse changes in barriers over time in the context of a discussion around sample composition, because these changes are unlikely to be driven by changes in sample composition. Rather, changes in the extent of barriers are more likely to reflect impact of the program in reducing barriers and broader societal changes in the communities studied.[48] It is for this reason that we do not specifically report the intersection of barriers and characteristics here, and refer readers back to the SOMGEP-T baseline report, where girls' personal, household, and community contexts were examined for the full sample of cohort girls, in an effort to understand the extent of their marginalization at the outset of the program.

3.3 Project Activities and Relevance

SOMGEP-T project activities are directed to achieve four main goals: (1) improving access to post-primary options, (2) fostering supportive school practices and conditions for marginalized girls, (3) promoting positive shifts on gender and social norms, and (4) enhancing the ability of MoEs to deliver quality education. While improved access to post primary options is specific to out of school girls, each of these goals speak to many of the kay barriers present for sample girls.

Specifically, SOMGEP-T's focus on teacher training directly addresses a number of the in-school barriers girls face (see the first table in the previous section). These include caregivers' negative perceptions of teachers, students' ability to approach teachers with questions, and girls' relationship with their teachers – i.e. if they are afraid of them. Intervention activities such as teacher training on improved delivery of literacy and English language, teacher training on improved delivery of numeracy, and teacher training to provide structured

[48] For example, later in this report, we document shifts in the share of caregivers who believe it is unsafe for girls to travel to local schools in their area. This share has fallen over time. This shift is not a function of a change in sample composition, because the shift occurs even when we limit the analysis to the panel sample of girls re-contacted through successive rounds (i.e. the same girls contacted in each round). Rather, shifts in perceived safety likely reflect either program impacts or changes – real or at least perceived to be real by caregivers – in the communities being studied.

remedial support, are particularly relevant to these barriers. Teachers' gender biases are also discussed as an in-school barrier in the previous section. These biases emerge as different treatment of boys and girls in the classroom and discrepancies in the frequency and nature of questions teachers ask girls versus boys. To meet this challenge, the SOMGEP-T intervention includes activities to strengthen gender departments' capacity to improve girls' education outcomes through trainings, development of action planning, and provision of incentives to retain gender focal points – especially in gender areas.

SOMGEP-T's activities are addressing a number of technical and infrastructural barriers as well. The intervention includes constructing additional classrooms in remote primary schools, building water facilities in new secondary schools, providing solar chargers for mobile devices and devices, and providing sanitary pads to schools. This is of relevance to student barriers, as computers and learning materials are chronically undersupplied in schools. The provision of WaSH infrastructure is specifically relevant to addressing girl students' inability to use toilets and water facilities at school. Also within this realm, SOMGEP-T is equipping boarding schools for girls with furniture and learning materials and promoting girls' enrolment.

School management – directly related to girls' perception of safety and security at school – is addressed through SOMGEP-T's work in strengthening CEC's and in providing support to Regional Education Officers (REOs) and District Education Officers (DEOs). Likewise, family support for girls' education, chore burdens, and general autonomy of girls and young women connects to the intervention's emphasis on engaging community-level stakeholders including religious leaders, women's groups, men, and boys.

While many of the characteristics of educational marginalization discussed above are not directly addressed through intervention activities, engagement with the community has the potential to, in the future, minimize the portion of child mothers and child brides. Similarly, providing adult literacy classes, non-formal education for mothers, and entrepreneurships skills for girls touches upon the strains often felt in female headed households and offers potential avenues to navigate to economic and livelihood challenges.

4. LEARNING

4.1 Overall Learning Outcomes

In this section we look at learning outcomes for the cohort girls. Throughout this section comparisons are being made using the difference-in-differences approach, by contrasting the changes in learning scores between girls in the intervention and comparison groups. The learning assessments that are being used for evaluating the learning outcomes of this project include an assessment on numeracy, Somali and English literacy and financial literacy.

The numeracy assessment measures skills such as: filling in missing numbers, single-digit addition, double digit addition, single digit subtraction, double-digit subtraction, as well as multiplication and division with single and double digits. The assessment also includes word problems which test the applied use of all of these skills. The Somali literacy assessment consist of six sub-tasks which measure reading fluency, reading comprehension with texts of different difficulties as well as writing skills. The English assessment, likewise, captures reading fluency and reading comprehension at different levels as well as writing skills. The assessment also measures the more fundamental skill of letter identification, and is composed of seven sub-tasks. The financial literacy assessment consists of 11 sub-tasks which measure the applied use of mathematics

in everyday problems through word problems. The skills measured in the test include subtraction and addition, division and multiplication. In addition, the assessment requires an understanding of the concept of profit and the difference between gross and net profit. All the assessments scores have been converted into a score out of 100 percent. This facilitates comparison between assessments.

In this section of the report we now present the aggregate changes in learning outcomes between Baseline and Midline Round 2. While exploring the aggregate learning results, different panels of respondents were used to investigate potential differences in learning outcomes. For the benefit of the reader, most of this analysis has been moved to Annex 18 of the report. Annex 18 provides the analysis of learning results using the cross-sectional sample which includes the substitutions that were made at Midline round 1 and Midline round 2.[49] A further section focusing on out-of-school girls as well as one looking specifically at those girls who have been enrolled since baseline can also be found in the annex.

In this section we present the analysis using what we call the 'true panel' – that is the group of individual respondents who have taken part in the evaluation since the baseline.

TRUE PANEL - BASELINE TO MIDLINE ROUND 2

The first results presented in terms of learning are for the true panel of those respondents who overlap between Baseline and Midline round 2. We call this the true panel as the overlap in respondents is full, and as such the program's impact should be able to be detected most clearly as the analysis is based upon comparing the exact same individuals over time.[50]

This panel includes 1,084 respondents, both in-school and out-of-school girls that have been re-contacted at Midline Round 2. This group does not include any substitutes made for missing respondents but is a direct

[49] The distinction between a "panel" and "cross-sectional" sample, please see the discussion in Section 2.3, especially the discussion around Table 9.

[50] This assumes that wash-out of the sample from both the intervention and the comparison schools is roughly similar. We have no reason to assume this not to be the case.

longitudinal analysis of the same individuals. There are 575 individuals in intervention schools and 509 in the comparison group.

FIGURE 5: AVERAGE LEARNING SCORES FROM BL TO ML2 – TRUE PANEL

As the results in the above graphs depict, the development of the learning scores in both the intervention group and among the comparison individuals are largely similar for both literacy assessments. For Somali literacy, the intervention girls' average score rose from 37.4 percent at Baseline to 56.7 percent at Midline 2, a difference of 19.3 percentage points. Meanwhile, however, the comparison group also improved their score by 18.2 percentage points. For Somali literacy, thus, the evidence for program impact is extremely limited, as the difference-in-differences estimate is not statistically significant. For English literacy the results are very similar, with the difference in differences of 1.0 and likewise this estimate lacking statistical significance.

TABLE 21: DIFFERENCE-IN-DIFFERENCES, LEARNING SCORES AMONG THE COHORT PANEL

	Intervention schools	Comparison schools	Difference in Differences

	BL	ML2	n	BL	ML2	n	DID	p
Numeracy	33.2	52.2	575	27.8	43.3	509	+3.5	0.175
Somali literacy	37.4	56.7	575	31.0	49.2	509	+1.2	0.652
English literacy	11.2	20.8	575	6.4	15.0	509	+1.0	0.727
Financial literacy	4.2	23.5	575	4.7	15.6	509	+8.4	0.007**

However, when we look at financial literacy, the intervention cohort fares significantly better than the comparison group. Program impact for financial literacy seems to be 8.4 percentage points. At baseline the two groups were almost identical, with intervention at 4.2 percent and comparison schools at 4.7 percent average score. This entails that at baseline on average the respondents would have been able to correctly answer fewer than one of the eleven questions in the financial literacy assessment. At Midline Round 2, the average score in intervention schools has increased to 23.5 percent, with an increase of 19.3 percentage points. Meanwhile, in comparison schools the average increased by 10.9 percentage points, meaning that the difference in differences is 8.4. The difference in difference in differences estimate is statistically significant, meaning that that it is unlikely that the results would have occurred due to random error.[51]

Learning results in numeracy are also overall better than in the comparison group, whereby intervention girls improved their scores between Baseline and Midline 2 by on average 3.5 points more than those in comparison areas. Yet this difference is not quite statistically significant.

In sum, however, the results conform to our theoretical expectations regarding program impact, showing positive effects, even if they are too small to be distinguishable from a null result. Indeed, the difference in differences estimate is positive for all of the different assessments, implying that girls in intervention schools have improved their learning outcomes more on average than those in comparison schools. Moreover, the results remain very similar when using weights.[52]

Importantly, when looking at those individuals who overlap between Baseline and Midline Round 2, the students in the intervention schools have systematically improved their financial literacy scores much more than those in comparison schools. It seems, thus, that SOMGEP-T is having an impact in the field of financial literacy.

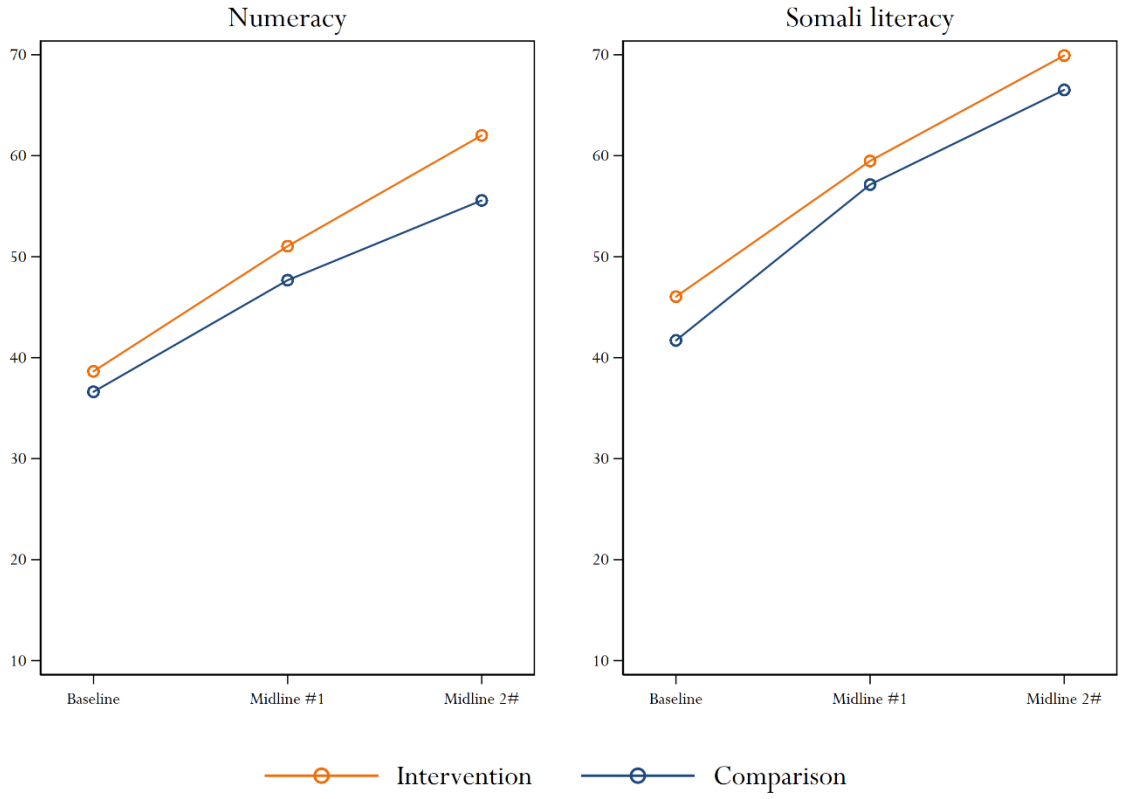
TRUE PANEL (ISG) - INCLUDING MIDLINE ROUND 1

Next, in order to see changes not just between Baseline and the current moment, we looked at learning outcomes over the three waves of evaluation. Assuming that the program's effect may take some time to show, it is important to discern the trend over the three moments evaluated. As Midline round 1 no financial

[51] Alpha level of 1 percent (p=0.007)
[52] Difference in differences estimate for numeracy, +3.3, p value=0.210
Difference in differences estimate for Somali literacy, +1.3, p value=0.613
Difference in differences estimate for English literacy, +0.5, p value=0.873
Difference in differences estimate for Finacial literacy, +8.5, p value=0.006

literacy or English literacy assessment were conducted, we are presenting only findings for numeracy and Somali literacy.

FIGURE 6: AVERAGE LEARNING SCORES ACROSS ALL 3 ROUNDS, IN-SCHOOL GIRLS ONLY



The panel used, is again is the pure overlap with the same panel of respondents over time. It should be noted, that the sample excludes all out-of-school girls since they were not re-contacted at Midline Round 1. Essentially, thus, this analysis is of in-school girls in the three waves of data collection.[53] The sample size for this panel is 529, which is composed of 286 girls in intervention schools and 243 individuals in comparison schools. This panel, as it is only composed of in-school-girls, will help further nuance the analysis conducted with the previous panel of girls enrolled since baseline as it has respondents who have been most exposed to the teaching activities.

For both numeracy and Somali literacy, we can see that the lines start to separate at Midline round 1. The gains that have been made are coming almost exclusively from the period following the first round of Midline. For example, if we look at numeracy, the overall difference between intervention and comparison groups is

[53] The sample also excludes some in-school girls who were included in our earlier analysis of BL-to-ML2 learning changes, because not all girls were successfully contacted during ML1. The "full panel" sample analysed here includes only those girls who were successfully contacted in all three rounds. Girls in schools that were temporarily excluded during ML1, and girls who could not be located at ML1 but were contacted at ML2, are included in the BL-to-ML2 panel sample, but are excluded from the "full panel" sample.

4.4 percentage points in favour of intervention groups. The difference has grown by 3.1 percentage points since the first Midline. A cautious interpretation could suggest that the effect of the program is starting to show over time, and we can expect the gap to become wider as time goes on.

Similarly, Somali literacy in fact developed more positively for the comparison group between baseline and the first round of Midline. Yet, at Midline round 1, this development has almost been reversed since the intervention group performed better than the comparison group between the two Midlines.

TABLE 22: DIFFERENCE-IN-DIFFERENCES IN LEARNING SCORES ACROSS ALL 3 ROUNDS

	Intervention schools				Comparison schools				Difference in Differences		
	BL	ML1	ML2	n	BL	ML1	ML2	n	BL - ML2	ML1 - ML2	p
Numeracy	38.6	51.0	62.0	286	36.6	47.7	55.6	243	4.4	3.1	0.134
Somali literacy	46.0	59.5	69.9	286	41.7	57.1	66.5	243	-0.9	1.1	0.772

The fact that the learning results have started to improve more markedly for the intervention schools since Midline Round 1 bodes well for the future. While the results are not statistically significant, the trend is clearly a positive one. Should it continue, the program should be able to show impact beyond financial literacy going into endline.

RESULTS OF OTHER ANALYSES

The analysis using different panels, in Annex 18 of this report, is fundamentally in line with the analysis conducted using the true panel. When looking at the cross-sectional panel which includes the substitutions for those cohort girls who could not be re-contacted at midlines, the results have improved slightly more for intervention girls in each assessment, but only for financial literacy this difference is marked enough to be significant at the 5 percent level.

Results are similar when focusing on those girls who have been enrolled since baseline. When it comes to literacy outcomes, the program does not appear to show impact. In fact, for Somali literacy, the scores in the intervention group have risen on average slightly less for those who have remained enrolled through the program. Meanwhile, as with all panels that were looked at, the most positive findings can be observed when looking at the development of financial literacy scores. The increases among the intervention cohort are on average 8.1 percentage points higher. The difference in differences estimate is marginally significant ($p=0.067$).

CONCLUSIONS FROM AGGREGATE RESULTS

Overall, three observations can be made based on the analysis above.

Firstly, the findings deriving from the analysis of different panels of data seems to suggest that the program has had an impact on financial literacy. When using the pure longitudinal panel of all the individuals who overlap between Baseline and Midline Round 2, the impact is a substantive 8.4 percentage in favour of

intervention schools. The impact is nearly of the same magnitude when looking only at the girls who have been enrolled since the baseline, the group most exposed to the teaching done through the program.[54]

Secondly, while aggregated numeracy outcomes are yet to be statistically significant in terms of the difference in improvement vis-à-vis the comparison group, increases in numeracy outcomes are systematically higher than in the comparison group. Assuming that the impact of the program in this regard may take some time, it is particularly encouraging that the outcomes have markedly improved since Midline Round 1. When looking at the true panel of the respondents who are the same across the three waves of data collection, the gap in learning between intervention and comparisons schools has almost entirely occurred between the two Midlines. This bodes well for the Endline because it seems that the program's effect is starting to show. Moreover, the panel consisting of girls who have been enrolled since baseline, and as such most engaged with the teaching, shows that the intervention cohort has improved their results on average by 4.6 points more than the comparison group since Baseline.[55] We can, thus, cautiously expect that this gap between intervention and comparison schools to grow as we go into endline and more girls have been exposed to the program for a longer time.

Thirdly, while the program has shown impact in terms of financial literacy, and potentially will do the same for numeracy moving forward, it has struggled to gain results in literacy outcomes. It is possible that the program's inclusion of financial literacy – where comparison schools have no such similar inclusion of the topic in their curriculum – is driving greater gains in financial literacy, compared to other learning outcomes. That is, the gap between intervention and comparison schools – in terms of the time and energy spent teaching a topic – is arguably greatest in financial literacy, and less in numeracy, English, and Somali. In the area of literacy, the results have thus far not been visibly different from comparisons schools and in fact in some cases negative.

LEARNING TARGETS

In terms of the program's targets in learning, the following tables detail progress thus far. Note that, for English literacy and financial literacy, comparisons are made from baseline to ML2, using the panel of in-school girls who appear in both rounds. For Somali literacy and numeracy, the subsample analysed includes all girls who appeared in all three rounds of the evaluation (i.e. the "full panel").

TABLE 23: PROGRESS TOWARD LEARNING TARGETS (ENGLISH LITERACY AND FINANCIAL LITERACY)

Objective	Indicator	Baseline	Target for ML2	ML2 outcome	Difference in differences	Target met
English literacy improvement	Percentage increase in SEGRA scores	13.3% intervention group SD: 17.3 7.9%	0.25 SD increase (4.993 points) relative to	25.9% intervention group 21.4%	-0.90	No

[54] Despite a much smaller sample size, the difference of differences estimate of 8.1 is marginally statistically significant.
[55] This finding is one of the closest to statistically significant, with a p- value of 0.13 .

		comparison group	comparison group improvement	comparison group		
Financial literacy improvement	Percentage of change in financial intelligence scores	4.7% intervention group SD: 14.0 6.0% comparison group	0.5 SD increase (7.0 points) relative to comparison group improvement	27.0% intervention group 20.3% comparison group	8.0	Yes

For financial literacy and English numeracy there are no results for ML1 as they were not measured at the time. The change observed since Baseline for English is actually more substantial for the comparison group, with the difference of difference estimate being negative 0.9. As such the learning target for English literacy – an 8.7% increase over the comparison group – has not been met. Meanwhile, the target for financial literacy has been met as the intervention group has improved their average score by 8 percent more than the comparison group.

TABLE 24: PROGRESS TOWARD LEARNING TARGETS (SOMALI LITERACY AND NUMERACY)

Objective	Indicator	Midline 1	Target for ML2	ML2 outcome	Difference in differences	Target met
Somali literacy improvement	Percentage increase in SEGRA scores	59.2% intervention group SD: 29.7 58.0% comparison group	0.25 SD increase (6.49 points) relative to comparison group improvement	69.3% intervention group 68.2% comparison group	-0.1	No
Numeracy improvement	Percentage increase in SEGMA scores	51.3% intervention group SD: 26.5 49.6% comparison group	0.25 SD increase (6.2 points) relative to comparison group improvement	61.1% intervention group 57.1% comparison group	2.3	No

For Somali literacy and numeracy, we have data for Midline Round 1, and as such progress is reported for the period between ML1 and ML2. As the above table illustrates, neither learning target has been achieved, however. For Somali literacy no difference can be observed between the groups while for numeracy the 2.3 percentage point difference of difference is not enough to meet the target of improving 6.4 percent over the comparison group.

While the program has not been able to achieve all of its targets in learning, it does seem to be based on valid assumptions about the ToC. This apparent paradox is further explored in section 4.9.

Progress toward the program's learning targets were also assessed in a regression framework, in line with most of the analysis above, which employed difference-in-differences regression models. The results for English literacy, Somali literacy, and numeracy are presented in three tables below. As with the remainder of our learning analysis, results pertaining to English literacy are based on a panel of in-school girls tracked from baseline to ML2, excluding ML1; results pertaining to Somali literacy and numeracy are based on a panel of in-school girls tracked across all three rounds from baseline to ML2.

The direction and magnitude of the results are consistent with those reported elsewhere in this section. Specifically, we observe the strongest positive effects of the program in the context of numeracy results. We find that girls in intervention schools improved by 3.1 percentage points *more* than girls in comparison schools from ML1 to ML2. When we expand that period to look at baseline-to-ML2 comparisons, the size of the effect goes up to 4.4 points – again, consistent with earlier findings. Neither effect is statistically significant after taking into account the clustered nature of the data, however, nor did the program meet its targeted level of improvement.

Somali literacy showed a smaller positive improvement, and scores in English literacy actually declined in intervention schools, relative to comparison schools. Somali literacy scores in intervention schools increased by 1.06 points from ML1 to ML2 – vis-à-vis comparison schools – and showed even less improvement from baseline to ML2, where the net change relative to comparison schools was negative. Targets were met for either Somali or English literacy.

Note that the regression models utilized here take into account the clustered nature of the data, with clustering at the school level, which reduces the power of statistical tests and increases the size of the standard errors in the regression models. No other adjustments are made in these models, except by restricting the sample size in line with earlier discussions in this section. In Annex 18, we report additional regression models to check the robustness of our results. The nature of difference-in-differences models means that control variables which are fixed over time do not substantially impact the regression coefficients of the models but *do* affect the standard errors and can lead to greater precision. More importantly, we also test the inclusion of grade-specific weights in Annex 18.

TABLE 25: ACHIEVEMENT OF SOMALI LITERACY TARGETS USING CORE REGRESSION MODEL

Result	Details	Comments
Somali Literacy Baseline – Midline # 1 - Midline #2 (ML1 to ML2 coefficients reported)	Beta = 1.06 Two-tailed p-value = 0.72 Target = 7.3-point increase above comparison group Performance against target = 14.5%	Results based on a model including clustered standard errors at school level, without additional controls, and without weights to adjust for differential sample sizes across grade levels. Sample includes in-

		school girls tracked from baseline, through ML1, to ML2.

TABLE 26: ACHIEVEMENT OF NUMERACY TARGETS USING CORE REGRESSION MODEL

Result	Details	Comments
Numeracy Baseline – Midline #1 - Midline #2 (ML1 to ML2 coefficients reported)	Beta = 3.1 Two-tailed p-value = 0.29 Target = 6.4-point increase above comparison group Performance against target = 48.4%	Results based on a model including clustered standard errors at school level, without additional controls, and without weights to adjust for differential sample sizes across grade levels. Sample includes in-school girls tracked from baseline to ML2, excluding ML1.

TABLE 27: ACHIEVEMENT OF ENGLISH LITERACY TARGETS USING CORE REGRESSION MODEL

Result	Details	Comments
English Literacy Baseline - Midline #2	Beta = -0.72 Two-tailed p-value = 0.85 Target = 8.7-point increase above comparison group Performance against target = -8.3%	Results based on a model including clustered standard errors at school level, without additional controls, and without weights to adjust for differential sample sizes across grade levels. Sample includes in-school girls tracked from baseline to ML2, excluding ML1.

4.2 SUBGROUP OUTCOMES - COHORT GIRLS

In this section we look at learning outcomes for various key groups among the respondents. We will begin by briefly looking at the scores for the sub-groups to see which ones are associated with lower or higher learning scores at Midline Round 2. This, however, will be somewhat brief as the barriers to learning have largely been identified at Baseline and Midline Round 1. Much of the section, thus, will focus on the program's impact among these different sub-groups by looking at the intervention schools in contrast to the comparisons in terms of the changes in learning outcomes. The objective is thus to see if the program is particularly successful among specific groups, or if the converse is true – and some groups are not seeing improvements in their learning levels. In this section we focus on sub-groups that have at least 50 girls in both intervention and comparison groups in order to avoid over-interpreting findings that may be disproportionately influenced by extreme cases.[56]

The tables below present the Midline Round 2 learning scores for each of the sub-groups.

[56] The analysis in this section has been conducted using the true panel of 1,084 individuals who overlap between Baseline and Midline Round 2. This consists of both in-school and out-of-school girls. We do not use the "full panel" that includes ML1 data, because doing so reduces the sample for analysis dramatically (from 1,084 girls to just 529). Given that the subgroup analysis involves parsing the sample into ever-smaller groups, it is important to start with the largest sample size available.

TABLE 28: LEARNING OUTCOMES BY SUB-GROUP

	Numeracy	Somali literacy	English literacy	Financial literacy	n
Overall	48.0	53.2	18.1	19.8	1,084
Geography					
Somaliland	43.5	49.7	12.4	16.2	698
Galmudug	52.2	43.5	36.8	22.5	53
Puntland	56.8	62.0	27.0	26.9	333
HH demographics					
Female-headed household	48.4	52.8	19.4	20.7	476
Lives without parents	45.8	49.2	17.3	15.7	102
Part orphan	49.9	56.5	18.7	23.2	128
HoH has no education	45.8	49.9	17.3	18.8	706
HoH has no formal education	45.8	49.9	17.3	18.8	706
CG has no education	46.0	50.7	17.5	18.8	797
CG has no formal education	47.3	52.2	18.2	19.3	880

As it pertains to geographic variation, Puntland students score the highest in everything apart from English where Galmudug has a higher average. This is not entirely surprising given that Puntland has a longer history in governance than Galmudug and has an advantage over Somaliland which continues to manifest the impact of the neglect that the region suffered during Siyad Barre's regime and the continued low levels of resources for education. Indeed, Somaliland scores lowest in everything apart from Somali literacy where Galmudug has the lowest score. However, it should be noted that the small sample size in Galmudug makes it vulnerable to skews from extreme cases.

Meanwhile, household demographic variables are not associated with any dramatic differences in leaning levels. Parents' low education level is slightly negatively associated with learning assessment scores as is not living with one's parents.

TABLE 29: LEARNING OUTCOMES BY SUB-GROUP

	Numeracy	Somali literacy	English literacy	Financial literacy	n
Overall	48.0	53.2	18.1	19.8	1,084
HH Assets and economics					
HoH pastoralists	38.4	39.5	9.9	15.1	120
HoH no occupation	48.9	53.4	18.9	19.6	431
Seasonal migration	35.3	39.1	9.5	17.5	60
Owns camels	37.9	38.6	13.3	16.0	111
Owns medium livestock	45.5	50.6	15.0	18.4	667
Owns a mobile phone	48.3	53.4	18.3	20.0	1032
Has regular access to water	52.2	57.3	20.3	21.3	554
Owns land alone	49.2	54.6	17.9	20.7	766
Owns land	48.8	54.1	17.9	20.6	862
HH poverty					
Poor quality roof	40.6	44.4	11.7	15.5	356
Gone without enough clean water most days	44.1	49.9	15.6	18.4	328
Gone without medicines of medical intervention most days	45.4	49.9	14.8	18.8	457
Gone without cash income most days	43.9	48.4	16.3	19.6	206
School quality					
Caregiver believes travel to school is unsafe for girls	42.6	42.5	8.1	17.8	77
Difficult to move around school	60.1	69.2	25.5	27.5	122
Doesn't use drinking water facilities	57.7	64.0	21.4	21.2	165
Doesn't use toilet at school	60.7	70.3	25.2	25.3	176
Doesn't use areas where children play/socialise	56.2	64.3	22.8	21.1	275

Teachers treat boys and girls differently in the classroom	58.3	66.0	22.7	21.6	254
Agrees teachers often absent from class	54.3	62.5	22.0	18.3	225
Afraid of teacher	55.3	65.6	20.6	21.9	358
Teacher punishes wrong answers	58.1	66.3	22.8	23.0	489
Teacher uses physical punishment	58.7	69.0	22.2	27.1	173
No computers at school	58.2	67.2	23.3	23.0	582
Cannot use learning materials at school	58.2	68.7	25.3	23.6	153
Not enough seats for all students	61.6	72.0	30.2	24.0	162
Girl says teacher asks girls/boys more questions (not equal)	59.3	70.1	25.0	19.1	64
Girl says teacher asks girls/boys harder questions (not equal)	54.5	66.5	16.3	19.0	73
Community attitudes					
High chore burden (whole day spent on chores)	36.9	36.9	11.2	13.6	253
Girl attends school most days	57.7	66.5	23.6	24.3	672
PCG and family member not involved in CEC	55.8	62.7	20.5	19.4	83
Girl feels no choice whether to attend or stay in school	57.7	66.8	24.2	23.3	539
Girl feels family makes schooling decisions for her	42.8	47.0	13.7	15.2	295

When looking at household assets and livelihoods, the variation in learning levels is intuitive. As has been identified in education programs in Somalia, pastoralism is negative predictor of learning outcomes. Pastoralist children score substantially lower for in each learning assessment compared to others in the cohort. Similarly, other variables associated with pastoralist livelihoods such as owning livestock and migrating seasonally also correlate with lower scores on each learning assessment. This is a well-understood barrier to education as pastoralist children are subject to frequent migrations, leading to high levels of attrition and disjointed educational experiences. The MoEs and CARE are trying to encourage further measures to better incorporate pastoralist children in education. However, this proves to be a challenge that has not yet been matched, as will be discussed in the sustainability section of the report.

Meanwhile, on indicators of household affluency, having regular access to water is positively correlated with learning levels. This may be because cities and other areas where water access is regular also have more established educational systems.

When looking at indicators of household poverty, they are overall negatively associated with learning outcomes, as predicted. Poor quality roof seems to be the most fit measurement for household poverty. It is associated with numeracy scores 7.5 percentage points lower and Somali literacy levels 8.7 percentage points lower than the average in the sample. Similarly, English is 6.4 percentage points lower and financial literacy 4.3 points lower than average. The measures of household poverty correlating negatively with learning outcomes is understandable. The theory there is that poorer households are more likely to face a lack of quality education in their area by being in the most under-resourced areas. Children of families that are struggling economically are also more likely to not be able to attend consistently or having to drop out entirely in order to support their family in making ends meet.

Curiously, various measures of challenging school environment – a girl saying it is difficult to move around school, saying that she does not like using drinking facilities or toilets or playing areas at school, or the girl saying that teacher treats boys and girls differently – are associated with higher learning assessments scores on average. Similarly, various other measures to indicate poor teaching quality (physical punishment, absenteeism, girl being afraid of teacher and the girl saying that teacher punishes wrong answers) are also in fact positively correlated with learning assessment performance. It thus seems that these measures are less apt in capturing the challenges of the school environment. They may in fact act as a kind of proxy measure for a girl's ability to think critically.

TABLE 30: LEARNING OUTCOMES BY SUB-GROUP

	Numeracy	Somali literacy	English literacy	Financial literacy	n
Overall	48.0	53.2	18.1	19.8	1,084
School characteristics					
School has reliable electricity	55.9	64.7	35.3	29.7	144
School has electricity at all (ever)	55.4	62.3	28.6	27.2	283
School feeds students	50.6	56.9	19.9	21.4	644
Language of instruction not Somali	62.9	68.6	35.1	39.1	85
School has a roof	48.1	53.2	18.3	19.9	1069
School has all cement floors	46.3	51.2	15.5	19.7	766
School has access to water	50.2	55.7	19.5	21.5	819
School has access to water within 1 km	47.6	54.6	15.7	21.0	563
School provides sanitary towels	51.4	56.4	15.1	23.8	146
At least one female FT/PT teacher	52.1	58.3	21.3	22.4	560
At least one female FT teacher	52.5	58.6	21.4	23.1	532
Short instructional time (less than 5 h)	42.0	47.7	12.0	16.7	312
Few contact hours (less than 30 h/week)	44.9	51.3	15.1	18.4	433

Class size greater than 30	49.7	56.8	20.4	21.1	370
Teacher absenteeism	44.0	49.2	14.7	16.4	266
School has toilets	47.9	53.0	17.0	19.9	910
School has toilets for girls	48.3	53.6	18.2	20.2	1000
Toilets that have privacy walls for girls	47.9	53.1	16.9	19.7	941
School has an active CEC	58.5	66.8	24.5	24.7	481
Girl can bring learning materials home	45.7	48.3	12.6	16.8	407
Disability					
Disability: mental health	46.9	49.0	21.6	19.6	52
Disability: any	45.0	48.1	20.7	18.2	66

Indeed, when we look instead at measures of a school's resources and characteristics using the data coming from the head teacher survey the learning levels corresponding with the indicators, the story that emerges is a more intuitive one. Indicators that are supposed to reflect the school's level of resources are positively correlated with learning assessment scores of the students. For example, the school having electricity, frequent or intermittent, is a positive predictor of learning outcomes of its students. School feeding and the school's access to water are positively associated with learning assessment scores. Similarly, if a school has female teachers, its girls are also on average doing better than those in schools where there are no female teachers.

Simultaneously the negative indicators of short instruction time as well as high levels of teacher absenteeism are negatively correlated with learning outcomes, as corresponds.

One finding, however, is slightly counter-intuitive. Namely, big class size is in fact associated with higher learning scores. Yet, when considering this, it may very well be that bigger class sizes are found in cities, which is where some of the other factors may positively influence learning outcomes. In fact, many of the indicators of school resources undoubtedly correlate closely with urbanicity of the schools, such as access to electricity, water, prevalence of female teachers, and so on.

Meanwhile, an active CEC is associated with substantially higher learning scores. It is associated with scores more than 10 percentage points higher on average for numeracy, 13.6 points higher for Somali literacy, 6.4 for English literacy and 4.9 for financial literacy. This is intuitive as CECs monitor the schools, including teaching quality and teacher absenteeism. Qualitative interviews have revealed that this puts a pressure on the school to perform better. The CECs also pay teacher salaries, addressing thus a potential shortage in teaching staff and motivating the teachers to do their best. They also substitute tuition fees for some girls and may seek out those who have not attended. All of this can positively contribute to the learning experience. Yet, it is good to see that active CECs in fact correlate with higher learning scores as the program as well as the MoE do much of their activities through the body.

In terms of disability, the only indicators that have a substantive enough sample size are mental health disability and having any disability at all. Both are somewhat negatively associated with assessments scores, apart from

financial literacy, which is where these groups in fact do better than others. However, the sample sizes are very small (52 and 66, respectively), and it would not be advisable to read too much into this.

The report will now turn to the program's impact among different sub-groups of girls in the sample to determine whether the program has had more or less of an influence among different demographic groups.

4.3 Sub-group Learning Outcomes – Program Impact

Having looked at the learning levels for each of the different sub-groups that are of relevance to this intervention, we now turn our attention to the program's impact on learning among the girls within these sub-groups. We compare the difference in the learning results at Baseline to those at Midline Round 2. We compare the changes between those in the intervention cohort to those from comparison areas. The program impact in each category, then, is the difference in the changes in the learning outcomes between the two waves between intervention and comparison groups.

It is important to do this analysis in order to assess whether there are any specific groups among the cohort who are contributing disproportionately to the improvement (or lack thereof) in the learning outcomes. Similarly, it is important to assess whether any specific groups significantly lag behind or indicate any signs of being left out of any progress observed. This will inform the program of any need for specific targeting in the future.

Geography

First, we look at learning outcomes by geographic region. The table below, as all tables in this section, presents the difference in differences of the learning score development between intervention schools and comparison schools for the specific group in question. For example, we can see that in Galmudug, the girls in intervention schools have on average improved their numeracy learning assessment scores by an impressive 24.5 points more than their comparison school counterparts. The same applies to Somali literacy scores in Galmudug where the difference of differences (DID) estimate is nearly as high, at 19.2. Both of these estimates are statistically significant at the 5 percent level, as is indicated by the asterisk following the p value of the DID estimate.

It should be noted, of course, that the sample size in Galmudug is rather small, with only 14 individuals currently in the comparison group. Yet, the program impact for numeracy and Somali literacy is significant and substantial. Moreover, the panel used for this analysis is that which has full overlap in the individuals in the sample. As such the results being affected by variation in the kind of individuals in the sample can be ruled out.

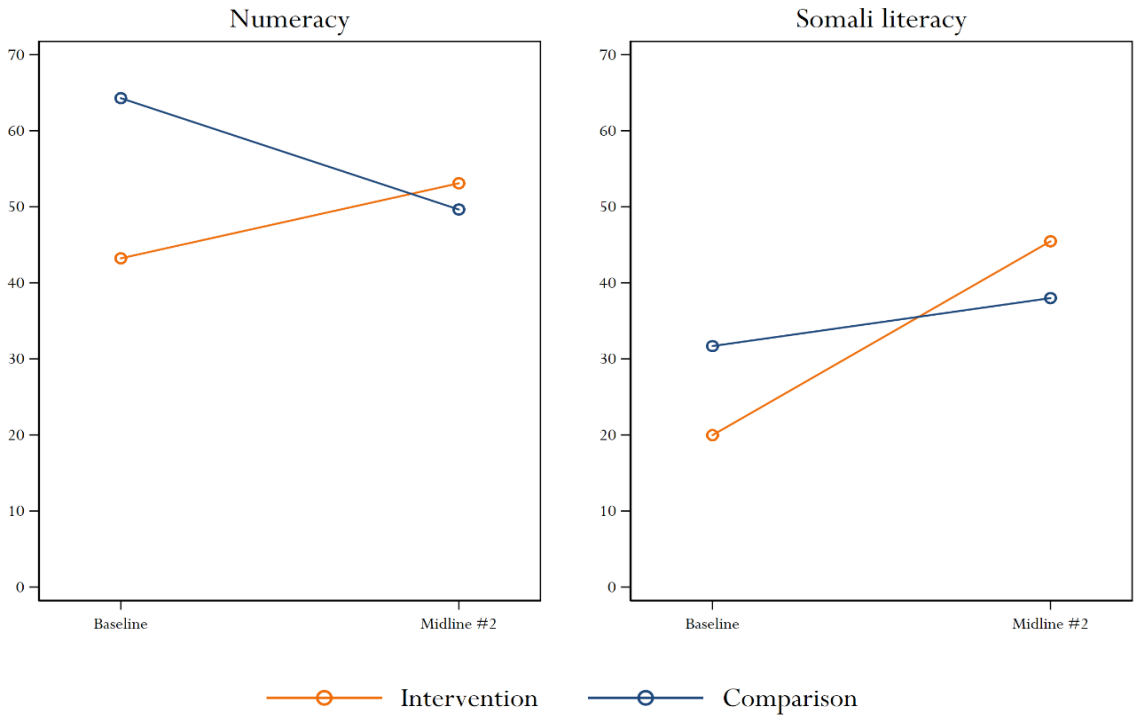
Table 31: Program Impact by Geography

	Numeracy		Somali literacy		English literacy		Fin. literacy		Sample size	
	DID	p	DID	p	DID	p	DID	p	Intervention	Comparison
Somaliland	2.6	.383	0.3	.905	0.5	.844	5.9	.116	690	706
Galmudug	24.5	.015*	19.2	.035*	1.7	.377	9.1	.341	78	28
Puntland	4.7	.372	0.9	.870	-2.7	.677	12.7	.040*	382	284

Further, in Puntland, financial literacy improved significantly more in intervention schools than in comparison schools. Improvement from Baseline to Midline Round 2 12.7 percentage points more, on average, than among those in comparison schools.

Overall, in terms of geographic variation, the findings are positive. Only English literacy assessments in Puntland had improved more among the comparison group, everything else was in favour of the intervention schools. As has been seen elsewhere in this report, the financial literacy trends are the strongest, and can be observed in all regions. The rather high DID estimates in Galmudug should be taken with a grain of salt as the sample size for the region is rather small (78 in intervention and 28 in comparison schools).

FIGURE 7: NUMERACY AND SOMALI LITERACY SCORES IN GALMUDUG



GIRL CHARACTERISTICS

In terms of the personal characteristics of the girl – whether she is married, whether she has ever been married or if she is a mother – the sample is too small for a rigorous investigation of the program's impact in this regard. There are only 28 girls in the intervention schools who are married. Only four in intervention and six in comparison schools are mothers. The limited number does not lend itself to rigorous statistical analysis.

HOUSEHOLD POVERTY INDICATORS

TABLE 32: PROGRAM IMPACT BY HH POVERTY INDICATORS

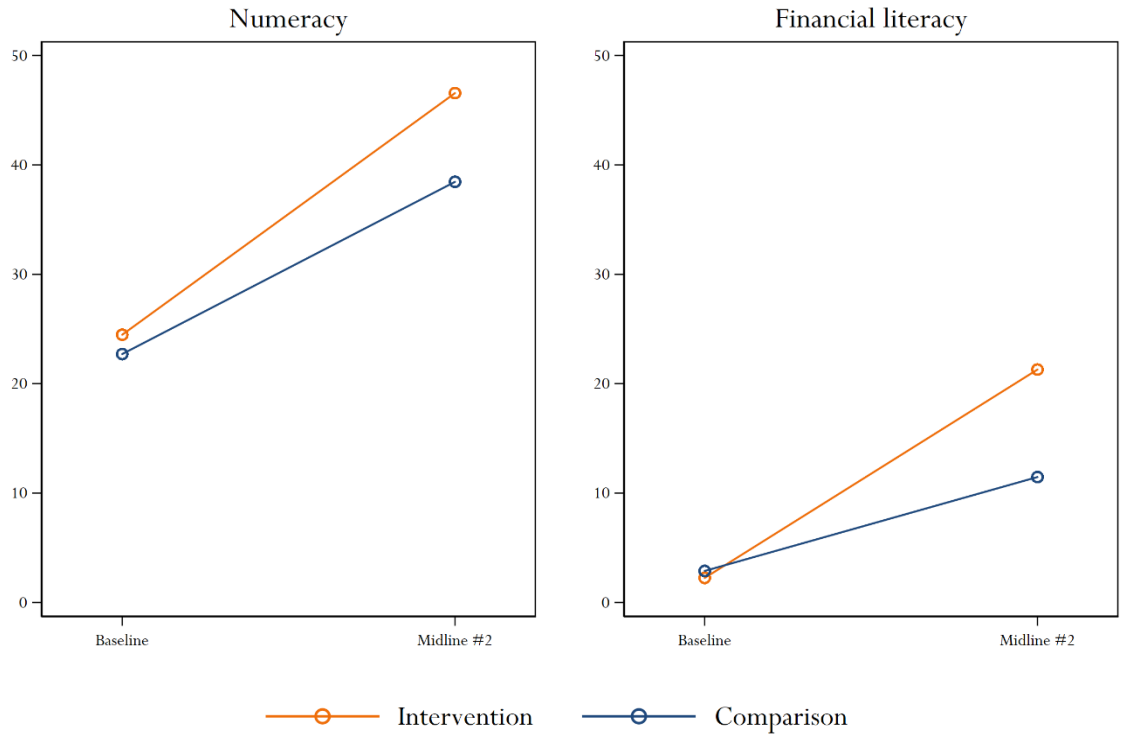
	Numeracy	Somali literacy	English literacy	Fin. literacy	Sample size

	DID	P	DID	P	DID	P	DID	P	Intervention	Comparison
Poor quality roof	6.8	.045*	-0.8	.846	-4.3	.141	6.3	.033*	340	372
Gone without enough clean water most days	-1.6	.639	-2.2	.513	-2.7	.432	5.2	.308	336	320
Gone without medicines of medical intervention most days	3.0	.393	3.5	.256	-0.9	.764	8.1	.052	480	434
Gone without cash income most days	-2.1	.506	1.0	.808	-0.4	.926	5.5	.344	204	208

As has been discussed earlier in this section, a household having a poor roof seems to be a good measure of poverty, at least insofar as it is reflected in overall lower levels in learning assessment scores. As it pertains to the program's impact within those whose families seem to be in an economically more disadvantaged position – measured through having a roof of poor material – the program seems to have impact in numeracy and financial literacy.[57] Girls in this sub-group who go to intervention schools have improved their numeracy scores by 6.8 points more, on average, than girls in comparison schools. For financial literacy, likewise, the program impact is 6.3 percentage points.

[57] This is the only poverty indicator that yields a statistically significant DID estimate.

FIGURE 8: PROGRAM IMPACT FOR GIRLS OF POOR HHS



Indeed, the qualitative interviews repeatedly mention the financial challenges that hinder learning and school attendance. The MOE respondents speak of various initiatives to cover the cost of education in terms of providing support for uniforms, fees and books.

In terms of the other indicators, no statistically significant differences are detected. It should be noted, though, that much in line with previous findings, financial literacy has grown more in every sub-group intervention cohort than in the comparison group. Meanwhile, English literacy has only negative DID estimates, indicating that, overall, the comparison schools are faring better in this regard when looking at poverty indicators, yet these differences, again, are not statistically significant.

DIET AND LEARNING

A number of dietary variables were introduced to the survey this Midline. This means that we cannot look at learning in relation to these factors over time. However, we can look at their effect on overall learning scores at this Midline to see if the variables are significantly correlated with learning outcomes.

First, we look at girls from households who have had to reduce their food intake in the past three months and those who had had to go without food for an entire day in the past one month.

TABLE 33: LEARNING SCORES FOR THOSE WITH REDUCED FOOD INTAKE IN THE LAST THREE MONTHS

	Mean score – reduced food intake	Mean score – no reduced food intake	Difference

Numeracy	48.5	47.9	0.6
Somali	53.2	53.3	-0.1
English	21.3	16.3	5.0
Fin. Lit.	20.7	19.4	1.2
n	392	681	

TABLE 34: LEARNING SCORES FOR THOSE WHO WENT WITHOUT FOOD AN ENTIRE DAY IN THE LAST MONTH

	Mean score – went without food	Mean score – did not go without food	Difference
Numeracy	50.2	47.6	2.6
Somali	54.4	53.0	1.4
English	23.6	16.6	7.0
Fin. Lit.	23.6	18.9	4.7
n	226	847	

The findings here do not seem intuitive as girls from households that have this condition present in fact seem to score higher than those girls where food security seems more stable. However, most of the measures do not have a statistically significant relationship with either variable. The relationship with learning outcomes and these variables is only significant when looking at their correlation with the English language assessment is this relationship statistically significant. Most likely these variables are thus not very good measures of food insecurity or the variable co-varies with an omitted variable that would explain the results.

TABLE 35: GIRLS FROM HHS WHO DO NOT HAVE A SOURCE OF PROTEIN IN THEIR DIET THE LAST 24 HOURS

	Mean score – no protein in diet	Mean score – protein in diet	Difference
Numeracy	42.7	49.4	-6.7
Somali	44.0	55.5	-11.5
English	16.0	18.6	-2.7
Fin. Lit.	16.7	20.6	-3.9
n	218	866	

Next we looked at the performance of girls from from HHs who reported not having a source of protein in their diet the last 24 hours. This is where the caregivers of the girl had said "no" to all major sources of protein (meat, eggs, dairy, lentils/nuts/legumes, fish). As the above table shows, for all learning assessments these girls have a lower average score than those who have had protein in their diet. The relationship for the variable

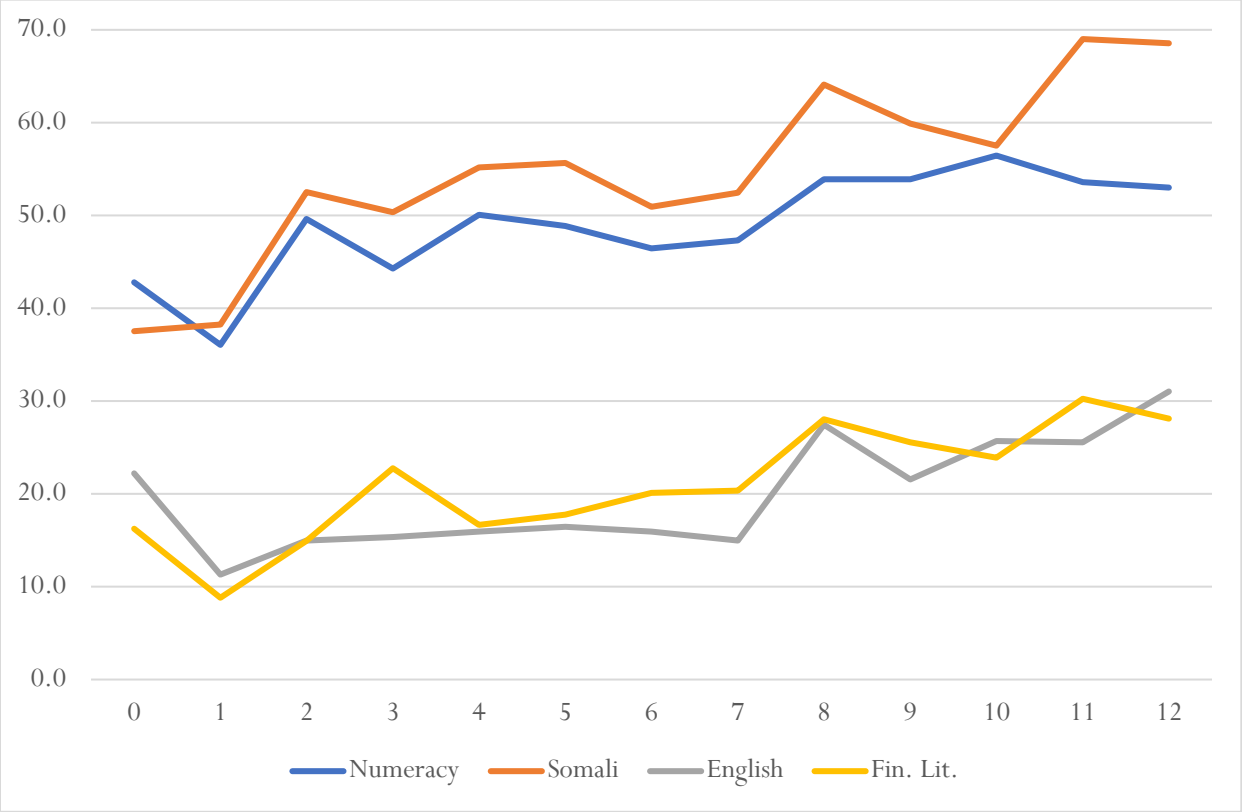
and the learning scores is statistically significant at the 99 percent confidence level. The absence of protein in the girl's diet is associated with substantially lower scores in every learning assessment with the effect being the most substantial for Somali literacy where girls who do had not had food containing protein in the past 24 hours had an average score 11.5 points lower than those that had had protein in the same timeframe.

TABLE 36: LEARNING SCORES FOR HHS THAT REDUCED THE NUMBER OF MEALS THEY ATE IN THE LAST MONTH

	Mean score – reduced meals	Mean score – no meals reduced	Difference
Numeracy	47.8	48.3	-0.4
Somali	52.0	53.9	-1.8
English	20.4	17.0	3.4
Fin. Lit.	19.6	20.0	-0.4
n	335	738	

We also looked at girls from households that had had to resort to reducing the number of meals they have consumed in the past month. Apart from the English assessment, the findings here are consistent with the hypothesis that reduced food intake would hinder learning outcomes. For numeracy, Somali and financial literacy, girls from these households score lower than their counterparts, although the effective is not very substantial.

FIGURE 9: EFFECT OF DIETARY DIVERSITY ON LEARNING



Finally, the dietary diversity measure, illustrated above is a variable that tallies the number of food groups the respondents had consumed in the last 24 hours.[58] As the above graph illustrates, the overall effect of dietary diversity on learning seems to be a positive one – high dietary diversity scores are associated with high learning assessment scores and lower scores on dietary diversity also tend to imply lower learning scores. However, this relationship appears not to be linear as the graph shows – girls with very limited diet seem to score better in the learning assessments that those with some, albeit limited, diversity in their food intake.[59]

HOUSEHOLD DEMOGRAPHICS

TABLE 37: PROGRAM IMPACT BY HH DEMOGRAPHICS

	Numeracy	Somali literacy	English literacy	Fin. literacy	Sample size

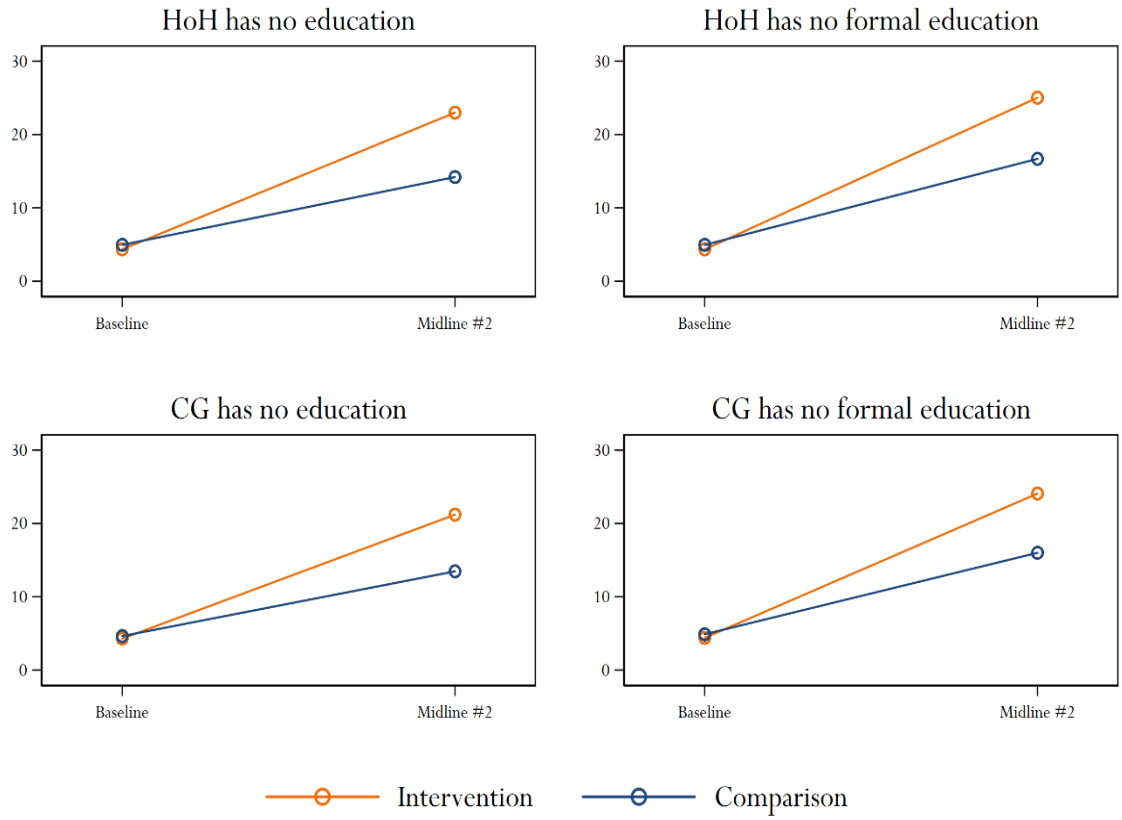
[58] The food groups, loosely, include loosely: cereal/grains, tubers/root vegetables, other vegetables, fruit, eggs, meat, legumes/lentils/beans, fish, dairy, fats/oil, sugar/sugary things, condiments/herbs.

[59] For all but English the relationship between dietary diversity and the learning assessments is a positive one which is statistically significant at the 99 percent confidence level. For English the coefficient is still positive but is not statistically significant (p=0.087).

	DID	p	DID	p	DID	p	DID	p	Intervention	Comparison
Female-headed household	4.0	.180	-0.2	.949	0.7	.835	4.7	.235	520	432
Lives without parents	5.5	.357	4.7	.427	7.2	.252	4.6	.404	84	120
Part orphan	3.2	.492	-1.4	.795	1.8	.735	8.3	.184	118	138
HoH has no education	3.9	.180	0.2	.954	1.2	.686	6.4	.028*	762	650
HoH has no formal education	3.9	.180	0.2	.954	1.2	.686	6.4	.028*	762	650
CG has no education	4.4	.091	1.6	.553	0.7	.821	7.3	.011*	822	772
CG has no formal education	3.3	.183	0.9	.733	1.3	.645	8.5	.004*	934	826

As it pertains to measures of household demographics and indicators of potential marginalisation, the program fares rather well. For numeracy, likewise, while the findings are not conclusive as none of the estimates are statistically significant, the overall trend is positive. For both assessments of literacy, the story is somewhat more ambiguous as the DID estimates are closer to 0 indicating a lesser program impact. Again, for various sub-groups the learning happening in financial literacy seems to be significantly faster than in the comparison cohort as is illustrated in the graph below. In particular, for every indicator of low levels of education in the household the intervention girls have grown a gap in financial literacy between them and those in the comparison schools. It thus seems that the program is able to turn the poor starting point around much more effectively than other schools.

FIGURE 10: PROGRAM IMPACT AMONG THOSE WITH UNEDUCATED PARENTS



HOUSEHOLD ECONOMICS

TABLE 38: PROGRAM IMPACT BY HH ECONOMICS

	Numeracy		Somali literacy		English literacy		Fin. literacy		Sample size	
	DID	p	DID	p	DID	p	DID	p	Intervention	Comparison
Head of HH is pastoralist	4.8	.473	-3.4	.722	-2.0	.593	3.5	.497	128	112
HoH has no occupation	4.7	.167	4.6	.165	2.3	.507	11.0	.007*	460	402
Seasonal migration	-4.5	.401	-9.3	.237	-1.9	.061*	-0.2	.979	54	66
Owns camels	-5.3	.220	-8.8	.141	-8.8	.107	7.0	.112	114	108

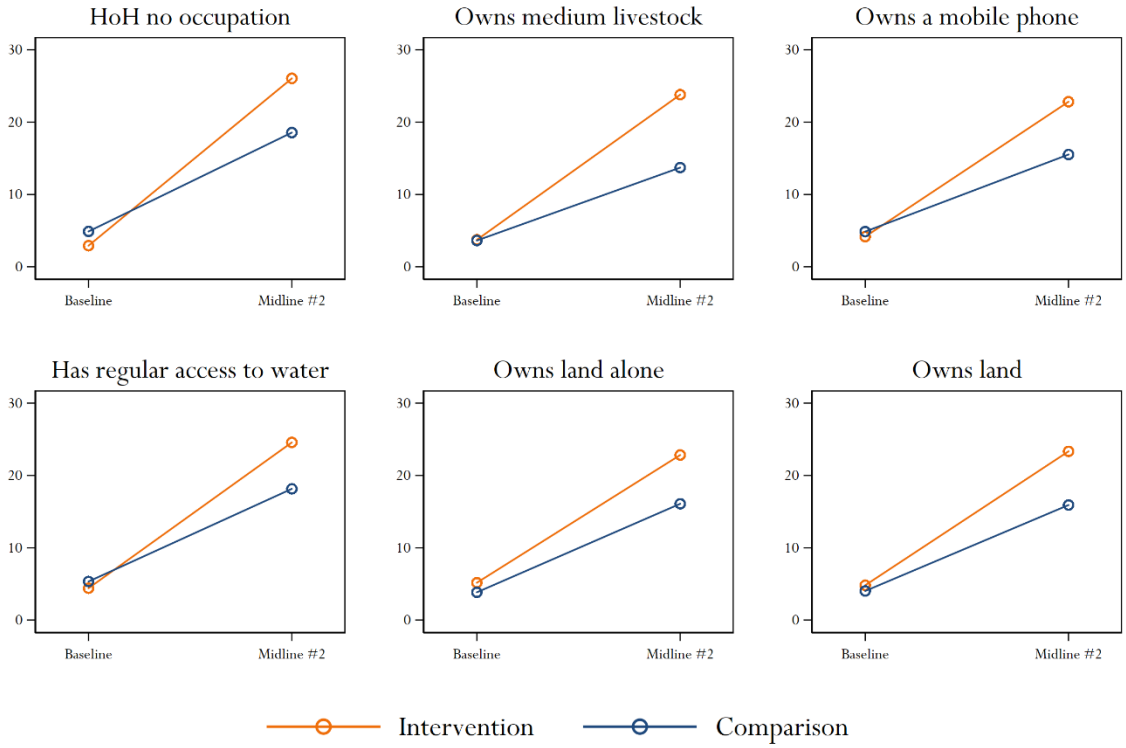
Owns medium livestock	6.3	.029*	2.1	.566	0.1	.961	8.6	.021*	648	686
Owns a mobile phone	3.3	.208	1.0	.708	0.4	.883	8.5	.008**	1096	968
Has regular access to water	-0.1	.965	-3.4	.385	0.4	.913	1.6	.008**	586	522
Owns land alone	3.9	.158	0.9	.759	2.4	.488	1.0	.003**	778	754
Owns land	3.9	.143	1.1	.684	2.8	.373	9.7	.003**	896	828

Next, we turn to some measures of household economic situation. Some of the indicators reflect a position of relative affluence – such as having regular water access, owning land or owning a mobile phone – whereas others are measures for the risk of marginalisation: head of household being involved in pastoralism, owning livestock (another indicator of pastoralism), family migrating seasonally, or the HoH having no occupation at all.

Encouragingly, it seems that whether we look at measures of affluence or marginalisation, the findings are similar in that when the difference in differences estimate is statistically significant it indicates that program has been more successful in improving financial literacy. This is encouraging because it indicates that the program's success in improving financial literacy does not seem to require a specific context as it is seemingly successful with both the socio-economically advantaged and the disadvantaged. For example, where the HoH does not have an occupation, program impact for financial literacy is as high as 11 percentage points. Yet, where the girl's household owns land, the program impact is almost as high, at 9.7.

Again, for literacy the program impact is not clear cut. In fact, the only statistically significant DID estimate has the comparison school improving their English literacy by 1.9 percentage points more than intervention school girls.

FIGURE 11: PROGRAM IMPACT AND HOUSEHOLD ECONOMICS



The challenges associated with reaching the pastoralists for education are well documented in the qualitative data:

> During the rainy season you can teach them though mobile team-teaching system, but during droughts period everyone evacuates. Therefore, the education of pastoral community is very poor and low, and nothing has been done. – MOE Respondent, Puntland

It is thus good to see that the program seems to have made an impact among this difficult cohort.

For numeracy, the findings are not as clearly positive as they have been for some of the sub-groups above. In two of the pastoralism-related sub-groups the DID estimate is negative. These groups incidentally have the smallest sample size and are thus the most vulnerable to skew. And the measures are not statistically significant. Most sub-groups, conversely, have improved relative to their comparison girls, but in only one sub-group this difference is statistically meaningful. That is, among those who have small livestock, the program impact on numeracy is 6.3 points.

COMMUNITY ATTITUDES

TABLE 39: PROGRAM IMPACT AND COMMUNITY ATTITUDES

	Numeracy		Somali literacy		English literacy		Fin. literacy		Sample size	
	DID	P	DID	P	DID	P	DID	P	Intervention	Comparison
High chore burden (whole day spent on chores)	2.0	.589	-2.2	.666	0.7	.861	1.6	.010*	244	262
Girl attends school most days	3.1	.359	-0.2	.944	-1.3	.742	8.5	.042*	742	602
PCG and family member not involved in CEC	7.6	.266	15.4	.028*	4.4	.380	8.8	.159	96	70
Girl feels no choice whether to attend or stay in school	2.7	.458	-1.1	.766	0.6	.877	8.9	.025*	620	458
Girl feels family makes schooling decisions for her	2.4	.498	0.4	.898	1.2	.704	7.0	.132	280	310

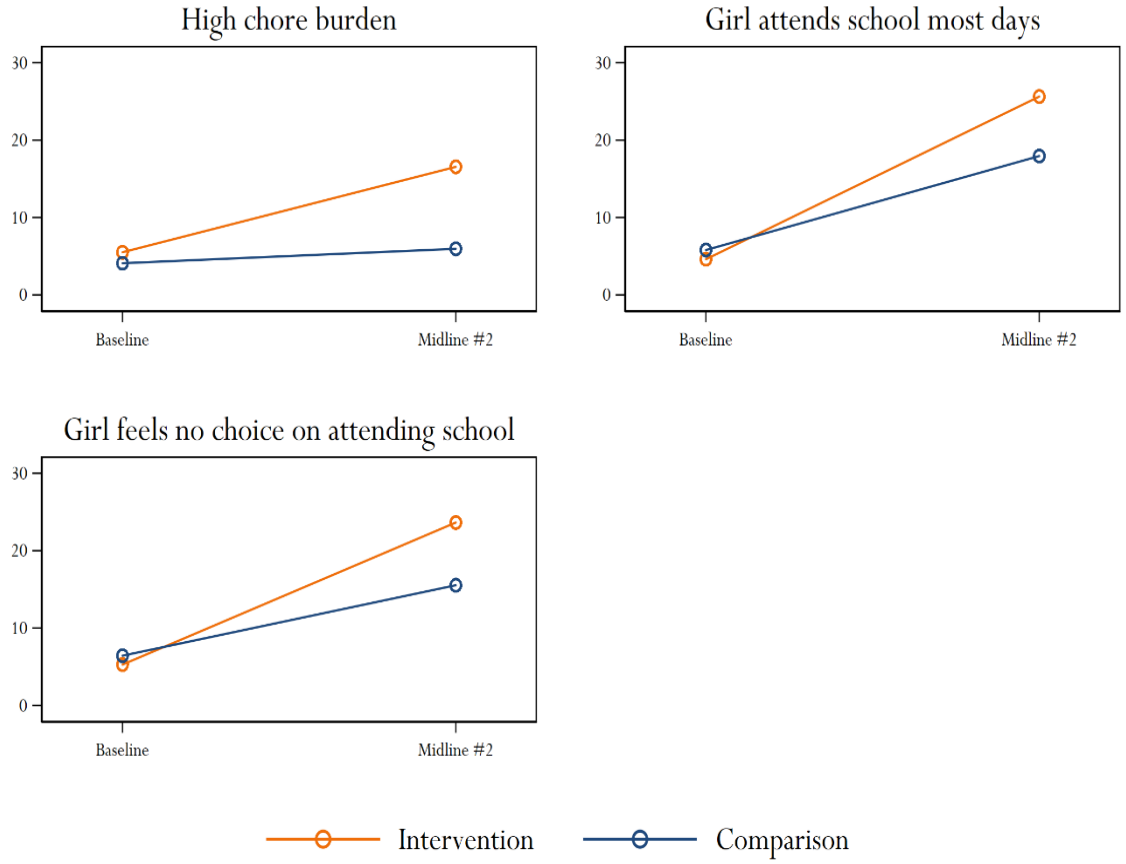
We turn to measures of community attitudes towards girl's education. Most of the measures are negative indicators for support for girl's education, while only the indicator on whether girl attends school most days can be viewed as a positive indicator of community support.

The findings are quite straightforward. Among none of the sub-groups is the comparison group faring better. All DID estimates for numeracy are positive, indicating that improvements in the intervention schools are more substantive than in comparison schools. However, none of these measures are statistically conclusive. Much like above, the literacy findings are again ambivalent. Among some groups such as those whose family members are not in the CEC, are showing much more improvement in Somali literacy in the intervention schools compared to other schools. Meanwhile, among some other sub-groups the intervention girls have improved their Somali and English assessment results less than the comparison cohort.

Financial literacy, again, is improving significantly more among the intervention cohort. Importantly, this is true whether we look at the negative indicators – i.e. those girls who most likely lack the support for education – or the positive indicator, reflecting likely support for girls' education. It seems that the program is having an impact in terms of financial literacy regardless of whether we look at groups that have a less favourable starting point or a more supportive one.

The below graph illustrates the changes between Baseline and Midline Round 2 for girls in these sub-groups.

FIGURE 12: PROGRAM IMPACT AMONG GIRLS WITH SUPPORTIVE OR NON-SUPPORTIVE COMMUNITY ATTITUDES



DISABILITY

The girls in the sample who have a disability are extremely few. Only two groups have enough respondents to make some cautious observations – those who have a mental health disability, and an aggregate variable for having any disability at all. However, as the table below shows, no differences at all can be detected between intervention and community groups when it comes to the learning outcomes of disabled girls. None of the DID estimates for numeracy, Somali and English come close to being statistically significant. However, as is the trend consistently, financial literacy is bordering on significance with substantially high differences between intervention groups and comparison girls. It is, again, encouraging to see that even among this group, which in the qualitative interviews is seen as very neglected and in many ways beyond the resources of the schools and the program to fully include, we can see that the improvement in learning seems much more rapid than in comparison schools.

TABLE 40: PROGRAM IMPACT AND DISABILITY

	Numeracy	Somali literacy	English literacy	Fin. literacy	Sample size

	DID	P	DID	P	DID	P	DID	P	Intervention	Comparison
Disability: mental health	-1.7	.800	0.7	.924	-3.0	.714	12.1	.134	60	44
Disability: any	0.7	.926	1.5	.825	-3.5	.668	9.4	.166	84	48

SCHOOL RESOURCES AND CHARACTERISTICS

In the first part of this section of the report we looked at various measures of school quality and resources reported by the caregiver or the girl. These measures did not correspond with learning levels in an intuitive manner as they were largely positively correlated when they were supposedly indicating the absence of resources or poor teaching quality. They might in fact, thus, be more correlated with self-assuredness and critical thinking rather than actual school resources or the quality of education at the learning institution. As such, when looking at school's resources we will focus on the indicators derived from the head teacher survey data.

The below table shows the developments in learning scores when comparing the intervention schools to the cohort schools among the sub-group.

TABLE 41: PROGRAM IMPACT BY SCHOOL RESOURCES

	Numeracy		Somali literacy		English literacy		Fin. literacy		Sample size	
	DID	P	DID	P	DID	P	DID	P	Intervention	Comparison
School has reliable electricity	-8.3	.195	-2.6	.680	-0.9	.928	0.8	.935	138	150
School has electricity at all (ever)	-3.7	.402	-2.1	.628	-6.3	.402	7.0	.331	354	212
School feeds students	8.1	.039 *	1.3	.717	2.9	.392	10.6	.010 *	798	490
Language of instruction not Somali	3.9	.142	1.3	.606	1.3	.647	8.6	.007 *	1120	1018
School has a roof	5.4	.084	1.2	.673	1.3	.670	7.5	.045 *	822	710
School has all cement floors	1.3	.617	0.0	.997	0.9	.795	8.3	.035 *	990	648
School has access to water	2.6	.449	1.4	.693	-2.5	.483	9.5	.074	684	442
School has access to water within 1 km	1.7	.670	2.5	.518	-2.4	.641	0.8	.859	786	334
School provides sanitary towels	1.7	.682	2.7	.493	-2.5	.635	1.8	.703	730	334

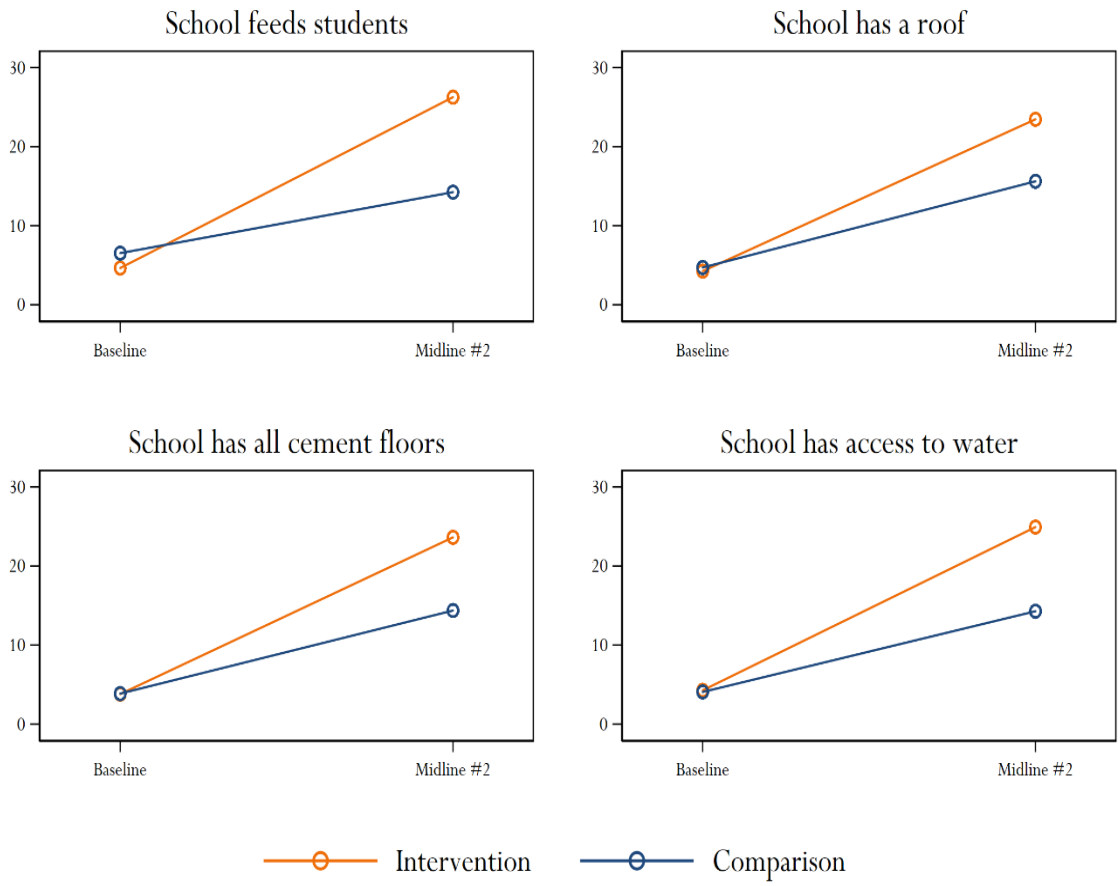
At least one female FT/PT teacher	7.6	.128	-0.5	.901	-7.1	.110	14.3	.007 **	296	328
At least one female FT teacher	7.4	.058	3.3	.409	0.1	.988	15.1	.001 **	438	428
Short instructional time (less than 5 h)	2.2	.540	2.6	.520	1.8	.723	9.1	.142	300	440
Few contact hours (less than 30 h/week)	5.3	.258	-0.9	.835	3.2	.503	13.5	.009 **	268	264
Class size greater than 30	6.1	.034 *	1.4	.638	2.9	.306	10.0	.005 **	1082	738
Teacher absenteeism	3.9	.156	0.2	.939	0.7	.830	8.8	.008 **	1112	888
School has toilets	3.7	.156	0.0	1.000	1.4	.652	9.6	.005 **	1078	804
School has toilets for girls	4.2	.281	3.1	.429	1.9	.626	10.8	.015 *	582	380
Toilets that have privacy walls for girls	5.5	.125	3.7	.323	-0.4	.890	13.3	.009 **	530	284

The findings are much in line with what has been observed already. Some sub-groups have seen a significant improvement in numeracy. For example, among schools with a feeding program, the intervention school students have improved their numeracy outcomes by 8.1 percent more than those in the comparison schools. Meanwhile, in schools with big class sizes, the program impact for numeracy has been 6.1 percent. The program impact for numeracy, importantly, is positive in almost all the sub-groups indicating that the success in this regard is not closely tied to other factors. This is encouraging as it implies that the program does not seem to be leaving behind any fundamental groups that could have been ignored.

As seen many times above, the results for literacy, both Somali and English, are more bifurcated whereby some groups have improved more than those in the comparison cohort, while others have not. For Somali the balance is slightly on the positive side, while for English the reverse is true. Yet, no statistically significant differences are detected to suggest the contrary – i.e. the comparison schools are not improving their literacy outcomes substantially better either.

As has emerged above time and again, most positive changes can be seen when looking at the improvements in financial literacy. The following graph shows the gap grown between intervention and comparison students in well-resourced schools.

FIGURE 13: PROGRAM IMPACT IN SCHOOLS WITH GOOD RESOURCES



The same impact can be observed in financial literacy in schools that have girl-friendly spaces.

FIGURE 14: PROGRAM IMPACT AND GIRL-FRIENDLY SPACES

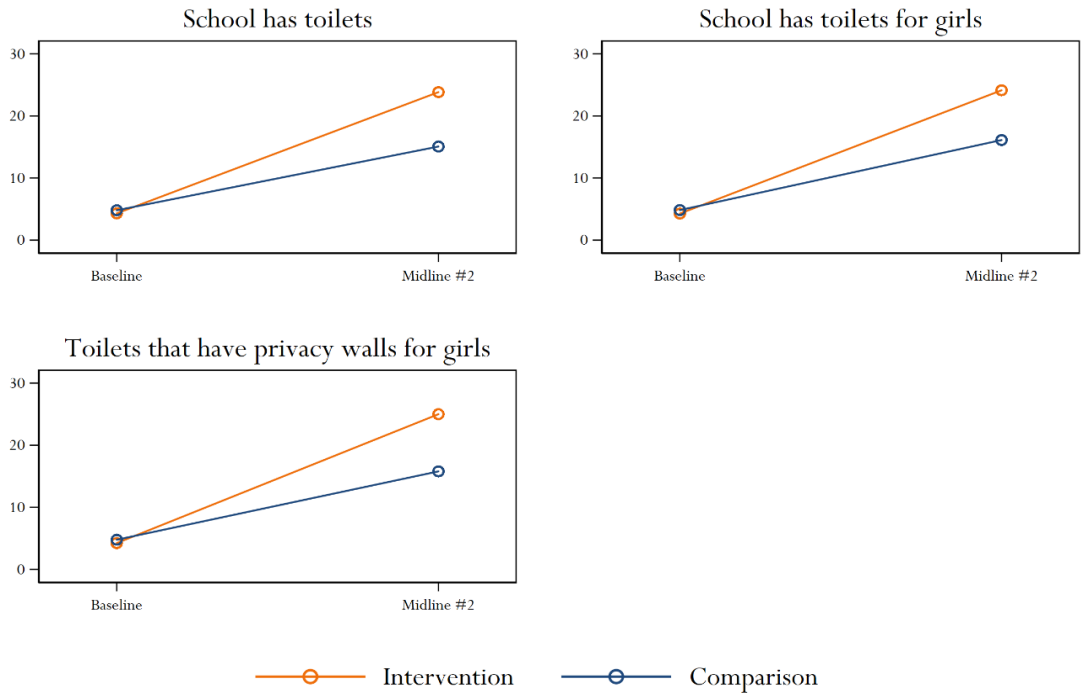
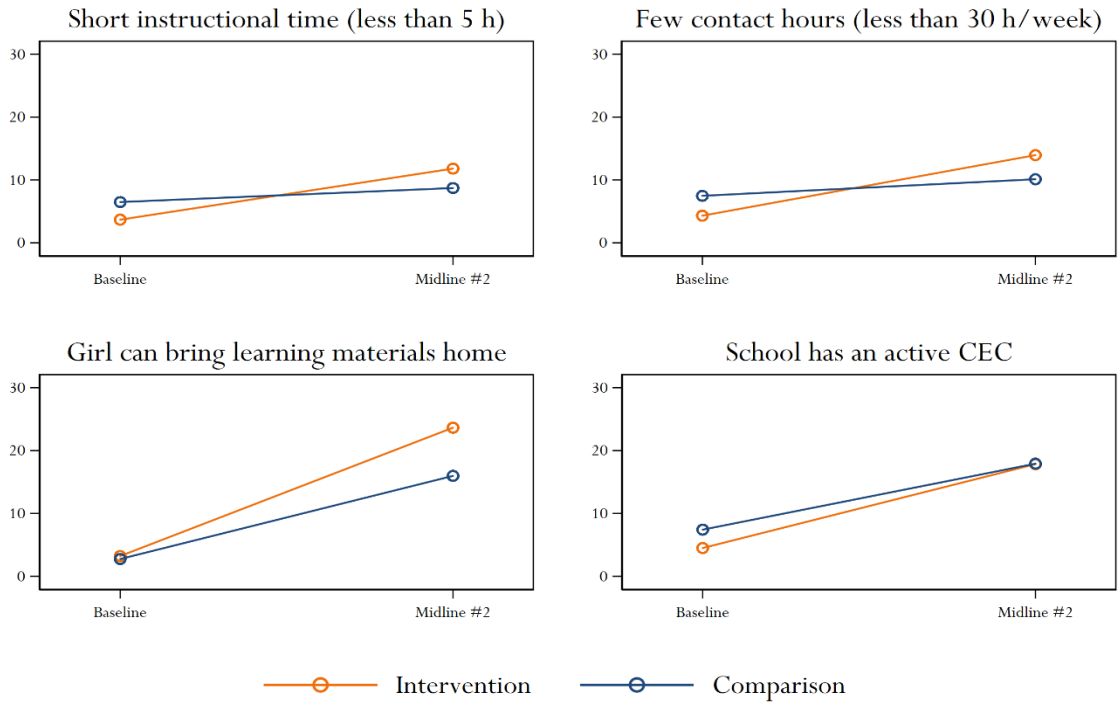


FIGURE 15: PROGRAM IMPACT AND SCHOOL QUALITY



Finally, the financial literacy effect can also be observed in schools that indicate having poorer teaching quality and resources. The graph above illustrates that students in intervention schools with short days and few contact hours have clearly improved their scores while comparison schools have seen very little change in their performance on financial literacy. Meanwhile in schools with a slightly more favourable learning environment due to student being able to take books home, or the school having an active CEC the intervention cohort has again either removed a gap that existed at baseline or created one between them and comparison school cohort, respectively.

FINDINGS FROM SUB-GROUP ANALYSIS

Some key trends emerge from looking at the program impact by different sub-groups.

Firstly, wherever statistically significant differences in differences are detected, they are positive, indicating that the intervention cohort is improving their learning outcomes more relative to the comparison group. As such, we can relatively confidently state that as it pertains to all the assessments, the comparison group is not showing more progress than the intervention cohort.

Secondly, as has been observed above when analysing the aggregate results, many sub-groups have shown a significant increase for numeracy outcomes when comparing to their comparison counterparts. Moreover, in very few instances is the difference in differences estimate for numeracy negative. This indicates that the success in this regard is not closely tied to other factors. This is encouraging as it implies that the program does not seem to be leaving behind any fundamental groups that could have been ignored.

Thirdly, as has already been discussed in the section on aggregate findings, very little program impact can be discerned in literacy outcomes. Indeed, in some sub-groups there have been improvements, and in others not. Where the pattern is clearly positive for financial literacy, and to a lesser extent, for numeracy, the same cannot be argued for literacy outcomes.

Finally, there is a clear and positive pattern as it pertains to financial literacy. Many sub-groups, regardless of whether they capture positive or supporting factors or the opposite, show a clear improvement among the intervention cohort vis-à-vis the comparison girls. This indicates, again, that the program's impact in improving financial literacy levels among the intervention cohort does not seem to vary significantly depending on whether the learning environment is one that support or creates challenges. The same seems to be true when looking at different kinds of students. Both in groups where you would expect to see learning difficulties as well as among groups where you would expect the students to learn with relative ease, the program is showing an emerging gap between intervention and cohort groups.

4.4 GRADE LEVEL ACHIEVEMENT

This section reports the share of girls achieving each grade level of literacy and numeracy, based on an analysis of the school curricula or syllabi of Somaliland and Puntland. A number of observations were made at baseline that continue to apply. Firstly, as noted at baseline, the SOMGEP-T schools fall under the direction of Ministries of Education in multiple jurisdictions (Somaliland, Puntland, and Galmudug) and in principle, they adhere to different curricula. Secondly, while at baseline there was no curriculum for Federal Government of Somalia (that would apply to Galmudug), the curriculum recently developed by the FGS is rather general

and ambiguous, thus not lending itself to detailed assessment of competencies specific to grade levels.[60] Thirdly, the curricula continue to focus almost exclusively on mathematics and English learning only, while the targeted learning outcomes for specific grade levels in Somali are either not given or very loosely specified. Moreover, the learning assessments were developed for the purposes of the program and not necessarily to measure grade level competency, the skills tested in the assessments do not always line up neatly with skills in the curricula.[61]

Given the above, the decisions made at Baseline continue to provide the best approach to measuring grade level achievement at Midline Round 2. Firstly, grade level achievement was assessed using the more well-defined curricula of Somaliland and Puntland. Given that Somaliland and Puntland schools are subject to different standards, we attempted to adjudicate between the two as fairly as possible, being slightly conservative with respect to the grade at which children should achieve a specific skill.[62] Where skills were not specifically listed in either curriculum, we used our judgment to match them to specifically-enumerated skills in terms of their perceived difficulty.[63] The occasionally subjective nature of our mapping of skills to grades should be borne in mind. Secondly, the grade level achievement can only be gauged for English literacy and mathematics, due to the lack of information on Somali learning outcomes.[64]

The following table outlines the standards for the two subjects for each grade.

TABLE 42: GRADE LEVEL STANDARDS FOR MATHEMATICS AND ENGLISH LITERACY

Grade Level Achieved	Mathematics Skills	English Literacy Skills
1	• Addition without carrying numbers (portion of subtask 2) • Subtraction without borrowing (subtask 3)	N/A
2	• Addition carrying one number (portion of subtask 2) • Addition with 3 digits, carrying up to 1 number (subtask 4)	• Letter identification (subtask 1)

[60] It is possible that in the future the curriculum of Puntland will also come under the FGS one as Puntland officially belongs under the authority of the Federal Government.

[61] To illustrate, consider mathematics at grade levels 5 and above: none of the specific skills tested in the learning assessments are indicative, specifically, of the achievement of Grade 5 (or 6) mathematics performance. As a result, although the learning cohort includes children in grades 5 and 6, the available data do not allow us to distinguish between students who achieve a 4th grade level and children who achieve a 5th grade level of performance in the subject.

[62] For instance, if a child in Somaliland was expected to achieve a skill in Grade 3 and a child in Puntland was expected to achieve the same skill in Grade 4, we would classify the skill as being at a grade 4 level.

[63] As an example, neither curriculum available specifies when a child should learn to construct the negative form of a sentence. We consider this skill on par with the difficulty of constructing future tense sentences, which is a skill expected to be developed in Grade 6.

[64] It is also important to note that the available curricula are focused on primary-level outcomes. Given the evaluation's focus on students in grades 3-6, this does not pose a problem for the baseline evaluation. However, a fuller understanding of grade level achievement at the endline may necessitate a deeper review of curricular materials – where available – for secondary schools.

	• Subtraction carrying one number (portion of subtask 5) • Addition and subtraction word problems with simple underlying arithmetic (subtask 6) • Multiplication of 1-digit numbers (subtask 7) • Division of 2-digit number by 1-digit number (subtask 9)	
3	• Subtraction carrying two numbers (portion of subtask 5) • Multiplication of 2-digit numbers (subtask 8) • Word problems with simple multiplication and division (subtask 11)	• Identification of basic words, e.g., classroom objects, foods, animals (subtask 2)
4	• Identifying number patterns (subtask 1) • Division of 3-digit number by 2-digit number (subtask 10)	• Reading simple sentences (subtask 3 and portion of subtask 4)
5	N/A	• Reading low-medium difficulty sentences (subtask 5; portion of subtask 4)
6	N/A	• Reading medium-difficulty sentences (subtask 6) • Filling in missing words with medium-difficulty words (subtask 7) • Converting to negative form (subtask 8) • Converting to future tense (subtask 9)

Some changes have occurred since baseline. These will be discussed in more detail in Annex X of the report which contains the full description of the curricula of Puntland and Somaliland. The table above, describes the standards developed by the evaluation team for grade level achievement in mathematics and English literacy. In cases where no standard is described, the learning assessments utilised in the evaluation did not include a skill specific to that grade level. In cases in which a subtask is distributed across two grade levels, we distinguish between less and more difficult portions of the subtask, on an item-by-item basis. In order to achieve a given grade level, a student must achieve a score of approximately 80 per cent on subtasks (or

relevant, grade-specific portions of a subtask) for that grade, and those for the preceding grades.[65] To elucidate, let us consider an example of a student being assessed for grade 1-level numeracy. They would need to achieve scores of approximately 80 per cent or higher on subtask 3. They would also need to achieve a score of approximately 80 per cent on the grade 1-level portion of Subtask 2, which tests addition that does not require "carrying" numbers. They would *not* need to achieve a passing score on the grade 2-level portions of subtask 2, which tests addition that requires carrying numbers. A student being assessed for grade 2-level competency would need to complete each of the subtasks specified for grade 1 *and* those specified for grade 2.[66]

TABLE 43: ENGLISH GRADE LEVEL ACHIEVED BY INTERVENTION (AND COMPARISON GIRLS IN PARENTHESES), BY GRADE – AT MIDLINE ROUND 2

Grade Level Achieved	OOS	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7
Below Grade 2	86.6% (95.4%)	100% (89.2%)	88.1% (90.7%)	79.8% (80.1%)	71.6% (76.1%)	61.0% (72.2%)
2	8.3% (2.6%)	0% (6.9%)	9.5% (4.7%)	11.5% (9.3%)	3.9% (11.9%)	11.6% (12.8%)
3	1.7% (1.3%)	0% (3.9%)	2.4% (2.3%)	4.8% (4.0%)	10.8% (6.0%)	11.6% (6.4%)
4	1.7% (0.7%)	0% (0%)	0% (2.3%)	3.9% (5.3%)	9.8% (6.0%)	12.6% (7.5%)
5	1.7% (0%)	0% (0%)	0% (0%)	0% (0%)	2.9% (0%)	2.1% (0%)
6	0% (0%)	0% (0%)	0% (0%)	0% (1.3%)	1.0% (0%)	1.1% (1.1%)

TABLE 44: ENGLISH GRADE LEVEL ACHIEVED BY INTERVENTION (AND COMPARISON GIRLS IN PARENTHESES), BY GRADE – AT BASELINE

Grade Level Achieved	Out-of-School	Grade 3	Grade 4	Grade 5	Grade 6
Below Grade 2	80.2% (90.3%)	74.1% (87.2%)	81.2% (77.0%)	83.5% (76.5%)	72.4% (73.6%)
2	15.7% (8.2%)	21.7% (12.0%)	17.0% (21.6%)	13.0% (15.7%)	23.5% (21.1%)
3	4.1% (1.5%)	4.2% (0.8%)	1.8% (1.4%)	2.6% (7.8%)	4.1% (5.3%)
4	0.0% (0.0%)	0.0% (0.0%)	0.0% (0.0%)	0.9% (0.0%)	0.0% (0.0%)
5	0.0% (0.0%)	0.0% (0.0%)	0.0% (0.0%)	0.0% (0.0%)	0.0% (0.0%)
6	0.0% (0.0%)	0.0% (0.0%)	0.0% (0.0%)	0.0% (0.0%)	0.0% (0.0%)

TABLE 45: MATHEMATICS GRADE LEVEL ACHIEVED BY INTERVENTION (AND COMPARISON GIRLS IN PARENTHESES), BY GRADE – AT MIDLINE ROUND 2

Grade Level Achieved	OOS	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7

[65] Where there are many items in a subtask, we follow the 80 per cent rule. Where there are four items, we allow one wrong answer (75 per cent correct) without disqualifying a student from achievement of a grade level.
[66] Note that the grade-level analysis includes only the in-school girls cohort – girls who were enrolled at baseline.

Below Grade 1	49.6% (52.5%)	32.2% (58.6%)	35.7% (32.5%)	9.6% (20.0%)	8.8% (13.3%)	6.3% (20.1%)
1	43.8% (44.1%)	61.3% (37.9%)	64.3% (51.2%)	68.3% (66.7%)	65.7% (67.2%)	54.7% (47.9%)
2	3.3% (2.0%)	6.5% (3.5%)	0% (14.0%)	14.4% (8.0%)	13.7% (10.5%)	24.2% (18.1%)
3	0% (0.7%)	0% (0%)	0% (2.3%)	4.8% (5.3%)	5.9% (9.0%)	7.4% (9.6%)
4	3.3% (0.7%)	0% (0%)	0% (0%)	2.9% (0%)	5.9% (0%)	7.4% (4.3%)

TABLE 46: MATHEMATICS GRADE LEVEL ACHIEVED BY INTERVENTION (AND COMPARISON GIRLS IN PARENTHESES), BY GRADE – AT BASELINE

Grade Level Achieved	Out-of-School	Grade 3	Grade 4	Grade 5	Grade 6
Below Grade 1	78.4% (80.6%)	54.5% (59.1%)	48.2% (45.9%)	35.6% (36.2%)	31.7% (32.8%)
1	19.8% (18.9%)	44.1% (40.1%)	44.6% (48.7%)	55.7% (55.0 %)	45.9% (57.9%)
2	1.2% (0.5%)	0% (0%)	2.7% (0%)	2.6% (3.9%)	7.1% (5.3%)
3	0.6% (0.0%)	0% (0%)	1.8% (0%)	1.7% (2.0%)	1.0 % (0%)
4	0.0% (0.0%)	1.4% (0.8%)	2.7% (5.4%)	4.4% (2.9%)	14.3% (4.0%)

The above tables show the breakdown of achievement of grade levels for both English and mathematics. The first percentage for each grade is the share of cohort girls who meet the grade level. The percentage in parenthesis are for the same share among comparison group girls.

The first observation emerging from the tables is that virtually nobody is able to reach their grade level achievement. Ideally, we would expect girls to achieve performance at their grade level. However, for example, in grades 3, 4 and 5, none of the intervention cohort are able to reach the grade level achievement for their grade in English literacy. In Grade 6, only 1 percent can. Meanwhile, some, very few, comparison schoolgirls are able to. Similarly, for mathematics competencies, no girls in either comparison or intervention cohorts are able to reach their expected grade level in grade 3 or 4.

It seems that the situation regarding English grade achievement has in fact worsened somehow since Baseline, particularly in the lower grades. At Baseline, more than 21 percent of intervention girls in grade 3 were able to achieve grade 1 competency. At Midline Round 2 only none can. A similar tendency can be observed for grade 4 English. Meanwhile, the situation for grade 5 and 6 has somewhat improved with more girl edging closer to their own grade level while not yet quite reaching it. This begs the question of whether the intervention has been particularly successful with one age or grade group that is advancing through the grades.

For mathematics, conversely, the development is more straightforward in that more girls in their age grades are reaching higher grade levels while still not quite reaching their own grade level. The development is also clearly more marked for the intervention cohort. Again, this is perhaps starting to reveal the programs impact on numeracy outcomes, which is predicted to fully reveal itself in the coming Endline.

4.5 ALP GIRLS LEARNING OUTCOMES

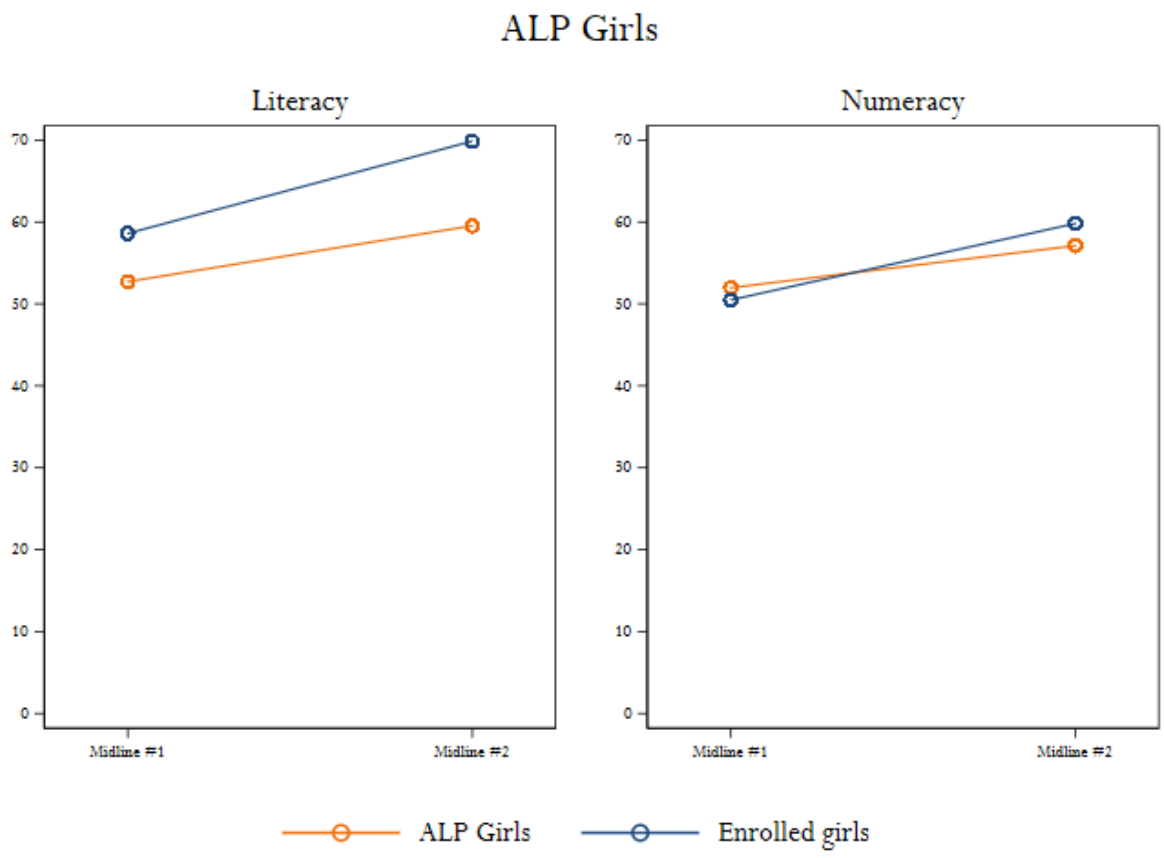
In this section, we analyse the learning performance of girls who were recruited, during ML1, into a sample of Alternative Learning Program (ALP) girls. At that time, 365 girls were recruited for the sample, and 336

girls from the sample were successfully re-contacted at ML2. The analysis in this section, because it focuses on learning outcomes that we expect to be impacted only through girls' participation in ALP activities, looks exclusively at the set of ALP girls who were successfully re-contacted at ML2 and who remained enrolled in ALP at ML2. There are 189 girls who were in ALP at Midline Round 1 and who remain in ALP programming at Midline Round 2. This section presents the learning outcomes for this group.

Assessment	ML1		ML2	
	Mean score	Standard deviation	Mean score	Standard deviation
Numeracy	52.0	26.1	57.1	24.1
Somali literacy	52.7	32.1	59.5	34.0

The table above results the learning scores for the ALP cohort. Both learning scores, naturally, have increased since ML1. This might be a result of maturation as girls get older and are more exposed to education. Let us compare the ALP girls to other girls at school.

As the below graph demonstrates, when comparing the ALP girls to all girls who have been enrolled between Midline Round 1 and Midline Round 2, while both groups have improved their learning scores, the enrolled girls have improved theirs slightly more in comparison.



A comparison between ALP girls and out-of-school girls would have been ideal but this was not possible due to the fact that at Midline Round 1 only in-school-girls were interviewed.

4.6 Boys Learning Outcomes

In this section we briefly turn to the learning outcomes for the boys sampled through the evaluation. The reader should note that the data used here is not part of a longitudinal sample like the girls' cohort as the boys interviewed at baseline have not been re-contacted at subsequent phases of data collection. Rather, at each wave of data collection, boys have been randomly selected through the households of the cohort girls that are part of the sample.

When comparing the boys to the girls we are using on the one hand the panel of girls which has full overlap with the individuals at Baseline and Midline Round 2 (i.e. the "baseline panel"). This sample contains girls who are both in school and out of school. The boys' sample, likewise, contains in-school and out-of-school individuals. Boys were sampled during each evaluation round through the household survey that was already being conducted with households of cohort girls – those households with a boy in the eligible age range were recruited into the sample of boys. However, in order to ensure maximum comparability between the boys at Baseline and now, we have excluded those boys who were sampled through the households of the ALP girls as they were not interviewed at Baseline. This is because it is reasonable to expect that given the nature of ALP programs, it is possible that households with children who attend ALP program can be systematically different to other households, and as such this course of action is preferable.

This results in the following sample distribution:

Table 47: Boys sample

	Intervention	Comparison	Total
Baseline	247	219	**466**
Midline Round 2	160	108	**268**

We have chosen not to conduct analysis at the grade level because the number of sampled boys per grade is not sufficient for rigorous estimation of program impact, and analysing findings with a very small group of individuals would be misleading.

Table 48: Grade distribution of boys

	Baseline	Midline Round 2
Out of school	157	48
Grade 3	91	8
Grade 4	66	13
Grade 5	85	37
Grade 6	67	44
Grade 7	0	54
Grade 8	0	40

The gender disparities motivating this program to begin with continue to be observed at this Midline. The table below presents the aggregate results for boys and girls across all sampled areas (inclusive thus of both intervention and comparison schools). Only in Somali literacy the gap between boys and girls seems to be

disappearing in both comparison and intervention schools. In numeracy the boys still beat girls by more than 12 percentage points. In English the difference is 7.6 percentage points in boys' favour. In financial literacy, the gap is 6.4 points.

TABLE 49: BOYS AND GIRLS LEARNING SCORES

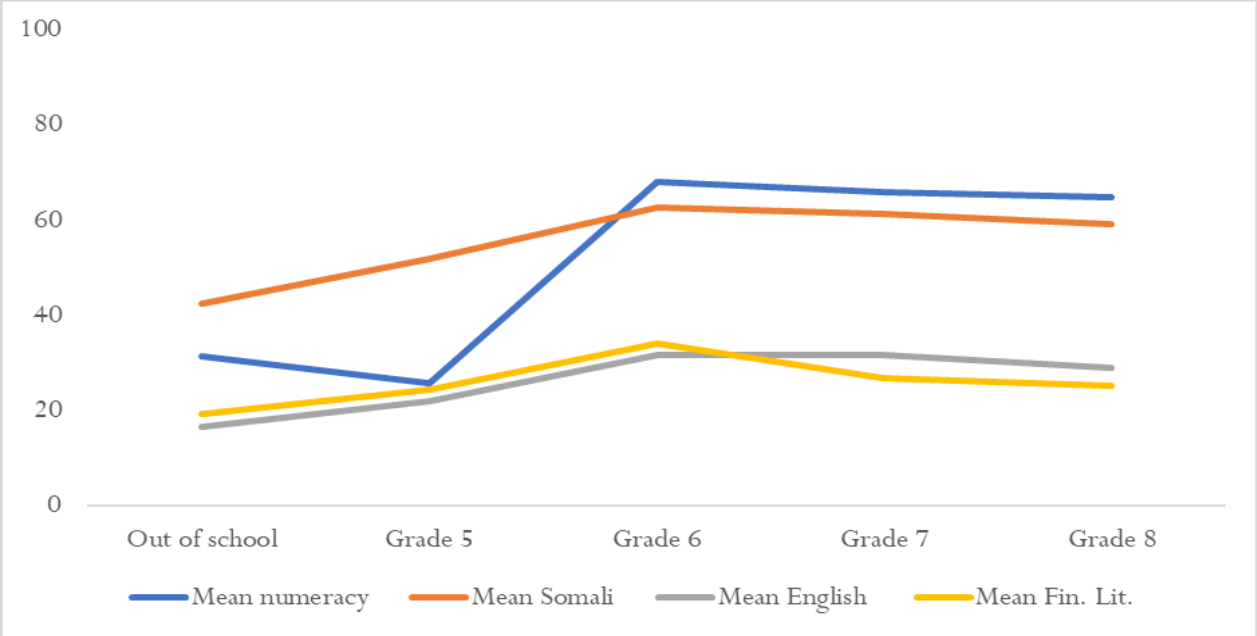
	Girls		Boys	
Assessment	Mean score	Standard deviation	Mean score	Standard deviation
Numeracy	48.0	28.8	60.1	28.3
Somali literacy	53.2	36.5	55.7	37.0
English literacy	18.1	25.1	25.8	30.6
Financial literacy	19.8	28.0	26.2	31.8

When we look at learning trends among the boys, we can observe that overall grade correlates positively with the learning assessment scores – higher grades correspond with higher learning assessment scores. However, the relationship does not seem to be a linear one as the table and graph below demonstrate. In some cases that boys in grade 3 score higher than those in grade 4 and 5. As such, unlike was observed last year, boys' scores do not increase monotonically with grade. However, this is possibly due to the very small sample sizes that are vulnerable to skew from extreme cases.

TABLE 50: BOYS LEARNING SCORES BY GRADE

Assessment	Mean numeracy	Mean Somali	Mean English	Mean Fin. Lit.	n
Out of school	31.5	42.6	16.7	19.2	48
Grade 3	20.5	71.3	38.7	28.6	8
Grade 4	25.5	50.9	15.8	23.8	13
Grade 5	25.9	52.0	22.1	24.3	37
Grade 6	68.2	62.8	31.8	34.1	44
Grade 7	66.0	61.4	31.7	26.9	54
Grade 8	64.8	59.1	29.1	25.2	40

FIGURE 16: BOYS LEARNING SCORES BY GRADE



We looked at program impact among boys by contrasting results from comparison schools to those in intervention schools. Despite SOMGEP-T's focus on girls, it is possible the boys in and around intervention schools would have been exposed to some of the program activities and as such could exhibit improvement in learning outcomes. However, no such impact can be detected as we look at the results, presented in the table below. While none of the measures are statistically significant, i.e. we cannot say for sure that the boys in interventions schools are learning faster than boys in comparison schools, the boys are improving overall more in the intervention schools than in the comparison ones when it comes to Somali, English and numeracy scores.

Interestingly, as it pertains to financial literacy the trend observed with girls in the preceding section seems to be reverse for boys in intervention schools. Where cohort girls in intervention schools are improving their financial literacy scores significantly more than their comparison school counterparts, the reverse seems to be true for boys. One possible explanation is that given the financial literacy activity in the Girls' Empowerment Forum, boys in intervention schools shy away from financial literacy as it becomes associated with girls.

TABLE 51: BOYS' LEARNING SCORES AT MIDLINE ROUND 2 – PROGRAM IMPACT

	Intervention			Comparison			Difference in Difference	P value
Assessment	BL	ML	Difference	BL	ML	Difference		
Numeracy	41.8	61.4	19.6	41.7	58.2	16.5	3.0	0.615

Somali literacy	43.4	58.8	15.4	41.8	51.2	9.4	6.0	0.421
English literacy	10.9	29.5	18.6	6.6	20.5	13.9	4.7	0.382
Financial literacy	13.6	28.8	15.1	3.6	22.1	18.5	-3.3	0.662

When comparing the learning levels of girls and boys in the interventions schools we can see that a gap continues to be very wide in numeracy where the girls at Baseline scored an average of 33.2. They have increased their score in the past two years by 19 points, to 52.2 percent. However, the boys at the same time have increased their average score by slightly more, leaving their average at Midline Round 2 at 61.4 which is nearly 10 percentage points higher than the girls.

The converse is true for Somali literacy where the gap between boys and girls has nearly disappeared. A similar trend can be observed when it comes to financial literacy as the initial gap of 9.4. percent between boys and girls at baseline has been reduced to 5.3 percentage points. The girls are thus catching up to the boys in both Somali literacy and financial literacy. Meanwhile, for English literacy the gap is widening substantially. This is concerning, and in line with what has been observed above in the aggregate learning section of this report. Where girls in fact had a slightly better learning score for English at Baseline, the boys are now in a nearly 9 percent point lead. The difference, moreover, is statistically significant. This is thus slightly concerning from a program impact perspective as we near the end of the project.

One possible explanation to the improved English results for boys derives from the way in which the boys have been selected for the intervention. Having been selected through the households of cohort girls who are part of the program, the boys have perhaps also experienced a boost in their education given that their households are now involved in a program that places a high importance on education. This boost might then be reflected in their improvement in English literacy outcomes. In the Somali context boys are more likely to continue onto further education which is often in English. It is possible that once education becomes more of a priority, boys are being prepared for further education and the need for mastering the English language becomes more pertinent. Meanwhile, the same boost has not contributed to improvements in financial literacy levels as this is being taught more in the GEFs. Similarly it could be argued that for cultural reasons, girls' education can take a more practical orientation that would prepare them to work in the market and in small shops where women are disproportionately represented compared to other sectors of the economy. To work in the market requires precisely the financial literacy skills that the girls are making strides with.

Note from the project: Regarding boys' English skills, it is more likely that boys leave the village more often than girls, as well as potentially learning some English in preparation for migrating to urban areas/ overseas. While this is speculative – as actual data on migration is limited and 'illegal migration' is underreported by households – anecdotal information indicates that boys migrate further away than girls. It is also worth noting that this difference may be related to expectations of moving into tertiary education (mostly offered in English). Among those enrolled in tertiary education across Somalia, only 31% are girls (Education Sector Analysis, p.299). While the majority of the SOMGEP-T students are unlikely to reach higher education due to the costs and limitations involved (including the fact that service provision is mostly limited to urban areas), the gendered enrolment pattern at this level should also be noted.

TABLE 52: COMPARISON OF BOYS' AND GIRLS' LEARNING SCORES IN INTERVENTION SCHOOLS

	Girls			Boys			Difference in Difference	P value
Assessment	BL	ML	Difference	BL	ML	Difference		
Numeracy	33.2	52.2	19.0	41.8	61.4	19.6	-0.6	0.861
Somali literacy	37.4	56.7	19.3	43.4	58.8	15.4	3.9	0.376
English literacy	11.2	20.8	9.6	10.9	29.5	18.6	-9.0	0.009
Financial literacy	4.2	23.5	19.3	13.6	28.8	15.1	4.1	0.251

Finally, to provide a point of contrast, we looked at girl's learning relative to the boys in comparison schools. Here we can observe, interestingly, that in comparison schools the trend for financial literacy is reversed – boys are in fact increasing the gap in financial literacy levels. This finding strongly suggests the impact of the program's financial literacy training. The program has three primary activities targeting financial literacy, all of which are oriented around direct training of adolescents: training through the GEFs; training through their boys' counterparts, the BEFs; and training through participation in ALP centres.67 The fact that – in intervention communities – girls outstrip boys in terms of gains in financial literacy could be attributable to the fact that GEFs have been in existence for longer and likely have higher participation rates. The typical girl in our sample is more likely to be in an empowerment forum than is a typical boy in our sample, and therefore more likely to receive financial literacy training. The fact that girls have improved their financial literacy more quickly than boys – while the opposite holds true in comparison schools – suggests that the program's activities are directly responsible for the improvements.

In most other ways, the results that compare boys and girls in comparison schools are similar to those for intervention schools. In numeracy the gap is rather wide and remains that way. For English literacy the boys are improving more and in Somali literacy the girls are gaining, just like in the intervention group. The key point to take home here is that the program impact in financial literacy, observed in the preceding sections is supported by the findings here. This effect seems to be taking place only among the program girls. The boys in intervention schools are improving their financial literacy scores less than those in comparison schools. Meanwhile, the girls in intervention schools not only have improved their scores more relative to the boys in their schools but also relative to the girls in the comparison schools. The program's activities on financial literacy, thus, seems to be having an impact among the female beneficiaries.

TABLE 53: COMPARISON OF BOYS' AND GIRLS' LEARNING SCORES IN INTERVENTION SCHOOLS

	Girls			Boys			Difference in Difference	P value
Assessment	BL	ML	Difference	BL	ML	Difference		

[67] Notably, ALP girls are not included in this analysis, so the trends observed stem exclusively from activities administered through the GEFs and BEFs.

Numeracy	27.8	43.3	15.5	41.7	58.2	16.5	-1.1	0.768
Somali literacy	31.0	49.2	18.2	41.8	51.2	9.4	8.7	0.068
English literacy	6.4	15.0	8.6	6.6	20.5	13.9	-5.3	0.097
Financial literacy	4.7	15.6	10.9	3.6	22.1	18.5	-7.6	0.063

4.7 Foundational Skill Gaps

The discussion of learning scores thus far has largely focused on establishing the extent of program impact on learning, either in the form of the core intervention at primary schools or in the form of the ALP centres. In this section, we largely move past estimates of changes in learning scores, and focus instead on identifying patterns in learning scores that may prove useful for adjusting programming.

This section analyses subtask-specific outcomes for girls enrolled in school at the time of midline data collection. We place girls into one of four categories based on their performance on a given subtask:

- Non-learners (0%) – these girls are entirely unable to perform the skills captured in a subtask
- Emergent learners (1-40%) – these girls may have a basic grasp of the skills required, but are unable to apply those skills widely
- Established learners (41-80%) – these girls understand the skills required but cannot apply those skills in all cases or to the more complex questions in a given subtask
- Proficient learners (81-100%) – these girls have achieved relative mastery of a particular skill

The idea underlying this analysis is to identify patterns in girls' performance that reveal something systematic about the ways in which they do or do not learn. For instance, if girls tend to underperform on word-based mathematics problems but can perform the arithmetic operations that underlie those word problems, it suggests that they understand the mathematical principles, but lack the ability to practically apply mathematical skills to real-world problems. The analysis can also reveal the levels at which most girls begin to struggle, splits in the sample between girls who perform very well and those who cannot complete a subtask at all, and so forth. These insights can help guide programming insofar as they provide evidence of where teachers may need to focus their efforts, or where developing specific pedagogical tools targeting a given skill (e.g., 2-digit arithmetic) may be helpful for overcoming roadblocks in learning.

We begin by focusing on girls who were enrolled in formal school at the time of midline data collection. We limit the sample to exclude girls who were enrolled in ALP or ABE programming, as we wish to consider the foundational skills gaps observed for ALP girls separately, below. In the table below, we report subtask-specific performance for the 11 numeracy subtasks, reporting the share of girls who fall into the four performance categories on each subtask. The top panel reports results for enrolled girls who reside in intervention communities, while the bottom panel of the table reports the same analysis for the full sample of enrolled girls (i.e. including both intervention and comparison communities). The purpose of reporting analysis using the aggregate sample is to maximize the sample size available, and to study whether inclusion of comparison girls alter our conclusions in any way.

The first finding that stands out in the table below concerns simple addition and subtraction problems, captured in subtasks 2 and 3. The majority of girls have achieved proficiency in these skills and, to the extent that some girls have not mastered the skills entirely, many fall into the "established learner" range. This seems to suggest that basic addition and subtraction are skills that almost all girls enrolled in primary school have gained. What is less clear from the table is whether the girls who could not perform these tasks have fallen behind their peers while being enrolled in school, or if they have recently enrolled in school and that explains their lack of facility with addition and subtraction. In fact, of the girls who fall into the non-learner category for subtasks 2 and 3, many were not enrolled at the baseline and have since enrolled in school. For instance, 64.3 percent of the non-learners in subtask 2 were out-of-school at the baseline, suggesting that girls who have remained in school more consistently have achieved proficiency at these tasks beyond what the results below suggest.

TABLE 54: FOUNDATIONAL NUMERACY SKILLS AND SUBTASKS

Subtask	1	2	3	4	5	6	7	8	9	10	11
Skill Assessed	Missing number	Addition (Level 1)	Subtraction (Level 2)	Addition (Level 2)	Subtraction (Level 2)	Word problems (add/sub)	Multiplic. (Level 1)	Multiplic. (Level 2)	Division (Level 1)	Division (Level 2)	World problems (mult/div)
Intervention Schools											
Non-Learner (0%)	1.6%	2.5%	7.8%	10.4%	20.8%	11.8%	20%	63.6%	37.8%	74%	49.3%
Emergent Learner (1-40%)	23.5%	1.4%	1.8%	7.2%	5.5%	6.7%	4.9%	7.1%	8.1%	6.4%	0%
Established Learner (41-80%)	42.2%	7.1%	5.8%	18.6%	15.4%	30.6%	14.1%	12.4%	18.2%	8.5%	15.4%
Proficient Learner (81-100%)	32.7%	89%	84.6%	63.8%	58.3%	50.9%	61%	17%	35.9%	11.1%	35.3%
All Schools											
Non-Learner (0%)	2.7%	3.8%	8.8%	13.9%	23.7%	14.1%	23.5%	65.7%	41.5%	75.9%	51.3%
Emergent Learner (1-40%)	23.4%	1.6%	2.4%	7.7%	5.9%	6.1%	4.8%	6.7%	7.3%	6.3%	0%
Established Learner (41-80%)	42.9%	7.9%	7.7%	18.9%	17.1%	30.9%	14.8%	12.2%	17.7%	8.2%	14.8%
Proficient Learner (81-100%)	31.1%	86.7%	81.2%	59.5%	53.4%	48.9%	56.9%	15.4%	33.5%	9.6%	33.9%

A second finding that emerges from the table is the fact that many girls show fluency in 1-digit addition and subtraction, but are less adept at 2-digit operations. This is not surprising, given the added complexity of 2-digit operations, which require understanding of specific rules, such as carrying digits from one "column" of a problem to the next. The finding shows, however, that a large proportion of the girls are not acquiring basic skills at the expected level for their grade.

Perhaps more surprising is the relatively small decline in proficiency that we observe from 1- to 2-digit operations of this kind. In addition, much of the decline in the share of girls who are proficient (scoring above 80 percent on the subtask) is driven by girls moving from proficiency on 1-digit operations to "established learner" status on 2-digit tasks. That is, many girls who are not proficient at these more difficult tasks are still able to complete them – and therefore understand the rules involved – but have not yet mastered these skills enough to consistently achieve true proficiency. Girls in this category are likely to require additional practice, rather than a more fundamental grounding in the rules and processes for 2-digit operations. Identifying girls who understand the procedures but cannot universally apply them, versus those who do not yet grasp the procedures for 2-digit operations is a key place where the use of formative assessments and tailored lesson design – or remedial instruction after school – could pay large dividends in terms of student learning.[68]

The third finding that emerges with regard to numeracy concerns the shift from 1- to 2-digit multiplication, from subtask 7 to 8. Here there is, as we would expect, a large decline in the number of girls who achieve proficiency. This is in line with earlier findings regarding 1- and 2-digit operations in this report, in previous SOMGEP-T evaluation reports, and even in other GEC-T evaluations in Somalia. What separates multiplication from addition and subtraction, however, is in the scale of the jump in non-learners from subtask 7 to 8. Where the share of non-learners rose between 7.9 and 13 points for addition and subtraction, respectively, the share of non-learners in multiplication rose from 20.0 to 63.6 percent, an objectively large increase.

Finally, we also note a marked decline in performance between subtasks 7 and 9, which cover 1-digit multiplication and division. Whereas girls' performance on addition and subtraction tended to track each other closely – a natural occurrence, since these operations are the inverse of one another – this appears less true for multiplication and division, where there is a marked decline in performance from the former to the latter.

In contrast to numeracy, we observe fewer sharp breaks in subtask-specific performance in the context of Somali literacy. The table below reports subtask-specific competency levels for the nine Somali literacy subtasks, among the same sample of girls enrolled in school at the time of midline data collection. While the numeracy results showed up- and down-swings in scores over the course of the test, in Somali literacy there is generally a more gradual movement from easier to harder subtasks (or, more precisely, from subtasks on which high numbers of girls achieve proficiency to subtasks on which fewer girls achieve proficiency).

[68] This point applies equally to the findings for 1-digit addition and subtraction, as the results show that very few girls fall into the middle range of performance on these tasks. When it comes to 1-digit operations, girls are generally either able to complete the tasks with proficiency or they are not able to complete them at all. Determining which girls fall into which categories and targeting those who are unable to complete these tasks at all for remedial learning should be a key goal of mathematics teachers in the program.

TABLE 55: FOUNDATIONAL SOMALI LITERACY SKILLS AND SUBTASKS

Subtask	1	2	3	4	5	6	7	8	9
Skill Assessed	Reading Words	Reading Comp (easy)	Reading Comp (medium)	Reading Fluency	Reading Comp (difficult)	Writing (fill blank)	Writing (negative form)	Writing (future tense)	Sentence completion
Intervention Schools									
Non-Learner (0%)	8%	13.8%	19.8%	12.2%	22.4%	22.1%	32.5%	39.8%	50%
Emergent Learner (1-40%)	7.6%	4.4%	6.7%	12.4%	9.9%	10.4%	4.6%	4.6%	6.9%
Established Learner (41-80%)	15.5%	26.1%	30.9%	30.6%	39.2%	19.3%	7.8%	8.3%	13.1%
Proficient Learner (81-100%)	68.9%	55.7%	42.6%	44.9%	28.4%	48.2%	55.1%	47.3%	30%
All Schools									
Non-Learner (0%)	8.8%	15.4%	21.4%	13.1%	24.4%	25.9%	33.8%	41.8%	51.3%
Emergent Learner (1-40%)	8.2%	4.4%	6.9%	13.9%	9.8%	9.9%	4.7%	3.7%	7.8%
Established Learner (41-80%)	17.2%	26.7%	32.2%	31.7%	38.1%	18.2%	7.6%	8.1%	12.2%
Proficient Learner (81-100%)	65.9%	53.5%	39.4%	41.3%	27.8%	46%	53.8%	46.4%	28.7%

Indeed, even those breaks that occur may be explained in part by the design of the assessment, rather than a fundamental divergence in proficiency in a tight window. Specifically, there is a substantial decline in performance from subtask 4 to 5, in terms of the share of girls who are considered proficient, from 44.9 to 28.4 percent. Scoring for subtask 4 results in a relatively continuous measure from 0 to the maximum number of words read in the passage (480) during a one-minute period, but capped a maximum score, in words per minute, of 100. Therefore, girls can achieve a score of any integer from 0 to 100, inclusive. Subtask 5 is a reading comprehension assessment, based on a passage the student read during subtask 4. Ssubtask 5 has just four reading comprehension questions, meaning that girls can exclusively achieve scores of 0, 25, 50, 75, or 100 percent. In practice, this means that a girl must answer all four reading comprehension questions correctly in order to be classified as proficient. If we include girls who scored 75 percent in the proficient category, 49.5 percent of girls achieve proficiency on this subtask – in line with trends in the subtasks that surround it.

A slightly more subtle trend that emerges in the Somali literacy results concerns subtasks 7 and 8, where girls begin to diverge into two sharply-distinguished groups. In the subtasks prior, a considerable number of girls fell into the middle ranges of performance – emerging and established learners. For instance, 29.7 percent of girls fell into these two middle categories on subtask 6, and 49.1 percent fell into these categories on subtask 5. In contrast, subtasks 7 and 8 had just 12.4 and 12.9 percent of girls in these middle ranges, respectively. These two subtasks are characterized by the most bimodal distribution observed in the Somali literacy assessment – the decline in the middle range of the distribution is not coupled with an overall sharp decline in performance, but in a bifurcation of the sample into a group of girls who are able to consistently perform the skills tested, and a group of girls who are entirely unable to perform the same skills.

The content of subtasks 7 and 8 are consistent with this idea, as they ask students to convert a sentence to its negative form and to the future tense, respectively. These are tasks that require knowledge of specific grammar rules but – once a student knows the basic rules – they can successfully complete many or most of the possible cases they might encounter in simple sentences. Therefore, seeing a sharp divergence in performance is not surprising, as girls either know the rules in question and how to apply them in many simple situations, or they do not know them and are unable to apply them to any sentences.

In general, there are no tangible differences between the results among intervention schools, exclusively, and the combined sample reported in the bottom panel of the table. This applies equally to the numeracy and Somali literacy results presented thus far, as well as to the English literacy results which we discuss below. The main finding that emerges between the intervention-only and combined samples is that girls in the intervention group perform slightly better than their comparison-area counterparts. However, this is not a finding specific to skill gaps, but to aggregate learning scores.

In the table below, we report results for English literacy. These results provide possibly the richest set of findings considered thus far. First, and most basically, a substantial share of girls lack basic English language skills. Just 30.7 percent of girls are able to recognize 80 percent or more of the 100 English letters presented to them. This result is even starker, for two reasons: first, many of the letters were repeated, so correctly identifying 10 letters may only mean that a girl can correctly identify 4-5 distinct letters; second, while girls may be unfamiliar with the upper-case or capitalized form of many letters, lower-case and upper-case letters were mixed in the assessment, and the median girl was only able to identify 42 of the 100 letters.

TABLE 56: FOUNDATIONAL ENGLISH LITERACY SKILLS AND SUBTASKS

Subtask	1	2	3	4	5	6	7	8	9
Skill Assessed	Letter Identification	Word Recognition	Reading Comp (easy)	Reading Fluency (medium)	Reading Comp (medium)	Reading Comp (difficult)	Writing (fill blank)	Writing (negative form)	Writing (future tense)
Intervention Schools									
Non-Learner (0%)	29.5%	35.7%	53.2%	50.9%	70%	77.6%	76.9%	85%	86.2%
Emergent Learner (1-40%)	18%	17.7%	7.1%	13.3%	7.6%	5.1%	4.8%	0%	0%
Established Learner (41-80%)	21.7%	20.1%	21.9%	17.1%	17.8%	11.8%	10.4%	4.1%	2.8%
Proficient Learner (81-100%)	30.7%	26.5%	17.8%	18.7%	4.6%	5.5%	8%	11%	11%
All Schools									
Non-Learner (0%)	32%	38.5%	58.1%	53.1%	73.1%	79.1%	78.3%	84.4%	85.7%
Emergent Learner (1-40%)	19.6%	18.5%	6.3%	13.9%	6.9%	4.9%	4.6%	0%	0%
Established Learner (41-80%)	21.9%	20.5%	19.3%	15.9%	15.1%	10.9%	9.9%	4.2%	3.6%
Proficient Learner (81-100%)	26.5%	22.5%	16.3%	17.2%	4.9%	5%	7.2%	11.3%	10.7%

Second, there is not a sharp decline in performance across subtasks 1, 2, and 3. This is surprising, insofar as the three subtasks require fundamentally different, and sharply increasing in difficulty, skills. Subtask 1 requires girls to recognize individual letters, while subtask 2 requires girls to identify entire words, a qualitative shift in difficulty. Subtask 3, extending this trend, requires reading complete sentences. While average performance on subtask 1 is poorer than we would expect, the fact that girls perform almost as well on subtask 2 is equally – if not more – surprising. It may also reflect a polarisation in the distribution of skills, with a subgroup of girls having acquired basic English skills – and therefore proceeding from task to task at a similar performance level – while another subgroup lacks basic decoding skills in English, or reading at a very slow pace and compromising comprehension.

Similarly, the gap in performance between subtask 2 and subtask 3 is not large. We find this initially surprising, though we note that some of the reading comprehension questions posed in subtask 3 are fairly simple, and can be answered correctly with minimal comprehension of the story read.[69] In light of this fact, the number of girls (53.2 percent) who were unable to answer any comprehension questions successfully is disconcerting.

Third, we observe a sharp decline in performance from subtask 4 to subtasks 5 and 6. The former subtask is a reading fluency assessment, in which the girl reads for one minute and is rated on the total number of words she read correctly (capped at 100); the latter subtasks assess reading comprehension, with subtask 5 based on the non-timed reading of the text used in subtask 4. In practice, the gap in performance between these subtasks is driven in part by the same problem that plagued the Somali literacy assessment: changes in scoring scales from subtask to subtask that reduce comparability. Whereas subtask 4 is scored on a continuous scale, subtasks 5 and 6 allow scores of just 0, 25, 50, 75, and 100 percent. If we include performance of 75 percent in the proficiency category, the share of girls achieving proficiency in both subtasks 5 and 6 increases to 10.1 percent.

Finally, we see a movement toward bifurcation into two groups – proficient and non-learners, respectively – in subtasks 8 and 9. In these two subtasks, the middle categories are hollowed out entirely with just 4.1 and 2.8 percent of girls falling in these middle categories. The substantive explanation for this trend is similar to that mentioned in the context of Somali literacy, as these subtasks ask girls to convert a sentence to the negative form and to future tense. These skills are simple to use with moderate proficiency once they are known, but girls are generally either able to apply the rules in most basic cases, or do not know the rules at all. The methodological explanation for this bifurcation is that there are only two test items in each subtask, meaning that girls can *only* achieve scores of 0, 50, and 100 percent, entirely eliminating the opportunity to fall into the "emergent learner (1-40 percent) category.

Skill Performance by Grade

One issue with the analysis above is that it aggregates across a disparate group of girls when reporting gaps in skills. While we have limited the analysis to girls enrolled in school at the time of data collection, these girls are spread across different grade levels. Some of the skills assessed may be uncommon for girls in Grades 2 or 3, but be readily possible for girls in Grade 6, for instance. In the set of graphs below, we adjust our approach to account explicitly for differential grade levels of girls in the sample, identifying the set of subtasks where different grade levels excel or struggle. Of course, the graphs below simplify the data in a different way, by reporting the mean score on each subtask, rather than categorizing learners into four different levels.

[69] For instance, one question asks the name of Fadumo's friend, and the correct answer is Hamda. But Hamda is the only other proper name used in the story, so it may not require true comprehension to successfully answer this question.

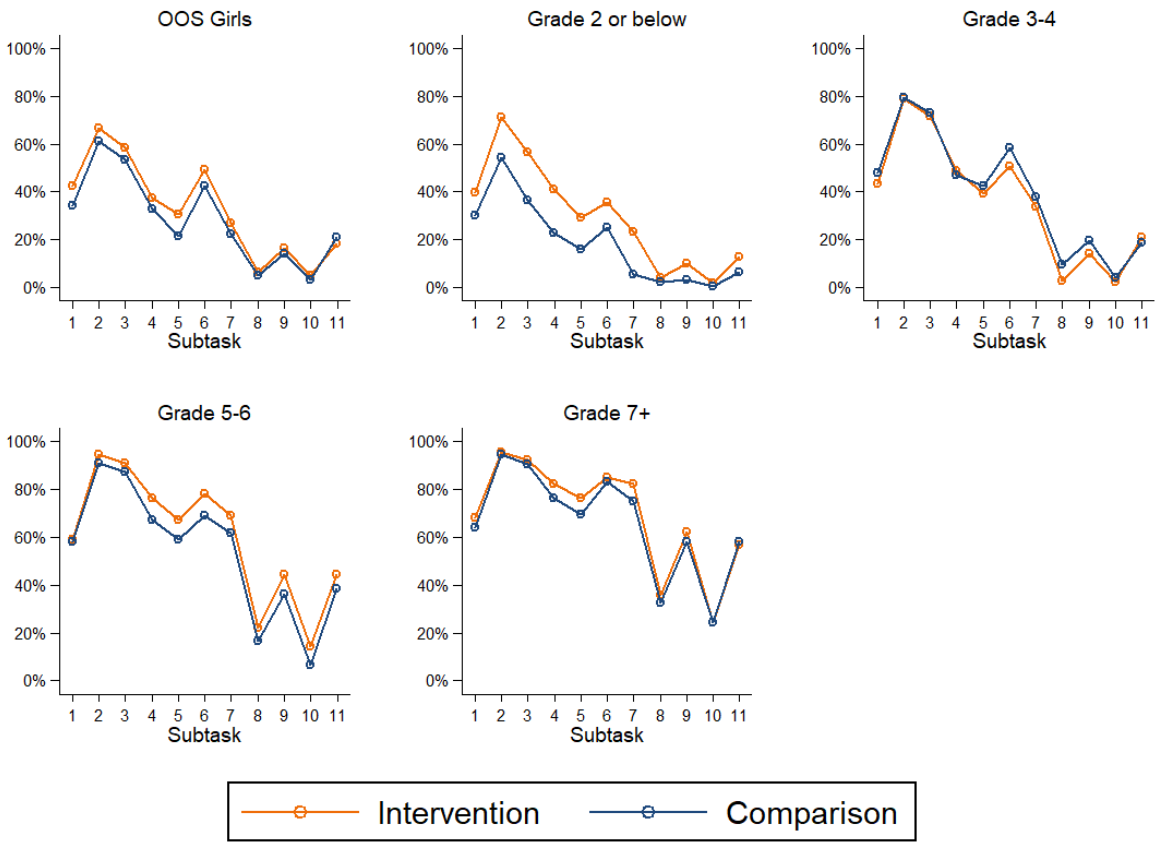
This tradeoff is necessary in order to make clear the skills girls in different grades have achieved, on average, even if use of the mean obscures some degree of nuance.

The graphs below report mean scores for numeracy subtasks, broken down by grade level and intervention versus comparison schools. In most cases, patterns in intervention and comparison schools mirror one another, so we only comment on areas where there are interesting differences.

One interesting pattern that emerges is between OOS girls and those who are enrolled in the lowest grades, primarily grades 1 and 2. Both groups of girls achieve moderate scores identifying simple number patterns and completing 1-digit addition and subtraction, though girls in the lowest grades show a sharper decline from 1-digit addition to 1-digit subtraction than do OOS girls. This may be due to OOS girls' prior exposure to education or because OOS girls are older, on average, and have learned addition and subtraction skills through practical use.

OOS girls also outperform girls in grade 2 and below in terms of addition-based word problems and 1-digit division. While both groups see an improvement in their scores on subtask 6 (addition and subtraction word problem) and subtask 9 (1-digit division) relative to surrounding subtasks, these improvements are much stronger in OOS girls. This suggests that OOS girls have either mastered addition and subtraction to a degree that early-grade girls have not, or that they are used to applying addition and subtraction skills in applied scenarios, such as those encountered during a word problem.

FIGURE 17: MEAN NUMERACY SUBTASK SCORES, BY GRADE LEVEL



Moving up in grade level, it is clear that girls' performance on a subtask-by-subtask basis improves as they advance in grade. By the time girls reach grade 5 or 6, they have achieved mastery of 1-digit addition and 1-digit subtraction to a degree that their younger counterparts have not. The drop in performance from 1-digit to 2-digit addition and subtraction is something we have noted previously in our discussion of skill gaps. But this decline remains even in grade 7 and 8 girls, shown in the lowest-right panel of the figure; even in this group of girls, performance on 2-digit subtraction is 18.2 percentage points lower than performance on 1-digit subtraction (73.4 percent versus 91.7 percent). While some decline is a natural result of the slightly higher complexity of the questions, the fact that even girls who will soon graduate primary school – and, indeed, some girls who *have* graduated primary school – are unable to reliably complete 2-digit subtraction is problematic.

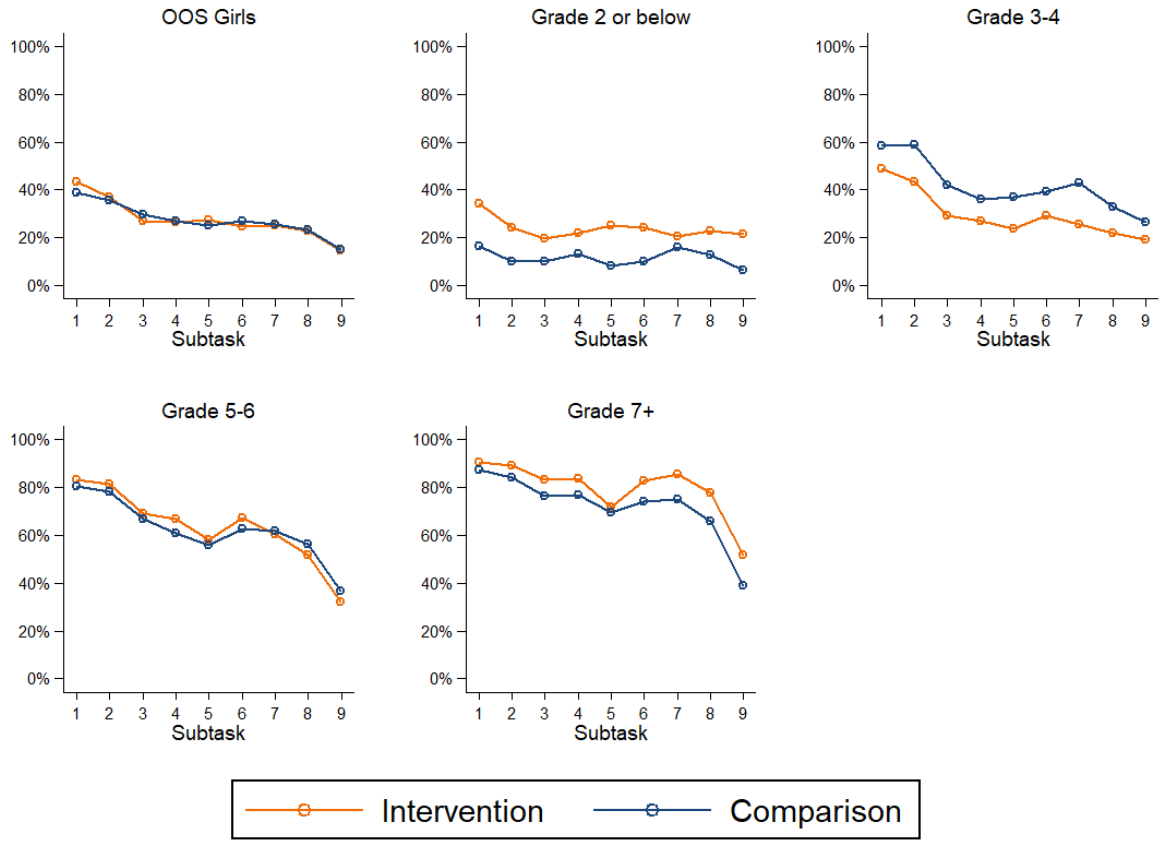
Girls in this age range also still exhibit significant declines in performance when moving from 1- to 2-digit multiplication and division. The data show that girls in grade 3-4 start to see improvements in 1-digit multiplication, and these improvements continue among grade 5-6 girls and are especially pronounced among girls in grade 7 and above. However, even for this group of girls who are the most advanced in terms of grade level, 2-digit multiplication and division are still substantially more difficult than the same operations in less complex, 1-digit applications.

In contrast to numeracy, where there are a number of skills – such as basic addition – that even younger and OOS girls have gained in moderate-to-large numbers, this is much less true for literacy in Somali. Our results for numeracy also showed sharp distinctions in performance across subtasks, among every grade cohort of girls. This is also fundamentally different in the context of Somali literacy, with a few small exceptions.

The first two panels in the figure below have a flat profile, with little differentiation in mean scores from subtask to subtask. Put bluntly, this is an unusual result, as we naturally expect performance to decline as subtasks increase in difficulty over the course of the assessment. We observe this trend in the aggregate, where mean scores decline from 54.4 percent on subtask 2 to 25.1 percent on subtask 9.[70] In contrast, among in-school girls enrolled in grade 2 or below, the mean score on subtask 2 is 18.3 percent, while the mean score on subtask 9 is 15.3 percent. This unusual result is explained by a small cohort of girls, around ten in total, who performed well on all Somali literacy subtasks, achieving total scores of 90 percent or higher. Among this cohort, there was no meaningful decline in performance from subtask 1 to subtask 9. Among the other girls in this grade level, there was also no large-scale decline in performance, but these girls achieved very low scores across the board, ranging from 6.2 percent to a high of 19.9 percent on a subtask-specific basis. In essence, the set of girls in grade 2 and below are relatively neatly divided into two groups: those who can read in Somali well, who perform well on all subtasks, and those who cannot read in Somali at all, and perform poorly in terms of even relatively basic skills.

[70] By aggregate here we are referring to the entire set of girls interviewed during midline round 2, including ABE girls, ALP girls, girls enrolled in school, and OOS girls.

FIGURE 18: MEAN SOMALI LITERACY SUBTASK SCORES, BY GRADE LEVEL



It is important to note that the "flat" nature of subtask scores cannot be explained entirely by this split into two groups, as the pattern is observed more weakly in other grade cohorts. For instance, among girls in grade 7 and above, there is not a significant decline in subtask scores until the last subtask of the assessment is reached, as shown in the lower-right panel of the graph.

We also urge caution in interpreting the decline in performance in grades 5 and upward on subtask 9 as especially meaningful regarding a lack of fundamental literacy skills. The reason for this caution is that this subtask in Somali literacy asks girls to fill in missing words in a sentence, such that the sentence – and the broader story in which it appears – makes sense. However, girls were not given a list of possible words from which to select, which expands the number of possible approaches girls could take. Moreover, judgment of whether the girl successfully completed the task is subjective, in the sense that enumerators were not given an unambiguous statement of what constituted the correct answer, and had to judge whether the story the girl constructed by filling in words was correct.[71]

Given our general concern regarding subtask 9, the pattern of subtask scores – excluding subtask 9 – across all grade levels are consistently flat, with small deviations. Our suggested interpretation of this finding is that literacy is inherently less subject to wild swings in performance across subtasks, when compared to numeracy.

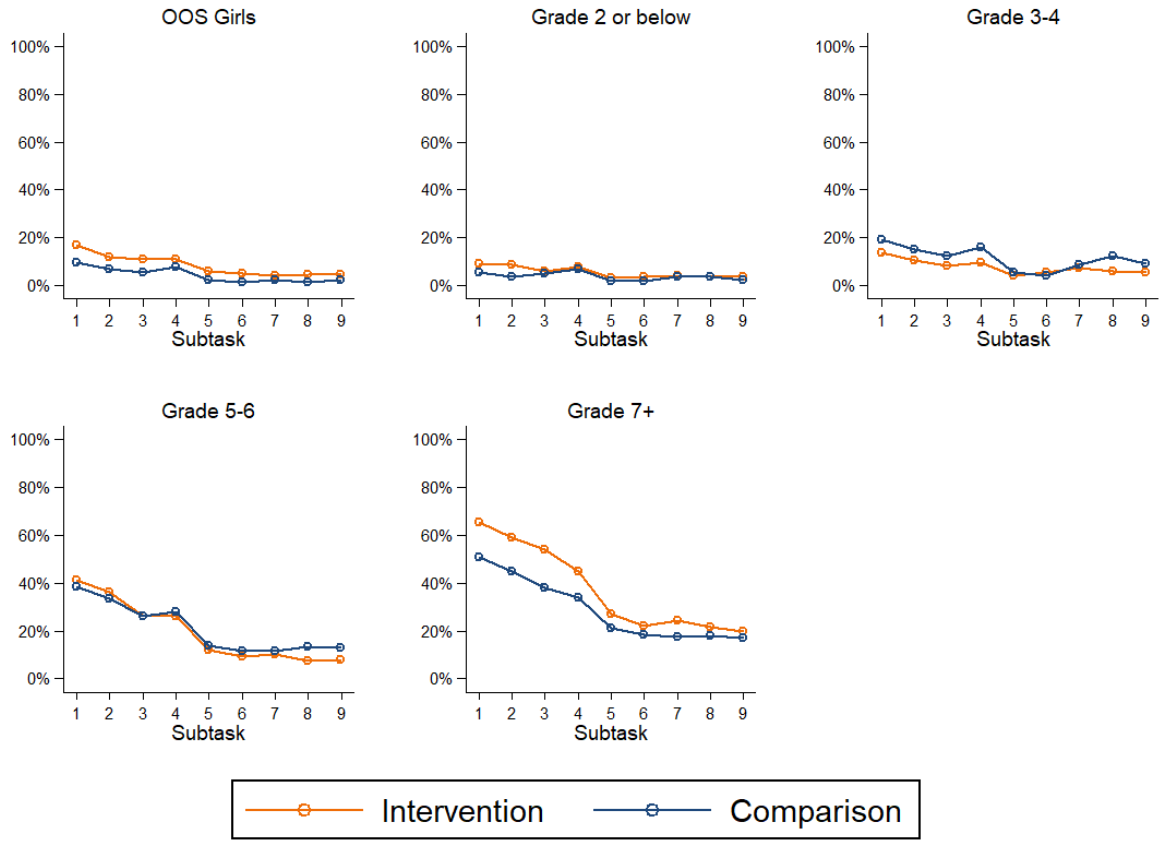
[71] This issue was raised during the training session for field staff, and was discussed extensively. The evaluation team attempted to provide clarity on what *types* of answers should be accepted, but it is unclear whether all enumerators understood these instructions and it is unlikely that they all approached grading for this subtask in the same way.

The theoretical underpinning of this argument is that mathematics is highly rule-based, where understanding of a given rule (i.e. the method for solving a problem) allows one to excel on a specific set of problems, but may not transfer readily to other problems. For instance, knowing how to carry digits in addition does not make it more likely that one will intuitively understand how to handle remainders in division. However, once one knows how to handle remainders, they can solve most division problems with a high degree of proficiency.

In contrast, reading and reading comprehension are less subject to such rules. Girls with a limited fluency in Somali reading can still understand basic stories and ascertain – based on their limited understanding and contextual clues – the correct answer to a few reading comprehension questions. They will not score very highly on reading comprehension, but they will "muddle through" at a low-moderate level of performance in many cases. Moreover, knowing a few rules does not improve literacy performance very much: while understanding how to convert a sentence to negative or apply a specific tense will help a child write much more clearly, a *lack* of this skill will not necessarily inhibit a child from reading short stories and intuiting their meaning. In short, if a child is able to identify Somali words and read *some* short sentences, they can often achieve moderate scores on reading comprehension and other tasks, resulting in relatively flat score trajectories across the range of subtasks on the SOMGEP-T learning assessments. This is clearly less true of numeracy, where wider swings in scores from subtask to subtask are observed.

This hypothesis is not necessarily confirmed by the results for English literacy, but these results do not specifically contradict our interpretation either. The figure below reports subtask-specific English literacy scores for each grade level. Among OOS girls and girls in lower grades (4 and below), scores are both extremely flat and extremely low, which is consistent with the fact that English is not taught systematically in early grades of most Somali schools.

FIGURE 19: MEAN ENGLISH LITERACY SUBTASK SCORES, BY GRADE LEVEL



One notable finding is that girls in grade 3 and 4 achieve lower scores on subtasks 5 and 6, which assess reading comprehension, than they do on any other subtasks, including those related to writing. However, the gap in performance across subtasks among this grade level is fairly small.

These same subtasks are the point at which even higher-grade girls begin to see very significant declines in average performance. Among all girls in grades 5 and up (the two graphs in the lowest panel of the figure below), scores are highest in the first subtask and decline steadily – though not necessarily monotonically – through subtask 5. Among these girls, subtask 5 represents a dividing line between varied proficiency on subtasks 1 through 4, and flat, relatively limited proficiency on subtasks 5 through 9.

4.8 WORKING MEMORY

The final learning assessment completed by cohort girls as part of the ML2 evaluation was a simple test of their working or short-term memory.[72] Each of the cohort groups were presented with a set of 19 images, depicting various objects with which they would be familiar from daily life, such as a camel, a bird, an egg, and so forth. Each child was instructed to review the set of images and try to remember as many as they can; enumerators were instructed to limit the review time to no more than 5 minutes, though most children –

[72] We do not take a specific stand on whether the assessment tests working or short-term memory, as there are subtle differences between these two concepts.

based on discussions with team leaders – reviewed the document for 1 to 2 minutes or less. At the conclusion of the review period, the papers with the images were removed, and the child was asked to list the images they saw. Their score was adjusted to a 0-100 scale based on the share of the images they recalled.[73]

Note from the project:

The working memory test is designed to mirror classic working memory span tasks developed by Alan Baddeley in the 1970's and later on adapted by Nelson Cowan. Those tasks typically involve non-sequential digit, word and sentence recall. Such tests are commonly used in clinical psychology nowadays, using computer-based platforms. A useful snapshot of common phonological span working memory tasks can be found in this review by Cabbage et al. (2017) (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5608376/).

Due to the literacy and numeracy levels of respondents, the test was designed to use images instead of word or digit lists, thus eliminating the risk of a potential misinterpretation (for instance, not knowing if a child is unable to recall numbers due to lack of numeracy skills or limited working memory). The use of nonwords, while possible, might pose a challenge in a multilingual setting. Therefore, this test was designed as a version of old adaptations of clinical working memory batteries, which used image cards instead of words and digits.

Classic working memory tests generally include three components – phonological span, visuospatial tasks, and central executive tasks, thus testing the three components of Baddeley's working memory model (phonological loop, visuospatial sketchpad and central executive). The task included in this study assesses phonological span only, as it intends to assess the extent to which working memory affects the acquisition of reading skills.

On average, children remembered 52.6 percent of the images, or approximately 10 of 19 possible images. The most common score was 9 out of 19 images, and scores appear to mirror a normal distribution closely, as the figure below shows. Surprisingly, 48 girls were able to recall all 19 images: our suspicion is that these girls employed some form of mnemonic device to aid their memory.[74]

The figure below plots memory scores for three groups of girls in this evaluation round:

- In-school girls
- Out-of-school girls and ABE girls
- ALP girls

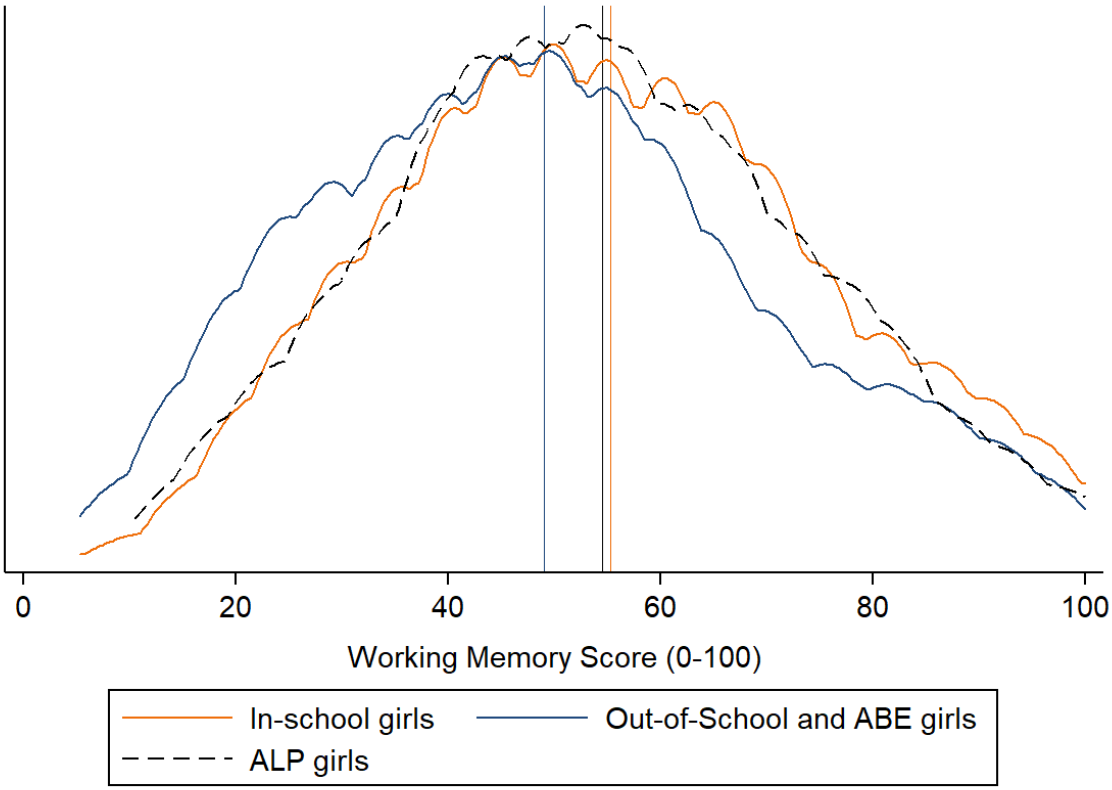
We opted to combine OOS and ABE girls into a single category for the sake of visualizing the data more clearly, and because these groups have relatively similar experiences of education, on average. ABE girls recruited at midline were generally not enrolled in school prior to their enrolment in an ABE program, and had relatively minimal exposure – at the time of data collection – to the program. Therefore, they are similar

[73] The mean scores for working memory we report here are useful for understanding variation within the sample, and linkages between demographic characteristics, working memory, and other learning outcomes. They will also be useful for understanding the impact of the program on working memory at endline. However, there are no international benchmarks against which we can compare SOMGEP-T girls, because the format of this test (image-based) is different from conceptually similar tests employed in developed countries, which use word recall. CARE had previously used this test successfully in East Timor, but it has not been used widely enough to generate benchmarks or metrics for valid comparison.

[74] The set of enumerators who oversaw memory assessments in which a girl achieved a perfect score suggests that enumerators were implementing the memory test correctly. For instance, many different enumerators administered perfect-score tests, and these tests were not disproportionately clustered at the start of fieldwork, when mistakes are most common.

to OOS girls in that they have very little formal education. The figure compares this composite group of girls to in-school girls (girls who were enrolled in school at baseline) and girls enrolled at ALP centres.[75] The vertical lines represent the mean scores for a given group, plotted over the full score distribution.

FIGURE 20: WORKING MEMORY SCORES, BY COHORT GROUP



As the graph shows, OOS and ABE girls achieve systematically lower scores than in-school and ALP girls. In-school and ALP girls achieve very similar scores, with mean scores of 55.3 and 54.5 percent, respectively. In contrast, OOS girls score just 46.3 percent and ABE girls score 51.2 percent, on average. The gap between OOS girls and ABE girls, on one hand, and all other girls in the sample, is statistically significant at the 1 percent level.[76] Finally, although we do not conduct a full analysis of differences between intervention and comparison schools in terms of average working memory, a cursory analysis suggests that girls in intervention schools – limiting the analysis to girls who were enrolled at the time of midline data collection – score 1.9 points higher than those in comparison schools.

While we do not have strong *a priori* expectations regarding the impact of age and educational attainment on working memory, we would broadly expect memory to improve with age, among adolescents, and with exposure to education. For instance, girls with more exposure to education may have learned strategies for

[75] Note that these categorizations do not reflect girls' current enrolment status; they reflect their enrolment status at the time of their recruitment into the sample (e.g., at baseline for in-school and OOS girls; at the time of the first midline round in the case of ALP girls).

[76] Comparing OOS girls to all in-school and ALP girls, OOS girls score lower with $p < .001$. ABE girls, on the other hand, are statistically different from in-school and ALP girls, but only marginally so ($p = 0.10$).

increasing short-term memory, even if they are implicit; education may also prompt more frequent use of short-term memory for completing schoolwork tasks, which can spur improvements in working memory. Somewhat surprisingly, the table below does not show a particularly meaningful relationship between age and memory. While there is a small uptick in memory scores from age 12-13 to age 14-15, this improvement is very small, and is either offset by a stagnation in later years – among in-school and ALP girls – or a decline in later years – among OOS and ABE girls.

TABLE 57: WORKING MEMORY SCORES AS A FUNCTION OF AGE

Age Group	In-School and ALP Girls	OOS and ABE Girls
12-13 years	53.3% (n = 326)	48.7% (n = 394)
14-15 years	55.6% (n = 390)	51.7% (n = 221)
16-17 years	55.8% (n = 281)	48.6% (n = 117)
18+ years	55.7% (n = 260)	46.1% (n = 118)

In contrast, there is an approximately monotonic increase in working memory as a function of grade level, among girls enrolled in formal schools. Girls in the first two years of primary school achieve a mean score of 45.8 percent, increasing continuously to a mean score of 58.1 percent among girls enrolled in grades 7 and 8. This suggests that, although memory does not improve with maturation alone, exposure to education likely improves working memory.[77]

Boys recruited through the household survey were also administered the working memory assessment. In the aggregate, boys outperform girls in terms of working memory by a notable degree, scoring 57.6 percent, compared to 52.6 percent for girls. This is equivalent to recalling approximately one more image out of 19 possible. However, it is possible that the set of boys interviewed is not precisely comparable to the sample of girls. We do not have any specific reason to suspect this is a problem, beyond the fact that the sample of girls includes ABE and ALP girls who were recruited through their respective learning centres, as opposed to through a random household survey.[78] It is also possible that random sampling variation between the samples of boys and girls could result in systematic underlying differences between the two groups, for instance if the sample included more boys that are older, on average, than the girls interviewed.

To guard against the possibility that the gap between boys and girls is driven by differences in sample composition, we took increasingly restrictive approaches to controlling for possible confounders, in the form of age, enrolment status, grade level, and type of learning centre in which they are enrolled. We also estimated a linear regression which employed a fixed effect or binary variable for each school in the sample, controlling for possible structural differences in memory scores across villages. In all of these models, boys still outperformed girls in terms of working memory by a considerable degree, ranging from 2.2 to 5.2 points.[79] This is not an extremely large difference substantively, but may be worth additional investigation,

[77] In practice, working memory scores are moderately correlated with higher numeracy, Somali literacy, and – to a lesser extent – English literacy scores. However, it is not clear whether this correlation arises because working memory is useful for learning or because working memory and learning scores are both driven by exposure to education.

[78] In-school and OOS girls, in contrast, were recruited through a random household survey at baseline, and should be equivalent to boys.

[79] In each of our regressions, the difference between male and female scores is statistically significant; at its largest, the p-value was 0.054.

as explaining the source of the gap could reveal an unobserved structural barrier to girls' educational performance.

4.9 TESTING THE THEORY OF CHANGE FOR LEARNING

In this section we test the relationship between various intermediate outcome measures and the aggregate learning score.[80] The section is divided into two parts. In the first part we look at individual level variables, such as attendance, self-confidence and attitudes of the caregiver. These are variables that are specific to a single girl (or, at most, her household). The second part looks at intermediate outcome variables that are measured at the school level. These are things that apply to all girls in the school and are not specific to any individual girl. The graphs below present regression coefficients for all the relevant variables. Each regression coefficient in the graph is from a separate regression.

INDIVIDUAL-LEVEL FACTORS

The analysis for individual level variables was conducted in a number of different ways. First, we made a regression model which controls for most important factors that have a bearing on learning.[81] This model was ran using both the cross-sectional sample which contains every cohort girl at Baseline and Midline Round 2 as well as using the panel sample, which contains only those individuals who are part of both waves of data collection, excluding replacement respondents. We then ran the regression model using a fixed effects approach whereby the analysis uses the same panel sample, but it is much more aggressive about controlling for other confounding factors that might explain changes in learning over time.[82] We feel that this is the most rigorous and most justifiable approach for assessing the individual level variables, as it best eliminates rival explanations.

[80]This measure aggregates the three main learning outcomes (English, S omali, numeracy) into one score. This is then adjusted to be on a 0-100 scale.

[81]These include age, grade level, treatment/intervention status, round of data collection (change over time), female-headed household, household hunger, disability status (any disability), and fixed effects for each school.

[82]This approach essentially creates a dummy variable for each girl in the cohort (716 girls). This way we do not have to control for girls' innate characteristics – or things that do not change much over time – anymore, because we are controlling for all her fixed characteristics automatically with the dummy variable. Some things do, however, change over time. Thus, the household hunger variable remains part of this model. This model also drops the school level fixed effects which are automatically captured through the girl's fixed effects. For example, if a school is poorer, the school fixed effect that captures that fact is simply the aggregation of individual-level poverty that is captured by the girl's own fixed effects.

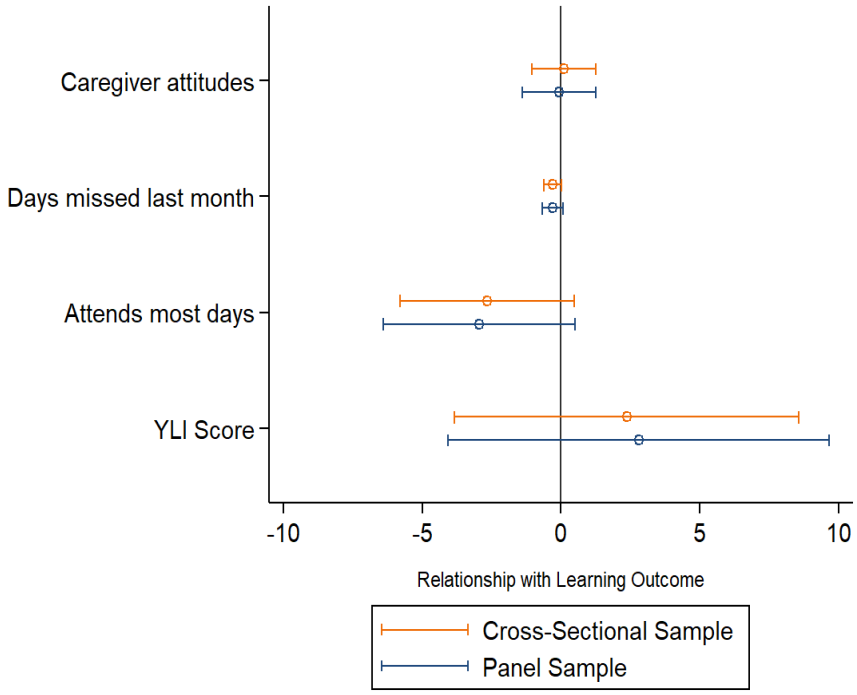
The first measure we focus on captures caregiver's attitudes towards girl's education. The measure aggregates three attitudinal statements: "girls' education is worth investing in even when funds are limited"; "girls are just as likely as boys to use their education"; "I listen to [girl] when making decisions about her education". In line with the ToC, we expect positive answers to these questions to correlate positively with learning outcomes. Indeed, the measure is statistically significant and positively correlated with the girl's learning assessments scores. It seems that the more positive the attitudes of the caregivers, the higher the learning assessments of the girls. While being conscious of any omitted variable bias, it does seem that the logic of the ToC in this regard is supported by the data.

Next, we assess the effect of missing school. The measure looks at the number of days of school missed in the last month. The variable is capped at 10 in order to avoid any bias coming from outliers where girls had missed 20+ days last month. We expect, naturally, that the more school is missed, the lower the learning levels. Again, the measure is statistically significant, and it runs in the expected direction – a girl who misses school is likely to fare worse in the learning assessments than her counterparts. However, the effect, albeit negative, is not substantial.

Meanwhile, looking at the girls who regularly attend there seems to be no clear relationship with learning outcomes as there are both girls who have a lower than average and higher than average learning assessment score in this group. The effect, overall, is negative, but this relationship is not statistically significant. Consequently, it seems that while not attending school can explain some of the negative outcomes in learning, attendance alone is not enough to produce positive learning outcomes.

As has been discussed elsewhere in this report, the YLI score as a measure of self-esteem, leadership skills and life skills in general is positively associated with learning – the higher the YLI score, the higher the aggregate learning results tend to be. However, a note of caution is perhaps in order here. It is not possible for us to determine the directionality of the correlation. In other words, we cannot say whether the attributes behind a higher YLI score have *caused* higher learning scores. It is entirely possible, rather, that positive learning

experiences in fact lead to higher levels of self-confidence and other personal attributes captured by the YLI measure.



When we run the same analysis using a standard regression and various control variables that can change over time – age, grade level, coming from a female-headed household, going without food, having a disability – the results are in line with the above. The relationship between the independent variable and the learning outcomes remains positive for caregiver attitudes and the YLI while remaining negative for both missing school and attending regularly. The results do not vary much between using the panel of the exact same individuals over time (panel sample) or when we use the cross-sectional sample that includes replacement girls introduced as substitutes for those cohort girls who could not be re-contacted at Midlines. Yet, this time, the variation within the sub-samples is much higher, producing limited statistical significance. In fact, none of the measures that were significant using the fixed effects regression analysis above are significant using this approach.[83]

In addition to the individual-level factors reported in the figure above, we also investigated the impact of membership or participation in the Girls Empowerment Forums (GEFs) as a driver of learning scores. This

[83] We have presented the results of the standard controlled regression for the sake of replicability and transparency. However, we feel that the fixed effects regression results are a more valid representation of reality. This is because the fixed effects regression is more able to control for omitted variables. For example, let us consider above the caregiver attitudes toward girls' education. The controlled regression results on the relationship between caregiver attitudes and learning scores could very well be explained by factors that we have not been able to control for. For instance, caregiver attitudes could be very positive because they are themselves well-educated, they have stronger inherent intellectual abilities, and they have passed that to their child. Then their attitudes may not be the cause of higher learning scores. In this case, the fixed effects regression is much better suited for purpose. We are studying the relationship between caregiver attitudes and girls' learning *within the same girl*. The variation in caregiver attitudes that we are using is *within each girl*. Consequently, our approach is able to control for all meaningful and potentially intervening variables insofar as they remain fixed over time.

analysis was motivated by the very strong relationship between GEF participation, on one hand, and YLI scores, on the other. We document this finding in more detail in Section 7.4, which shows that nearly all of the improvements made by the program in terms of girls' self-esteem and life skills are driven by girls who at least occasionally participate in GEF activities.

In light of the relationship between GEF participation and YLI scores, we also investigated whether GEF participation is associated with higher learning scores. The results suggest that participation in a Girls' Empowerment Forum seems to have a dramatic positive effect on learning outcomes. The fixed effects regression shows that a girl who has participated at least once in a GEF activity has, on average, a learning score 16.3 points higher than other girls.[84] As with our discussion of GEF participation and YLI scores, we are cautious regarding the interpretation of this finding. It is possible that highly motivated and higher-performing girls are more likely to join the forums; in other words, we cannot rule out bias stemming from self-selection into the GEFs and a form of reverse causation – that girls who already score well join GEFs.

Although it is not immediately obvious why GEF participation would contribute to higher learning scores, the causal mechanism almost certainly operates through some form of social/support network effects and improvements in self-esteem (again, if self-selection is not the operative mechanism). As noted above, there is a fairly strong correlation between GEF participation, YLI scores, and self-esteem. Our hypothesis is that GEF participation impacts learning scores by providing girls with a stronger support network, encouraging more frequent attendance and greater participation in the classroom. In addition, girls can receive additional help with schoolwork, either through the GEF directly or through friendships made at the GEF.

In practice, there is some evidence for this "social capital" or support network theory. When girls were asked about the role the GEF plays in their school and community, they focused on four main activities:

- Raising awareness of the importance of education within their communities
- Supporting other girls, helping them with schoolwork and encouraging them to attend school
- Seeking help for disadvantaged girls, especially with school fees
- Cleaning or otherwise improving the school

The second point supports a theory of social capital and – outside of awareness-raising – was the most common GEF action cited by girls in our FGDs. For instance, one girl described the activities of the GEF as encouraging girls who are struggling with their education.[85] Other girls made reference to the fact that the GEF provides a way for girls to receive extra help, often citing the GEF in the context of hypothetical stories about girls who were struggling in school, and highlighting the fact that they could receive help from other girls in the GEF.[86] Even leaving the GEFs aside, girls often cited the importance of receiving help on assignments or notes from class from other girls, which also supports the idea that participation in GEFs could provide girls an avenue to improve their support network and, by extension, their ability to keep up in class.

The social capital mechanism does not appear to operate solely through providing support on assignments, however. Some GEFs appear to proactively encourage attendance of other girls at school. One mother

[84] This regression is statistically significant at the 1 percent alpha level ($p<0.01$)
[85] Vignettes FGD with girls, Puntland, Int. 242.
[86] Vignettes FGD with girls, Puntland, Int. 242; Vignettes FGD with girls, Somaliland, Int. 141; Vignettes FGD with girls, Somaliland, Int. 142.

participating in an FGD indicated that girls who are part of the GEF will call girls who have not been attending school and encourage them to attend.[87]

The quantitative data appear to support these two arguments. Beyond the joint correlation between GEF participation, YLI scores, and learning outcomes, there is also a correlation between GEF participation and school attendance. Among girls who have participated at least once in a GEF activity, 92.9 percent reportedly – according to their caregivers – have attended school most days this year. Among girls who have not participated in a GEF activity, the share is 85.7 percent. GEF participation is also associated with girls reporting that they have trusted friends they can talk to when needed and girls reporting that their friends ask *them* for advice. Both of these outcomes are indicative of girls who have a stronger network amongst their peers, in line with the argument outlined above.[88]

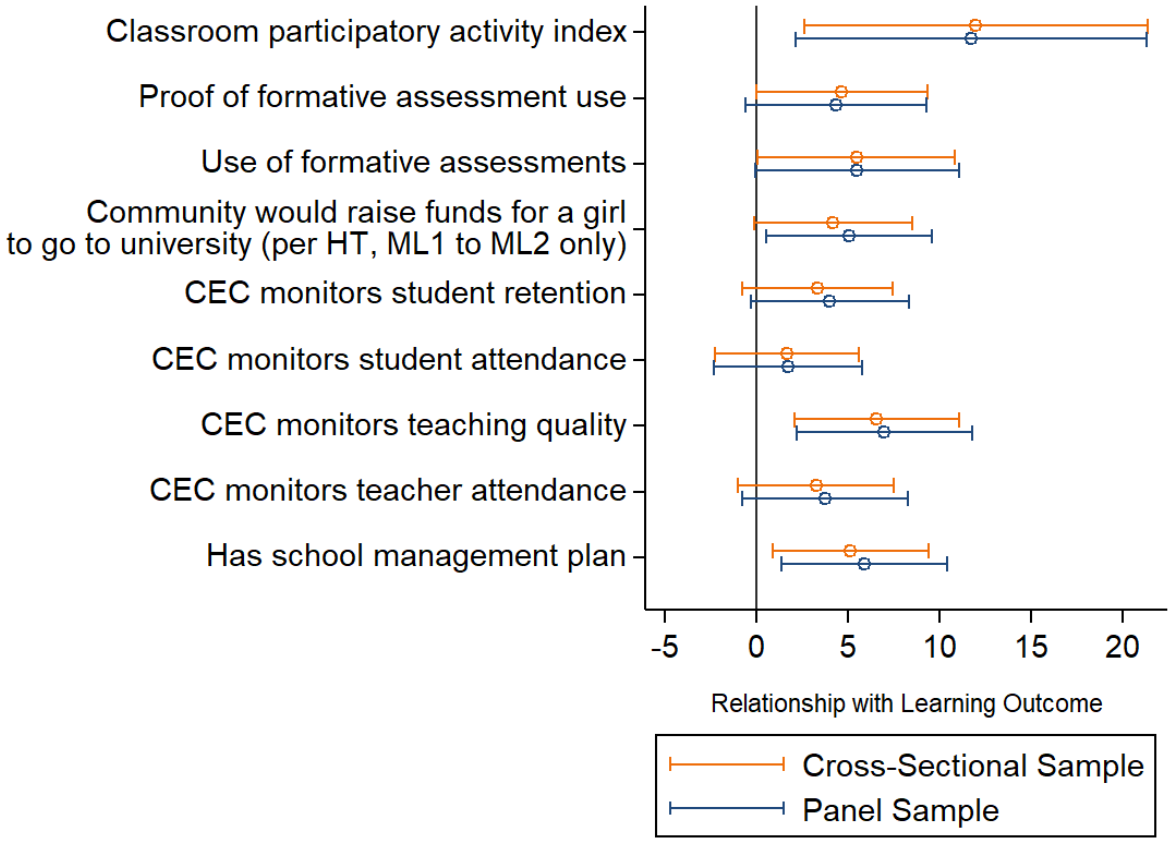
It is important to note that the associations reported here are not conclusive evidence that GEF participation *causes* higher learning scores, improvements in self-esteem, higher attendance, or a stronger social support network. As mentioned previously, there are theoretical reasons to believe that some portion of the relationship between GEF participation and self-esteem is driven by self-selection of motivated girls into GEF activities. A similar argument could be made for other correlations reported here: for instance, girls who are more social and have a stronger social network to start may be more willing to engage in social activities and join the GEF. Nonetheless, there are also good theoretical reasons to believe that participating in a GEF can facilitate all of these downstream outcomes. And, given the substantial positive relationship between GEF participation and learning – and the related support for a theory of GEFs improving social capital among girls – this appears to be an encouraging finding for this portion of the program's activities.

SCHOOL-LEVEL FACTORS

We now turn our attention to various factors at the school level.[89] The first measure concerns teaching quality as it looks at the use of participatory practices by teachers in class. The classroom participatory activity index is informed by the classroom observations where the researcher looked at whether:

1) Students do not spend most of the time copying from the blackboard
2) Students do not spend most time repeating teacher's words aloud
3) Teacher used student-centered games
4) Students instructed each other
5) Teacher asked open-ended questions
6) Teacher elicited student opinions
7) Teacher sought to include students who were not participating
8) Students worked together in a group

[87] FGD with Mothers, Galmudug, Int. 311.
[88] To illustrate: among girls who have not participated in a GEF activity in the last month (many have not participated at all), 53.2 percent report that their friends ask them for advice "almost always" or "most of the time." Among girls who are active and frequent participants in GEF activities or meetings, 59.3 percent report this outcome.
[89] Unlike with individual level variables, for school-level models we do not use individual fixed effects, because doing so does not make analytical sense. In cases where the "treatment" (e.g., CEC monitoring activities) occurs conceptually at the school level and applies to all girls in the school, individual fixed effects are not necessary. Instead we incorporate school fixed effects to account for underlying differences in learning outcomes across schools.



The index is thus a very comprehensive measure of participatory practices in classroom. And it turns out to be positively correlated with learning outcomes. It seems that girls in schools where teachers use overall more participatory approaches to teaching tend to score higher in their learning assessments.[90]

Next, as a further measure of teaching quality, we assessed the relationship between the use and recorded use of formative assessments and learning. Both measures are marginally significant and positive, implying that when teachers make use of formative assessments, their students are learning better. Of course, it is possible that the use of formative assessments is itself a reflection of the kind of resources available at the school, rather than the direct cause of improved learning outcomes. However, when we look at this relationship using the fixed effects regression approach, the reported use of formative assessments is still significantly and positively correlated with learning while controlling for the individual circumstances of the girls.

Another measure coming from the head teacher survey effectively serves as a robustness check for the community attitudes questions looked at above. The head teachers were asked if the community would raise funds to send a girl into university. It is interesting to look at this variable as it is possible that the above factors reflecting caregiver attitudes towards girls' education might themselves be affected by girls' positive learning outcomes. In other words, it is possible that caregivers appreciate girl's education more, because their girls are performing well – instead of girls performing well because their parents are supportive. In addition, this question is more of a measure of community attitudes as it is not linked to a specific girl. The question was

[90] Please note that there were only two classroom observations conducted in each school and the analysis has not specifically linked an observation to a girl. As such, the classroom observations should be viewed as a reflection of the overall situation of teaching in a school, but they do not necessarily capture the within-school variation that exists.

only asked at Midline Round 1 and Midline Round 2, so the effect presented here is for that timeline. Encouragingly, the results here lend support to the idea that positive community attitudes support learning. Whether looking at the panel or the cross-sectional sample, the relationship between this measure of community attitudes and aggregate learning outcomes is marginally significant and positive.

Finally, various measures of school management were used. Given the CEC's role in this area, many of the variables focus on the CEC's activities. We looked at the CEC's monitoring of retention, attendance, teaching quality and teacher attendance. We also looked at whether the school having a management plan would affect learning outcomes. All variables behave in the expected manner in that they are positively correlated with learning. Monitoring of student retention is marginally significant while monitoring attendance is not statistically significant. Meanwhile having a management plan and monitoring teaching quality are both positively linked to aggregate learning outcomes. Finally, monitoring teachers' attendance is marginally significant and also positive. In sum, it seems that there are reasonable grounds to make the argument that improved school management will produce better learning outcomes.

In sum, testing the relationship between various intermediate outcome measures and the aggregate learning levels has produced results that are largely intuitive and supportive of the program's ToC. It seems that positive community attitudes support learning, as does self-confidence and leadership skills. Meanwhile, not attending school seems to have a negative impact on learning outcomes. Teaching quality measures are also positively correlated with learning as are measures of goof school governance. The only variable that seems to produce counter-intuitive results is that of attendance. Regular attendance in fact seems to be associated with lower learning levels on average, although this relationship is not statistically significant to be conclusive. We can perhaps thus say that attendance alone does not produce learning. Overall, the results here lend strong support to the program's focus on teaching quality, community attitudes as well as school management as the evidence provided supports the hypotheses.[91]

In addition to testing the intermediate outcome measures, we also looked at the relationship between some of the outputs and learning. Namely, we looked at the effect of the primary caregiver being a member of a savings group. When running the fixed effects regression, we find that the primary caregiver being a member of a savings group is associated with a learning aggregate score on average 5.5 percentage points higher.[92] This is encouraging given CARE's focus on building savings groups in the areas of their program activity as it does indeed seem to positively contribute to learning.

Indeed, it seems that according to our analysis, the ToC is valid in that when looking at the assumed causalities in learning the variables assumed to be associated with improved learning outcomes are indeed correlated with higher learning outcomes. Yet, as can be seen in Section 4.1 of the report, the program has largely been unable to meet its learning targets. This apparent paradox is perhaps attributable to the fact that when we look at intermediate outcomes – the ones explored in this section – the program has not made a clear discernible impact. For example, as it pertains to the CEC activity, as section 7.2 of the report shows, none of the variables measuring CEC activity show program impact in a statistically significant manner. Similarly, as regards community attitudes, section 7.5 of the report illustrates that attitudes toward girls' education appear to be improving across all communities, not just those which have received benefits from SOMGEP-T programming. We assume that the ToC should apply equally in comparison schools and intervention schools. As such, when comparison communities see improvements in a given intermediate outcome, we would also expect them to improve in terms of learning. The fact that intervention schools seem not to have

[91] Please note that we also tested Somali, English and Numeracy separately instead of an aggregate measure. The analysis therein did not produce substantial divergencies from the results already presented.

[92] This regression is statistically significant at the 5 percent alpha level (p=0.04)

shown more rapid improvement as regard these intermediate outcomes, we would not expect them to show systematic improvements in learning either. It seems, thus, that while the program is based on seemingly valid assumptions about learning, program activities have not yet set the intervention schools apart from the comparison schools as no clear pattern of improvement can be observed for the intermediate outcomes.

Note from the project: The affirmation above should be qualified by noting that the intermediate outcome of life skills acquisition has been largely met – and is directly related to improved learning outcomes. It is also important to note that the similar rates of progress on intermediate outcomes such as participatory management (CECs) and quality of teaching in comparison as well as treatment schools does not mean that the intervention is not having an impact – but rather than parallel interventions such as those operated by the GPE-funded ESPIGs are having an impact at a similar level as SOMGEP's in those areas.

5. TRANSITION

Transition is the second top-level outcome for all GEC-T projects, including SOMGEP-T. The operationalization of transition reported on here is fairly complex, representing the wide variety of possible life pathways that Somali girls may go down. Transition is not merely a measure of continued enrolment in school; rather, it incorporates the possibility of alternative education, vocational training, and even employment. Further, successful transition is defined not just by the end result that a girl achieves; successful transition is defined in part depending on a girl's "starting point". We discuss the varied pathways in greater detail below, but it is useful to note – at the outset – the complexities of measuring and analysing transition outcomes.

Our discussion in this section, as with learning, is focused first on estimating the program's impact on transition rates over time. Transition rates were measured for all cohort girls at the baseline, which allows us to study the program's impact from baseline through this evaluation round using the quasi-experimental methods used extensively in our learning analysis. In general, we analyse transition rates among subgroups of girls that are defined by their "starting point" – i.e. their enrolment or employment status. For instance, we analyse the change in transition rates among girls who were out of school at the time of the baseline, looking for differential rates of re-enrolment, enrolment into alternative education programs, or movement into the formal labour market, across intervention and comparison communities. Our core analysis is divided between OOS girls, who we study from baseline to ML2, and the cohort of girls who were enrolled at baseline, who we study from ML1 to ML2. Finally, we also disaggregate successful transition into its constituent pathways, as girls starting from a position of non-enrolment have different pathways – such as enrolment into alternative education – that the program considers a successful outcome. Following on the pathway-specific analyses, we also report a broader set of within-subgroup estimates of program impact, looking at the change in transition rates associated with the intervention among a specific group, such as girls living in economically marginalized households.

5.1 TRANSITION PATHWAYS

SOMGEP-T takes a broad view of what constitutes successful transition, tailoring the goal outcomes to match the realities in which rural Somali girls live. The table below describes the pathways that the program considers successful or unsuccessful transition outcomes, depending on the girl's "starting point." To be clear, transition as an outcome is measured over time – at baseline, successful transition was measured by comparing girls' enrolment and employment status at baseline to their status in the year prior to baseline. In this round of data collection, we define transition based on a comparison of enrolment and employment status at ML2 backward to baseline.

To illustrate this idea, consider a girl who was enrolled in primary school at baseline. If such a girl drops out of primary school and enrolled in informal education at ML2, this is considered a negative transition outcome, because she has dropped out of school and moved downward in the hierarchy of educational institutions. In contrast, a girl who was out-of-school at baseline who was enrolled in informal education at ML2 would be considered a positive transition case, as she had moved toward what the program considers a more positive outcome. This distinction between two girls who reach the same ending point at ML2 is what we mean when we refer to the importance, for operationalizing transition, of the girl's "starting point." The starting points

referred to in the table are: the year prior to baseline, for assessment of baseline transition rates; baseline for assessment of ML2 transition rates.[93]

TABLE 58: TRANSITION PATHWAYS

	Starting Point	Successful Transition	Unsuccessful Transition
Lower primary school	Enrolled in Grade 3, 4	• In-school progression • Drops out but is enrolled into alternative learning program	• Drops out of school • Remains in same grade
Upper primary	Enrolled in Grade 5, 6, 7, 8	• In-school progression • Moves into secondary school • Moves into ALP	• Drops out of school • Moves into work, but is below legal age (under 15 years)
Secondary school	Enrolled in Grade F1, F2, F3, F4	• In-school progression • Enrols into technical & vocational education & training (TVET) • Gainful employment • Moves into ALP	• Drops out of school • Moves into employment, but is unpaid or otherwise exploited
Out of school	Dropped out	• Re-enrol in appropriate grade level in basic education • Enrol in alternative learning program • Engages in wage/self-employment	• Remains out of school

While much of the table above is self-explanatory, we do note how we define different employment outcomes. Girls are classified as entering under-age employment if they are employed outside the home, in

[93] We use a girl's baseline status as the starting point for defining her transition success at ML2, rather than using retrospective questions that identify her status in the year prior to ML2 (two years passed between baseline and ML2). The logic of using a girl's baseline status as the starting point for defining her transition success during ML2 is two-fold: first, the data collected during baseline may be more reliable than retrospective questions asked during ML2, which are subject to recall bias. The data is also more reliable because baseline data collection was more detailed than retrospective data collection at ML2. Second, retrospective transition-related data was only collected for a subset of girls – new girls selected into the sample as replacements – rather than all girls in the cohort sample. As we note elsewhere, this decision has consequences for the trend in transition rates from baseline to ML2, because baseline transition rates are measured over one year, whereas ML2 transition rates are determined by changes in a girl's status over two years, producing lower transition rates. However, this trend does not influence the valid estimates of project impact, because the change in measurement and the resulting downward trend affect both intervention and comparison groups equally.

any capacity, but are under the age of 15 years. Gainful employment refers to employment among girls who are 15 years or older and who have a permanent, formal, paid position. Girls who are in "casual" or "informal" employment, even if they are of the proper age, are classified as being non-gainfully employed. Only gainful employment is considered a positive transition outcome for SOMGEP-T.

In comparison to the ML1 evaluation report, the analysis in this section represents a fuller picture of transition outcomes. At ML1, data collection occurred exclusively among in-school cohort girls – those girls who were enrolled in school when they entered the sample at baseline. Therefore, the ML1 evaluation reported dropout rates from baseline to ML1, drawing conclusions based exclusively on this transition pathway. There is nothing inherently wrong with this approach, and it reflected the best use of the available data. However, the ML2 transition analysis is more complete, as it includes as many girls from the baseline cohort as could be located, and accounts for the full set of possible transition pathways outlined in the table above. The distinction between ML1 and ML2 also has consequences for interpreting the comparisons we make below, because, in some samples, transition is defined over a two-year period instead of a one-year period. We highlight this issue where relevant below, especially in the context of the cohort of OOS girls.

5.2 Overall Transition Outcomes

Table 58: Transition Targets

Objective	Indicator	BL/ML1	Target for ML2	ML2 outcome	Difference in differences	Target met
Transition - ISGs	Percentage of girls, enrolled at baseline, who have transitioned into the next stage of education, into training or gainful employment	**82.9% (83.1% comparison) at BL**	**10-point improvement (5 points per round)**	**89.2% (84.4% comparison)**	**+5.1 points**	**No**
Transition – OOS girls	Percentage of girls, out-of-school at baseline, who have re-enrolled in school, moved into training, or into gainful employment	11.6% (15.8% comparison) at BL	24-point improvement (12 points per round)	60.5% (52.0% comparison)	**+12.7 points**	**No**

The analysis in this section focuses on transition outcomes over the two primary cohorts tracked for transition. Unlike the learning analysis – much of which focused specifically on in-school girls, who had the greatest exposure to the project's more learning-centred activities – the transition outcome is genuinely expected to impact all girls in intervention communities. The sample design employed at the baseline, which drew girls from intervention and comparison communities using a random walk method, was used specifically to ensure that OOS girls were represented in the cohort of girls to be tracked over time. The emphasis on including OOS girls in the baseline reflects the fact that the project actively sought to encourage positive transition outcomes among *all* girls in intervention communities, via enrolment into primary schools, alternative learning tracks, or other pathways.

At the same time, it is important to emphasise that this section does not include a single aggregate analysis of all girls tracked from the baseline to ML2. Rather, our discussion is divided into two disparate analyses: in-school girls and out-of-school girls, defined by their status at baseline. The in-school girl cohort was tracked from baseline to ML1 and eventually to ML2. Importantly, the sample of in-school girls we study is limited to those girls who were successfully tracked and re-contacted through all three evaluation rounds thus far. They therefore represent a true panel – without replacements or attrition of any kind – with three years' worth of data on transition status. In total, the panel includes 529 distinct girls, all of whom were enrolled at the baseline. The decision to focus on this panel was made because it maximizes comparability across rounds, by excluding girls who were replaced and their replacements (in line with our focus on the panel sample for learning, in the previous section). In addition, this sample was tracked over three rounds, allowing us to understand how transition rates change evolved over time in intervention and comparison communities, rather than strictly from baseline to ML2 or over the shorter time period from ML1 to ML2. Note, however, that the panel available for analysis from baseline to ML2 – excluding ML1 – is larger, which generates some inferential benefits. Therefore, in Annex 18 we provide supplemental analysis of this larger cohort.

The out-of-school cohort constitutes the second half of our primary analysis. This group was tracked from baseline to ML2, and we include only those OOS girls – those girls who were OOS at the time of baseline recruitment – who were successfully re-contacted at ML2. None of these girls were sought out at ML1, for logistical and budgetary reasons. As with the in-school girl sample, our analysis of OOS girls utilizes a true panel, without including replacement girls or those they replaced. It is important to keep in mind that OOS girls are being assessed over a period of two years of project implementation, which could drive outsized impacts among this cohort, vis-à-vis in-school girls who were tracked over single year intervals.

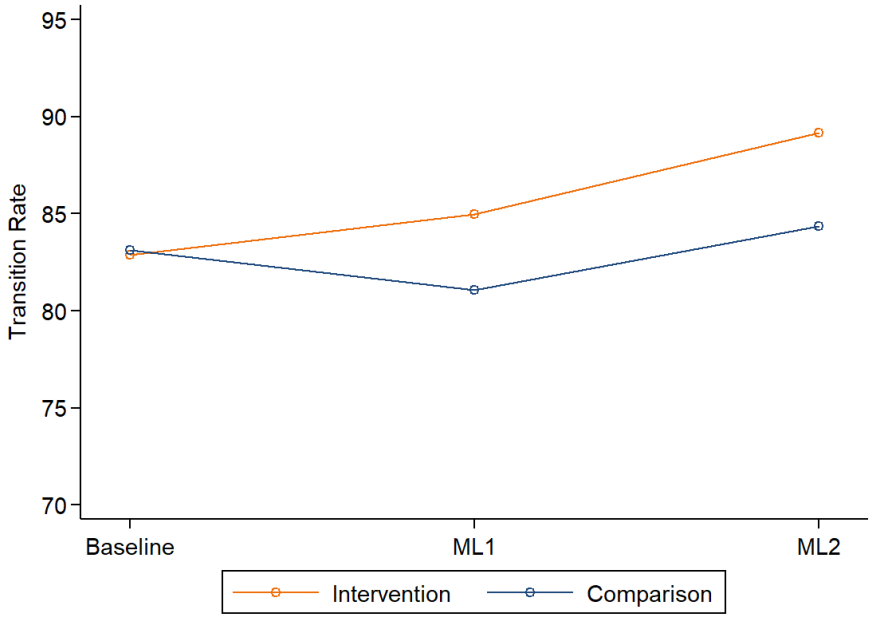
In total, we track 368 OOS girls from baseline to ML2, divided between intervention and comparison communities. OOS girls experienced the highest attrition rate from the sample – just 53.8 percent were re-contacted successfully – thanks to the fact that two years passed between contact attempts. Additionally, OOS girls tend to have lower re-contact rates even in comparable samples, possibly because they are more likely to migrate away from their communities for marriage or work. In both the in-school and OOS girl cohorts, relying on the "true panel" of respondents who were successfully interviewed in all relevant rounds of data collection ensures the most rigorous possible comparison between intervention and comparison groups over time, as the set of girls in both rounds is identical.

IN-SCHOOL GIRLS TRANSITION OUTCOMES

We begin with the in-school girl sample. The figure below reports the topline trends in transition rates among this cohort, by intervention status and round. The results follow the same 529 girls over three rounds of data

collection. In this analysis, we define transition as a binary outcome, with positive outcomes including all those identified as positive transition pathways in the pathways table above, including in-school progression through grades, re-enrolment in school, and transition into either gainful employment, vocational training, or alternative education, if it is age-appropriate. In practice, the girls in this sample generally follow a narrower set of pathways than this: all of the girls at baseline had either enrolled since the previous year, been held back a grade, or advanced a grade. At ML1 and ML2, girls either advanced a grade, were held back, dropped out, or – in rarer cases – dropped out and entered informal education or non-gainful employment.

FIGURE 21: TRANSITION RATES, ISG COHORT, BY INTERVENTION STATUS AND ROUND



The figure makes clear that intervention and comparison communities began from similar positions in terms of transition rates, a fact that is confirmed in previous analysis from the baseline. The baseline report itself, on the basis of a much larger sample of girls, concluded that intervention and comparison communities had similar transition rates (50.8 versus 50.5 percent) at the outset of programming. In this sample, baseline transition rates are higher, as we are considering only girls who were enrolled at the baseline; nonetheless, baseline rates are very similar in the two types of communities.

After just one year, however, transition rates had increased somewhat in intervention communities, while declining in comparison communities. In ML1, the gap between intervention and comparison communities had risen to 3.9 points, widening to 4.8 points by the latest evaluation wave. The differential trend between baseline and ML1 is especially stark, but there are consistent improvements in intervention communities – both relative to the communities against which we are comparing and in absolute terms – over the two years since the baseline.

While the results in the figure are relatively straightforward to interpret, the results are descriptive, rather than inferential. That is, we have reported positive findings in intervention communities, but without a measure of uncertainty around those findings, or an indication of whether the results are statistically

distinguishable from a null effect. To formally assess whether the program has had a positive impact on transition rates, we estimated a series of difference-in-differences models, testing whether the differential trends are sufficiently large to be distinguished from a null result. Our primary model replicates the figure precisely, without incorporating additional control variables; in alternative models, we also control for age, grade, and any latent school-level characteristics that might influence transition rates, by incorporating school fixed effects – i.e. a dummy variable for each school in the sample, allowing each school to exert an independent impact on transition rates, without biasing our results.

The results are calculated relative to baseline, and we estimate a program impact of 4.2 points at ML1 and 5.1 points at ML2. To be clear, the results at ML2 are *relative to the baseline*, implying that the majority of gains since baseline occurred at ML1, and the gains were merely consolidated and extended slightly at ML2. Despite the substantial effect size estimated, the results are not statistically significant ($p = 0.45$ and $p = 0.32$ for the two rounds, respectively). Inclusion of age, grade, and school-level fixed effects attenuate the results slightly but do not substantively change them or our conclusions about the impact of the program on transition. The program appears to have had a clearly positive impact on transition rates, but the effect size is too small to distinguish from zero in our small or moderately-sized sample.[94]

Positive changes in overall transition rates are not the only consideration when judging the program's impact. In addition to our binary coding of transition, there are important distinctions to be drawn between transition pathways. Consider the distinction between being held back a grade, while remaining enrolled, and dropping out of school entirely. While neither is considered a positive transition outcome in this case, the former is clearly preferable to the latter, insofar as a student who has been held back has remained in school, will continue to learn, and may continue to higher levels of education; a child who has dropped out of school may re-enrol and end up in the same position, but the prospects of this occurring are lower.

In the table below, we report the share of students – all of whom were enrolled at baseline – who achieved different transition outcomes at ML1 and ML2. As the table shows, the most common outcome in both periods was successful transition in the form of in-school grade progression. However, over 10 percent of students at ML1 and around 5 percent of students at ML2 were held back a grade, while there were also high dropout rates in some groups.

TABLE 59: TRANSITION PATHWAYS FOR IN-SCHOOL GIRLS, FROM BASELINE TO ML1 AND ML2 (N = 529)

	Intervention	Comparison
Midline 1		
Advanced grade	85.0%	81.1%
Held back	11.9%	11.5%
Dropped out, into informal education	0.4%	0.4%
Dropped out	2.8%	7.0%
Successful transition rate	**85.0%**	**81.1%**
Remain-enrolled rate	**96.9%**	**92.6%**
Midline 2		

[94] While a sample size of 529 girls is sufficient to detect large effects, it is important to remember that the sample is clustered and loses power as a result of this fact. At the same time, the decision to focus on a smaller sample that includes all three rounds of data collection does not fundamentally alter the findings – in larger samples that we analyse elsewhere in this report, the results are also statistically significant, and there are not large gains in power in these alternative samples.

Advanced grade	89.2%	84.4%
Held back	5.2%	4.1%
Dropped out, into informal education	0.4%	2.5%
Dropped out	5.2%	9.1%
Successful transition rate	**89.2%**	**84.4%**
Remain-enrolled rate	**94.4%**	**88.5%**

An important finding emerges from the table when we look at the specific pathways across intervention and comparison communities. In both panels of the table – ML1 in the top panel, and ML2 in the bottom panel – we can see that intervention communities had higher successful transition rates than comparison communities, in line with the previous analysis. However, we can also see that the successful transition rate undersells the impact of the program slightly, because intervention communities had higher rates of "held back" students, whereas comparison communities had higher net dropout rates. While both outcomes are considered unsuccessful transitions, it is reassuring to see that nearly half of all cases of unsuccessful transition in intervention communities at ML2 were students who failed to advance a grade but remained in school. In contrast, almost three-quarters of unsuccessful transitions in comparison communities, at ML2, resulted in dropping out of school. If we consider being held back a result of intermediate value, the program's positive impact rightly appears larger than it does when considering transition as an exclusively binary outcome.

We observe a similar trend when we look at an overlapping, but different, sample of in-school girls, focusing on all those girls who were enrolled during ML1 data collection and were successfully re-contacted at ML2. We use this alternative sample as a robustness check, because it is somewhat larger, providing a sample of 634 girls who fall into the in-school girl cohort. These girls were either recruited into the in-school girl cohort at baseline successfully re-contacted at ML1, or they were selected into the in-school girl cohort at ML1 as replacements for in-school girls who could not be located. Again, we limit the sample to girls who were enrolled at the outset (ML1), and study transition pathways in the current round, as we report in the table below. Unsurprisingly, given that this sample includes many of the same girls as the full panel sample, we find similar results regarding overall transition rates, with rates 2.5 points higher in intervention schools at ML2. Consistent with our earlier findings, this result actually understates the difference slightly, as girls in intervention schools are more likely to remain enrolled but be held back, while girls in comparison schools are more likely to drop out of school entirely. Compared to a transition rate gap of 2.5 points, girls in intervention communities are 3.7 percentage points more likely to stay enrolled in school between ML1 and ML2. The findings should give us additional confidence in the main transition results reported earlier, because the findings hold in an alternative sample that focuses on a narrower timeframe.

TABLE 60: TRANSITION PATHWAYS FOR IN-SCHOOL GIRLS FROM ML1 TO ML2, ML1 COHORT (N = 634)

	Intervention	Comparison
Advanced grade	87.4%	84.9%
Held back	8.6%	7.4%
Dropped out, into informal education or non-gainful employment	0.6%	1.8%
Dropped out	3.4%	6.0%
Successful transition rate	**87.4%**	**84.9%**
Remain-enrolled rate	**96.0%**	**92.3%**

In-school girls in intervention schools are more likely to remain enrolled in school and more likely to advance a grade than their comparison community counterparts. After one year of program implementation, their rate of successful transition had risen, and it continued to rise – albeit at a slower pace – two years into programming. The evidence in favor of program impact is far from overwhelming, however. To supplement our analysis, we also assessed the impact of the program on transition rates among a wider sample of in-school girls, without excluding those that fell out of the sample at ML1. We analyze this sample in more detail in Annex 18, where we look at both in-school and OOS girls. Here we note that, among in-school girls only, the difference-in-differences estimate of impact in this sample is slightly negative, contrary to our other findings for in-school girls. While we do not take this as evidence as dispositive, it does undercut the broader positive finding, and suggests that the program's effect on transition rates among in-school girls is not particularly strong, and is sensitive to the particular sample and time period studied.

OUT-OF-SCHOOL GIRLS TRANSITION OUTCOMES

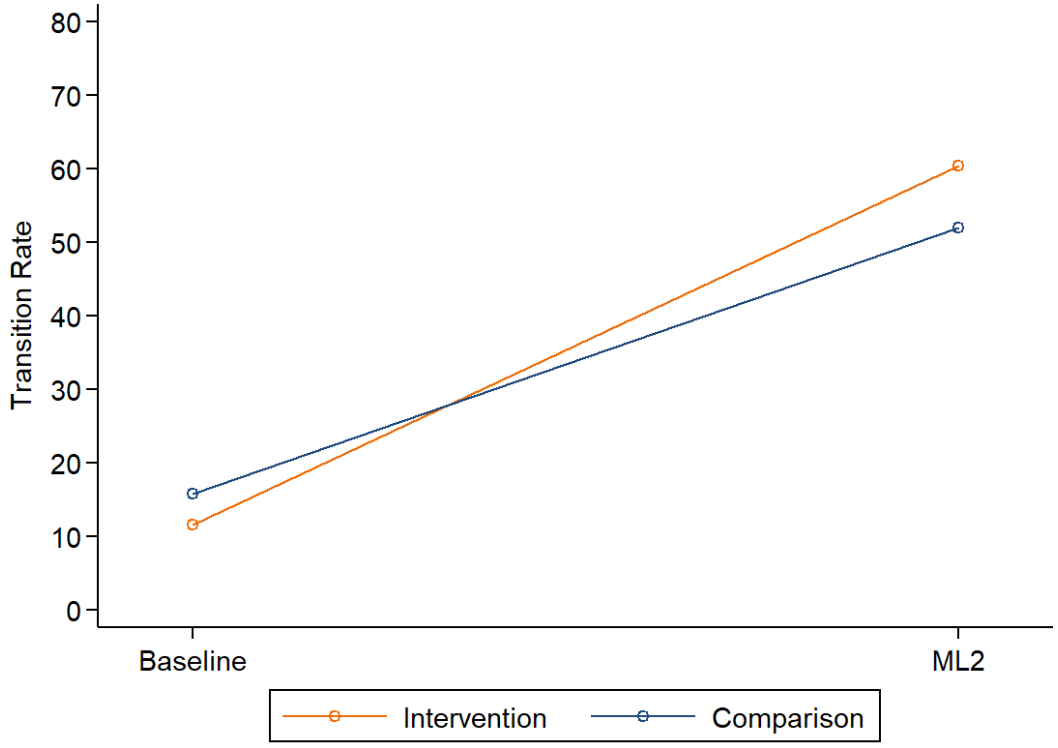
The second sample of girls we study are those who were out-of-school at the baseline. The baseline evaluation explicitly sampled both in- and out-of-school girls, drawn randomly from the communities selected, in order to provide a strong evidence base regarding transition outcomes. Because the program sought to encourage enrolment of OOS girls into school, alternative education, or vocational training, knowing the underlying enrolment rate was critical. This sampling approach gives us the opportunity to study transition rates among girls who were not enrolled at baseline, to assess how successful the program has been in bringing OOS girls into some school or alternative forms of education.

At baseline, enrolment rates in the overall sample were just 43.3 percent. Transition rates were slightly higher, as transition includes non-school outcomes, such as enrolment in informal education. Successful transition was approximately equal in intervention and comparison communities, at 50.8 and 50.5 percent, respectively. The fact that so many girls were not enrolled in school or any form of alternative education underlines the importance of the program's focus on bringing OOS girls into school.

The sample we analyze to understand transition rates among this cohort consists of all girls who were not enrolled when they were recruited into the study at baseline and who were subsequently re-contacted successfully at ML2. As noted above, this construction of the sample is important due to the high attrition rate among OOS girls. The resulting sample includes just 368 OOS girls, with data collected from two evaluation waves, spaced two years apart (BL and ML2). The sample is relatively evenly divided between intervention (n = 172) and comparison (n = 196) communities.

The figure below shows the aggregate findings among this cohort. Consistent with their enrolment status when they were recruited, transition rates at baseline in this cohort were very low, just 11.6 percent in intervention communities. The sharp rise in rates between baseline and ML2 occurs even in comparison communities, which may seem surprising. However, it is not altogether surprising, because successful transition outcomes for girls who were not enrolled previously include re-enrolment – as opposed to girls who were already enrolled, who must advance a grade. Despite this slightly lower bar for successful transition, the stark increase in transition rates is still dramatic.

FIGURE 22: TRANSITION RATES AMONG OOS GIRLS, BY INTERVENTION STATUS AND ROUND



Importantly, the gains in intervention communities clearly outpace those in comparison communities, illustrated by the steeper slope from baseline to ML2 for intervention communities. All told, transition rates increased by 12.6 points from baseline to ML2 among girls who were out-of-school at baseline.

To assess whether this increase is statistically significant, we estimated a difference-in-differences model in a regression framework. In a simple model, without control variables, the results are not statistically significant ($p = 0.24$), though the findings are suggestive. Adding control variables, including age, grade, and school fixed effects does not appreciably change the results, including the p-values estimated. We do note that the sample size available for analysis is very small, given the clustered nature of the sample. With a moderately-large sample (e.g., 800-1000 respondents, split between intervention and comparisons groups), we would expect an effect with this magnitude, 12.6 points, to be statistically significant. Difference-in-differences models require higher sample sizes, all else equal, than single-period cross-sectional models, and this also factors into the non-significant findings, as does the fact that the sample is highly clustered. In other words, while it is important to exercise caution when making positive inferences about non-significant results, we do feel the sheer magnitude of the effect is compelling.

The findings are also broadly supported by analysis of the transition pathways that OOS girls followed between baseline and ML2. First, among this group of girls, successful transition rates were lower in intervention than comparison communities at baseline. While enrolment rates were very similar, girls in comparison communities were more likely to have recently moved from being out-of-school entirely to joining an informal education program. The interesting finding with regard to specific transition *pathways* is that this trend flipped at ML2, as shown in the table below. Among girls who were OOS at baseline, there was no difference between intervention and comparison communities in terms of re-enrolment rates, shown

in the first row of the table. However, girls in intervention communities were much more likely to move from apparent idleness – not working, not in informal or alternative education – into informal educational programs. Overall, they were nearly twice as likely to move into informal education, and much less likely to remain out-of-school entirely.

Note from the project: It is important to highlight that the accelerated basic education at primary level (G1-4/ ABE) was introduced months before the ML2 data collection. Since the majority of the OOS girls – 75% as per the baseline results – had had no previous access to education, this component was the main intervention catering to their needs. Therefore, the sharp improvement at ML2 is to be expected.

TABLE 61: TRANSITION PATHWAYS AMONG OOS GIRLS, FROM BL TO ML2

ML2 Outcome	Comparison	Intervention	Difference
Re-enrolled	42.4	42.4	0.0%
OOS, moved into informal education	9.7	18	+8.3%
OOS, moved into non-gainful employment	1.5	1.7	+0.2%
OOS, moved into underage employment	1	0	-1.0%
Remained OOS	45.4	37.8	-7.6%
Total	100.0%	100.0%	
Successful Transition Rate, OOS Girls	52.1%	60.4%	8.3%

Given the high rate of attrition in this sample, efforts were made during data collection to capture transition information even among girls who could not be contacted successfully. When a girl could not be located at ML2, enumerators were asked whether they were able to identify anyone in the community who knows the girl and can provide information about her whereabouts and enrolment status. If so, they recorded enrolment information, but no other details about her grade level, employment status, or other transition activities.[95] This approach gives us a slightly larger sample to work with, keeping an additional 207 girls in the panel. We limited inclusion to those girls for whom we collected information from a parent, another family member, or a teacher, and excluded those whose information was drawn from a friend or neighbour, for the sake of maintaining reliability.

Unfortunately, there are also problems with this approach. The coding of transition in this sample is coarser. We do not know, for example, whether girls who are still enrolled progressed to the next grade, or whether girls who are out-of-school have enrolled into alternative education programs. For the purposes of the analysis, we code *all* girls in the sample using coarser categories, which ensures the girls are comparable to one another but simultaneously obscures nuance in the data. The four outcomes we are able to identify for this group are:

- Enrolled at baseline, remain enrolled at ML2 (positive)

[95] More fine-grained data could be collected, but would be of questionable reliability in many cases, given that our informants were often head teachers or extended family members.

- Enrolled at baseline, dropped out at ML2 (negative)
- Out-of-school at baseline, enrolled at ML2 (positive)
- Out-of-school at baseline, remain out-of-school at ML2 (negative)

Among the OOS girls who fell out of the overall sample but for whom we were able to collect enrolment information (n = 153), the vast majority had remained out of school (94.8%) had remained out of school. We were able to locate family members but unable to ascertain enrolment status for a handful of girls, typically because the family member was not sure whether the girl was enrolled. Just 5.2 percent of girls had re-enrolled, with information provided from a relatively reliable source.

The findings in this sample are slightly less positive than in the main sample; while both intervention and comparison communities show improvement over time, the trends are very similar, implying that the project did not have any appreciable impact. We emphasize the limitations of this approach, as it can easily obscure important differences between baseline and ML2, especially in the context of girls who remain enrolled in school but who do not advance to the next grade. Additionally, girls who were out-of-school in previous years but have moved into acceptable transition paths – such as non-formal education – are included in the sample, but their transition status is mischaracterized in the data as negative (non-enrolment) rather than positive (enrolment in non-formal education), because the data collected from family members of these girls is insufficiently granular. In general, our preference is to rely on the sample of girls who were successfully re-contacted, primarily because it allows for finer-grained distinctions in transition status to be drawn, which is essential to understanding the program's true impact.[96]

While we believe any attempt to limit attrition – both to maintain statistical power and to prevent bias from different re-contact rates – is valuable, we view the results that incorporate girls who could not be re-contacted with skepticism. The strength of the effect among OOS girls who were successfully tracked from baseline to ML2 is compelling, and this finding emerges from the more rigorous analysis, in which we are more confident that enrolment status and other transition details were recorded correctly. Second-hand information from family members should be viewed more cautiously, especially when it necessitates a coarser coding of transition status. Overall, we conclude that the program has likely had a strong effect on transition rates among OOS girls, but that these results should be interpreted cautiously, in light of the relatively small sample size on which they are based and the level of uncertainty surrounding our estimates.

5.3 ALP Girls Transition Outcomes

The final cohort for whom the evaluation has been tracking transition outcomes are referred to as Alternative Learning Program (ALP) girls. ALP girls were brought into the sample for the first time at ML1, when they were recruited from 35 ALP centers established by the program in intervention communities. It is important to note that ALP girls do not have a logical comparison group. While we compare them to in-school and OOS girls in the analysis presented in the previous section to gain a sense of whether ALP girls are learning at a rate faster than they would if they were not in school at all (i.e. natural "growth effects" that we observe even in girls who are not enrolled at all), this approach makes less sense in the case of transition.

[96] Many of the transition pathways targeted by the program are more nuanced than simply enrolment in formal school. It is also worth noting that the bias from excluding girls who fell out of the sample – the vast majority of whom were out-of-school in previous rounds – is likely less than the bias that stems from coarsening the transition status of all girls in the sample, in an effort to include "attrition girls" also.

ALP girls can follow one of several transition pathways, including moving into formal school, transitioning into vocational training, or entering the labor market. It is important to note that ALP girls are older than in-school or OOS girls, whose mean age was 13.6 years at ML1. In contrast, ALP girls' average age was 17.0 years at ML1, and the range was from 11 to 21 years. Therefore, transition into the labor force is an acceptable outcome for a larger share of ALP girls than the other cohorts.

The table below reports transition outcomes for ALP girls at ML2, one year after their recruitment into the sample. At ML1, all girls were enrolled in ALP, and 73.5 percent remain enrolled in ALP one year later. An additional 8.9 percent of girls were enrolled in some other type of non-formal education, either an ABE program (in just one case) or a somewhat ambiguous "community-based education" described by their caregivers. In practice, we suspect that some of the girls cited as being in community-based education are in ALP, with their caregivers not identifying it clearly as such. Finally, 3.9 percent of girls had transitioned into formal school one year later. On the other hand, a small set of girls – 12.8 percent – had dropped out of school and were not engaged in any other activity.

TABLE 62: TRANSITION RATES AMONG ALP GIRLS

Transition Outcome at ML2	Share of ML1 ALP Girls
Enrolled in ALP	73.5%
Out-of-school altogether	12.8%
Enrolled in ABE or Community-Based Education	**8.9%**
Enrolled in formal school	3.9%
Gainful employment	0.4%
Non-gainful employment	0.4%

It is difficult to draw firm conclusions about ALP girls based on this transition information. The vast majority of ALP girls remain in some form of education and have therefore transitioned successfully from year to year, if we ignore the possibility that they have been "held back" – a fact which is not captured well, as grades or levels within ALP centers are not entirely clear. It is also unclear because girls recruited at ML1 from the ALP centers had only recently entered – that year – ALP, so few of them have had a chance to move into other pursuits at this point. The impact of ALP centers on transition outcomes will likely become more clear at the endline, as two years will have passed since they were first located at the ALP center, and we would expect more girls to have moved from ALP into other pathways.

5.4 SUBGROUP TRANSITION OUTCOME – PROGRAM IMPACT

The results presented thus far, especially in the context of in-school and out-of-school girls, have focused exclusively on aggregate impacts across the entire sample. As Section 5.2 showed, intervention communities have experienced an uptick in positive transition outcomes for both in-school and out-of-school girls, *relative to comparison communities*. Further, among unsuccessful transition outcomes, intervention communities have experienced a shift toward "less bad" outcomes, such as being held back a grade but not dropping out entirely, relative to those same comparison schools. In this section, we narrow our focus to specific subgroups of girls and estimate the impact the program has had on the likelihood that they experience positive transition outcomes.

The purpose of this section is to understand differences in transition rates *and* differences in program impact across subgroups of girls. As the baseline and ML1 reports documented, there are meaningful gaps in transition rates across *types* of girls. To take a single example, the baseline evaluation found that girls who were responsible for a significant share of their household's chores on a consistent basis were less likely to transition successfully. Differential transition opportunities are also embedded in the program's Theory of Change and design, as the program specifically targets pastoralist girls, those from marginalized households, and so forth.

Importantly, our primary interest at this stage in program implementation is in differential impact within subgroups, rather than understanding underlying differences in transition rates across subgroups. To be clear about this distinction, we acknowledge that there are considerable differences in transition rates between groups (e.g., girls with an educated caregiver versus those whose caregiver has no education). While this gap continues to be of interest to the program, our focus in this section is on estimating program impact within subgroups. If the program has hypothetically improved transition rates by 5 points, we are interested in whether there is heterogeneity in this estimate across subgroups – has the program had a particularly outsized impact on girls from female-headed households, or has the program had its main impacts among the least marginalized girls?

Assessing this question is similar, in terms of methodology, to assessing aggregate program impact. For each subgroup of interest, we formed a subsample of the data composed exclusively of girls with that characteristic. For that subsample, we estimate a difference-in-differences model to establish the program's impact *within that subgroup*. For instance, when we assess program impact within Somaliland, we are studying the relative change between intervention and comparison schools over time, but only for those intervention and comparison school located in Somaliland. The main methodological limitation to this approach is the fact that it produces very small sample sizes in many cases. We are often unable to study subgroups that are of programmatic or theoretical interest because of the small number of girls who fall into the subgroup and who were simultaneously tracked for transition outcomes. In the results below, we limit our attention to subgroups with sample sizes that are large enough to be indicative, though we still note cases where our analysis is extremely tentative due to the number of observations.

The table below reports program impact based on difference-in-differences models for both in-school girls (left panel) and out-of-school girls (right panel). As in the aggregate analysis, our sample of in-school girls consists of girls who were enrolled at the time of the baseline and who were successfully tracked from baseline through each subsequent round of data collection. In this sample, the reported impact at ML1 is the difference-in-differences identifying the program's impact at ML1 relative to baseline. At ML2, the impact is also the program's impact at ML2 *relative to baseline*. Note that impact is not cumulative across ML1 and ML2 – where we report aggregate impact of 4.2 points at ML1 and 5.1 points at ML2, this does not mean that transition rates have increased 9.3 points from baseline to ML2; rather, it means that most of the gains in transition rates (4.2 points) occurred between baseline and ML1, and rates increased only 0.9 additional points between ML1 and ML2. The analysis of OOS girls here also mirrors that of our aggregate OOS analysis described previously. We study the set of OOS girls who were recruited at baseline and successfully re-contacted at ML2, with a total sample of 368 such girls.[97]

The table below reports findings from an extensive set of difference-in-difference models. Note that overall impact for in-school and out-of-school girls is 5.1 and 12.6 points at ML2, respectively. This aggregate finding

[97] We do not estimate impact at ML1, as OOS girls were not contacted at ML1. The only impact estimate is that of impact from baseline to ML2.

provides a benchmark against which we can judge the relative magnitude of impacts among particular subgroups.

Among in-school girls, three broad findings stand out. The first is that girls who have lost one parent ("partial orphans") saw particularly large gains in transition as a result of the program. Despite the fact that this sample includes just 60 girls, the program's impact on this subgroup is huge – 23.2 points at ML2. The second broad finding concerns girls in households that are occupationally marginalized. Pastoralist households show especially large program impacts, but we are cautious about leaning too heavily on these results, given the small sample size and the fact that the results are not statistically significant. However, we also find that girls living in households in which the head-of-household does not have an occupation (i.e. the head of household specifically indicated they had no occupation, after reviewing a list that included farming, pastoralism, and small-scale trading, among other occupations) also experienced outsized program impacts. Among this subgroup of 212 girls, the program increased transition rates by 10.3 points between baseline and ML2.

TABLE 63: SUBGROUP-SPECIFIC PROGRAM IMPACTS ON TRANSITION RATES

	In-School Girls			Out-of-School Girls		
	Impact, ML1	**Impact, ML2**	**N per round**	**Impact, ML1**	**Impact ML2**	**N per round**
Overall	4.2	5.1	529		12.6	368
Somaliland	-0.5	5.4	342		14.8	253
Puntland	14.2	1.9	173		14.0	95
Female-headed household	1.8	-2.6	222		13.1	158
Lives without parents	0.0	-2.8	34		-1.9	50
Part orphan	21.0*	23.2**	60		-8.0	41
HoH has no education	4.4	5.2	317		12.6	272
HoH has no formal education	4.4	5.2	317		12.6	272
CG has no education	-0.5	2.1	367		13.9	298
CG has no formal education	1.5	3.9	417		12.2	314
HoH pastoralists	14.2	22.2	43		32.7	58
HoH no occupation	9.9	10.3	212		8.3	160
Seasonal migration	-2.8	8.3	21		24.8	36
Owns a mobile phone	4.0	3.3	507		12.2	346
Has regular access to water	6.4	7.2	292		18.6	156
Owns land alone	2.1	8.4	388		14.7	253
Owns land	3.6	8.8	427		12.5	289
Poor quality roof	4.7	14.5	167		19.3	135
Gone hungry many days in past year	10.0	15.0	40		10.0	59

Gone without enough clean water many days	5.2	9.2	151		2.8	132
Gone without medicines many days	10.2	9.8	212		9.3	175
Gone without cash income many days	-1.9	2.1	205		9.2	171

The third finding concerns girls that live in relatively poor households. Poverty in the program's context is not always simple to classify, because urban and rural households have different markers of poverty. However, the consistency of this finding suggests that – regardless of how we classify households – those that are economically marginalized have experienced particularly large gains as a result of the program. Households with a poor quality roof, households whose members have gone hungry many (more than 10) days in the last year, and households that have experienced similar deprivation in terms of access to clean water and essential medicines all experienced larger improvements in transition rates as a result of the program than their less-marginalized counterparts. While none of the effects among these subgroups are statistically significant, the estimated program effects are substantively large and consistent across multiple measures; the results are also consistent with the program's design, insofar as the program sought to reduce economic barriers to girls' education by providing bursaries and sponsoring VSLAs.

Among OOS girls, we find both reinforcing and contradicting themes. Recall that the aggregate impact estimated among OOS girls is much higher – 12.6 versus 5.1 points from baseline to ML2 – in this sample, when compared to the in-school girl sample. Even within this environment of higher aggregate impact, pastoralist girls stand out – the program appears to have increased transition rates by 32.7 points among pastoralist girls, above and beyond whatever increases were also observed in comparison communities.[98]

The results reported above can come across as relatively abstract, especially because transition outcomes among OOS girls rose dramatically in both intervention and comparison communities from baseline to ML2. To illustrate the extent of the program's impact, consider the probability of successful transition among pastoralists, shown in the table below. In the top panel, we highlight the impact among the ISG sample from baseline to ML2. The table makes clear that in-school grade progression was much higher for pastoralists in comparison communities at baseline – in essence, pastoralist girls in comparison communities were less marginalized, or at least less likely to be held back a grade, than equivalent girls in intervention communities. However, by ML2, this gap had entirely closed, and transition rates among pastoralist girls were nearly identical across the two community types (68.0 percent in intervention and 66.7 percent in comparison communities).

TABLE 64: TRANSITION PATHWAYS AND OUTCOMES AMONG PASTORALIST GIRLS

	Baseline		ML2	
	Intervention	**Comparison**	**Intervention**	**Comparison**
In-School Girls Sample				

[98] Despite the fact that the OOS girls sample is smaller overall than the in-school girls sample, there are actually more pastoralists in the former sample, because pastoralist girls were much less likely to be enrolled in school at baseline than their non-pastoralist peers.

In-school grade progression	56.0%	88.9%	68.0%	66.7%
Re-enrolled	12.0%	0.0%	0.0%	0.0%
Held back a grade	32.0%	11.1%	16.0%	22.2%
Dropped Out			16.0%	11.1%
Successful Transition	**68.0%**	**88.9%**	**68.0%**	**66.7%**
OOS Girls				
Re-enrolled			52.0%	39.4%
Dropped out	16.0%	9.1%	0.0%	0.0%
OOS, moved into informal education	16.0%	30.3%	24.0%	18.2%
Remained OOS	68.0%	60.6%	24.0%	42.4%
Successful Transition	**16.0%**	**30.3%**	**76.0%**	**57.6%**

Among OOS girls, presented in the bottom panel of the table above, pastoralist girls were relatively more marginalized in intervention communities, where they were less likely to have recently moved into informal education when we surveyed them at baseline. But the two keys to the program's impact among pastoralist girls in the OOS girl cohort comes from their re-enrolment in school and their movement into non-formal educational opportunities. These girls, who were out of school at the baseline, are much more likely to have re-enrolled in school or into informal education if they live in intervention communities – these facts represent the tangible impact of the program on pastoralist girls who were not in school at baseline.

Many of the same themes explored above also emerged in the qualitative data around girls' transition and barriers to enrolment. Despite the substantial gains made among pastoralist groups, respondents (including parents, CEC members, and education professionals) continue to describe the nomadic lifestyle as incompatible with the education system. Many communities find, when a pastoralist family relocates, their child's continued education is contingent on the ability and willingness of a community member to host the girl. Such instances are few and far between, as neither CEC's nor community members have the resources to support pastoralist students in their families' absence. In fact, though the quantitative data indicates strong positive impact pastoralist student enrolment, the qualitative data suggest the uptick in enrolment may be owed – at least in part – to environmental conditions. Recent droughts had catastrophic effects on livestock owners and many families lost their entire herd. As a result, the pastoralist population in Sanaag town spiked, in turn increasing enrolment of girls from pastoralist families.

In addition to logistical obstacles around family relocation, pastoralist students are confronted with many of the same obstacles as other girls permanently based in the target communities. Respondents closely associate economic factors with a child's likelihood to remain enrolled in school. Low income students, a group often inclusive of pastoralists, are more likely to come to school if it helps them meet their basic needs. Both teachers and mothers assert school feeding programs help keep children of limited means in school. Family financial constraints may also pull students out of school to help contribute to livelihoods, such as livestock rearing within pastoralist communities. More frequently, school fees prohibit students' continued enrolment. In these cases, the CEC is sometimes able to step in to raise funds for the girls. Schools may also waive fees for students in need.

> *"The CEC successfully brought back some students who had dropped out of school. These students are now back to pursuing their education. The CEC worked extremely hard to bring students back to school."*[99]

In fact, the CEC is reported to actively monitor attendance and intervene in the event of a student drop-out. Girls may be kept home to tend to domestic responsibilities. In these cases, the CEC reaches out to the family and works with them to get the girl back into school. Several committee members cite specific success stories where they have re-enrolled girls who left school through this outreach strategy. While efforts are also made to reach girls who leave school to marry, respondents convey more mixed perceptions of the CEC's ability to bring these girls back into the education system. Several speak conclusively about marriage marking the end of a girl's education. It appears the introduction of alternative education options has mitigated this challenge somewhat. Nonetheless, domestic responsibilities and early marriage continue to be among the most cited drivers of girls drop-outs.

> *"There will be no education, she will drop out because she will be busy with life and her husband."*[100]

> *"There also may young girls who get married and do not come back to school. However, with help of many international organizations , we now have a school for the married girls and adults and they are currently enjoying studying."*[101]

There is broad agreement amongst qualitative respondents that children with disabilities have remained the most neglected group in terms of improving access to education. Most reveal nothing has been done to break down barriers preventing disabled children from enrolling. Teachers do not have sufficient training, needs-based curricula have not been developed, and assistive technologies – such as wheelchairs or glasses – are unavailable. Targeted schools report no or few students with disabilities enrolled and little done to ensure their right to access education.

> *"I can say, nothing has been done for the children, the first activity we are planning is to training couple of teachers."*[102]

Note from the project: This affirmation is inaccurate and does not reflect the current situation. Teacher training on inclusive education took place in parallel with the ML2 data collection, led by the Somaliland National College of Education (SNCE) in Somaliland and the Garowe Teachers Education College (GTEC)

[99] FGD with Mothers, Somaliland, Int. 113.

[100] FGD with Mothers, Galmudug, Int. 311.

[101] FGD with CEC members, Somaliland, Int. 101.

[102] KII with MOE representative, Puntland, Int. 261.

in Puntland, working in collaboration with the respective MOEs. The provision of assistive devices has been planned and delayed due to COVID-19. Last but not least, it should be noted that this comment seems to refer exclusively to CwDs with physical disabilities, completely excluding the project's activities with those facing mental health issues (which have a direct impact on transition).

In some regards, the quantitative and qualitative data produce a consistent narrative in which programme impacts have been clear and positive. This is the case for economically marginalized populations – or those who have poor quality roofs and/or have experienced prolonged periods of hunger over the past year. These students are reported in the qualitative data to have benefitted from the CEC efforts to raise funds for school fees and school flexibility in accommodating families with financial constraints. However, the quantitative data also suggests substantial programme impact on occupationally marginalized groups, such as pastoralists. While this group may have also reaped the benefits of a more active and capable CEC, there is also evidence in the qualitative data of external forces influencing nomadic enrolment rates. Multiple distinct interviews in Sanaag describe an influx in pastoralist groups into town, as herds were lost to the recent droughts. In fact, it is likely this phenomenon worked in confluence with community outreach efforts by the CEC to increase pastoralist families' inclination to send girls to school. There is also some evidence of CEC activities succeeding in both preventing and reversing family decisions to remove girls from school to tend to domestic responsibilities. However, chores and marriage continue to prevail as the main threats to girls' transition. And, despite CEC efforts to reach financial marginalized children, their lack of resources is frequently cited and school fees are a leading cause of school drop-outs – among both boys and girls.

5.5 Testing the Theory of Change – Transition

SOMGEP-T is designed to target the most marginalized girls, as discussed in the previous section, where we investigated differential program impacts across subgroups of respondents. The design of SOMGEP-T is based on a series of assumed connections between program activities and learning and transition outcomes, through activities' effect on teaching quality, community attitudes, attendance rates, school management, and girls' life skills. This Theory of Change was hypothesized prior to the baseline, and has been studied in each evaluation wave since that time.

In this round, we analyze whether the program's Theory of Change are supported by the changes we have observed in intervention communities since the program's inception, with specific reference to transition outcomes. It is important to note the methodological approach that we take, which mirrors – in some ways – the method used in our earlier analysis of learning outcomes. First, we identify a set of variables that capture the key aspects of the program's intermediate outcomes. In some cases this is simple, because a single variable captures the core of the intermediate outcome, as in the case of YLI scores and their measurement of life skills. In other cases, we employ more complicated formulations. The variables we focus on as predictors of transition outcomes are:

- Community attitudes – Strong agreement of caregivers with each of three statements – whether it is worth investing in girls' education even when funds are limited; that girls are just as likely as boys to use their education; and whether they listen to their girl when making decisions about her education. This measure of community attitudes is scaled from 0 to 3, where the highest possible score represents caregivers who strongly agreed with all three statements
- Life Skills – The Youth Leadership Index (YLI), for in-school girls. Because the YLI was not completed by out-of-school girls at baseline, we employ the life skills index for separate analysis of transition rates among OOS girls

- Attendance – We employ two measures. The first is binary, identifying whether a girl's caregiver stated that she has attended most days of school that year. The second is a continuous measure of the number of days missed in the last month, as reported by caregivers (right-censored in the data at a maximum of 10 missed days to avoid unnecessary leverage from outliers)

As these variables make clear, our analysis focuses heavily on individual-level factors.

Assessing the relationship between intermediate outcomes and transition is not always straightforward, thanks to the changing nature of the sample we use for transition analysis. Our approach is to estimate a regression predicting transition outcomes at ML2, based on a girl's characteristics at baseline. We take this approach because a girl's transition outcome at ML2 influences our measurement of attendance and life skills in that round. Girls who have dropped out of school at ML2 may have experienced a decline in self-esteem *because* they dropped out, rather than a decline in self-esteem driving their decision to drop out. In addition, it does not make conceptual sense to assess the relationship between attendance and transition rates in a single cross-sectional sample at ML2, because girls who have dropped out of school are, by definition, not attending school.

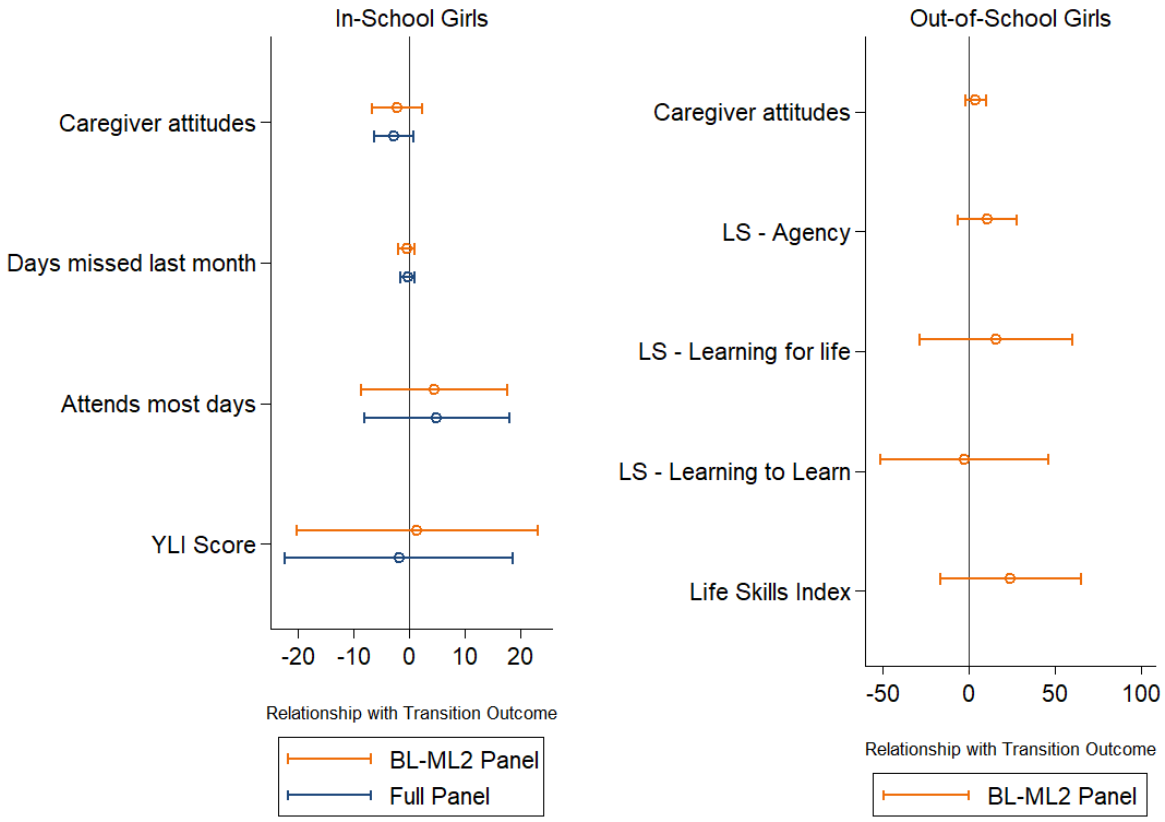
Conceptually, the question we want to answer is whether higher attendance, more self-esteem, and more positive community attitudes influence later transition outcomes. In the case of in-school girls, is a girl more likely to remain in school and progress through the grades if she attends school regularly and has greater self-esteem? In the case of girls who were out of school at the baseline, are they more likely to enrol in school by ML2 if their parents or caregiver are pro-education? These questions motivate our approach. By predicting ML2 transition outcomes on the basis of past characteristics, we can understand whether these characteristics drive decisions to remain enrolled, re-enrol, and so forth.

Note from the project: "Life skills" or "leadership skills development" should not be narrowly defined as 'self-esteem' as articulated in the paragraph above. CARE does define leadership skills as the combination of vision, voice, decision-making, organization and self-confidence. While increasing one's self-esteem may be beneficial, it may be insufficient in the absence of clear vision, planning, ability to voice one's interests and to negotiate for them.

We study the in-school and out-of-school cohorts separately, because their different starting points drive very different underlying transition rates at ML2. The figure below reports the results of a series of regression models. The left panel is restricted to the in-school girl cohort – those girls who were in-school at baseline. We report two sets of results: one utilizes all in-school girls who were successfully re-contacted at ML2, regardless of whether they were included in the ML1 sample. We refer to this sample as the BL-ML2 Panel (n = 716); the alternative is the sample of in-school girls who appeared in all three rounds of data collection, called the Full Panel (n = 529). For each intermediate outcome listed in the left panel, we estimated a separate linear regression, controlling for age, grade, disability status, female-headed households, treatment status, and experiences of household hunger over the previous year. In addition, we employ a full set of school fixed effects to account for school-specific factors that might influence transition rates.

The figure reports the regression coefficient representing the association between a given intermediate outcome variable and transition rates, where transition is a binary outcome. The bands on either side of the point are the 95 percent confidence interval for that coefficient; where these bands cross the vertical line, it means that the association is not statistically significant.

FIGURE 23: PREDICTORS OF SUCCESSFUL TRANSITION



The results do not support the program's Theory of Change in the context of transition, at least among in-school girls. For instance, a 1-point improvement in caregiver attitudes (measured on a 0-3 scale) was associated with a -2.1 point or -2.7 point – depending on the sample used – reduction in the likelihood that the girl would transition successfully at ML2. In short, more positive caregiver attitudes at baseline do not seem to translate into a higher transition rate for their girls. In other cases, such as YLI score, the effect size estimated is very small, and the uncertainty around the estimate is so large that we cannot draw any firm conclusions about its association with transition. Finally, we note that the one finding that is consistent – in direction, though not in magnitude – with the Theory of Change concerns attendance. Girls who reportedly attended most school days at baseline are more likely to transition successfully in later rounds. We view this finding as an indication that attendance and some transition outcomes actually operate on a spectrum, where poor attendance is actually a segue into non-enrolment in many cases. If this is true, it is not surprising that higher attendance rates are correlated with higher downstream transition rates.

As in the left panel, the right panel of the figure above does not offer much evidence in favor of the program's Theory of Change. The right panel studies the out-of-school girl cohort, with a sample size of just 368 girls tracked from baseline to ML2. Because none of these girls were enrolled in school at the baseline, they do

not have attendance records and we cannot estimate the impact of attendance on the likelihood of their transition.[103]

Among OOS girls, both caregiver attitudes and life skills are positively associated with higher transition rates, though neither relationship is statistically significant. A 1-point change in caregiver attitudes is associated with a 3.9-point higher likelihood of successful transition at ML2. Similarly, a 1-point change in a respondent's aggregate life skills score is associated with 0.24-point increase in the likelihood of transition at ML2. Note that we employ the life skills index in the analysis of OOS girls because the YLI score was not captured for this cohort at baseline. While the life skills index and YLI score are only moderately correlated (rho = 0.27 in a large sample of girls who completed both), the findings of the life skills section suggest that the program has affected them similarly.

Finally, we emphasize the relatively large substantive effect sizes estimated here, despite their non-significance in statistical terms. Among OOS girls from the baseline, successful transition rates are just 56.0 percent at ML2. A relatively minor improvement in caregiver attitudes is associated with a shift in transition rates from 56.0 to 59.9 percent, an unambiguously large impact.

The intermediate outcomes studied in this section seem to confirm, with significant caveats regarding the strength of the findings, the idea that caregiver attitudes, self-esteem and higher attendance rates are associated with more positive transition outcomes down the road. The evidence also seems to suggest that specific program outputs may drive higher transition rates. We performed additional analysis of the relationship between participation in a Girls Empowerment Forum (GEF) and later transition rates and the results are in line with our discussion, elsewhere, of the importance of GEFs. As the table below shows, GEF participation at baseline is associated with higher ML1 *and* ML2 enrolment rates; likewise, GEF participation at ML1 is associated with slightly higher ML2 enrolment rates. The correlation described here is not a function of different "starting points," as both GEF participants and non-participants in this analysis were enrolled in school at the time GEF participation was captured (either baseline, in the top panel, or ML1, in the bottom panel). This suggests that girls who participate in GEF activities are more likely to remain enrolled 1-2 years later.[104]

TABLE 65: DOWNSTREAM ENROLMENT RATES AS A FUNCTION OF PAST GEF PARTICIPATION

	GEF Participant (at Baseline)	Not a GEF Participant
ML1 Enrolment Rate, Full Panel	96.7%	95.5%
ML2 Enrolment Rate, Full Panel	96.7%	91.7%
ML2 Enrolment Rate, BL-ML2 Panel	94.4%	89.9%
	GEF Participant (at ML1)	**Not a GEF Participant**

[103] While many of the girls were enrolled when contacted for ML2 data collection, their attendance cannot be a predictor of successful transition, as it occurred *after* their successful transition. Moreover, we would only be able to measure attendance for girls who had re-enrolled.

[104] The finding is consistent with evidence presented elsewhere in this report regarding the importance of GEFs for girls' self-esteem and learning. The discussion in Section 4.9, above, highlights the fact that GEF participation may increase girls' support network at school. If girls make new friends or strengthen relationships with other girls and their teachers through participation in the GEF – as it appears they do – it would not be surprising that they would be more likely to remain enrolled in school as a result.

ML2 Enrolment Rate, ML1-ML2 Panel	91.8%	89.3%

Outside of GEF participation, another program output that we would expect to influence enrolment and transition rates is the provision of bursaries to students. SOMGEP-T provided scholarships to many students, seeking to encourage enrolment by reducing the financial barriers households face when attempting to educate many children. We investigated the relationship between bursaries provided at baseline and downstream enrolment decisions by the same girls. In contrast to GEF participation, there is no evidence that bursaries provided at baseline increased ML1 or ML2 enrolment rates for the girls who received them. However, this negative finding is almost certainly an artifact of how girls were selected to receive a bursary. The program targeted relatively disadvantaged girls, such as pastoralists and those coming from relatively poorer households. Insofar as these girls have lower underlying transition rates, it is not surprising that there is not a large positive correlation between bursary receipt and successful transition. Unfortunately, it is not possible to disentangle the effect of bursary provision on transition rates from the underlying lower level of transition observed in this group.

5.6 Conclusions

The transition analysis reported in this section sought to shed light primarily on the program's impact on transition rates since baseline. The evidence presented is strongly suggestive, but not entirely conclusive. Among girls enrolled at baseline, the program has had a small but meaningful positive effect on transition rates, relative to comparison schools, but this finding is not statistically significant. Among girls who were not in school at the baseline, the program has had a more substantial impact, raising transition rates by 12.6 points in total, but the small sample size available for analysis means that this result, too, is not statistically significant. Nonetheless, the two results, combined with the fact that intervention schools have seen an increased share of students remain enrolled even if they are held back a grade, certainly implies the program has influenced transition rates meaningfully for the better.

Moreover, while the analysis of subgroup-specific impacts is tenuous, the evidence seems to show that the program has had a particularly outsized impact on some of the most marginalized girls in the communities where they work. Gains in successful transition are greatest among girls from pastoralist households and girls whose households have experienced various types of economic privation over the past year. The program's outputs also seem to have specifically been associated with higher transition rates, and there is some mild evidence in favor of the program's Theory of Change in terms of transition. Our uncertainty regarding the program's impact on transition rates is not driven by contradictory findings, but by the relatively small sample size and noisy outcome studied, which means that even substantively large effect sizes are often statistically insignificant. However, based on the weight of the evidence across multiple analyses, the program appears to have had a considerable positive effect on transition rates in intervention communities.

6. SUSTAINABILITY

This section presents an analysis of project sustainability at each of three levels: community, school, and system. The results of this analysis are presented in narrative form in the subsections that follow. The triangulated analysis was used to generate a qualitative sustainability score (ranging from 0 to 4) for each of the key sustainability indicators identified in the SOMGEP-T Logframe. Sustainability scores for each indicator can range from 0 to 4, in line with the FM's MEL Guidance:[105]

- 0 – Negligible
- 1 – Latent
- 2 – Emerging
- 3 – Becoming established
- 4 – Established

Scores for indicators in the same level – community, school, or system – are aggregated into a single level score; in turn, those levels are averaged to produce a single sustainability score. These scores, for both baseline and midline, are reported in the scorecard table below.

While additional rationale for each indicator's score are provided in the more detailed narrative sections that follow, it is important to note that the scores are admittedly subjective. We assigned scores based on the totality of information available – in some cases, this was limited to a few qualitative interviews, while in other cases it included quantitative data from multiple respondent groups and a range of qualitative data. As a broad rule of thumb, where noticeable but not dramatic changes have been observed since the baseline, we tended to score indicators as one grade above the baseline. Where two indicators in the same level (community, school, or system) both had marginal improvements, we split the difference, again relying on our best, but subjective, judgment.

TABLE 66: SUSTAINABILITY SCORECARD

	Community	School	System
Indicator 1:	***Percentage of CECs actively engaged in mobilizing for girls' education through fundraising for payment of additional teachers' salaries and school supplies***	***Percentage of project target schools adhering to implementation standards for ALP, ESL, Numeracy and Remedial classes.***	***Inclusion of ALP in the national non-formal education frameworks***
	Baseline Status:	**Baseline Status:**	**Baseline Status:**
	9.5 % of CECs provide support for teacher salaries (18.8% intervention; 0% comparison).	39.7% of teachers in intervention schools reported using formative assessments. This assessment is limited to the same sample of schools employed at midline, to facilitate comparisons. (Comparison schools: 62.5%)	Not applicable because activities not established yet

[105] See "GEC-T MEL Guidance Part 2," pages 46-52, distributed May 2017.

	31.4% of parents report CECs provided financial support of some kind (fundraising, infrastructure, buying materials, financial support to students) to schools in the last 12 months (31.3% intervention; 31.6% comparison).	32.8% of teachers in intervention schools could show records of using formative assessments. (Comparison schools: 44.6%)	
Indicator 1:	**Midline round 1 Status:**	**Midline round 1 Status:**	**Midline round 1 Status:**
	30.2% of CECs provide support for teacher salaries (25.8% intervention; 34.4% comparison).	69.0% of teachers in intervention schools reported using formative assessments. (Comparison schools: 72.4%)	Some mention of ALP in REO interviews. Awareness among Ministry officials is clearly growing with increased discussion of alternative learning opportunities.
	The panel for household survey at Midline round 1 is exclusive of out-of-school girls. At Midline round 1, the wider panel between BL and ML2 is thus used.	34.5% of teachers in intervention schools were able to show records of the use of formative assessments. (Comparison schools: 24.1%)	
	47.6% of CECs, per head teachers, raise funds for school improvements (59.4% intervention; 35.5% comparison).		
Indicator 1:	**Midline round 2 Status:**	**Midline round 2 Status:**	**Midline round 2 Status:**
	15.9 % of CECs provide support for teacher salaries (21.9% intervention, 9.7% comparison)	91.2% of teachers in intervention schools reported using formative assessments. (Comparison schools: 91.2%)	Talk of ALP is very limited in interviews with MoE respondents. Many mentions of special efforts for including the nomadic population and disabled children. But most respondents do not view these as possible at this time, due to financial constraints and other operational challenges.
	54.9% of parents report CECs provided financial support of some kind (fundraising, infrastructure, buying materials, financial support to students) to schools in the last 12 months (57.9% intervention; 50.5% comparison).	35.1% of teachers in intervention schools were able to show records of the use of formative assessments. (Comparison schools: 22.8%)	
	46.3% of CECs, per head teachers, raise funds for school improvements (56.3% intervention; 35.5% comparison).		
	Percentage of parents in intervention schools indicating that CECs are functional	***Not applicable***	***No. of MOE departments engaged in support of girls' education from National to regional and district levels.***
	Baseline Status:	**Baseline Status:**	**Baseline Status:**

	71.9% of parents in intervention communities report a functional CEC, compared to 58.5% of parents in comparison communities.	***Not applicable***	Interviews reveal a systemic lack of funds. Local schools do not have enough money to maintain facilities and pay their staff partly because CECs do not have the funds to support their schools. The CECs do not have funds because little to no financial support is given to them by the MoE. Some in-kind support from MOEs was provided. 31% of schools had CEC Management plan, taken as proxy of MOE involvement.
Indicator 2:	**Midline round 1 Status:**	**Midline round 1 Status:**	**Midline round 1 Status:**
	The panel for household survey at Midline round 1 is exclusive of out-of-school girls. At Midline round 1, the wider panel between BL and ML2 is thus used.	***Not applicable***	Two out of six REOs explicitly confirmed the existence of gender units. Ongoing difficulties recruiting and retaining female teachers and high-quality teachers. Mixed evidence on mainstreaming of improved teaching practices, but these appear to be emergent at this phase. Minimal evidence of MOE efforts to encourage female pre-service graduates.
	96.8% of schools have a functioning CEC (head teacher survey) (96.9% of intervention schools, 96.8% of comparison schools)		
Indicator 2:	**Midline round 2 Status:**	**Midline round 2 Status:**	**Midline round 2 Status:**
	66.5% of parents in intervention communities report a functional CEC, compared to 59.2% of parents in comparison communities.	***Not applicable***	Gender units seem to be more active than at previous Midline. Their financial sustainability, however, cannot be assessed. Various activities were detailed relating to teaching quality and addressing retention and transition concerns. However, lack of funding complicates the efforts and their uniformity cannot be assessed at this stage. When looking at quantitative measures of CEC activity as a proxy for MoE activity, all signs are positive.
	According to the head teacher survey 96.8% of schools have a functioning CEC (100% of intervention schools, 93.6% of comparison schools)		
Baseline Sustainability Score*	**1**	**1**	**0.5**

Overall Sustainability Score**		**0.8**	
Midline round 1 score*	**1.5**	**2**	**1.5**
Midline round 1 Overall score**		**1.7**	
Midline round 2 score*	**1.5**	**3**	**2.0**
Midline round 2 Overall score**		**2.2**	
*Range 0-4	**Average of the three level scores		

INDICATOR 1 – COMMUNITY LEVEL – PERCENTAGE OF CECS ACTIVELY ENGAGED IN MOBILIZING FOR GIRLS' EDUCATION THROUGH FUNDRAISING FOR PAYMENT OF ADDITIONAL TEACHERS' SALARIES AND SCHOOL SUPPLIES

Community Education Committees (CECs) are local structures that oversee the operation of a primary school in a given community. The committees typically consist of the school's head teacher, community leaders, and parents. CECs support the operation of schools in a number of ways, by monitoring student and teacher attendance, promoting enrolment and attendance, raising awareness of the importance of education, liaising with religious leaders and other individuals of influence in the community, and providing material support to the school, among other tasks.

The importance of the CECs for school governance and management has already been discussed elsewhere in this report as the committees are involved in monitoring enrolment, attendance and teaching quality. In terms of the sustainability of project efforts, the activities of the CEC in terms of securing material and in-kind support for the schools that they operate in. CECs are expected to raise funds from their own resources, by mobilizing community members to donate, and by seeking support from outside organizations, including NGOs and the government. It is safe to say that the sustainability of schools following the conclusion of SOMGEP-T requires continued material support from CECs.

In order to assess the level of CEC activity in funding the schools, various questions were looked at. The indicators at Midline round 1 and 2 are slightly different from the baseline evaluation, because midline data collection did not include surveys of teachers. At the baseline, teachers were asked to assess the level of support provided to the school by the CEC and the community more broadly over the past year. As at Midline round 1, the data collection at the second Midline did not include surveys with teachers. Instead, head teachers were asked about the CEC's contribution to teacher salaries, student scholarships and their efforts to secure funding for improvements of school facilities. In addition, the household survey with the caregiver

and the head of household of the cohort girls included questions on the role CECs play in fundraising, improving school infrastructure, supporting students financially, and purchasing learning materials.[106]

CEC Contribution to teacher Salaries

The first outcome we analyse is the extent to which CECs raised money and paid the salaries of teachers in their schools. Teacher salaries are an area of particular need, because teachers are often paid late; over time, this contributes to discontent among the teaching staff, increasing absenteeism and even prompting teachers to seek other opportunities. As an REO in Puntland articulates:

> *There are challenges and difficulties; when the number of students increase and the number of teachers also increases, places far from the cities cannot afford paying teacher's salary and these teachers depend on incentives provided by the ministry and some from the world bank. One of the main challenges we have is when you train a teacher, he seeks better jobs and you cannot convince him to stay in such schools as located in the rural village.*

Indeed, the problems with salary payments for teachers were very frequently mentioned by the REOs interviewed. As such, addressing this issue in a sustainable manner is of utmost importance. To assess the extent of CEC support for teacher remuneration, head teachers were asked to indicate the share of male and female teacher salaries that CECs paid over the previous year. As shown in the table below, the overall share of CECs who provide support for teacher salaries is very low.

TABLE 67: CEC SUPPORT FOR TEACHER SALARIES, BY ROUND

CEC provides support for teachers salaries	BL	ML1	ML2
Intervention	18.8%	34.4%	21.9%
Comparison	0%	25.8%	9.7%
Overall	9.5%	30.2%	15.9%

While a sharp increase in salary support was observed between Baseline and ML1, at ML2 the share of head teachers reporting school CECs paying a portion of either male or female teachers' salaries has declined to 15.9 percent. At ML1, the same share was 30.2 percent. The decline in CEC support for salaries is somewhat similar across intervention and comparison schools whereby support in comparison schools decreased by 16.1 and in intervention schools by 12.5 percentage points.

In terms of gender differences of salary support, while overall more male teachers are still being supported at ML2 – 12.7 percent of head teachers say the CEC supports male teachers – this seems to be a reflection merely of the higher prevalence of male teachers in general. Indeed, when limiting our attention to schools that have female teachers (28 out of 63 schools at ML2), CEC support for female teachers is in fact more common. In these schools, 17.9 percent of the head teachers report support for female teachers while 10.7

[106] The former measure, from head teachers, was only collected at the midlines. The latter measure has the advantage that parents in both the baseline and midline were asked to describe the activities of their CECs, providing the opportunity to study improvements in CEC engagement over time.

percent report support for male teachers.[107] Support for female teachers continues to be more prevalent among intervention schools, where 4 out of 18, or 22.2 percent, of the schools that had female teachers had CEC support for them. Meanwhile, in the comparison group, only 1 out of the 10 schools with female teachers seemed to have CEC financial support for the female teachers. As such, there is evidence that the program is supportive towards female teachers. This is a very important element of the girl's education project as the prevalence of female teachers is directly linked to girls' enrolment, as will be discussed below when considering the MoE's involvement in the program.

CEC raising funds for school improvements

Beyond paying a portion of teacher salaries, CECs also take primary responsibility for raising funds for school improvements or repairs. While this element of their activity was not captured at baseline, the level of CEC activity in raising funds between Midline round 1 and 2 is largely similar and remains very high. At Midline round 1, head teachers in 56.3 percent of the intervention schools reported that the CECs had raised funds for school improvements of some kind, according to head teacher reports, compared to 35.5 percent of CECs in comparison areas. No significant change can be detected in this regard in the last year.

CEC provision of scholarships

Overall at Midline round 1, fewer girls are receiving scholarship support than at the previous midline. While at the last Midline, 58.7 percent of the headteachers reported that some of the girls at their school were receiving scholarship support, at Midline round 1, only 34.9 percent did.

TABLE 68: SHARE OF GIRLS WHO RECEIVE SCHOLARSHIP SUPPORT

	BL	ML1	ML2
Intervention	53.1%	87.5%	56.3%
Comparison	3.2%	29.0%	12.9%
Overall	28.6%	58.7%	34.9%

The decline can be observed in both intervention and comparison schools, though decline in intervention schools is actually more drastic, with over 30 percentage points compared with the decline of 16.1 percent in comparison schools. Similarly, the decline seems to have taken place equally across different sources of funding. The CEC share remains at approximately one fifth of all scholarship support.

TABLE 69: SHARE OF CEC IN SCHOLARSHIP SUPPORT

	BL	ML1	ML2
Intervention	14.3%	22.2%	17.4%

[107] In our view, this is the relevant distinction: when female teachers are present in a school, their salaries tend to be supported more by CECs than those of male teachers, though the gap between support for female and male teachers is not particularly large. However, many schools do not have any female teachers, and female teachers are vastly outnumbered by male teachers in every school in the sample. Therefore, the fact that CECs support female teacher salaries at even similar rates – and higher rates where at least one female teacher is present – is actually a testament to the fact that CECs are supporting female teacher salaries. The broader problem is, arguably, one of recruitment and hiring, given the overall underrepresentation of female teachers.

Comparison	33.3%	25.0%	30.8%
Overall	18.9%	22.7%	20.3%

It has already been discussed at the last round of evaluation that much of the CEC support to students may go uncaptured as it does not officially qualify as scholarship support for the formal and long-term connotations of the word. For example a discussion with a CEC in one school at Midline Round 2 first had the discussants say that the CEC had collected money to provide free tuition for some of the students, only to then argue that no scholarships had been provided at the school.[108]

However, assuming that the understanding of the word has not changed much since the last data collection, the picture that emerges is that when looking at CEC support for schools and their level of involvement, the trend is currently negative. This is potentially concerning for the sustainability of the project. Indeed, between ML1 and ML2, the changes overall have been negative when looking at CEC financial activity. The changes between baseline and ML2 have been marginal. Further, in fact, as the table below shows, the scholarship support overall has seen more positive developments in comparison schools, while the positive changes in intervention schools in terms of CEC share in scholarship support are negligible.

TABLE 70: IMPACT OF PROGRAM ON CEC FINANCIAL SUPPORT FOR EDUCATION

	Intervention				Comparison				Difference in differences
Indicator	BL	ML1	ML2	Difference	BL	ML1	ML2	Difference	
Girls who receive scholarship	53.1%	87.5%	56.3%	**+3.2**	3.2%	29.0%	12.9%	**+9.7**	**-6.5**
Share of CEC in scholarship support	14.3%	22.2%	17.4%	**+3.1**	33.3%	25.0%	30.8%	**-2.5**	**+5.6**
CEC raises funds for school improvements	N/A	59.4%	56.3%	**-3.1**	N/A	35.5%	35.5%	**+0.0**	**-3.1**

Indeed, at *prima facie* it thus seems like the sustainability of the program is on less robust ground this year compared to Midline round 1. However, it is important to note the following. Firstly, some of the REOs interviewed highlighted the CEC's involvement in other fund-raising activities including seeking contributions from the diaspora. Indeed, when looking at the head teacher survey data, the share of schools that have received outside support has steadily increased since baseline. At baseline 27.0 percent of the schools reported receiving support from outside groups whereas at Midline round 2 38.1 percent do.

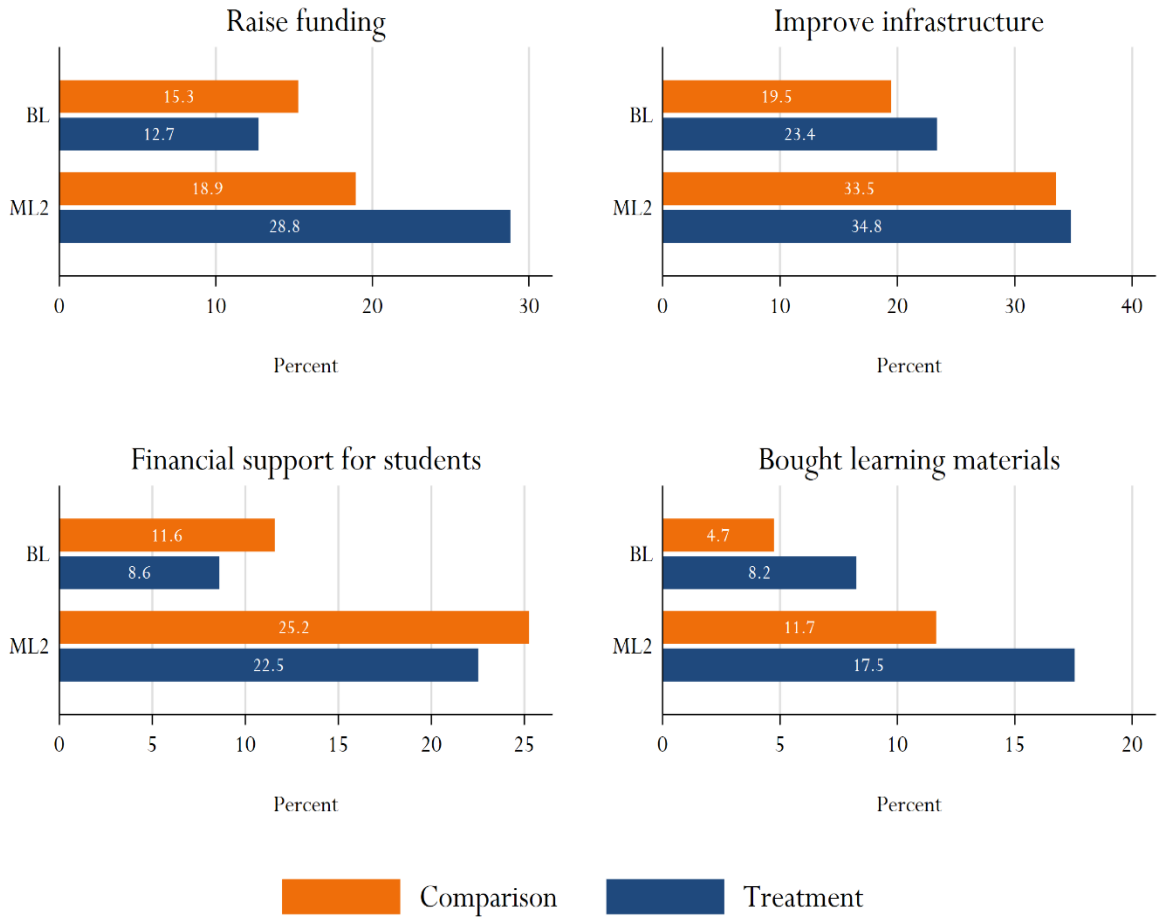
Scholarships are generally used to cover the cost of education. The costs in the education for the children, in turn, are largely due to the need to cover teacher salaries. When looking at the development of scholarships, it is thus necessary to see if schools continue to charge tuition fees. When looking at the cost of education, it turns out that fewer schools in the sample are charging feeds for enrolment. At Baseline, 34.9 percent of the schools charged for enrolment. The same share of schools charged enrolment fees at Midline round 1. However, at Midline Round 2 only 14.3 percent of the schools did this. As such, it is very possible that given

[108] FGD CEC, Puntland, 201.

the increased funding from outside sources, fewer schools need to charge their students fees. This might similarly reflect on the decreased need for CECs to provide support for teacher salaries.

Indeed, at this point it might be worthwhile to triangulate some of the findings using the household survey data. The results there contradict somewhat the picture emerging from the head teacher surveys.

FIGURE 24: FINANCIAL SUPPORT OF SCHOOLS BE CECS OVER THE LAST YEAR, ACCORDING TO PARENTS



When we look at the parents' perceptions of CEC activities, all indicators suggest that all activities have increased since baseline. The graph above and the table below present largely the same information. When looking at parents' responses to whether the CEC raises funds, improves infrastructure, provides financial support for students or has bought learning materials for the schools, findings for all four areas of financial support suggest a clear increase from Baseline to Midline round 1.[109] The findings show an increase across both intervention and comparison schools. Only as regards CEC raising funding the difference between intervention schools and the comparison group is statistically significant – intervention schools seem to have

[109] Midline round 1 is excluded from this analysis in order to have the largest panel of data as Midline round 1 did not include out-of-school girls or their parents.

increased their CEC fund-raising activities much more than comparison schools. For sustainability of the project, the fact that increases are perceived across all communities is encouraging.

TABLE 71: IMPACT OF PROGRAM ON CEC ACTIVITY LEVELS

	Intervention			Comparison			Diff-in-Diff
	BL	ML2	Difference	BL	ML2	Difference	
Raise funding	12.7%	28.8%	+16.1	15.3%	18.9%	+3.6	+12.5**
Improve infrastructure	23.4%	34.8%	+11.4	19.5%	33.5%	+14.0	-2.6
Financial support for students	8.6%	22.5%	+13.9	11.6%	25.2%	+13.6	+0.3
Bought learning materials	8.2%	17.5%	+9.3	4.7%	11.7%	+7.0	+2.3

When looking at the qualitative data, discussions with CEC members revealed that approximately half of them were involved in taking on some of the financial burdens of tuition of some of the more marginalized children. In such cases the CEC would raise the funds and either cover the teachers' salaries that way, or substitute the tuition fees of the poorer children.

> *We have also become aware that many parents don't send their daughters to school due to school fees and we finally waived the school fees for these families.*[110]

> *We raise funds; for example, if three students come from a house we ask to pay additional $2 per student. So we raise this, and support the teachers with it. We support students who can't afford the full fee, and pay if the salary of the teachers is late.*[111]

However, in about half of the schools the CECs said they had no resources to support the least well-off either through paying fees or giving scholarships. It can be assumed that this is the case with most of the more resource-stripped and rural communities.

> *There is nothing that we do for other people because when people are better than each other financially is when they help each other.* [112]

> *Residents of the village are poor and have no money.*[113]

Focus group discussions with CEC members reveal that the committees are involved in a variety of activities. Yet, most of the discussants complained about the lack of resources which in all schools meant that there was

[110] FGD CEC, Somaliland, 101
[111] FGD CEC, Puntland, 201.
[112] FGD CEC, Somaliland, 103.
[113] FGD CEC, Somaliland, 106.

very little tangible that the CECs were doing in terms of addressing barriers for nomadic and disabled populations. Approximately half of the CECs substituted the education costs of marginalized students or provided scholarships.

In terms of increasing the enrolment of disabled or pastoral children, the CECs detailed some awareness-raising activities, but by and large the committee members were of the opinion that addressing barriers for these groups was beyond their means.

> *We help them [pastoralists] if they come to us, but we don't go to them.*[114]

All discussants talked about the importance of supporting the disabled, but beyond some limited accounts of supporting in any way that is possible, most CECs said that the school had no disabled children. Only one discussion in Somaliland mentioned a child who had received supporting equipment by CEC to be able to study. In Somaliland one discussion references various organisations and government coming to ask about disabled children in the community, but these children were not enrolled at school and the CEC was not involved in any follow-up action. It is important to note, however, that references to children with disabilities may still be very much focused on physical disabilities affecting mobility. Indeed, addressing the issue was largely seen as requiring a level of resources that the CECs do not currently have:

> *As CEC we believe that we couldn't able to support them because it needs big tangible support, the disabled students need facility equipment in the school, also they need a car to take them to and from school.*[115]

In general, while there was evidence from about half of the schools that the CEC was involved in some fund-raising activities, there was no evidence that this financial support would have increased in the past year. In fact one community said that they found it more difficult in the past year. The majority of the CEC discussions indicated that the financial situation in the communities was largely the same as the year before, and as such the CEC has not had an opportunity to raise more funds than before.

Broadly speaking, the analysis here leaves a rather inconclusive view of the CEC financial support. The head teacher surveys indicate that while CEC share of teacher salaries increased between Baseline and the first Midline, it has since declined. Moreover, the level of scholarship support across schools is currently more than 20 percent lower than at Midline round 1. The CECs' share of the scholarships has also declined very slightly.

However, it is possible that this can be explained through increases in other sources of funding for the schools. Indeed, at baseline 27.0 percent of the schools reported receiving support from outside groups whereas at Midline round 2 38.1 percent do. It could be that CECs are providing less teacher salary support and fewer scholarships if other sources of funding are available.

Similarly, both qualitative interviews and household survey data in fact suggest that CECs have been more active in raising funds and in other forms of financial support since the baseline. For all the indicators apart from fund-raising the changes are happening across both intervention and comparison schools. Given the discrepancy in the different data sources, we have to conclude that the findings regarding the sustainability of

[114] FGD CEC, Puntland, 201.
[115] FGD CEC, Puntland, 107.

this indicator at the community level are inconclusive.[116] In other words, we cannot say conclusively that the intervention schools are becoming more sustainable than the comparison schools. Yet, for most of the variables that we have looked at, the trend is positive indicating that indeed both sets of schools are becoming more sustainable. For example, community members in both types of communities reported that their CECs were more active in fund-raising, improving school infrastructure, providing financial support for some of the students, as well as acquiring learning materials. Thus, the trend is positive indicating increased sustainability. Simultaneously it is doubtful whether the schools are yet in a situation where they could survive without outside funding. For example, as it pertains to sustainability coming from schools securing their funding from the CECs, particularly through teacher salaries, the current situation is not very encouraging as only approximately a fifth of the schools have CECs contributing towards teacher salaries.

INDICATOR 1 – SCHOOL LEVEL – ADHERENCE TO IMPLEMENTATION STANDARDS FOR ALP, NUMERACY AND REMEDIAL CLASSES

This evaluation did not collect direct measures focused on implementation of ALP, numeracy and remedial class teaching standards. Our assessment of teaching quality tended to focus on teachers' demeanour and interactions with students, the level of classroom participation, the use of corporal punishment, and so on, rather than adherence to specific standards, such as those set out in CARE's teacher training programs. One area of implementation that the evaluation did capture was the use of formative assessments, which is a system of continual evaluation of student comprehension using a wide range of data points – e.g., interaction in the classroom, short assignments to gauge comprehension, or listening in on group work conversations. Formative assessments are widely considered a critical component of improved learning outcomes, as they help teachers identify students who are falling behind more quickly, and identify topic areas of particular concern. As such, they form part of the package of teacher practices that would contribute to positive learning outcomes as well as retention due to the classroom being accommodating to more students.

CARE's teacher training programs emphasize the use of formative assessments. Their use is also further explored in the teaching quality section of this report. The focus on formative assessments in this section should be viewed as a proxy measure of increased teacher professionalism in schools. The assumption is that it would lead to improved learning outcomes in the long-term as well as would contribute positively towards transition and retention which are both linked to teaching quality.

TABLE 72: IMPACT OF PROGRAM ON REPORTED USE OF FORMATIVE ASSESSMENTS

	Intervention				Comparison				Diff-in-Diff
Indicator	BL	ML1	ML2	Difference	BL	ML1	ML2	Difference	
Report using formative assessments	39.7 %	69.0 %	91.2 %	+51.5	62.5 %	72.4 %	91.2 %	+28.7	+22.8
Records of formative assessments shown	32.8 %	34.5 %	35.1 %	+2.3	44.6 %	24.1 %	22.8 %	-21.8	+24.1

As the table above shows, more teachers report using formative assessments now compared to baseline and Midline round 1. Across both intervention and comparison schools, 91.2 percent of the teachers said that

[116] It should also be noted that the questions on different financial activities of the CEC in the household survey were only asked if the community members had indicated that there is an active CEC in their community. Fewer community members did at Midline round 1 compared to the Baseline.

they used to formative assessments. This is up by 51.5 percent in intervention schools and 28.7 percent in comparison schools since baseline. The difference in differences is 22.8 percentage points in favour of the intervention schools.[117] It thus seems that the use of formative assessments has become much more prevalent in intervention schools than in the comparison schools.

As was noted at Midline round 1, teachers may overstate their use of formative assessments, either because they do not fully understand the term and its meaning, or because they are aware that the use of formative assessments is desirable, and they do not want to admit they do not use them. To then better capture the actual use of the assessments, researchers also asked teachers whether they could provide evidence they used formative assessments, typically in the form of example short exercises or notes taken on student performance. When looking at the cases where the teachers are able to show the researcher their formative assessments, the advances made are much less visible. The share of teachers who can show their use remains largely the same as at Midline round 1 in intervention schools and has increased slightly since baseline. Meanwhile, for the comparison schools the share has actually decreased markedly since the baseline, going from 44.6 percent to 22.8 percent.[118]

As such, at Midline round 1 the evidence does not conclusively support increases in the use of formative assessments. While the recorded use of formative assessments is more prevalent in intervention schools, and has in fact declined in comparison schools, this difference is not statistically significant. Much emphasis should thus be put on this element in the late stages of the project.

INDICATOR 1 – SYSTEM LEVEL - INCLUSION OF ALP IN THE NATIONAL NON-FORMAL EDUCATION FRAMEWORKS

Next, in terms of system level sustainability we looked at the qualitative data as coming from interviews with REOs of the Ministry of Education in the different project regions. The focus here is on the Ministries' inclusion of ALP and in general the provision of learning to those who require specific efforts – in particular the pastoralists and the disabled. The assumption here is that for education for the marginalised to be more sustainable, systemic changes are required that would allow for these children to be part of the education system without additional interventions.

As at Midline round 1, the recognition of ALP among the ministry staff seems very minimal. As in the last round of evaluation, only one REO explicitly mentioned the ALP program by name:

> *If the girl is older than the school age, we have set up a recovery program called ALP, so we have done two parts of the ABE that the girl can adapt, and we have its curriculum. A part of the ALP education is that people who had left school will be taught for a year and then tested, to check the suitable level...*[119]

[117] The difference in differences estimate is marginally significant at the 5 percent level (p value 0.061).
[118] The difference in differences estimate, however, is not statistically significant. However, the dataset used for the analysis has a very small sample size as only up to 2 classroom observations were conducted per school. As such, the threshold for statistical significance is relative difficult to reach.
[119] REO, Sool, Somaliland

The same respondent viewed ALP very much in relation to education for disabled children saying that "the SOMGEP project has found that many children at home have a problem, and now the agency has made a commitment to do something about it."

Indeed, while ALP is not widely cited by respondents, most of the respondents talked about different efforts to expand the educational system's reach to previously excluded groups such as pastoralists and disabled children. While those quotes do not necessarily refer to ALP, which is not designed as a program to support children with disabilities, they reflect the awareness raised by the program regarding the possibilities to support this subgroup of the population. There are some accounts of particular activities that may have been successful. The REO in Galmudug for example, spoke of the ministry training 100 teachers in sign language to be able to teach deaf children. REOs in Somaliland mentioned that the ministry now has a dedicated program for disabled children, but not much information was provided about its activities and reach. Overall, it seems that awareness of the disabled is growing among the ministry staff, but the actions to address their exclusion are lagging behind.

> *Nothing is changed [in education for disabled children] since there is no strategic plan from the government.*[120]

> *I can say, nothing has been done for the [disabled] children, the first activity we are planning is to train a couple of teachers.*[121]

An REO in Puntland explained that the MoE is very weak when it comes to addressing special needs. He argued that there should be special schools for disabled students, yet he said only one such school exists, and it is severely under-resourced. The school caters for 40 students, all with hearing disabilities.

An REO in Somaliland captures the complexity of the problem with addressing the question of including disabled children in education:

> *We have been distracted, we have slacked and forgotten about the disabled, we were busy building the schools, and finding teachers. We haven't gotten time for the disabled. Today, I think that we can find time for them and we have now started raising awareness, to sensitize the community that if this person learns he can add something to the community and can study and improve, because if the community is aware a lot can improve. They are being held in the houses so we should break down the doors because the parent may say that the able one should first work hard and improve, while the disabled one is will be dealt with later and lock him up in the house[…]. We have talked about this with CARE and all the other organizations that we raise awareness from house to house and share it through the media. What comes next is to build infrastructure that makes learning easier for the disabled, like at the entrance of the classes, before steps they should at least have their own door that makes it easy for even the wheelchair to pass through. The next point is that*

[120] REO, Mudug, Galmudug
[121] REO, Karkaar, Puntland

> *teachers should be trained in teaching the disabled children and knowing the ways of doing so.*[122]

According to another REO in Somaliland, the MoE is currently constructing a school for disabled children. Yet, while the above quote highlights the issues around mobility and the challenges of education for those who have a physical disability, it is unclear if efforts to include disabled children are taking into account those with other types of disabilities, for example in cognition or mental health.

The inclusion of pastoralist communities is challenging mainly in terms of retention:

> *The main challenge is the one from nomads because there are many nomads. We have seen a village that had 100 children enrolled in school. After two months, you will see that only 30 of them are here.*[123]

When it comes to pastoralists, the ministries talk of some practical solutions that they have implemented in order to partially address the problem. An REO in Puntland explained that pastoralists are taught through a mobile teaching system in the rainy periods, but during droughts everyone still migrates, consequently stopping the education there. In this regard, another REO outlines a practical solution that is being implemented in his region of Somaliland:

> *If the student moves, we ask the principle to send with him the list of where the student was at so that they will then take that list to the village and the student can continue education from where he left off. If the student moves again from that place, they are given another list so that they don't miss out on education.*[124]

In general, the interviewees lament the lack of funding that makes it difficult for the ministry to successfully bring disabled children and pastoralists into the educational system, and all agree that the efforts to address special needs groups are wanting.

In sum, talk of ALP specifically is very limited in interviews with MoE respondents. Many mention efforts for including the nomadic population and disabled children. But most respondents do not view these as sufficient at this time, due to financial constraints and other operational challenges.

INDICATOR 2 - COMMUNITY LEVEL - CEC FUNCTIONALITY AND ACTIVITY LEVELS

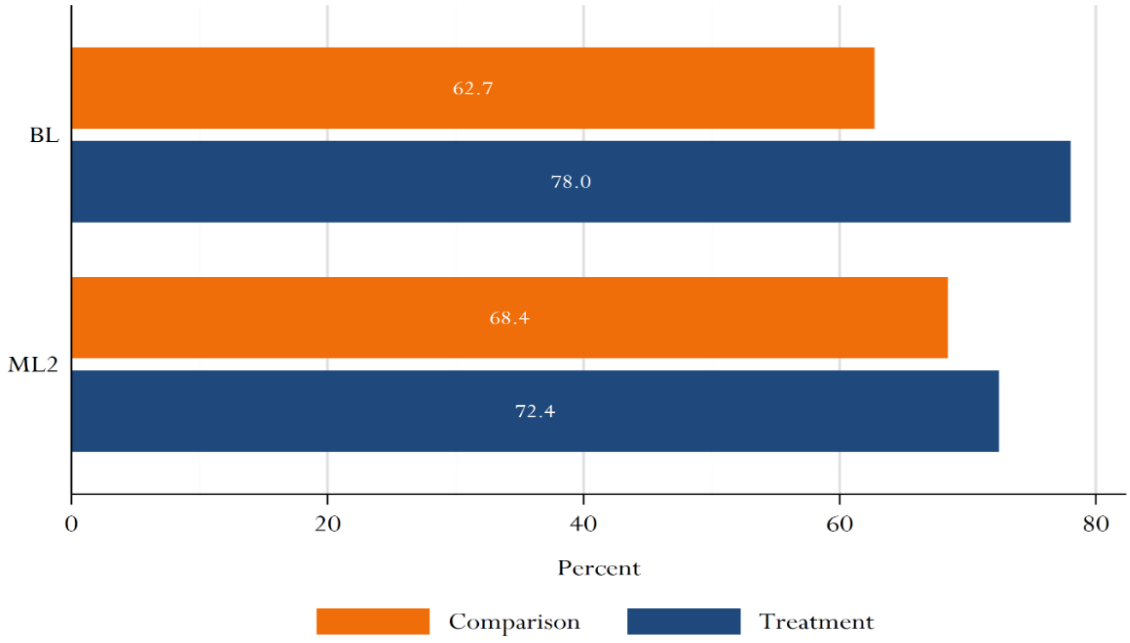
The second community-level sustainability indicator focuses on the extent to which CECs are perceived as functional by parents and community members. In line with the previous sustainability indicator, the CECs are a fundamental part of ensuring the continued sustainability of education efforts as they provide financial support and other activities that are important for school management, and act as a go-between connecting the school to the Ministry and to the community. Given that CECs are responsible for a significant share of

[122] REO, Somaliland
[123] REO, Somaliland
[124] REO, Sanaag, Somaliland

school oversight and management, parents should be broadly familiar with their activities, particularly if they are fulfilling their expected role.

When looking to capture this indicator through the household surveys, the measure is rather direct, whereby parents are asked whether their child's school has "a CEC that helps with school-related matters."

FIGURE 25: SHARE OF SCHOOLS WITH CEC



The indicator presents a fairly low bar, in that it does not specify an activity level that constitutes "active" or "functional"; rather, it asks whether a CEC exists and whether they help with school-related matters, large or small.

The findings at this Midline seem to reflect negative developments when looking at household survey data. The share of parents reporting a functioning CEC has declined by 5.4 percent in intervention schools and increased very slightly in the comparison group. The share of respondents who report a CEC in their community remains at about two thirds of those in intervention communities, and 59.2 percent in the comparison group. The findings in the community survey are not conclusive to say that the CEC is becoming more active in the community. In fact, the share of community members saying that there is a CEC at the local school has decreased in intervention communities, bringing it closer to the comparison sites where the share has very slightly increased.

It is possible that at the initial beneficiary selection phase of the program the CEC saw more interaction with the community and their activity has been less visible to the parents since then. It is also possible that the community members are genuinely not aware of what happens at the schools.[125] At the Midlines this information was thus further substituted by asking the head teachers if their schools had a functioning CEC in place. This information can be considered relatively reliable as the head teachers themselves tend to be part

[125] We also checked these findings with only parents of in-school girls. The findings were much in line with what is presented here.

of the Community Education Committee, and as such should be able to accurately reflect the situation. It should be noted, however, that the head teachers are likely aware that the program deems the establishment of the committees desirable, which makes their responses somewhat prone to desirability bias. However, given that the parents might not always know the exact situation regarding the existence of the CEC, it is advisable to look at both data sources to get a good reflection of the situation on the ground.

TABLE 73: IMPACT OF PROGRAM ON CEC FUNCTIONING, PER CAREGIVERS AND HEAD TEACHERS

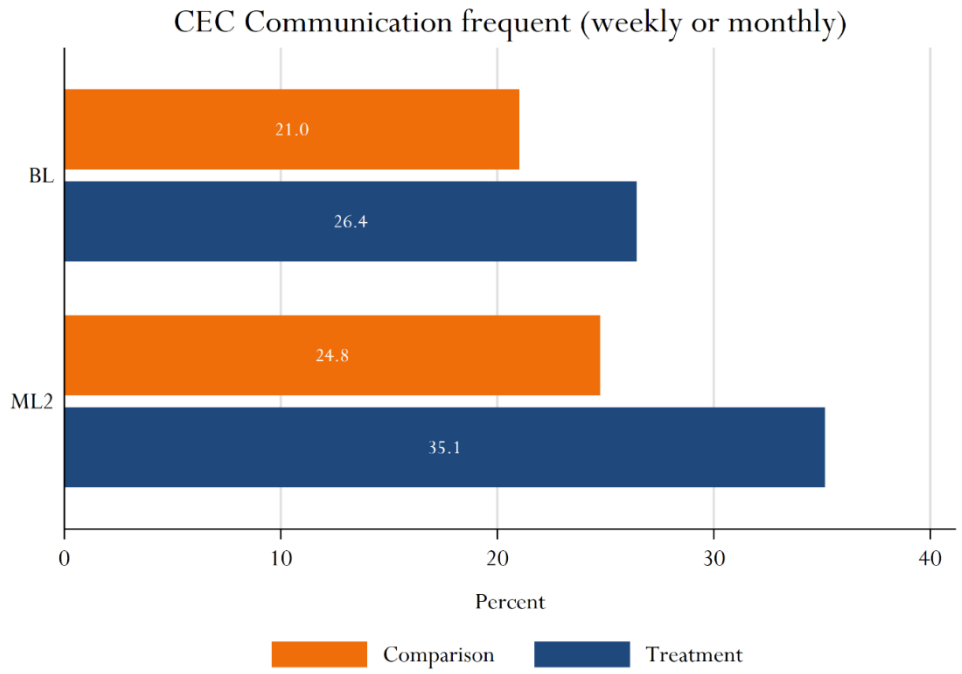
	Intervention				Comparison				
Indicator	BL	ML1	ML2	Difference	BL	ML1	ML2	Difference	Diff-in-Diff
School has functioning CEC - HH survey	71.9%		66.5%	-5.4	58.5%		59.2%	+0.7	-6.1
School has functioning CEC - Head teacher survey		96.9%	100.0%	+3.1		96.8%	93.6%	-3.2	+6.3

The findings from the household surveys are the reverse of the head teacher data. While community members around intervention schools still are more likely to report that there is a CEC at the school, the share has actually declined from baseline by 5.4 percentage points, while the share of CECs reported among comparison schools has largely stayed the same, and slightly lower than intervention schools.[126] Meanwhile, according to the head teacher surveys, between Midline round 1 and 2, the number of schools with a functioning CEC has increased by one school in the intervention group and declined by one in the comparison group. Currently all intervention schools have a functioning CEC whereas in the comparison group two schools do not have the structure in place.

Other measures of CEC activity

To get a more comprehensive view of the CECs, we also asked parents about CECs communication with them. While the number of CECs reported active by community members has decreased, when respondents were asked how frequently the CEC communicated with them when the CEC was in fact in place, the activity level of those CECs had increased. The share of parents in intervention areas who said their CEC communicated with them monthly or weekly rose from 26.4 percent to 35.1 percent. The development is also positive in comparison communities but less marked. While the share of those saying that the CEC communication is frequent in their communities has increased somewhat in the intervention communities while remaining the same in the comparison locations, this difference is not statistically significant.

[126] The difference in differences is negative, and marginally statistically significant (p=0.073) indicating that intervention schools have seen a negative development in CEC activity. The findings are largely similar even if the analysis is conducted only with data from parents of in-school girls, who are arguably more aware of what is happening at the local school.



We also looked at the head teacher reports on whether the school has a management plan or whether the CEC has conducted monitoring visits at the school. These findings are also considered under the systemic changes of this indicator and will thus only be mentioned shortly here. However, the findings there are clearly positive.

An increased share reported, at midline round 1, that their schools had a School or CEC Management Plan. This was true for both intervention schools (from 43.8 at Baseline to 71.9 percent at Midline round 1) and comparison schools (from 12.9 to 41.9 percent). Similarly, the share of head teachers who report that a CEC member has visited the school for monitoring purposes in the last year increased from 59.4 to 87.5 percent from baseline to midline round 1 in intervention schools and 22.6 percent to 67.7 percent in comparison schools. These positive changes have been sustained since Midline round 1, remaining exactly the same for both groups of schools at Midline round 1.

TABLE 74: IMPACT OF PROGRAM ON MEASURES OF SCHOOL MANAGEMENT

	Intervention				Comparison				
Indicator	BL	ML1	ML2	Difference	BL	ML1	ML2	Difference	**Diff-in-Diff**
School management plan	43.8%	53.1%	71.9%	+28.1	12.9%	38.7%	41.9%	+29.0	-0.9
Monitoring visits by CEC	59.4%	87.5%	87.5%	+28.1	22.6%	67.7%	67.7%	+45.1	-17.0

The qualitative interviews with REOs overall revealed that they believe in the importance of the CECs, and view them as very active and an important partner in the relationship between the Ministry, community and school.

It [CEC] raises awareness in every area. It encourages and motivates the students. They are also aware of the students' absence from the school. Writing those things down has a big role, it helps in motivating the school members. They repair what's broken. So, the committee has a big role in quality education.[127]

While REOs overall viewed the CECs as active, there were no accounts explicitly stating that the Committees would have become more active since baseline, in fact some argued that the CECs are more active in big cities.

Some have come to build buildings and the CEC have been involved in contributing to the school. The CEC decided not to be a dysfunctional committee. Some built classrooms, some built halls, some built new schools; and we now manage them.[128]

I cannot say all the committees are performing effectively, but for those CECs that have a chance and are in the big cities, they perform well and they focus because others might see what they are up to.[129]

Across the board, REOs called for more training for members of the Community Education Committee, arguing that it would improve their work.

The achievement of the CECs is good but they need trainings. The person should be understanding what he's being told.[130]

The CEC members themselves also called for training. They also detailed various activities that they are involved in at the school level, including monitoring, awareness-raising, fund-raising and work against corporal punishment as well as for improving teaching quality. In one case, the respondents indicated that the CEC had been involved in resolving a violent clan conflict that had been taking place in and around the school.[131]

In particular, in all discussions with CEC members, the discussants indicated that monitoring at schools was very frequent. Some respondents even said the CECs monitors school activities every day. Most said that they monitored school activities weekly or twice a week. A member of the committee explains their approach:

We write down the names of students who trouble the teachers, and we follow-up with the teachers that don't teach students well. We keep him if he improves. And if not, we replace him and get another one. We do it [monitoring] three times a month and we share it with the principal and the parents.[132]

[127] REO, Somaliland
[128] REO, Somaliland
[129] REO, Puntland
[130] REO, Somaliland
[131] FGD CEC, Somaliland, 107
[132] FGD CEC, Puntland, 201

CEC members also indicated that they do a lot of work reaching out to the community for awareness-raising for girls' education and trying to mobilize disabled children and pastoralists. However, CEC activities tend to be limited to what the committee can do without additional funding. As explained above in relation to Indicator 1, CEC members in half of the schools where focus group discussions were not able to levy any funds from their respective communities.

In general, the discussions with CEC members did not show clearly that their activities would have increased in the past year, but neither was there clear evidence to the contrary. Indeed, considering all the different data points available for gauging CEC activity, we cannot convincingly make the case that the CEC activity would have increased or decreased since Midline Round 1. Further, when looking at CEC functionality and level of activity, the different sources of data paint opposing pictures. Head teachers report nearly universal functional CECs, school management plans are in 87.5 percent of the schools, and monitoring visits by the CEC have increased across the sample. Conversely, community members reported fewer active CECs while reporting some increased frequency in the communications of existing CECs. The findings on the prevalence and activity levels of the CECs are thus somewhat inconclusive. And indeed, if the CECs are as prevalent and as active as the head teachers would have us believe, this is not reflected in more CEC visibility in the communities. Given that the body's role would be to connect the school to the community and vice versa, the conditions for doing this seem not to be in place as of this Midline.

INDICATOR 2 – SYSTEM LEVEL – MINISTRY OF EDUCATION ENGAGEMENT IN GIRLS' EDUCATION INITIATIVES

Much of the project's sustainability in the long-term hinges on the Ministries of Education ability to carry it forward. The responsibilities of the MoE include training CECs, recruiting teachers, monitoring classrooms and fundraising.[133] The baseline identified the barrier to girl's education to relate largely to the practice of generally preferring sending boys to schools while girls remain at home to help their family with chores. SOMGEP-T's intervention aims to support MoE staff (local education officers) in the following four areas: (1) Strengthening the capacity of Gender Departments to improve girls' education outcomes through trainings, development of action planning and provision of incentives to retain the gender focal points, particularly in rural areas, (2) Providing support to Regional Education Officers (REOs) and District Education Officers (DEOs) to mainstream improved teaching practices and address retention/transition issues (3) Working closely with MoEs' TVET/NFE Units to explore opportunities for vocational training, and (4) Advocating for employment of female pre-service graduates in target schools.

An analysis of information provided by regional education officers (REOs) was conducted in order to understand MoE performance. For the midline, we again conducted KIIs with REOs, and explored the performance of MoEs across the four dimensions mentioned above.

As at Midline round 1, the REOs interviewed identified various challenges to their work. For example, an REO said only 50 of the 500 teachers are getting paid, and their salaries being dependent upon donations and whatever the CEC can collect. The resulting scarcity of teachers means that students are mixed up in large groups of different grades where learning becomes very difficult. Lack of salaries for teaching staff is a frequently mentioned problem. Moreover, MoE representatives across the board complain about the lack of finances and resources. REOs in Somaliland and Puntland alike say that they do not have the vehicles necessary

[133] KII – REOs – Puntland, Midline round 1

to reach more remote areas to make their monitoring visits more frequent. The MoE staff then rely on INGO staff to be able to reach some of these areas.

REOs also talked about the lack of money to pay teachers' salaries and the consequent brain-drain as teachers trained by the Ministry seek alternative employment:

> *One of the main challenges we have is when you train a teacher, he seeks better jobs. And you cannot convince him to stay in schools located in the rural villages.*[134]

At this time, only one REO complained that the relationship with CARE was not good. Otherwise the employees of the MoEs thanked CARE for their support with the ministry and viewed the relationship in positive terms.

Despite the continued challenges the system level findings for indicator 2 are somewhat positive. Some evidence to suggest that the Gender units in the Ministries are more active than at Midline round 1. However, with the data available it is not possible to determine the funding sources and status of the units, and as such their financial sustainability. Meanwhile, REO respondents detail various gender-specific program activities that are being undertaken.

Similarly, the qualitative interviews detail MoE activities that are focused on improving teaching quality and addressing issues of retention and transition. In this regard the ministries are increasing their monitoring, seeking funding for teachers and training more female teachers. Specific activities have also been designed for pastoralists and disabled children. However, across all of the activities the lack of funding and the consequent lack of access to more rural areas continue to hinder the efforts. Moreover, it is difficult to determine to what extent the efforts constitute a concentrated and uniform strategy of assessing these questions, and to what extent the efforts have increased or improved since the last Midline.

While at this Midline there is no evidence of vocational training focus within the Ministries, the emphasis on training more female teachers seems to have increased across the three states.

Overall, the qualitative information is somewhat encouraging but inconclusive to showing consistent improvements across the board. Meanwhile when assessing CEC activity in the communities as a proxy for the MoEs' efforts, this seems to have increased visibly across the board. Schools are much more likely to have a management plan, and CECs are much more likely to conduct monitoring visits. These changes have taken place in both intervention and comparison schools which would suggest that the MoEs work can explain the positive changes, at least partially. Overall, thus, we cautiously interpret that some of the system level changes required to sustainability are coming into place. However, going into endline, it will be crucial to assess in more depth the funding sources for various CEC and MoE activities to gauge their true sustainability beyond this (and other) education project(s).

The different thematic areas of the systemic changes will now be discussed individually.

Establishment of a Dedicated Gender/Child Protection Unit Strengthening the Capacity of Gender Departments

One of the core areas through which the project is providing support to MoEs is in strengthening the capacity of Gender Departments to improve girls' education outcomes through trainings, development of action planning, and provision of incentives to retain the gender focal points, particularly in rural areas. At Midline

[134] REO, Puntland

round 1 two of the six REOs interviewees were able to explicitly confirm that there was a gender unit or department in their region. However, they described many activities that were in line with the unit's remit.

At this time, REOs in different regions of Somaliland and Puntland talk explicitly of a gender unit within their Ministries:

> *Yes, the Ministry has a gender department. Their activities include girls receiving free education in four schools, from form one to form four, and at the university level. So gender plays a vital role.*[135]

> *They train girls, they pay their salaries, even though the girls are in high school…gender unit brings menstrual hygiene products.*[136]

> *The gender unit provided training to girls, conducted awareness towards the importance of education both for the girls as well as parents. And if there are challenges that these young female students are facing, they advocate and resolve. They establish the girl's club. They also provide training to education officers and teachers to convince them of the importance of education.*[137]

It seems thus, that in the past year, the awareness of the gender unit and its activities have increased somewhat. However, from the data acquired at this Midline, it is not possible to determine the funding sources and situation of the gender unit and consequently their financial sustainability.

Overall, however, the REOs understanding of the challenges specific to girl's education seems reasonably good and comprehensive. Respondents talk of the need for girl-friendly spaces, the need for female teachers, addressing issues with facilities and resources for dealing with female hygiene issues, as well as the financial issues relating to the fees as well as girls earning a livelihood that might prevent them from enrolling and attending school.

> *We want female teachers to teach the girls, and also support them during challenges that they face in their education. We support them with their books, uniforms, and paying their fee if the girl is from a poor family. All of these factors will improve girls' education.*[138]

> *Female students do not have separate toilets and they are shy to share with the boys. Also, sometimes sanitation kits run out at the school it is also a challenge. In the community, one of the main existing challenge for young females is the traditional FGM which can lead health problems…and sometimes results in*

[135] REO, Somaliland
[136] REO, Somaliland
[137] REO, Puntland
[138] REO, Somaliland

> *not attending school over a long period of time. Primary schools should also make spaces friendly for girls which are very few.*[139]

The absence of spaces that are dedicated for girls at the school seems to be an issue that has received some attention in Somaliland. An REO in Somaliland explains some efforts that were undertaken to address this need for girl-friendly spaces:

> *I think that in my region in the city, 80% of the girls have private areas where they have their break time food, and that has allowed girls to have their own private space. We solved an obstacle that was standing in the way of girls' education. You see, in our Somali culture, our girls cannot eat, get dressed, talk or rest in front of the boys. When we thought about that, we decided to give the girls somewhere to rest. They have small chairs, a mirror and toilets inside.*[140]

Overall, thus, there is some evidence that the gender units are being established and are active in the Ministries of the states. Moreover, REOs demonstrated a good understanding of gender-specific barriers for education and described various activities undertaken to address them. It is not possible, however, to assess to what extent these activities are dependent upon external funding and as such their sustainability beyond the program in terms of finances.

REOs and DEOs Mainstreaming Improved Teaching Practices and Addressing Retention and Transition Issues

In addition to gender-specific concerns, the MoEs efforts in terms of addressing questions of teaching quality and issues in transition and retention are deemed important for the long-term sustainability of the project. In terms of teaching quality, some of the REOs showed a holistic understanding, highlighting the variety of inter-linked activities that are required for improving it:

> *We are currently focusing on how to improve the quality of education. I would say that quality needs much more work, because schools need to be funded on time. But quality has no end point; all the time you should keep improving. It needs constant monitoring from you every day and being aware of its status. You have to ensure the monitor is qualified, teacher is qualified, curriculum is appropriate, as well as repairing and fixing the building that you had built before…The teachers should also be aware and be given trainings and international teaching methods. So, the truth is, quality is a thing that is very wide, and its improvement needs a lot of planning and studying.*[141]

Various efforts were described in this context. The REO from Galmudug explains that the MoE has created uniform exams of the students to identify schools that require additional support. These schools then make an improvement plan. CECs send teachers to training, and aim to increase teacher salaries.[142] The MoE then

[139] REO, Puntland
[140] REO, Somaliland
[141] REO, Somaliland
[142] This is consistently identified as a key challenge for retaining good teachers, particularly in rural areas.

monitors schools to see their progress in relation to these issues. Similar efforts were described in Somaliland where REOs talked about continued monitoring of teachers and schools. Likewise, in Puntland the REO talked about the importance of supervision and monitoring of schools in improving the quality of education. He also placed an emphasis on the community awareness-raising activities in order to improve enrolment and retention.

In terms of learning and access, the efforts are quite visible. A Puntland REO mentioned a NORAD-funded program running until 2023, which focuses on teacher trainings, particularly that of new female teachers, as well as on community awareness-raising. Somaliland REOs mentioned tuition support, distributing sanitary kits as well as conducting awareness-raising activities in the communities. Indeed, REOs in the three states said the MoEs had in place bursary or scholarship programs for disadvantaged families.

There were also efforts specifically targeted at enrolling and retaining pastoralists and disabled children. These have been discussed in more detail earlier in this section, in relation to the roll-out of ALP at the national level. Nevertheless, while the efforts overall are not sufficient, according to the REOs, the MoEs are experimenting with some ways to improve the inclusion of these previously more excluded groups in the school system. For example, an REO talked of an information sharing system that will ensure that a pastoralist student can continue their education from the same level when they migrate to an area with a different school.

Also, as identified at Midline round 1, the hiring of female teachers is a way to encourage girls to stay at school. At the time, one REO directly linked dropping out to lack of female teachers: "Female students also feel shy to ask the male teacher to repeat the lesson if they do not understand it. This can cause the girls to drop out of school because they feel left out."[143] As will be discussed below, there is evidence of increasing emphasis on the hiring of female teachers.

Another factor that can be linked to encouraging retention is the work to remove corporal punishment from schools. The MoEs work actively in this. An REO from Puntland explains:

> *Back in the days, students used to be punished which sometimes resulted in some of the students being disabled in parts of their bodies, and some used to leave school. Therefore, we spoke to the teachers and trained them...During this year we have imposed that no teacher without this education should be allowed to work in the school.*

Similarly, two of the REOs in Somaliland went as far as to say that corporal punishment has completely stopped in their schools. While other sections of the report will talk in more detail about the prevalence of the practice, it is encouraging that this has been a focus are for the MoEs as it may lead to increased retention.

Other activities that are more specific to girls are also being conducted. For example, as explained in the previous section on the work of the Gender units, some MOEs are involved in ensuring that schools have girl-friendly spaces. For the REOs this directly relates to retention:

> *When we were building the area that the girls were resting in, our objective was to increase girls' enrolment and make sure they like learning more. So, after it is built, we should monitor and make sure that the girls have reached the*

[143] KII – REOs – Somaliland

number we wanted and whether or not they have stayed at school. So, if one of them goes missing, we will inquire on the reason that they haven't achieved what was intended.[144]

Given that the ministries are linking these constructions to retention and emphasising the importance of monitoring girls' attendance, this should ideally result in increased activity by CECs to follow-up on drop-outs and girls who attend infrequently. Indeed, as the sections above looking at the CEC levels of activity have shown, the CECs across the board are now much more active in monitoring visits, and many more schools have a school management plan. These are both factors that should reflect positively on teaching quality, as well as retention and transition.

Finally, it should be noted, however, that as at Midline round 1, it is difficult to assess to what extent there is a uniform approach and strategy in place to deal with the question of teaching quality, retention and transition across the board. There is certainly evidence of various activities being undertaken by the MoEs, but to what extent they are being implemented in different areas, particularly in the more rural regions, cannot be determined with robustness. Moreover, while there are accounts of various activities in this regard, it is difficult to gauge to what extent these activities have increased since Midline round 1.

TVET/NFE Units Exploring Options for Vocational Training

The project is also working closely with MoEs' TVET/NFE Units to explore opportunities for vocational training. While at the previous midline there was some evidence of MoEs involvement in design and implementation of vocational training to drop-outs and adults. At midline round 1, no REOs mentioned activities on vocational training at all. However, the research tool did not have specific questions on vocational training, and as such this omission does not necessarily imply that the activities have been halted.

Advocating for Employment of Female Pre-service Graduates

As a final element of MoEs support for girl's education, we looked at the ministry's advocation for the employment of female pre-service graduates in target schools. At Midline round 1 we reported that the evidence in this regard was sparse, and we concluded that recruiting female teachers did not seem like a priority issue for the MoEs at the time. At Midline round 1, conversely, there was more evidence of an increased focus on recruiting female teachers. The following quotes are from REOs in each of the three states:

These plans include to register the ministry payroll around 300 [female] teachers, 50 of them allocated for this region, so we are expecting to give a chance to female teachers.[145]

The biggest thing is that more female teachers are being hired…it is essential to get well-trained and paid female teachers in the ministry.[146]

[144] REO, Somaliland
[145] REO, Galmudug
[146] REO, Somaliland

The Ministry also trains many female teachers in order for other female students to be admired and motivate them to become female teachers.[147]

Indeed, in this regard, it seems that advances are being made as there is evidence from Somaliland, Puntland and Galmudug that the ministries are putting specific focus on hiring female teachers.

6.1 Overall Findings

Indicator 1

When looking at CEC financial support as a measure of community level sustainability, the findings are somewhat bifurcated. The head teacher surveys indicate that while CEC share of teacher salaries increased between Baseline and the first Midline, it has since declined. Moreover, the level of scholarship support across schools is currently more than 20 percent lower than at Midline round 1. The CECs' share of the scholarships has also declined very slightly. Yet, it is possible that this is due to increases in other sources of funding for the schools. Both qualitative interviews and household survey data in fact suggest that CECs have been more active in raising funds and in other forms of financial support since the baseline. Nevertheless, fewer community members viewed their communities as having a functional CEC. Given the discrepancy in the different data sources, the findings regarding the sustainability of this indicator at the community level are inconclusive.

At the school level, similarly, the evidence does not conclusively support that the use of formative assessments - a measure of overall teacher professionalism – would have become more prevalent since Midline round 1. While the reported use of the assessments had increased, this might be a result of social desirability bias rather than anything else – as teachers know they *should* use the assessments in their work, and report accordingly. The share of teachers who can show records of using them has not changed significantly since the last round of data collection. Yet, the differences are markedly more positive in intervention schools, which would indicate that the sustainability in them is increasing.

At the system level talk of ALP specifically is very limited in interviews with MoE respondents. Many respondents mention efforts for including the nomadic population and disabled children. But most respondents do not view these as sufficient at this time, due to financial constraints and other operational challenges.

As such, as it pertains to the first sustainability indicator, unfortunately no significant progress can be reported between the two rounds of Midline. In particular, if we are focusing on the CEC's ability to provide salary support to teachers, a problem widely documented in the qualitative interviews, it is concerning that only about a fifth of the CECs in intervention schools support teachers financially. If in the long-term the plan is for the schools to be financially independent from outside sources (or if the outside funding is acquired by the CECs), there is still a long way to go.

Indicator 2

In terms of sustainability indicator 2, at the community level, again when looking at CEC functionality and level of activity, the different data sources paint opposing pictures. According to head teachers there are more functional CECs now, school management plans are more common, and monitoring visits by the CEC have

[147] REO, Puntland

increased across the sample. Conversely, community members reported fewer active CECs while reporting some increased frequency in the communications of existing CECs. In sum, much as with the community level findings for Indicator 1, the prevalence and activity levels of the CECs are inconclusive.

At the system level, many activities of the MoEs were looked at. The findings there are somewhat positive. Some evidence to suggest that the Gender units in the Ministries are more active than at Midline round 1. MoE activities seem to be more focused on improving teaching quality and addressing issues of retention and transition. The ministries are increasing their monitoring, seeking funding for teachers and training more female teachers. Specific activities have also been designed for pastoralists and disabled children. The emphasis on training more female teachers seems to have increased across the three states.

Overall, the qualitative information for the system level changes at the MoE are somewhat encouraging but inconclusive to show consistent improvements across the board. When assessing CEC activity in the communities as a proxy for the MoEs' efforts, this seems to have increased visibly across the board. Schools are much more likely to have a management plan, and CECs are much more likely to conduct monitoring visits. These changes have taken place in both intervention and comparison schools which would suggest that the MoEs work can explain the positive changes, at least partially. Thus, it can be cautiously interpreted that some of the system level changes required for sustainability are coming into place.

Yet, the financial sustainability of these activities could not be established, and this ought to be a concern as the respondents continue to list many challenges relating to funding. At endline, it will be important to establish the source of funding for different elements of the MoEs' work that relate to the sustainability of this project.

Overall, thus, it seems that some gains in way of increasing the sustainability have been made when looking at the system and school level. However, the measures on sustainability at the community level are more inconclusive. Nevertheless, much work remains in order to ensure the sustainability of the program.

6.3 Sustainability – Project Response

TABLE 75: SUSTAINABILITY MATRIX

	Community	School	System
Change: what change should happen by the end of the implementation period?	CECs should not only be able to raise funds and support part of the school costs (including salaries), but also maintain a 'safety net' fund to enable communities to respond quickly, efficiently and equitably at times of great need CECs are able to support ultra-marginalised students to attend, waiving fees/ providing scholarships (within the possibilities of each community, considering that major crises, such as the ongoing COVID-19 outbreak, would affect their financial capacity) CECs should conduct monitoring activities in a consistent manner, across all measured areas of fidelity of implementation[148] Initial shifts in social norms regarding child protection and gender-based violence, such as those related to the use of corporal punishment against students and harassment in school	The majority of the teachers (70%) use formative assessments and adhere to fidelity of implementation standards for ALP and numeracy boost (80% FOI score).	Gender-focused positions funded and functional at MoEs Sector planning documents incorporate and prioritise accelerated education approaches for upper grades Critical learning from SOMGEP-T informs the Education Sector Analysis and Education Sector Strategic Plan, including the effect of girls' life skills development on learning and transition; the impact of functional CECs on learning; and the gaps in teacher capacity
Activities: What activities are aimed at this change?	CEC coaching (responding to specific findings on the level of fidelity of implementation) Strengthening linkages between VSLA members and schools to create a school support fund Coaching of head teachers	Tailored coaching of teachers based on fidelity of implementation assessment findings Coaching of head teachers Training of new teachers on the use of student-centred methodologies Support to REOs/DEOs to monitor schools	Technical and financial support to gender units/ focal points; engagement of focal points in joint advocacy, implementation of activities and monitoring ALP approval process, joint monitoring and advocacy for social change/inclusion of marginalised subgroups

[148] SOMGEP tracks CEC monitoring activities in the following areas: Child protection, student/ teacher attendance, follow-up on dropouts, corporal punishment, enrollment tracking.

	Reinforcing messaging on negative impact of harassment at BEFs/ Increased CEC awareness of the impact of violence and harassment on learning		Sharing SOMGEP-T ALP and ABE modules with a new USAID-funded initiative aimed at developing a national framework for accelerated education Disseminating SOMGEP-T findings to government partners, Education Sector Committee and consultants involved in the preparation of the Education Sector Analysis/ Education Sector Strategic Plan
Stakeholders: Who are the relevant stakeholders?	Community education committees, head teachers, REOs, DEOs, religious leaders, BEFs/GEFs	Teachers, head teachers, REOs/DEOs	Gender focal points, REOs/DEOs, MoEs' senior leadership
Factors: what factors are hindering or helping achieve changes? Think of people, systems, social norms etc.	Positive: Sector support initiatives (Horumarinta Elmiga, Education Is Light, ESPIG) supporting CEC mobilisation and training (including in comparison schools); support from REOs/DEOs (anecdotal reports of SOMGEP's approach being replicated). Some evidence of social norm change already taking place: Increased teachers' awareness of corporal punishment as unacceptable – not observed in class, but still reported by students. Negative: Limited capacity of head teachers to mobilize and engage CECs in some areas (in schools where the head teacher is female, traditional gender norms may hinder her authority/ mobilisation capacity); conflicts between head teachers and CECs, and/or local clan conflict undermining the impact of CEC activities; temporary school closure due to clan/border conflict. The emerging COVID-19 crisis will have an impact on the economic capacity of households and diaspora, reducing contributions to the CEC. Additionally, the COVID-19	Positive: Interest for new methodologies and high adherence in some schools, despite limited exposure to date. Strong results on financial literacy suggest that teachers are adopting the numeracy boost approach. Students reporting increasingly respectful behaviour from teachers. Negative: Teacher turnover (trained teachers moving to other schools/ grades); limited leadership and technical capacity of head teachers; limited capacity and logistical resources of REOs/DEOs; willingness of older teachers to accept new methodological approaches and break traditional relationships of power in class; temporary school closures undermining impact. The emerging COVID-19 crisis poses an added risk of loss of teachers (given the age group involved) and overcrowded classrooms.	Positive: Approval of ALP framework; sector plans focusing on equality; shift in social norms resulting in greater support for SOMGEP's intervention; donor push for accelerated education and inclusion of heavily marginalised communities. Development of the ESSPs in 2020 offers a major opportunity. Negative: High turnover of senior personnel at MoEs; traditional dynamics of power resulting in resistance to acknowledge and address issues affecting ultra-marginalised groups; ongoing conflict limiting coverage, access and impact; heightened tensions between states and Federal Government may affect distribution of resources and capacity to support REOs/DEOs.

	crisis may have a disproportional impact among CEC members due to their age range and limited access to healthcare. Given the prolonged school break and the combined burden of disease and economic downturn, it is likely that CECs will be overwhelmed by the number of students in need.		

There is clearly a mixed picture in sustainability findings. CECs have clearly become more active in ways that have a positive impact on learning; as noted under the learning findings section, students from schools with active CECs have a much higher difference-in-differences in relation to the comparison group across all measurements of learning (Somali literacy, English, numeracy and financial literacy). Additionally, the proportion of CECs monitoring child protection has increased considerably and seems to be having an effect in reducing the acceptance of corporal punishment (as noted in classroom observations), despite its use still being reported by students. On the other hand, CECs have reached a plateau in terms of performance in financial support to schools; support to marginalised students; and following up on attendance and dropout. In order to break this pattern, there will be a need to provide further guidance to VSL members on the use of the social funding in order to enhance financial support to CECs. Expectations for this strategy should be reasonable though. The conflicts experienced last year in 20% of the school communities and the locust swarms observed in late 2019/early 2020 are likely to have taken a toll on the financial capacity of CECs to support schools and mobilise financial support for students in need. It is likely that the ongoing COVID-19 crisis may further exacerbate this situation, as communities will face the combination of disease burden with a reduction in income due the financial impact on the diaspora and the likelihood of reduced revenue from livestock exports to the Gulf States and Saudi Arabia[149]. Therefore, the likelihood of sustainable CEC financial support to schools will need to be re-assessed in the aftermath of the ongoing crisis.

On the other hand, the ongoing crisis presents an opportunity for CECs to step up in tracking attendance and dropout upon re-opening of schools, and also to support communities to address school-related gender-based violence. The role of CECs in tracking dropouts and supporting their return to school will be fundamental to mitigate the risk of a spike in early marriage and cases of exploitation and abuse of minors whose parents / caregivers have passed away or become incapacitated as a result of the outbreak. CECs are also fundamental for social norm change processes and could play a major role in working with BEFs to reduce harassment in schools. Although it is unlikely that a dramatic social norm change on harassment may occur within the short period of time remaining in the life of the project, particularly in a context where violence has been normalised, initial levels of change can still be expected/ observed (as was the case with corporal punishment and support for girls with disabilities[150]).

At school level, the robust impact on financial literacy suggests that teachers are adopting improved methodologies such as the Numeracy Boost. There are indications of social norm change already occurring – more respectful engagement with students; support for GwDs (the impact on financial literacy is even higher among GwDs than the general average). Still, the use of formative assessments has reached a plateau, and the same was observed in the use of student-centred methodologies, potentially as a result of teacher turnover. While additional training / coaching will be necessary to reinforce the use of formative assessments and student-centred methodologies, it is essential, from a sustainability standpoint, to further engage REOs/DEOs/ QAOs in coaching teachers, as well as increasing the capacity of head teachers. To maximise sustainable impact, it is essential to focus on social norm change as well as methodological approaches; the findings in this report show that within a limited period of time, powerful shifts in norms have already taken place, which are likely to be sustained once community acceptance and expectations change. On the other hand, as discussed above, there is a likelihood that the ongoing COVID-19 outbreak may have a major impact on teachers, given their age range; this remains a threat to the sustainability of project-related changes as well as to the functionality of the broader education system.

[149] Considering the suspension of the pilgrimages to Saudi Arabia's Holy Sites. Somalia is a large exporter of live goats and sheep to the Gulf States and Saudi Arabia.

[150] As illustrated by the impact on financial literacy outcomes among GwDs exceeding the average DiD.

At system level, the results indicate gains in ownership of accelerated learning by REOs/DEOs, as well as heightened awareness of the need to support marginalised groups, including children with disabilities. There is a clear articulation of the issues faced by girls and the barriers to their attendance and participation in class, in contrast with previous evaluation rounds. While it is expected that ongoing funding issues will persist (despite investment from donors, in particular EU-funded projects, in funding REOs/DEOs' offices and logistical needs), the increased articulation of such barriers and potential solutions bodes well for sustainability, as all three Ministries are presently conducting their Education Sector Analyses and Education Sector Strategic Plans[151]. The buy-in from regional and district-level staff, as well as the presence of influential Gender Focal Points, represents an opportunity to inform sector investment in measures to address issues of equity, inclusion and quality in general. CARE is already engaged in the discussions leading to the development of the Education Sector Analyses and upcoming Education Sector Strategic Plans. Additionally, the onset of *Bar ama Baro*[152] / BAB – a USAID-funded investment intended to develop a general framework for accelerated education in Somalia – represents an opportunity to inform the design of this framework through SOMGEP-T's approach and results. CARE has been engaging with USAID and project implementers to share both the ALP/ABE modules as well as to disseminate findings and lessons learned.

While the scenario remains favourable to inform sector planning, it is important to note that the combination of the ongoing outbreak, along with upcoming elections and political conflict between states may have a detrimental effect on the ability of state actors to allocate resources for the proposed ESSP activities. Additionally, the major financial impact of the outbreak on the diaspora – one of the largest sources of funding for the education sector in Somalia – cannot be underestimated.

7. INTERMEDIATE OUTCOMES

7.1 ATTENDANCE

Outcome	Indicator	BL	ML Target	ML	Target achieved ? (Y/N)	Target for next evaluation point	Will IO indicator be used for next evaluation point? (Y/N)
Attendance	Percentage improvement in attendance rates (headcounts)	81.2%	7-point improvement (88.2%)	79.2	No	4-point improvement (92.2%)	Yes
Main qualitative findings							
• The qualitative data do not speak directly to changes over time in attendance rates, though anecdotal evidence suggests that mothers believe their children are attending school more regularly.							

[151] Galmudug falls under the FGS Education Sector Strategic Plan.
[152] https://www.usaid.gov/sites/default/files/documents/1860/Fact_Sheet_-_Somalia_BAB_February_2020.pdf

- The primary barriers – according to qualitative reports – to regular school attendance by girls appear to be domestic household work, and inadequate toilet/sanitation facilities at school. Other factors that emerge from the qualitative data are more closely tied to reduced transition outcomes (including seasonal migration and economic marginalization at the household level) though they may influence attendance as well, or in less extreme cases (e.g., girls whose attendance is reduced because the household experiences economic hardship, but keeps her enrolled).

Outcome	Indicator	BL	ML Target	ML	Target achieved? (Y/N)	Target for next evaluation point	Will IO indicator be used for next evaluation point? (Y/N)
Attendance	Mothers' support to adolescent girls' attendance	Mothers support education and show increasing appreciation for education as a means of obtaining better jobs; but most still prioritise domestic chores and pastoral work over attendance	Mothers prioritising adolescent daughters' attendance to ALP and formal school, as opposite to domestic/ pastoral work	Mothers appear to be aware of the trade-off between domestic work and schooling, but domestic workloads have not declined in intervention schools, nor are mothers reporting explicit efforts to change domestic/pastoral workloads	No	Mothers adopt work-sharing strategies to ensure that adolescent girls attend classes consistently	Yes

Main qualitative findings

- Mothers report encouraging their girls to enrol and attend school, including following up with teachers regarding their attendance and performance in school
- Mothers occasionally reference domestic work as a barrier to attendance; however, their implication of domestic work as a barrier to enrolment and attendance is much less common than occurs when speaking with girls themselves.
- While many caregivers are aware that domestic work keeps their girl out of school or occasionally prevents her from attending, many still report that their daughter completes a half-day or more of work but this workload does *not* prevent her from attending school, a contradiction that highlights the fact that domestic workloads are not universally seen as a barrier to girls' attendance, even in households where girls experience particularly heavy workloads.

In many ways, attendance is the most proximate intermediate outcome to learning and transition results. It is not entirely surprising that it has among the strongest theoretical support, in prior reports, for its

relationship to learning and transition. Indeed, in some ways transition is an extension of attendance, as girls often begin to attend irregularly before dropping out of school entirely, and maintaining regular attendance is almost certainly a key method for reducing dropout rates.[153] Further, regular attendance increases the number of contact hours between teachers and students; even in the context of moderate teaching quality, a larger number of contact hours and a larger amount of time spent engaged in reading, writing, and so forth, should improve learning outcomes, even if only marginally.

We employ three measures of attendance, each of which is drawn from a different source and some of which cover a slightly different time period. The three measures we employ are:

- Physical headcounts in classrooms – during fieldwork visits, team leaders recorded enrolment numbers from school records and took a headcount of students present in class, with attendance rates calculated as the share of enrolled students present in class.
- Caregiver estimates of attendance frequency – during the household survey, caregivers were asked whether their girl had attended school most days during the current school year.
- School records – schools record attendance daily in most cases, and this information was recorded for all cohort girls enrolled at the time of data collection

As the baseline and ML1 reports documented, there is typically a correlation between attendance outcomes derived from two different sources, but these correlations are not as strong as one might imagine. The correspondence across measures is likely imperfect for a number of reasons, including the different time periods being considered, but also because of poor record quality in enrolment records (which impact headcounts) and attendance records (which impact results based on school records). It is also the case that caregivers may not accurately recall the frequency with which their girl is absent, may misinterpret the question, or may give answers that are more positive than the truth, due to the perceived social desirability of a girl attending school regularly. Finally, headcounts may be subject to other forms of bias, such as higher attendance on the day of the visit, perhaps encouraged by the head teacher to improve the image of the school. To the extent possible, we investigate these possibilities below, but it is important to be clear about the relative limitations of the data, which motivate an approach based on heavy triangulation.

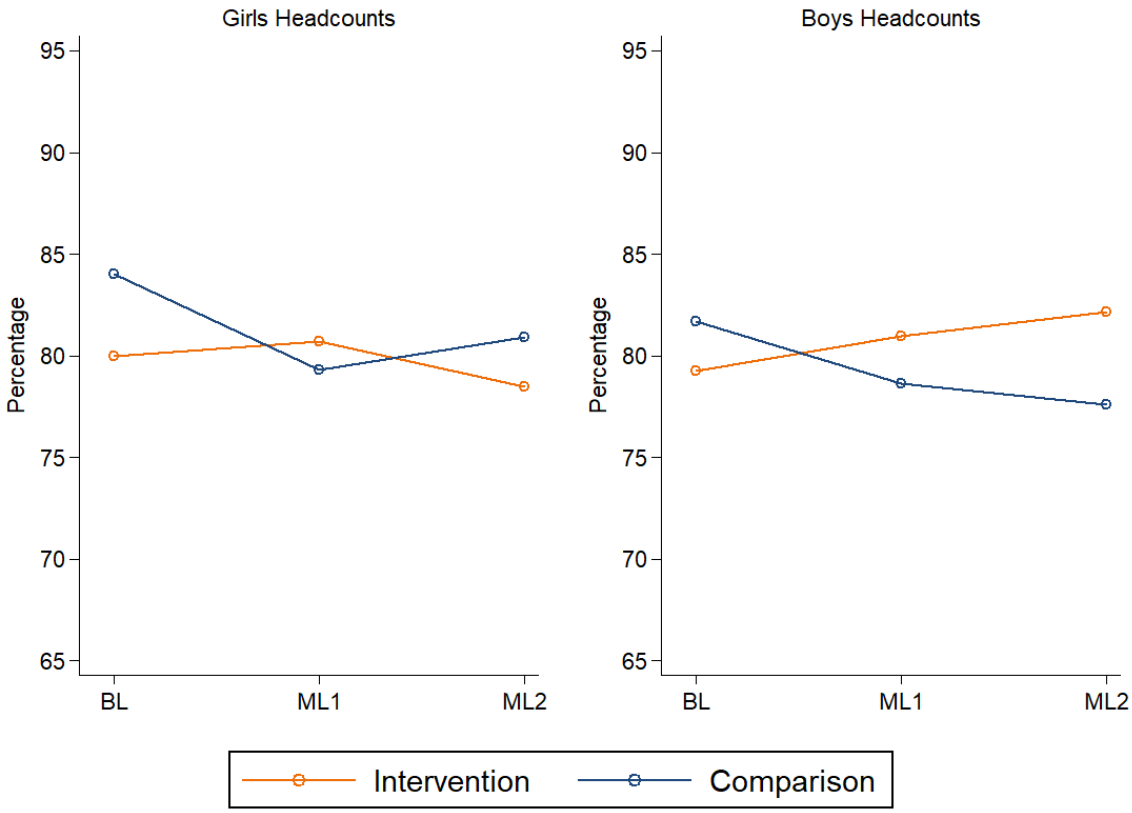
Attendance from Headcounts

We first report results based on physical headcounts in classrooms. Our sample includes 1,264 headcounts, spread across three rounds. Throughout this analysis, we limit the sample to the set of schools that appeared in all relevant rounds being analysed. For instance, when we make comparisons across all three evaluation waves, we utilize only the subset of schools – but not necessarily classrooms or grades – that appear in all three rounds. We apply the same rule to analyses of baseline to ML2 comparisons. In essence, we use the widest possible sample in each analysis, while still maintaining the greatest degree of comparability possible.

In the figure below, we report the mean attendance rate in classrooms, based on physical headcounts, across rounds and disaggregated by intervention and comparison schools. While this graph is straightforward to interpret, it is analogous to the more formal difference-in-differences analysis we perform later in this section. In both cases, we are interested only partially in the trends over time; rather, we are interested in the *difference* in trends over time between intervention and comparison schools.

[153] In some cases, there may be a fine line between attendance and transition – caregivers and girls may view a girl as being out-of-school if she rarely attends school, despite being enrolled. The household survey specifically asks caregivers whether the girls is enrolled in school, rather than whether she attends school, but respondent's may interpret these questions in different ways.

FIGURE 26: HEADCOUNT ATTENDANCE RATES ACROSS TIME AND INTERVENTION STATUS



This point is illustrated in the left panel, which documents girls' attendance based on headcounts. From baseline to ML2, girls' attendance declines in both intervention and comparison schools, which is worrisome. However, from the perspective of estimating program impact, the over-time decline is of less importance than the fact that comparison schools experienced a greater decline over this period than intervention schools. This is seen most easily by comparing the relative gap between intervention and comparison schools at baseline (i.e. the two points on the far left of the graph) to the relative gap between these groups at ML2, where the gap has narrowed. This implies that, relative to comparison schools, girls' attendance in intervention schools has improved, even if it has declined slightly in real terms. The logic described here may seem counter-intuitive, but it is the basis of the difference-in-differences design.[154]

The decline in girls' attendance rates over time is notable, however, and it is worth considering why attendance rates may have fallen, across all schools, over time. One difference between rounds of data collection concerns the timing of fieldwork, as the ML2 data collection took place 2-3 weeks later in the year than either the baseline or ML1. Data collection for ML2 was completed in very late December, at which

[154] The existence of secular trends such as these is the entire justification for using a difference-in-differences approach, with an explicit comparison group, rather than a pre-post design without a comparison group. In a pre-post design, the small decline in girls' attendance rates from baseline to ML2 among intervention schools would be interpreted as a negative program impact. However, the graph strongly suggests that the program is not responsible for declining real attendance rates, because the same trend is witnessed, with greater severity, among comparison schools. This fact indicates that exogenous factors are driving attendance rates for girls lower in the area where the project is being implemented. By controlling for this underlying trend, the difference-in-differences design provides unbiased estimates of program impact in expectation.

point some schools had entered the December break. It is possible that attendance declined as the break approached. It is also possible that there is some other degree of periodicity in attendance rates that would produce lower attendance rates due to the later fieldwork.

However, this explanation does not explain why aggregate attendance rates – across all kinds of schools – declined from baseline to ML1. We also investigated this possibility empirically, by studying the relationship between attendance rates and the specific dates of fieldwork, and found no evidence that later fieldwork periods were associated with lower attendance rates. In fact, attendance rates tended to increase over the course of ML2 fieldwork, casting doubt on fieldwork timing as an explanation for declining attendance rates.

A more straightforward explanation is that flooding was more severe in many parts of Somalia during the ML2 evaluation than in the two previous years. Reports from team leaders indicated that flooding had caused delays and accessibility issues they were forced to work around. Flooding causes displacement, and if this were more severe from one year to another, it could impact attendance rates.

In terms of program impact, the figure shows a small relative improvement in attendance among intervention schools. More starkly, there is a dramatic improvement in boys' attendance, based on headcounts, in intervention schools, relative to comparison schools. This trend started between baseline and ML1, or even prior to these rounds, and continued into the ML2 period. It is not clear why boys' attendance has improved in intervention schools, nor why the program appears to have had a more positive impact – vis-à-vis comparison schools – on boys' attendance than on girls' attendance. To the extent that community attitudes have become more focused on equality of educational opportunity between boys and girls, this would suggest that girls' attendance should be increasing relative to boys' attendance, or at least maintaining par while both were rising.[155]

It is possible that, with improving attitudes toward girls' education and a general emphasis on the importance of schooling for girls, their enrolment rates have increased.[156] If the general trend is for more girls to enrol in school, and that means relatively marginalized girls – who would not have otherwise enrolled, and who can be expected to have lower attendance rates due to household demands or for other reasons – this could drive girls' attendance rates downward, by shifting the denominator of enrolled girls upward. This is one reason why we dedicate attention to the analysis of school attendance records for cohort girls later in this section – changes in attendance rates in that analysis are much less likely to be an artifact of increased enrolment rates.

To more fully assess changes in girls' and boys' attendance over time, we employed the formal difference-in-differences, estimated within a regression framework, as in previous sections of this report. The table below reports the mean attendance rates by round and intervention status for baseline and ML2, providing the difference-in-differences estimate in the right-most column. Using this metric, the program has increased girls' attendance by 1.5 percentage points since baseline, while improving boys' attendance by 5.9 points over the same period.

TABLE 76: DIFFERENCE-IN-DIFFERENCES, HEADCOUNT ATTENDANCE RATES

Group	Baseline	Midline #2	Combined

[155] Community attitudes, overall, are moving in this direction: since baseline, there has been 12.1 point improvement in the share of caregivers who strongly agree that a girl is just as likely to use her education as a boy.
[156] Declining attendance rates are a common phenomenon any time enrolment increases dramatically, such as when free education is implemented.

	Intervention	Comparison	Intervention	Comparison	Difference-in-Differences
Girls	81.2%	84.3%	79.2%	80.8%	1.5
Boys	80.8%	82.2%	82.3%	77.8%	5.9
Group	**Midline #1**		**Midline #2**		**Combined**
	Intervention	**Comparison**	**Intervention**	**Comparison**	**Difference-in-Differences**
Girls	79.7%	79.3%	77.2%	80.9%	-4.1
Boys	79.9%	78.6%	80.5%	77.6%	1.6

The results in the table above are based on a straightforward implementation of difference-in-differences, but it does not make full use of the available data. While there is no reason to think the results are biased, a multivariate regression approach can improve the quality and precision of our estimates of program impact. By incorporating additional predictors of attendance rates, we can improve the fit of the statistical model, though incorporating such variables is *not* strictly necessary to draw valid causal inferences from the design.

The table below reports results from a series of linear regressions. In the top panel, we report results for girls' attendance, based on increasingly saturated regression models, incorporating zone, and grade level of the classroom being observed. Both zone and grade level are correlated with differences in attendance rates; while controlling for these factors in the sense of a traditional regression is not necessary, incorporating them into the regression can improve the efficiency of our estimates. We are also interested in subgroup effects, such as the impact of the program in one zone specifically; we report zone-specific regressions, which provide valid estimates of program impact within Somaliland and Puntland alone, in then table below.

The coefficients in the second column are estimates of program impact.[157] Coefficients denoted with an asterisk are statistically significant at the 5 percent level; only one such coefficient appears in the table. We also report the 95 percent confidence interval for each estimate and the sample size employed for the regression, to make clear the limitations of the zone-specific analysis. The top panel results, focused on girls' attendance, confirm the conclusions drawn above: there is a small improvement in girls' attendance in intervention schools, as compared to comparison schools, but this improvement is not statistically significant, even when we account for additional predictors of attendance rates. The zone-specific regressions indicate that the positive outcome is driven by schools in Somaliland, while girls' attendance in intervention schools in Puntland has actually declined relative to their comparison schools.[158] However, neither effect approaches statistical significance.

[157] More precisely, this is the coefficient of the interaction term between intervention and round, where round is defined as 0 for baseline and 1 for ML2.

[158] Note that we do not report zone-specific regressions for Galmudug, as the sample size of headcounts was too small to allow analysis.

TABLE 77: PROJECT IMPACT ON HEADCOUNT ATTENDANCE RATES

Regression Details	Impact Estimate (Regression Coefficient)	95% Confidence Interval	Sample Size
Girls Headcounts			
No controls	1.5	-5.8 - 8.9	832
Control for zone	1.7	-5.6 - 8.9	832
Control for zone and grade level	1.4	-5.9 - 8.7	832
Somaliland only, control for grade level	2.3	-7.7 - 12.3	537
Puntland only, control for grade level	-2.4	-11.7 - 6.9	265
Boys Headcounts			
No controls	5.9	-2.7 - 14.5	836
Control for zone	6.1	-2.3 - 14.4	836
Control for zone and grade level	6.1	-2.4 - 14.5	836
Somaliland only, control for grade level	14.0*	2.3 - 25.6	539
Puntland only, control for grade level	-8.4*	-15.9 - -0.9	266

In the bottom panel of the table, we report the same analysis for boys' attendance rates. Our best estimate of the impact of the program on boys' attendance is that it has increased attendance rates by 6.1 percentage points from baseline to ML2. This result is not statistically significant, but it is sufficiently large that it should be considered a very tentative finding.[159] As with girls' attendance, the results seem to be driven exclusively by improvements in the Somaliland intervention schools; when we limit the analysis to Puntland, intervention schools appear to have regressed dramatically, vis-à-vis their comparison-group counterparts.

What explains the decline in attendance rates? Is it because enrolment numbers are going up, i.e. getting marginal girls enrolled but then they are less likely to attend? That doesn't seem to be the story, because per-classroom enrolment numbers seem to be declining, according to the headcount data (enrolment numbers in b7 and b12). But it's also possible that there are other explanations like periodicity in attendance that would affect both treatment and comparison schools, but the periodicity argument does not seem to hold water – schools we visited later actually have higher attendance rates, slightly.

Although we do not report these results in great detail, it is useful to understand the extent to which these results are robust to additional analyses. For instance, our primary analysis of headcounts did not employ weights to correct for differences in the number of headcounts across schools. When we use weights in the analysis, it does make a difference in the magnitude of the coefficients we estimate, increasing the

[159] In both regressions with the full sample of boys' attendance headcounts that control for zone, the p-value associated with the coefficient is 0.16.

improvement in girls' attendance, as a function of the program, from 1.7 points to 2.2 points, in regressions controlling for geographic zone. The impact of weighting on the boys' attendance findings is similar – a small increase in estimates of program impact, but no change in our substantive conclusions. We also used alternative models, limiting the sample to grades 1-8 (a few schools included headcounts with classrooms for grade 9-12) and then limiting the sample to the core grades of the SOMGEP-T intervention (grades 3-8). Neither approach yielded a change in findings.

Attendance from Caregivers

In this section, we turn to a discussion of girls' attendance reported by caregivers. As part of the household survey, caregivers were asked whether their girl attended school in "most" days the school was open during the current school year. If respondents stated that their girl did not attend most days, they were asked to clarify the approximate frequency of attendance. However, for our purposes, we focus on a binary distinction between caregivers who said their girl attended most days, versus those who said their girl did not attend most days. Surprisingly, a meaningful share of caregivers (13.4 percent) reported that their girl attended fewer than "most days" during this round of data collection.

In the figure below, we plot the share of caregivers who said their girl attended most days, over time. Our analysis is limited to the set of girls who appeared in both rounds of data collection: in the left panel, we use the panel of girls who appeared in the baseline and were successfully re-contacted at ML2; in the right panel, we use the panel of girls who appeared in the ML1 round and were successfully re-contacted at ML2.[160]

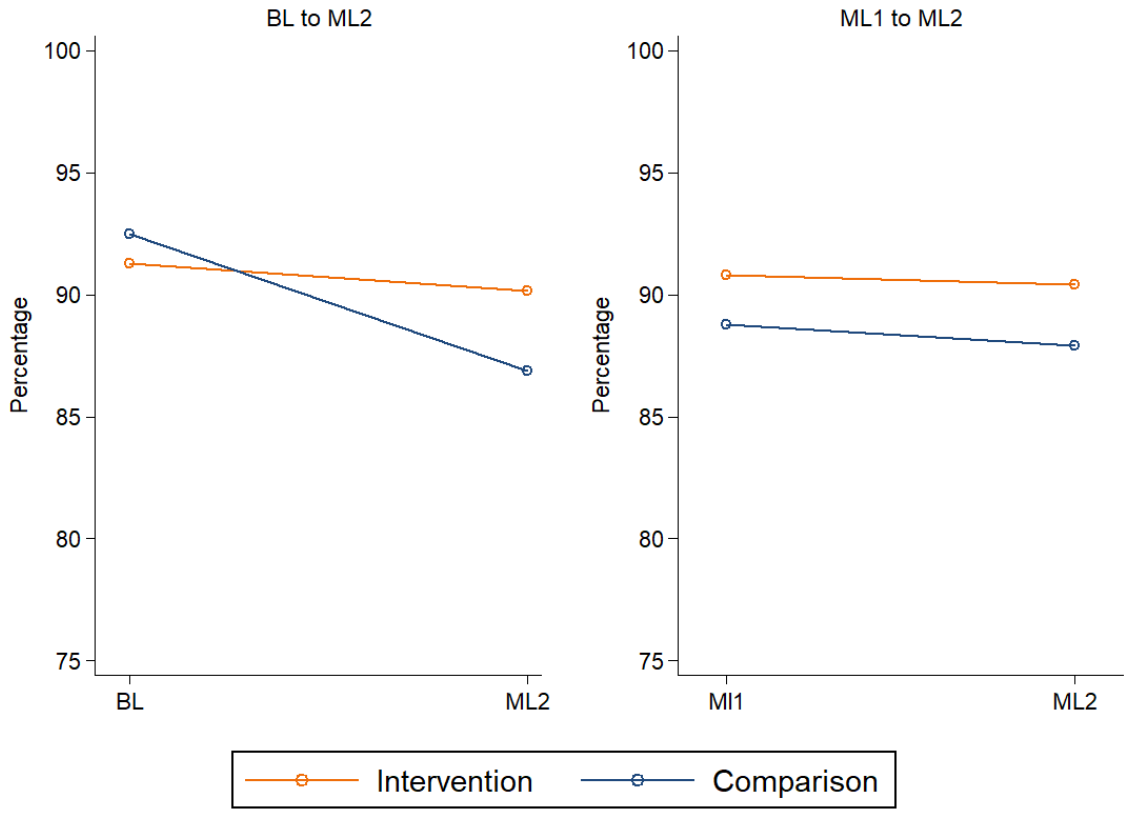
As the results in the left panel make clear, intervention schools performed significantly better than comparison schools on this metric over time. While girls in intervention schools experience a small decline in attendance over time (from 91.3 to 90.2 percent who attend most days), this decline is markedly bigger among comparison girls, whose rate declined from 92.5 to 86.9 percent. These findings are broadly consistent with the trend from baseline to ML2 among headcounts, though it would not be surprising if the results differed due to the fundamental differences in the measures and the time period they assess.[161]

The right panel reports the trend in attendance outcomes from ML1 to ML2, limiting the analysis to a shorter time period and a slightly different overall sample. In general, it appears that there is a small effect of the program between ML1 and ML2, while there is a larger effect between baseline and ML1. In contrast, the previous midline report, ML1, did not report a significant *relative* improvement in attendance from baseline to ML1 among intervention schools. While this disjuncture is surprising, it is also partially a function of the sample selected for analysis. When we limit our analysis to a panel of girls that appeared in all three waves, the main impact of the program is seen between baseline and ML1; however, other samples yield results in which the impact is spread across the two periods. For the purposes of reporting at ML2, the most important finding is that the trend is toward higher attendance in intervention schools, relative to comparison schools, since the baseline.

[160] Despite our focus on analyzing a comparable sample of girls over time, the sample does change slightly from one round to the next in this analysis, as some girls dropped out of school. Such girls did not have their attendance assessed by their caregivers, as they were not enrolled.

[161] As noted above, headcounts assess attendance rates on a single given day, whereas caregivers are asked to assess attendance over approximately half the school year.

FIGURE 27: SHARE OF CAREGIVERS REPORTING THAT GIRL ATTENDS SCHOOL MOST DAYS



A more formal analysis is reported in the table below. As with headcounts, we report a series of linear regression models, estimating program impact using the difference-in-differences estimator. The first model is the basic model, with controls for geographic zone, which suggests the program has increased the share of girls who attend school most days by 2.8 percentage points. The regression models that follow below apply additional control variables, including region, age, and grade level. None of the models employing the full sample are statistically significant; on the other hand, the results are consistent, with estimates of program impact between 2.8 and 3.1 percentage points.

The final two rows in the table report estimates within Somaliland and Puntland, respectively. Unlike the results based on physical headcounts, attendance does not appear to have declined in Puntland-area intervention schools, relative to comparison schools. However, as with the headcounts analysis, the program seems to have had a slightly more positive impact in Somaliland than in Puntland.

TABLE 78: PROJECT IMPACT ON ATTENDANCE, BASED ON CAREGIVER REPORTS

Regression Details	Impact Estimate (Regression Coefficient)	95% Confidence Interval	Sample Size
Control for zone	2.8	-3.7 - 9.4	1,218

Control for region	2.8	-3.8 - 9.4	1,218
Control for region and age	3.1	-3.6 - 9.9	1,218
Control for region, age, and grade	3.1	-3.8 - 10	1,218
Somaliland only; control for region and age	2.9	-6 - 11.7	746
Puntland only; control for region and age	1.1	-6.8 - 8.9	430

Before turning to additional analysis, we sought to judge the veracity of the findings regarding caregiver-reported attendance rates. Because attendance questions were only asked of girls enrolled in school, differential attrition from baseline to ML2 could explain a portion of our results. This is especially true if the girls who drop out of school have the poorest average attendance rates. Consider a situation with differential attrition, in which girls on the margin of dropping out – and currently attending infrequently – are more likely to drop out if they are in intervention communities. In that case, the sample becomes biased over time, as girls with poor attendance gradually leave the intervention sample but remain in the comparison sample. Such a process could explain the positive results reported above. However, a review of dropout rates in the sample suggests that this cannot explain the results, as dropout rates are actually higher in comparison schools, among this sample, than among intervention schools. To the extent that differential dropout rates influence attendance rates reported by caregivers, it would seem to bias our analysis *against* a positive estimate of program impact.

We also sought to isolate the impact of individual interventions, rather than a broad assessment of the entire SOMGEP-T package of interventions. One intervention that we hypothesized would impact attendance rates is the Girls Empowerment Forum (GEF). We expect girls who participate in this forum to feel more supported at school, to have a community of peers that motivates them to attend school, and generally provide support from a dedicated adult in the school. We identified girls who reported they had participated in a GEF activity in the past, over any time period, limiting our focus to girls who appeared in the true panel sample. Among the overall intervention group, we observe a 2.8 point improvement in caregiver-reported attendance, relative to the comparison group; in contrast, we observe a 6.1 point increase among this GEF-active subsample. It appears that the program's impact is concentrated among girls who participate – at least somewhat – in GEF activities. Among girls in intervention schools who *do not* participate in any GEF activities, the program's impact is dramatically smaller, just 0.6 points.[162] While some caution is warranted in interpreting these results, they do suggest that most of the program's impact on attendance is driven by girls who participate in GEF activities; even if that is a function of self-selection into GEF activities on the part of more motivated girls – or girls who are fundamentally different from their peers in another unobserved way – it suggests that the program has an impact on attendance, though the effect may be entirely among girls with the facility or motivation to take advantage of the opportunities SOMGEP-T provides.

[162] Neither effect is statistically significant. The 6.1 point effect estimated among girls participating in GEF activities is calculated relative to the full comparison sample, and has a p-value of 0.16. The estimate of intervention impact on girls who *do not* participate in GEF activities has a p-value of 0.87.

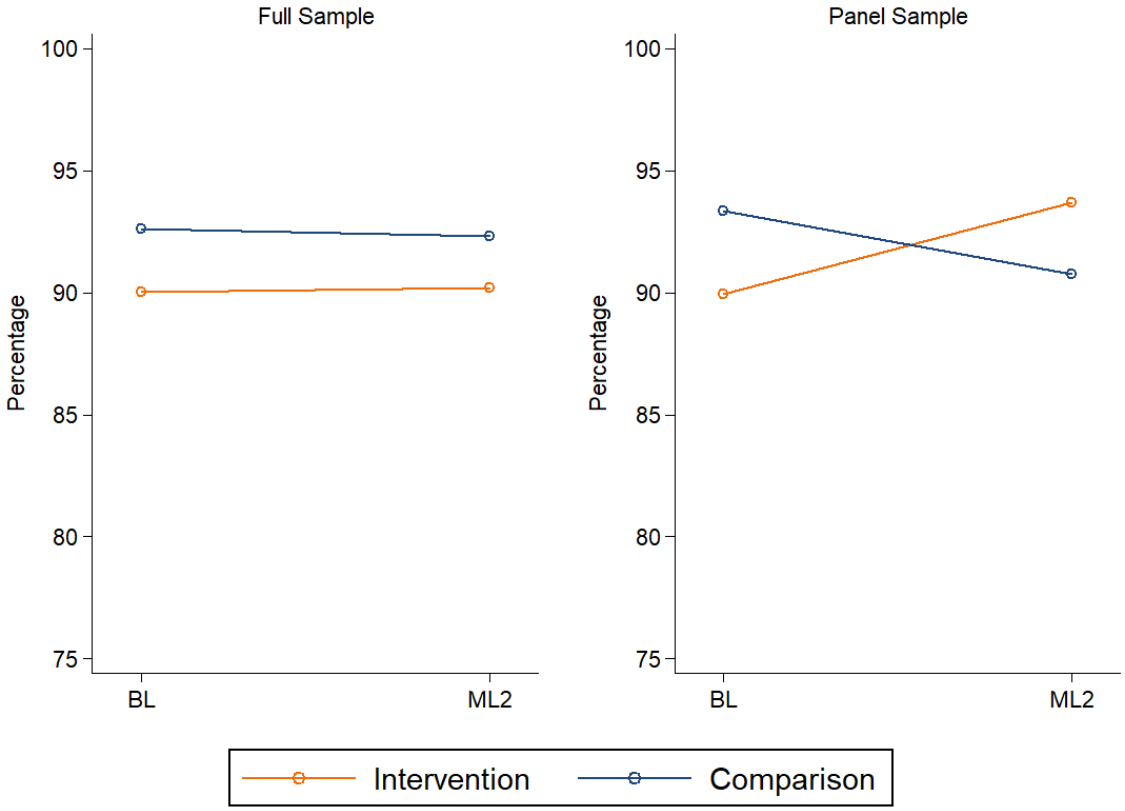
School Attendance Records

Our last major source of information about changes in attendance rates are school attendance records. As the baseline and ML1 reports documented – and as we discuss in more detail in the following section focused on triangulation – there are significant limitations to using school attendance records. The primary issues are the sheer number of girls for whom attendance data is not available, and concerns about the accuracy of record-keeping and data collection based on paper records. The former issue means that our analysis is based on a relatively small subsample of cohort girls (501 girls at baseline and 727 at ML2). The latter issue makes us hesitant to draw strong conclusions from the available data, as systematically inaccurate records in just a few schools could severely bias our results. For this reason, we consider the school records mainly as a method to triangulate our previous findings.

To clarify the nature of this data: at each formal primary school, team leaders reviewed school attendance records and recorded attendance information for each cohort girl over the course of the full year. Specifically, they recorded the number of days the girl attended and the number of possible days of attendance thus far that year. Therefore, our measure is the attendance rate for a given girl in the current year.

The figure below reports the attendance rate at baseline and ML2 among intervention and comparison groups. In both rounds, there are more girls with documented attendance records in intervention schools than in comparison schools – this may be a function of better record-keeping in intervention schools, or head teachers being more willing to provide access to the records in intervention schools. The left panel of the graph uses every cohort girl for whom attendance records were available in the two rounds, even if the set of girls changed over time. As the graph shows, there is very little change in attendance rates in either intervention or comparison schools over time, and the difference-in-difference estimate of program impact is very small, albeit positive.

FIGURE 28: CHANGE IN ATTENDANCE RATES OVER TIME, BY INTERVENTION STATUS



On the other hand, the right panel of the graph reports results using the panel of girls who have attendance records at both baseline and ML2. This is a much smaller sample, just 387 girls in each round, with a heavier skew toward girls in intervention schools. In this sample, the impact of the program is fairly dramatic, increasing attendance rates by 6.3 percentage points in intervention schools from baseline to ML2, relative to changes in the same time period among comparison schools.

We do not report a more detailed regression analysis at length in this section. However, we used a series of linear models, incorporating additional control variables, to ensure that the results from the graphs are not an artefact of the regression model. The results are highly sensitive to the sample employed, but additional control variables do not affect the results. In general, attendance among girls with records in both rounds went up over 6 points, vis-à-vis comparison schools, though this improvement is not statistically significant.[163]

Data Quality and Triangulation of Attendance Rates

The analysis thus far in this section has reported aggregate results showing small, mostly positive impacts of the program on attendance rates. However, the strength of that conclusion hinges on the source of data that one considers. In this discussion, we attempt to shed light on the quality of data underlying the different results, and use all three sources to draw conclusions regarding program impact.

As aforementioned, physical headcounts in classrooms are generally the most reliable metric available, but capture attendance on a single day. Moreover, attendance on that day could be biased by the presence of a

[163] $p = 0.14$.

data collection team associated with the project. Our first test of the veracity of headcounts was to consider whether attendance was impacted by the presence of data collection teams. A simple check of this was to compare the attendance reported by the teacher yesterday and today; if the school was encouraging girls to come to school on the day of a visit, we would expect higher attendance on the day of the visit, compared to the day before. In the table below, we report attendance rates reported for the two days. As the table shows, teachers actually reported higher attendance rates on the day *before* a visit than on the day *of* a visit, which suggests they are not inflating attendance on the day of the headcount.[164]

TABLE 79: GAP BETWEEN TEACHER-RECORDED ATTENDANCE AND HEADCOUNTS, DAY OF VISIT

Round	Register Attendance Today	Register Attendance Yesterday
Baseline	84.8	86.7
Midline #1	79.9	82.8
Midline #2	72.7	82.7
Total	81.6	84.1

As a second check on the quality of headcounts, we compare the attendance rates recorded by teachers on the day of the visit, versus our own headcount. Because both rates take class enrolment figures as their denominator, this comparison simplifies to a comparison of the count of girls our teams made versus the count made by teachers on the same day. While there were some cases of wide divergence between the two counts, they are not significantly different, either among intervention or comparison schools – in the aggregate, across rounds, the gap between the two counts was equivalent to just 0.4 percentage points of attendance. We employed two additional checks on headcount data quality that did not reveal any reasons for concern.[165]

In comparison to headcounts, school records are prone to more systematic and potentially severe biases, but are desirable for the length of time they cover. The issues with school record-keeping cannot be overstated, as our data collection teams routinely report on GEC-T and other education projects in Somalia. During ML2 data collection, team leaders were asked to rate the completeness of attendance records kept in each classroom; just 28.3 percent of classrooms were rated "extremely complete" and just 22.3 percent were considered "extremely neat and organized." Team leaders also recorded how many days – out of the last five – had attendance data recorded in the classroom register. This year, 27.8 percent of classrooms had not recorded all five of the previous five days of attendance data. Finally, these assessments of classroom-level

[164] It is possible that teachers are inflating reported attendance on the day prior, knowing that data collection teams will record attendance for that day. This would be an easier way to falsify attendance rates than encouraging girls to come to school on the day of a visit. However, such inflation has no effect on our headcounts, because it drives *yesterday's* attendance upward, without influencing the number of girls in school on the day of the field team's visit.

[165] The first approach was to compare average attendance rates, derived from headcounts, for the same schools across baseline and ML2. This process would allow us to identify any schools with a suspicious shift in attendance rates. While some schools did experience moderately large changes, none were sufficiently large to elicit concern. The second approach was to review average attendance rates across team leaders, to assess whether individual enumerators understood or filled the survey differently. While there were small differences in average attendance rates reported by the 11 team leaders who completed headcounts during ML2 – and in prior rounds – there were no systematic differences that were not explained by the location of the headcounts or other contextual details.

attendance record-keeping do not indicate whether classroom-level attendance is reported to and recorded by the head teacher, or otherwise kept systematically over the course of the full year.[166] Clearly, attendance record-keeping in schools is not always a reliable indicator of true attendance.[167]

Given these supplementary investigations, how should one adjudicate between alternative data sources? To recapitulate the results:

- Headcounts
  - Small, 1.5 point increase in girls' attendance in intervention schools since baseline
  - More substantial improvement in boys' attendance
- School records
  - Among the full sample of girls with records, no effect from baseline to ML2
  - Among the sample of girls appearing in both rounds, strong positive effect from baseline to ML2
- Caregiver reports
  - No effect from ML1 to ML2
  - Moderate positive effect from baseline to ML2

In data with contradictory results, we tend to favour the more rigorous option. In the case of the school records, our preference is for results based on the panel sample; in the case of caregiver reports – and most analyses in this report – we tend to favour baseline to ML2 comparisons, as they draw from a larger overall sample and capture the full program impact over time. Given these preferences, it is clear that the program has had a small or medium, but statistically insignificant, impact on attendance rates since baseline. All three preferred analyses point toward a positive effect, one which is substantively meaningful in the context, but which cannot be distinguished from a null result.

Subgroup Analysis of Attendance Rates

The analysis in this section is driven largely by the contents of the qualitative data. Given the number of data sources and possible individual- and school-level characteristics that are likely to be correlated with attendance rates, we chose to focus our quantitative analysis on the topics or issues that emerged from the qualitative data. In other words, we chose to analyse the relationship between attendance and other variables in the data based on whether they were specifically mentioned as barriers to attendance in qualitative interviews.

The first subgroup we consider are girls in schools that have been affected by conflict in the past 12 months. This analysis is limited to the ML2 round, because explicit questions regarding the experience of conflict were not included in previous rounds of data collection. At an individual and school level, we are interested in determining whether conflict is associated with lower attendance rates. We define conflict based on reports of caregivers; when we perform school-level analysis of headcounts, we code a school as conflict-affected if at least two caregivers in the community reported conflict.

[166] Some teachers may record attendance on a daily basis, but not record this information systematically or transmit the information to the head teacher.

[167] Unlike our review of headcounts, we did not directly compare attendance rates derived from school records across rounds for the same schools, because the set of girls available in each round often resulted in uninformative comparisons. Also, given the number of girls for whom this data was collected in each school, changes in just a few girls' attendance rates could dramatically shift the overall average, as field teams collected records for fewer than 10 girls in a considerable number of schools.

Overall, there does not appear to be a systematic relationship between conflict and attendance. In the case of headcounts, attendance rates are 3.1 points higher in schools affected by conflict than in schools that were not, after controlling for region. We find a similar pattern in individual-level attendance rates: girls living in households that report having experienced violent conflict missed fewer days (0.7 days fewer, on average) than other girls. When we consider the share of girls who have attended "most days" of the last school year, we also find that caregiver-reported attendance rates are higher in communities that have experienced conflict. It is possible that – before conflict tends to be geographically concentrated – our findings could be biased by broadly higher attendance rates in the same regions that are prone to conflict. On the other hand, it is also possible that any effects of conflict dissipate rapidly over time, and the conflicts in question have mostly or entirely ended.[168]

One of the most common themes in the qualitative data on transition and attendance was the role of domestic work in causing girls to miss school. When girls were presented with hypothetical stories about girls their age who had significant household responsibilities, they were more pessimistic – on average – than in the context of any other type of story. Characters in the story who struggled to learn or who were pastoralist were seen as more likely to continue their schooling than the girls who were presented as needing to help out a lot at home. For instance, when presented with a story about Barwaaqo, who is a high-achieving student but also helps out at home, some students fixated on the role of chores and indicated that she would leave school soon as a result.[169] Other stories prompted similar responses.[170]

The view that chores are a significant barrier to attendance was not exclusive to girls. Mothers occasionally cited household work as a reason that girls do not attend school or may drop out of school altogether, as did other types of qualitative interviewees.[171] One Ministry of Education official listed domestic work – alongside early marriage – as the two biggest challenges they face to ensuring girls' education.[172] The prevalence of chores as a burden for girls has been documented elsewhere in this report; however, it is worth noting, again, the number of girls – 66.4 percent at ML2 – who complete more than a half-day of chores per day, according to their caregivers.[173] Among caregivers themselves, 16.4 percent state that the chore burden their girl faces prevents them from attending school at times, or even enrolling in the first place. Based on the viewpoints expressed by girls in qualitative interviews, the majority opinion of girls is that domestic work is one of the top three reasons girls drop out of school or have poor attendance.

The emphasis on domestic work seems to be well-placed. Among girls facing a high chore burden (half or a full day of reported chores) just 84.5 percent were reported to attend school most days; among girls who complete a quarter-day or less of chores (including no chores at all), 90.0 percent attend school most days. This result excludes girls who have dropped out of school entirely – an additional potential effect of domestic work – in order to focus exclusively on attendance as an outcome. Girls facing a high chore burden missed

[168] A further limitation of our analysis is that it is strictly cross-sectional and includes only data from ML2. A stronger investigation would study the impact of the program on attendance within versus outside conflict-affected communities.
[169] Vignettes FGD with girls – Somaliland – Int. 142.
[170] Vignettes FGD with girls – Somaliland – Int. 141.
[171] FGD with mothers – Somaliland – Int. 115; FGD with mothers – Puntland – Int. 211; FGD with CEC members – Somaliland – Int. 104.
[172] KII with MOE official – Somaliland – Int. 161.
[173] Our expectation is that this is an underestimate of the time required, in most cases. Individuals have a tendency to underestimate the amount of time other people spend on a task, and caregivers have an incentive to underestimate the burden household work places on their daughter for psychic reasons.

2.8 days of school, on average, over the past month, compared to just 1.8 days for the group responsible for fewer chores.174

It is worth noting that the program itself does not appear to have reduced the chore burden of girls. While the relative chore burden of girls in intervention schools has declined slightly since baseline, a bigger decline has been witnessed in comparison schools. Indeed, if we were to estimate the impact of the program on the share of girls who complete a half-day or more of domestic work daily, the impact would be substantively negative, though not statistically significant.175 In short, the evidence suggests that chore burden is a significant predictor of lower attendance rates, but that the program itself has not contributed significantly to reducing girls' chore burdens.

Another factor that emerged from the qualitative data in this round – and in previous rounds – was the role of school sanitation facilities in shaping attendance rates for girls. The girls participating in qualitative interviews reported that the toilet facilities are often unclean, and cited the toilets as a source of anxiety and fear, because they might be spied upon.176 These accounts were confirmed, at times, by our fieldwork team leaders and even by MOE officials, who might not be expected to have as fine-grained of concerns about school infrastructure and its impacts on girls.177 While MOE officials tended to discuss schools that lacked separate toilet facilities altogether, the emphasis on toilets – whether their sanitation or their presence at all – is still noteworthy. Typically, a lack of suitable sanitation facilities means that girls will travel to somewhere else in the community, or their home, to use the bathroom, which reduces their instructional time. Inadequate facilities may also reduce attendance for entire days, if girls are less likely to attend school during their menstrual period because the toilets are unsuitable, or if they stay home when they are slightly ill, knowing that the facilities at school will be unpleasant to use.The data collected during this evaluation round are suggestive of a link between the adequacy of toilet facilities and girls' attendance rates. As the table below demonstrates, girls who attend schools where sex-separated toilet facilities are available missed fewer days of school in the last month, according to their caregivers. Their caregivers are also more likely to report that they attended school at least half of all days in the previous year, though both relationships are far too weak to be definitive.

TABLE 80: GIRLS' ATTENDANCE AS A FUNCTION OF TOILET AVAILABILITY AND QUALITY

	Girl will use toilet at school	Girl will not use toilet at school	
Days missed in last month	2.29	2.43	
Share of girls who attend half time or more	95.7%	94.9%	
		Separate Girls Toilets Available	Separate Girls Toilets Not Available
Days missed in last month	2.05	2.93	

[174] This result is statistically significant, after accounting for clustering at the school level, and a number of control variables, including school fixed effects in the most extreme case. The results are based on data from ML2 only, estimated in a cross-sectional regression.
[175] As with most of our estimates of program impact, we focus on the true panel of girls from the baseline re-contacted at ML2; we also studied the true panel of girls from ML1 re-contacted at ML2. In both cases, the results were substantively similar.
[176] Risk Mapping Exercises with Girls – Somaliland – Int. 123; Risk Mapping Exercise with Girls – Somaliland – Int. 121.
[177] Interview with Fieldwork Team Leader – Somaliland; KII with MOE official – Somaliland – Int. 161; KII with MOE official – Puntland – Int 261; KII with MOE official – Galmudug – Int. 262.

Share of girls who attend half time or more	95.7%	96.6%

At the school level, there is also a mild relationship between the attendance documented through classroom headcounts and the availability of separate toilets. Among schools with sex-separated toilets, headcounts documented attendance rates between 0.5 and 2.1 percentage points higher than in schools without such toilets.178 While a number of caveats apply to this finding, one compelling piece of evidence is the fact that the impact of sex-separated toilets on boys' attendance is much lower than that on girls' attendance. This is consistent with our expectations, as the lack of sex-separated toilets do not present as significant a barrier to boys' attendance as to girls'. This analysis also does not address a more pernicious – and unobservable in our available data – effect of toilet inadequacy: the reduction in instructional time when girls travel away from school to use outside facilities. Capturing this effect at endline is possible, if questions are integrated into the survey module and qualitative interviews targeting girls.

Disability status has also been a minor theme of the qualitative interviews in each evaluation wave, as qualitative interviewees report that girls with disabilities are both less likely to enrol in school and less likely to attend consistently. For instance, a frequent point of discussion was the effect that a lack of mobility – among girls with disabilities that impair their walking – on attendance and enrolment rates, because girls cannot reach school on time.[179]

The final factor influencing attendance that emerged clearly from the qualitative data was the role of CECs in monitoring student attendance. As one mother described the situation, "if [her children] are a little late, the principal calls me. They go regularly. Even the committee calls us and tells us that our children are not at school."180 Other mothers reported that head teachers and teachers reported to parents when students were absent from school and that active monitoring prompted better attendance from students.181 The general impression received from the qualitative data is supported by the quantitative data in each round of the evaluation: at baseline, ML1, and ML2, caregivers of girls with disabilities reported lower attendance rates for their girls, on average. In the most recent round, 90.5 percent of caregivers of girls without disabilities reported that their girl attends school most days, compared to 84.4 percent of caregivers of girls with a disability. This gap can be observed in both intervention and comparison schools. When caregivers were asked how many days their girl had missed in the last month – a question with a higher likelihood of accurate recall – the same pattern emerged, with girls with disabilities missing a disproportionate number of school days, on average.

Unfortunately, the quantitative data do not appear to support this view of CEC impact on student attendance. For instance, when we investigated the relationship between CEC monitoring of student attendance – as reported by head teachers – we found that schools in which this occurred had *lower* headcount attendance rates, all else being equal. The relationship in question was quite stark: where a CEC had monitored student attendance on their last school visit, girls' attendance was lower by 6.6 percentage points than schools where the CEC either did not visit or had visited but had not monitored student attendance. At the same time, this finding should be interpreted cautiously, as CEC monitoring of student attendance is not random – CECs

[178] The exact magnitude of the relationship is sensitive to modelling choices, such as the inclusion of zone as a control variable. Importantly, this is a cross-sectional regression model, limited to ML2 data, and we are unable to infer that sex-separated toilets *cause* higher attendance rates – a caveat that applies to most of the subgroup analysis presented in this section.
[179] FGD with Teachers, Galmudug, Int. 331.
[180] FGD with Mothers, Somaliland, Int. 111.
[181] FGD with Mothers, Somaliland, Int. 111; FGD with Teachers – Puntland – Int. 231.

overseeing schools with low attendance rates may be more likely to specifically monitor attendance, because they are aware of the problem in their schools.[182]

When we analyse the relationship between CEC activities and individual-level girls' attendance, the results are slightly less negative. First, it appears that the program has had an impact on CEC activity levels in the form of communication with parents. Using the same difference-in-differences framework with which we estimate program impact throughout this report, the data show that caregivers are 10.3-11.0 points more likely to report that their CEC communicates with them at least monthly, as a result of the program.[183] At the same time, CEC communication and activity levels have only a weak and inconsistent relationship with attendance rates, according to attendance figures reported by caregivers. While we find that caregivers who report more communication from their CEC also report better attendance in some regression models, this finding is extremely dependent on often arbitrary choices regarding model specification. In other models, the effect is reversed, reducing our confidence in the finding.

The subgroup analysis in this section was driven by qualitative reports of barriers to attendance that have been a common theme across at least three rounds of SOMGEP-T data collection. Toilet facilities, household chores, and – to a lesser extent – the role of CEC or head teacher monitoring of, and reporting on, student attendance have all been cited as impacting student attendance rates. The quantitative results, on the other hand, are less clear. There does appear to be a relationship between the availability of sex-separated toilet facilities and attendance – specifically girls' attendance. The finding is theoretically motivated, as well. However, a stronger research design would involve tracking student attendance before and after the construction or rehabilitation of toilet facilities, as it is difficult to separate the effect of toilet construction from the effect of other factors, including other interventions implemented by the program.

The strongest quantitative results concerned domestic chore burden. In a cross-sectional model, there is a strong relationship between higher chore burdens and lower attendance rates. Again, there are concerns about interpreting this as a causal relationship, but the relationship is sufficiently strong that it merits further research and consideration in program design. This is especially true because the program does not seem to have reduced the extent of domestic household work, at least as reported by girls' caregivers. It is possible that the program has reduced domestic work around the margins, or influenced girls' attitudes sufficiently that they are more likely to attend school *despite* having chores waiting for them at home. However, unpacking this relationship and others discussed in this section would require a more targeted research design than is possible as part of this broader impact evaluation.

[182] This is one instance of "Ashenfelter's Dip", in which individuals selected into a treatment (in this case, CEC monitoring of student attendance) may have been selected into that treatment – or self-selected into that treatment – as a result of lower-than-average performance on the outcome in the pre-treatment period. It is a common threat to drawing causal inferences in longitudinal research designs.

[183] To illustrate: at baseline, parents in intervention communities were 4.6 points *less likely* than those in comparison communities to indicate that their CEC communicated with them monthly or more often. By ML2, that trend had flipped, and parents in intervention communities were 5.7 points *more likely* to report the same, vis-à-vis parents in comparison communities. The sample composition is controlled for explicitly in this analysis – as in most of our difference-in-differences estimates – by using the sample of girls who were successfully contacted at both baseline and ML2.

7.2 SCHOOL GOVERNANCE AND MANAGEMENT

Outcome	Indicator	BL	ML2 Target	ML2	Target achieved? (Y/N)	Target for next evaluation point	Will IO indicator be used for next evaluation point? (Y/N)
School governance	CECs' perceptions of the importance of the retention of marginalised sub-groups, such as pastoralists	Limited number of CECs described paying pastoralist children's school fees and providing remedial courses for displaced students to catch-up. No engagement in retention.	CECs include retention of marginalised groups in school improvement plans	See below			Y

Main qualitative findings

- CEC members seem to be quite aware of the difficulties faced by these groups. However, most committee members viewed addressing barriers to their education as beyond their means due to the very severely limited financial resources at their disposal.
- There are many anecdotal accounts of efforts to reach out and include members of marginalised groups
- The efforts are, however, more focused on pastoralists and the disabled. Members of traditionally marginalised occupational minority groups are not receiving the same attention as interviewees often state 'there are no minorities here'.
- Findings from the head teacher survey indicate at least partial improvements as 78.3 percent of school management plans in the intervention group now include plans to follow-up with dropouts

Outcome	Indicator	BL	ML2 Target	ML2	Target achieved? (Y/N)	Target for next evaluation point	Will IO indicator be used for next evaluation point? (Y/N)
School governance	Percentage of CECs in ASLPs /	34.4% of CECs (interventio	80%	65.5% (interventio n schools)	No	85%	Yes

schools tracking girls' retention	n schools) monitoring attendance

Main qualitative findings

- CECs are clearly active in a lot of fields, such as monitoring the school for teaching quality, attendance and corporal punishment. Their activity, however, is constrained by the lack of training of CEC members as well as few financial resources to address more substantial barriers to attendance (transport issues, school fees in areas of a weak economic base etc.)
- The program does not seem to be having an effect in this regard. Intervention schools started out with more active CECs. The difference has since become narrower whereby the activity levels in intervention schools are subsiding while increasing in comparisons schools.

Improved school governance and management is the second SOMGEP-T intermediate outcome. It is not only important for the sustainability of the project, but is also an essential outcome for improvements in students' learning and transition. The purpose of this section is to assess the degree to which SOMGEP-T has realized its intended outcomes with respect to school governance and management affairs. A difference-in-differences (DID) approach is first used to establish program impact along various indicators, and tio assess the extent to which school governance and management has changed in intervention schools vis-à-vis the comparison group.

The analysis utilizes the same key indicators established in the quality of school governance baseline and the previous Midline. Since the baseline study no teachers' survey has been utilised, and as such the indicators are based the survey with head teachers and the household survey, completed by primary caregivers. Therefore, the results presented below fall into two main categories: the first category examines the extent of Community Education Committees' (CECs) establishment and performance from the head teachers' perspectives, while the second category assesses CECs and school management by asking the primary caregivers to rate how well the school is managed and how head teachers performed.

We then look at a number of qualitative data points, such as the focus groups discussions conducted with members of the CEC as well as those with teachers and the Key Informant Interviews (KIIs) with regional education officers (REOs).

As it pertains to school management and governance, it appears that progress in this regard has been somewhat wanting. The target was for 80 percent of the CECs to be monitoring attendance at ML2 – 68.8 percent did. Along many of the indicators the gap in favour of the intervention schools has been narrowing since baseline as CECs in comparison schools are improving their performance. This is cause for alarm and the program should put more emphasis on making the CECs more active in intervention schools. This is particularly important as they have a big role to play in making the intervention sustainable.

COMMUNITY EDUCATION COMMITTEES' ASSESSMENT FROM HEAD TEACHERS' PERSPECTIVES

This section focuses on the performance of the CECs as they are the most important body of school management at the local level. The committees consist of local volunteers including parents, religious leaders,

head teachers, and members of women's and youth groups.[184] The body has many responsibilities, including acting as a go-between and liaison for the relationship between the school and the local community. Members of the CEC also monitor various things at school such as teacher and student attendance and performance. The CEC is also involved in making the school development plans which include various elements that are of relevance to the sustainability of the SOMGEP-T project, such as plans for enrolling out-of-school girls. The CECs often also raise money from the community and elsewhere to cover teacher salaries or to make improvements in school facilities and resources. The material and immaterial support that the body provides is essential for the sustainability of the project. Moreover, an active and effective CEC may positively contribute to learning outcomes through its influence on teaching quality, practices and resources at the schools.

We first look at the level of CEC activity at the schools using the head teacher survey data. We have a variety of variables to look at. However, some of them were only introduced at the previous Midline, and as such do not provide a comparison between Baseline and Midline Round 2. Let us begin by looking at those where the comparison between Baseline and now is possible.

The below table presents the regression and difference in differences results for the variables in question.

TABLE 81: PROJECT IMPACT (DIFFERENCE-IN-DIFFERENCES) ON SCHOOL MANAGEMENT – BL TO ML2

	Intervention		Comparison		Difference in Differences	p value
	BL	ML2	BL	ML2		
School has management plan	45.9%	73.0%	12.5%	40.6%	-1.1	.933
CEC member has visited school	64.9%	86.5%	25.0%	68.8%	-22.1	.134
Monitored: teacher attendance	83.3%	68.8%	50.0%	72.7%	-37.3	.147
Monitored: facilities	50.0%	62.5%	25.0%	36.4%	1.1	.962
Monitored: teaching quality	54.2%	53.1%	50.0%	36.4%	12.6	.609
Monitored: student attendance	58.3%	68.8%	37.5%	59.1%	-11.2	.62
Monitored: student retention	37.5%	37.5%	25.0%	31.8%	-6.8	.766

Firstly, when looking at whether a school has a management plan, the most important tangible product of CEC activity, we can see that the share of schools that have the plan has increased significantly across the board. In treatment schools at baseline less than 46 percent had the management plan in place. At this Midline the share is 73 percent. When looking at the comparison schools, however, their increase in school management plans has in fact been very similar – increasing from a much lower starting point, 12.5 percent, to 40.6 percent.

Next, when looking at whether a CEC member has visited the school, the difference in differences is clearly favourable to comparison schools that have seen a dramatic increase from a quarter at Baseline to nearly 70 percent at this Midline. Meanwhile, intervention schools have seen an increase of more than 20 percent

[184] James H. Williams & William C. Cummings (2015) Education from the Bottom Up: UNICEF's Education Programme in Somalia, International Peacekeeping, 22:4, 419-434

points, to 86.5 percent. Thus, the intervention schools seem to maintain a much higher level of CEC activity in this regard, yet comparison schools have improved their performance markedly more since the Baseline as indicated in the negative DID estimate.

When it comes to CECs monitoring teacher attendance, the intervention schools are reporting lower levels at Midline Round 2 when compared to the Baseline (83.3 to 68.8 percent). Simultaneously comparison schools have gone from exactly half of the schools to 72.7 percent reporting CEC monitoring of teacher attendance. As per facility monitoring, data from head teacher surveys shows a very similar increase for both intervention and comparison schools while the level of activity for comparison schools continues to be much lower. Interestingly, monitoring teaching quality seems to have decreased for comparison schools while remaining largely the same for intervention schools.

Conversely, when asked about CEC monitoring student attendance, head teachers in comparison schools demonstrate a more substantive increase in comparison schools than those in intervention schools, with the DID estimate being -11.2. However, as for many of the indicators, the levels for intervention schools remain higher (68.8 percent vs. 59.1 percent).

Finally, the monitoring of student retention seems to have remained exactly the same for intervention schools while having increased somewhat for comparison schools. Here, again, the activity continues to be more prevalent in intervention schools.

In terms of changes between the two midline rounds, the following table shows the changes for the same variables above. While all changes since the previous round have been positive in favour of the intervention schools, only the variable measuring the CEC's monitoring of teaching quality has increased in a statistically significant manner in the intervention schools. This is largely due to the fact that this activity seems to have declined substantially in the comparison schools since Midline Round 1.

TABLE 82: PROJECT IMPACT (DIFFERENCE-IN-DIFFERENCES) ON SCHOOL MANAGEMENT – ML1 TO ML2

	Intervention		Comparison		Difference in Differences	p value
	ML1	ML2	ML1	ML2		
School has management plan	53.1%	71.9%	38.7%	41.9%	+16	0.332
CEC member has visited school	87.5%	87.5%	67.7%	67.7%	0	1.000
Monitored: teacher attendance	75.0%	67.9%	85.7%	76.2%	+2	0.891
Monitored: facilities	53.6%	57.1%	42.9%	33.3%	+13	0.406
Monitored: teaching quality	46.4%	53.6%	81.0%	38.1%	+50	0.017 *
Monitored: student attendance	64.3%	75.0%	52.4%	57.1%	+6	0.749
Monitored: student retention	32.1%	39.3%	52.4%	28.6%	+31	0.109

Before presenting some overall observations, we will turn our attention to those variables where our comparisons are only possible for the time between Midline Round 1 and Midline Round 2.

TABLE 83: PROJECT IMPACT (DIFFERENCE-IN-DIFFERENCES) ON CEC EXISTENCE AND ACTIVITY

	Intervention		Comparison		Difference in Differences	p value
	ML1	ML2	ML1	ML2		
Does this school have a functioning CEC?	96.9%	100.0%	96.8%	93.5%	6.4	.16
School management good or very good	90.3%	90.6%	86.7%	89.7%	-2.7	.79
CEC meetings once a month or more	83.9%	93.8%	63.3%	75.9%	-2.6	.84
Has a CEC member come to school to monitor facilities/teaching quality/attendance?	87.5%	87.5%	67.7%	67.7%	0.0	1.00
In past year, # visits by CEC member to monitor facilities/quality/attendance?	14.3	8.8	15.2	9.1	0.7	.88

The first step in effective governance of the schools by CECs is to establish the CEC itself. At Midline Round 2 now all head teachers in intervention schools report that the school has a functioning CEC. All but one comparison school head teacher reported the same. As per school management, the share of head teachers viewing it either very good or good remained very high, although comparison schools have increased their share slightly.

In terms of the CEC's level of activity, head teachers reported the committee meeting once a month or more frequently in more than 93 percent of the time in intervention schools (up 10 percent from ML1). This share remains lower for comparison schools, yet it has increased more in their case – from 63.3 to 75.9 percent of head teachers. Head teachers reporting CEC members visits to the school remains unchanged for both intervention and comparison schools, resulting in this remaining much more frequent in intervention schools (87.5 percent vs. 67.7 percent). Yet, the number of visits reported by head teachers has decreased across the board, from 14.3 to 8.8 times and from 15.2 times to 9.1 times average for intervention and comparison schools respectively. None of the difference in differences estimates are statistically significant.

TABLE 84: PROJECT IMPACT (DIFFERENCE-IN-DIFFERENCES) ON SCHOOL MANAGEMENT

	Intervention		Comparison		Difference in Differences	p value
	ML1	ML2	ML1	ML2		
School has management plan	53.1%	71.9%	38.7%	41.9%	15.5	.33

Does plan include info on monitoring school (attendance, teach practices, etc)?	88.2%	73.9%	75.0%	69.2%	-8.6	.67
Does plan include info on child protection policies or plans to improve CP?	82.4%	73.9%	50.0%	61.5%	-20.0	.40
Does plan include info on encouraging school enrolment?	82.4%	91.3%	83.3%	69.2%	23.1	.18
Does plan include info on plans to follow-up with dropouts?	76.5%	78.3%	83.3%	69.2%	15.9	.38

We now turn our attention to the school management plan. While at baseline we had asked whether the plan exists, at ML1 questions were added to inquire for different elements and contents of the plan. We can now compare with some more nuance how the plans are developing. Between ML1 and ML2, we can see that the share of schools that have the plan to begin with has increased much more among the intervention schools, moving from 53.1 percent to 71.9 percent, while the same figure has increased only very little among the comparison schools – from 38.7 percent to 41.9 percent. While the DID estimate is not significant – which is not surprising given the very small sample size – it is substantive, at 15.5 percentage points. As per the contents of the management plan, the findings are bifurcated. The prevalence of all of the elements remains higher in intervention schools. However, for the plans inclusion of monitoring information the share of schools with said information in their plans has decreased across the board, yet more so in intervention schools.

Similarly, the plans' inclusion of child protection policies or plans has increased among comparison schools while decreasing in intervention schools. Conversely, the school management plans in 91.3 percent of intervention schools now include information on encouraging enrolment (up from 82.4 percent) while this share decreased for comparison schools (from 83.3 percent to 69.2 percent). Something similar has happened with regards to plans for following up with dropouts. This is a positive finding, moreover, as the program logical framework has set the objective that "CECs include retention of marginalised groups in school improvement plans". 78.3 percent of school management plans in the intervention group now include plans to follow-up with dropouts.

The changes are most likely entirely attributable to the schools that have recently developed their school management plans. The overall share of schools with said plan, is much higher among the intervention group, which is encouraging. It also seems that the focus in the content of the plans is more on enrolment and follow-up on dropouts in the intervention group as that is where the gap is growing between intervention and comparison schools.

Yet, again, it should be noted that all elements of the school management plan were more prevalent among the intervention group.

TABLE 85: PROJECT IMPACT (DIFFERENCE-IN-DIFFERENCES) ON CEC MONITORING

	Intervention		Comparison		Difference in Differences	p value
	ML1	ML2	ML1	ML2		

Does CEC monitor student attendance records?	50.0%	65.6%	54.8%	51.6%	18.9	.17
Does CEC follow-up with or contact dropouts?	56.3%	71.9%	61.3%	67.7%	9.2	.53
Does CEC monitor teacher attendance?	65.6%	71.9%	71.0%	64.5%	12.7	.29
Does CEC take action against teachers who are not attending school regularly?	68.8%	68.8%	54.8%	51.6%	3.2	.83
Does CEC raise funds for school improvements?	59.4%	56.3%	35.5%	35.5%	-3.1	.83
Does CEC reinforce use of non-violent discipline instead of corp. punishment?	65.6%	68.8%	48.4%	58.1%	-6.6	.69
Does CEC address child protection issues?	75.0%	78.1%	58.1%	74.2%	-13.0	.39
Does CEC promote enrolment of OOS children?	75.0%	78.1%	54.8%	61.3%	-3.3	.84

Finally, we look at specific CEC activity and monitoring as reported by the head teachers. The findings are again somewhat divided. Half of the indicators show a positive difference in differences, while the other half shows a negative one.

When it comes to CEC monitoring student attendance records, or following up with dropouts, the developments are positive in that the intervention group head teachers report these practices to have increased more than in the comparison group. The same applies when looking at CEC monitoring of teacher attendance or taking action against absentee teachers.

However, when looking at raising funds, promoting the use of non-violent disciplinary methods, addressing child protection issues or promoting the enrolment of out-of-school girls, these activities have increased more in comparison schools that in intervention schools. Yet, all of the practices remain more prevalent among the intervention cohort, given the higher levels at start, at Midline Round 1. Consequently, the program meets its objective of 30 percent of CECs taking action against corporal punishment and other child protection issues as such the share of intervention schools reporting the CECs addressing child protection issues is 78.1 percent.

When looking at these variables overall, regardless of whether we focus on those where we can compare between BL and ML2 or those where comparisons can only be made between the two Midline rounds, some general observations emerge.

Notably. the difference in differences estimates do not run consistently in one direction. Some of them are positive, while others are negative. In fact, exactly half of the variables looked at here had a positive difference in differences estimate. Meanwhile the remaining variables show a negative difference in differences, indicating that the levels of reported CEC activity have in fact increased more for comparison schools. Yet, some factors should be noted.

Firstly, the data used here is from the head teacher survey which is limited to one per school. When looking at the sample between Baseline and Midline Round 2, this sample consists of 69 schools, with 32 in comparison, and 37 schools in the intervention group. The sample for comparing ML1 and ML2, similarly,

consists only of 63 schools – 32 in intervention and 31 in comparison. The sample is thus extremely small and consequently subject to fluctuations.

Yet, the findings for arguably the most important variables – i.e. whether the school has a functioning CEC or not, or whether the school has a management plan – show positive results. All intervention schools now have an active CEC, and nearly 72 percent of intervention schools have developed their management plans at ML2 while at ML1 only 53.1 percent had. The same increase is not witnessed in comparison schools where the schools that have a management plan remains nearly unchanged, at 41.9 percent.

Secondly, intervention schools tend to have a much higher baseline for most of the indicators, and in fact report a higher share for all but one indicator at Midline Round 2. It is difficult to tell whether this is due to some of the program activities at the time of the baseline, or due to randomness or indeed a result of some form of bias in the school selection. In any case, the higher levels of activity in intervention schools have two consequences of import here. Namely, it is possible that we are witnessing something like diminishing returns in the results. In other words, it is possible that seeing an increase of, say, 20 percent to 40 percent is possible with relatively limited efforts when comparing to an increase from 60 to 80 percent.

Indeed, for example when looking at the first table of this section, comparison schools had a baseline level of less than 30 percent for half of the indicators, and that of less than or equal to 50 percent for all of them. Meanwhile, intervention showed a level of more than or equal to 50 percent for five of the eight indicators. In addition, when considering why these variables are measured here, the question of program impact on these indicators becomes less important. To elucidate, while it may not be possible to attribute higher CEC activity to the program per se, this point is not entirely relevant. For sustainability of the program may in fact be better that the CEC activity is not dependent upon the program. The more independent the CEC, the more sustainable the program. As such, the fact that the intervention schools continue to have a much higher level of CEC activity as reflected in the responses to the questions analysed above, bodes well for the program.

COMMUNITY PERCEPTIONS OF SCHOOL MANAGEMENT

In addition to the head teachers who were interviewed about school management, the students' primary caregivers' were also asked about their impressions about the management of the schools that their children attend. The data presented here, is thus drawn from a much larger sample, consisting on 730 caregivers at Baseline and 802 at Midline Round 2.[185] At Midline Round 2, 348 of the respondents are parents to the girls going to comparison schools while 454 respondents' children attend interventions schools.

Given the larger sample, the analysis has more statistical weight than that conducted with the head teacher survey data. Yet, while the head teachers are presumably very well aware of what is happening at their schools, the same might not be true with caregivers. Nevertheless, with this caveat in mind, looking at both caregiver and head teacher reported data should provide a more balanced view of the situation in the schools.

The below table presents a first set of difference-in-differences results. Much like with head teacher data, we see again that the overall levels of different activity or satisfaction are higher for the intervention group. All indicators in the below table have a higher share in the intervention cohort for both Baseline and Midline Round 2 responses.

[185] As these questions are asked to those whose children are enrolled, the share is larger at ML2 as more girls have enrolled.

The primary indicator of whether the CEC exists seems to have declined somewhat in the intervention group – going from 78 percent at Baseline to 72.4 percent at Midline Round 2. The project does thus not meet the objective of ML2 for 90 percent of parents in the treatment sample indicating that CECs are functional.

Meanwhile, comparison group has seen an increase of a similar magnitude – from 62.7 to 68.4 percent. It is possible, as has been theorised elsewhere in this report, that the CECs were initially active in program activities, going into the communities to identify potential beneficiaries and reaching out for enrolment and other such activities that would be quite visible to community members. The absence of such activities from comparison schools could explain the initial distinction between them and the intervention schools. As the program has progressed, the CECs activities have been more focuses on their functions at the school and among those who are not enrolled. As such, the parents of enrolled girls might not be as aware of their activities. Meanwhile, as the sustainability section of this report has shown, the MoEs in each region have worked with the CECs, increasing their presence across the board, potentially explaining why the comparison schools are narrowing the gap to the intervention schools.

TABLE 86: PROJECT IMPACT (DIFFERENCE-IN-DIFFERENCES) ON SCHOOL MANAGEMENT – HH SURVEY

	Intervention		Comparison		Difference in Differences	p value
	BL	ML2	BL	ML2		
Does the school have a CEC that helps with school-related matters?	78.0%	72.4%	62.7%	68.4%	-11.3	.169
School managed extremely well	47.7%	65.9%	41.1%	54.6%	4.7	.447
School management improved	53.0%	65.4%	47.3%	57.0%	2.7	.597
School head teacher excellent	53.2%	67.1%	46.2%	54.8%	5.3	.288
Frequency of CEC communication (3=weekly, 0=never)	1.6	1.9	1.7	1.8	0.3	.183
The CEC initiatives improved the quality of schooling girl received	65.7%	78.3%	63.6%	77.7%	-1.4	.873

Encouragingly, all measures of caregivers' perception of quality of school management show positive difference in differences estimates despite having higher Baseline starting points for intervention schools. This indicates that caregiver satisfaction with school management continues to increase and the gap between intervention and comparison schools continues to grow. For example, currently 67.1 percent of caregivers in intervention schools feel that the head teacher of their daughter's school is excellent, compared to 53.2 percent at Baseline. Meanwhile, the same share for comparison schools is nearly 23 percentage points lower, at 54.8 percent (up from 46.2 percent). The DID estimates, however, are not statistically significant and as such should be treated as indicative rather than conclusive.

Next we looked at perceived frequency of CEC communications towards the community. The responses (weekly=3, monthly=2, annually=1, never=0) were coded so that a higher number corresponds with higher frequency. The results indicate that on average the frequency of CEC communications was seen to have increased slightly more in the intervention schools. Meanwhile, caregivers' positive perception of whether CEC's activities have improved the schooling of their girl had become more prevalent among both intervention and comparison schools, yet the latter shows a slightly larger increase in this regard.

TABLE 87: PROJECT IMPACT (DIFFERENCE-IN-DIFFERENCES) ON SCHOOL MANAGEMENT – HH SURVEY - ML1 TO ML2

	Intervention		Comparison		Difference in Differences	p value
	ML1	ML2	ML1	ML2		
Does the school have a CEC that helps with school-related matters?	89.6%	75.7%	77.3%	69.1%	-5.7	0.443
School managed extremely well	72.4%	76.6%	66.3%	72.4%	-2.0	0.769
School management improved	78.7%	76.2%	73.7%	75.7%	-4.5	0.384
School head teacher excellent	76.9%	79.4%	71.2%	72.0%	+1.6	0.790
Frequency of CEC communication (3=weekly, 0=never)	1.7	2.0	1.9	1.9	+0.32	0.217
The CEC initiatives improved the quality of schooling girl received	64.7%	74.7%	48.6%	82.5%	-23.9	0.012 *

However, as this table shows, the trend in all of the above variables since Midline Round 1 seems to be negative. Yet, only the variable on the perceptions of overall improvement of CEC initiatives on a girl's schooling is statistically significant at the 5 percent level. As such, while the indicators have overall improved since baseline, much of this improvement has come in the period leading to Midline Round 1 and seems to be reversing at the moment. The program should thus keep a close eye on the CEC-related activities to see if this can be turned around. Nevertheless, the relationships are mostly not statistically significant and should be treated merely as indicative rather than conclusive. In any case, should the program desire to show impact in this regard, it currently looks challenging.

TABLE 88: PROJECT IMPACT (DIFFERENCE-IN-DIFFERENCES) ON CEC ACTIVITY – HH SURVEY

	Treatment		Comparison		Difference in Differences	p value
	BL	ML2	BL	ML2		
CEC Monitored student attendance	52.6%	68.2%	54.2%	68.4%	1.4	.848
CEC Monitored teacher attendance	37.8%	58.3%	40.0%	53.4%	7.1	.419
CEC Raised funds	12.7%	28.8%	15.3%	18.9%	12.4	**.024 ***
CEC Improved school infrastructure	23.4%	34.8%	19.5%	33.5%	-2.6	.741
CEC Supported students financially	8.6%	22.5%	11.6%	25.2%	0.3	.974
CEC Bought learning materials	8.2%	17.5%	4.7%	11.7%	2.4	.641
CEC Promoted enrolment of out-of-school children	22.3%	27.2%	15.3%	27.7%	-7.6	.405
CEC Provided remedial support	6.5%	7.6%	2.6%	7.3%	-3.6	.351
CEC Reinforced the use of non-violent disciplines	4.5%	11.6%	4.7%	9.2%	2.6	.667

CEC Monitored student retention	13.4 %	15.9 %	11.6 %	13.6 %	0.5	.938

Turning to specific activities of the CEC, we see that the community members around intervention schools are more likely to report that the CEC is involved in raising funds. This share has grown much more in the intervention schools, with the difference in differences being 12.4 and statistically significant at the 5 percent level. This is a positive findings, as if we look at the list of variables in focus here, the CEC's raising of funds is perhaps one of the most visible to the community members, and as such one with which we can estimate the highest level of confidence. The fact that this activity seems to have increased substantially more in intervention communities is a good proxy for overall levels of increased CEC activity in intervention schools.

Meanwhile, none of the other variables have a statistically significant difference in differences estimate. The estimates, again, present a somewhat inconclusive picture. Monitoring both student and teacher attendance seems to have increased somewhat more in intervention communities. The converse is true for CEC working on improving school infrastructure, promoting OOSG enrolment and providing remedial support. Simultaneously the changes for monitoring student retention and supporting students financially are very similar in both intervention and comparison schools. Finally, the promotion of alternatives to corporal punishment and buying learning materials are elements of CEC work where the intervention school respondents' answers indicate that they have increased slightly more.

Indeed, when looking at the difference in differences estimates, only some of the variables show some program impact – i.e. the program improving significantly more than the comparison schools. And only one variable – CEC raising funds – has a level of statistical significance that allows to say that the difference is most likely not due to random error. However, when looking at all the data on CEC activity and performance that we have collected through the household survey the overall trend is one of improvement. Indeed, all but one of the variables have increased, some very substantially. While this has not happened much more markedly for the intervention schools, it is a positive finding for the sustainability of the project. The fact that the comparison schools are also improving along the same indicators is perhaps due to other programs or MoE activity, yet, does not make the intervention schools any less likely to be sustainable in the long-term.

When looking at these variables for trends since Midline Round 1, we can see that the findings are much in line with the overall trend observed since Baseline. Namely, across the board in both intervention and comparison schools the CEC activities seem to have become more prevalent. However, no systemic difference between the two sets of schools can be observed as none of the difference in differences estimates show a statistically significant deviation between intervention and comparison schools.

TABLE 89: PROJECT IMPACT (DIFFERENCE-IN-DIFFERENCES) ON CEC ACTIVITY – HH SURVEY – ML1 TO ML2

	Treatment		Comparison		Difference in Differences	p value
	ML1	ML2	ML1	ML2		
CEC Monitored student attendance	63.6%	65.7%	63.7%	70.4%	-4.5	0.578
CEC Monitored	48.4%	60.2%	53.8%	58.4%	7.1	0.410

teacher attendance						
CEC Raised funds	19.1%	28.7%	17.5%	22.4%	4.7	0.534
CEC Improved school infrastructure	26.2%	35.4%	13.8%	31.2%	-8.3	0.369
CEC Supported students financially	16.4%	23.8%	3.8%	23.2%	-12.1	0.095
CEC Bought learning materials	7.6%	17.1%	2.5%	12.0%	0.1	0.990
CEC Promoted enrolment of out-of-school children	20.0%	28.7%	15.6%	28.0%	-3.6	0.622
CEC Provided remedial support	7.6%	8.8%	1.9%	8.0%	-4.8	0.292
CEC Reinforced the use of non-violent disciplines	5.3%	11.6%	3.1%	7.2%	2.2	0.733
CEC Monitored student retention	9.3%	15.5%	10.6%	15.2%	1.6	0.738

SCHOOL MANAGEMENT IN QUALITATIVE DATA

Most CEC members indicate that their committee is very active in the monitoring front, and other activities that do not require many material resources such as doing awareness-raising or attitude change work with parents and other community members.

> *We write down the names of students who trouble the teachers, and we follow-up with the teachers that don't teach the students well. We keep him if he improves and if not, we replace him and get another one.*[186]

Most CEC member focus group discussions also mentioned their efforts to eradicate corporal punishment from the schools, and one CEC had been involved in successfully mediating a clan conflict that had broken out in their community. However, they all lament the lack of resources:

> *We deal with some of the children, but we can't deal with all nomadic children, because we are not yet financially stable.*[187]

> *There is nothing that we do for other people because when people are better than each other financially is when they help each other.*[188]

> *Residents of the village are poor and have no money.*[189]

Some also called for training as CEC.

> *We monitor the school regularly, and teachers are good, but we as the CEC don't get enough trainings as often as we require.*[190]

Indeed, this sentiment is echoed by some MOE respondents. For example, when a Somaliland REO was asked what was missing, he specifically mentioned CEC trainings, since "most of the committee members are not trained." He then continued: "the achievement of the CES's is good but they need trainings. The person should be understanding what he's being told."[191]

In general, the REOs interviewed at this Midline viewed the CECs as a fundamental part of their work and tended to view them as very active.

> *Some have come to build buildings and the CECs have been involved in contributing to the school. The CEC decided not to be a dysfunctional committee. Some built classrooms, some built halls, some built new schools, and we now manage them.*[192]

However, not all CECs are similarly active and well-resourced. Indeed, as the CEC member discussions have revealed, poor communities do not have the kind of resource base that can be used to fees and other forms of financial support to pay teacher salaries and support poorer families to enrol their children. These poor

[186] FGD CEC, Puntland, 201
[187] FGD CEC, Somaliland, 101
[188] FGD CEC, Somaliland, 103
[189] FGD CEC, Somaliland, 106
[190] FGD CEC, Somaliland, 101
[191] KII REO, Somaliland, 161
[192] KII REO, Puntland, 263

communities also tend to be rural where the issue of pastoralist children dropping out of school or attending intermittently is more marked. An REO from Puntland summarises:

> *I cannot say all the committee are performing effectively, but for those CECs that have a chance and that are in the big cities, they perform well, and are focused, because others might see what they are up to.*[193]

Further, when talking to teachers about the CEC and school management in general, the views are much in line with the other respondents. All but one discussion revealed that the CEC meets regularly. The teachers also viewed the committee as very important for the management of the school:

> *The school does not have a consistent donor, so the donor of the school is the CEC. If the school faces any challenges, the CEC helps and makes sure the school is running and functioning well.*[194]
>
> *The CEC mobilized the community and encouraged the parents to enrol their children to the school.*[195]

But the issue of lack of resources is raised by teachers also.

> *Both the committee and the community have the same problem. Generally, these people are poor, and everyone in the community thinks about their livelihoods. Someone who does not have a livelihood can only give you moral support and there is nothing else he can do.*[196]

Some call for the CEC to be paid for or for more financial assistance for them. Furthermore, like some of the CEC members and REO respondents above, many of the teachers in the FGDs call for CEC trainings:

> *The CEC should also receive a training as well in order to improve the work they are doing to the school.*[197]
>
> *The challenge is that they are people who do not understand education… and then the committee does not understand the work they are doing. So, once a committee is formed, they will leave and say 'what is in it for me?', so there is no organized committee that knows the structure of the school and what it needs.*[198]

When it comes to reaching out to marginalised groups, the evidence from the focus group discussions with CEC members is somewhat divided. In some schools the CEC members indicate a kind of apathy and resignation when it comes to addressing these issues.

[193] KII REO, Puntland, 261
[194] FGD Teachers, Puntland, 231
[195] FGD Teachers, Somaliland, 133
[196] FGD Teachers, Galmudug, 331
[197] FGD Teachers, Galmudug, 331
[198] FGD Teachers, Somaliland, 135

> *We don't support them, but we have done some awareness raising to get their children to school.*[199]
>
> *No, there are not any disabled students. And this is beyond our ability.*[200]

> *Yes, there are pastoralist children, but such people do not mix with our students, and we do not have a place to stay for them and provide support.*[201]
>
> *We can only support with our voices and we do not have any other thing to support them and we have encourage them only.*[202]
>
> *We help them [pastoralists] if they come to us, but we do not go to them.*[203]
>
> *All we can do is to tell them to come to school and learn.*[204]

Meanwhile, in other schools the CEC members outline some activities that they have undertaken to improve the situation. These include attitude-change activities across the board, both among the parents for them to send the girls to school, as well as among their peers in terms of anti-bullying campaigns.

> *We as the CEC call and convince students with disability students to come and pursue education. We motivate them often and we tell teachers to give then special care when they come to school. We also raise awareness about disability education and caution other students against insulting or discouraging them.*[205]

Bullying happens with both the disabled girls as well as the pastoralists:

> *the older girls from the rural areas face challenges like bullying from urban girls in the school, they say she is older than us and she cannot read and write.*[206]

In some instances, the CECs have supported the nomadic or disabled girls also in form of providing transport and school feeding, where possible. However, in most schools the reality is that these activities are very constrained due to the lack of money. Awareness-raising requires transport, which costs. Children who come a long way to attend school also require transport and food. In addition, some CEC members argue that the nomadic girls who are older do not attend their school because there are no female teachers. Hiring more teachers would also cost.

[199] FGD CEC, Somaliland, 104.
[200] FGD CEC, Galmudug, 301.
[201] FGD CEC, Galmudug, 301.
[202] FGD CEC, Galmudug, 301.
[203] FGD CEC, Puntland, 201.
[204] FGD CEC, Puntland, 201.
[205] FGD CEC, Somaliland, 101
[206] FGD CEC, Puntland, 202

It is perhaps telling that even in the schools where there have been efforts to enrol the disabled girls, none are currently attending the intervention schools. This CEC member summarises the challenges they face in further including the marginalised:

> *There are many things standing in the CEC's way. Because these people need funds and private teachers. And the CEC does not have funds and cannot support in anything. We cannot find cars to pick them up, because the car needs gas to get them to school and will have to take them back home as well. There are many that we cannot take, and, as a result, they stay home. There are also a lot that are mentally ill and will go crazier if they are brought to this noisy place. So, we have these challenges facing us.*[207]

As it pertains to members of the traditionally marginalised occupational groups such as the Madhibaan, Tuumaal and Gaaboye, the respondents typically argue that there are no marginalised groups in that regard, and that all groups are treated equally.

> *Nobody is discriminated, and such culture is not acceptable here. We have passed the period of ignorance where people used to discriminate, and education is going well.*[208]

However, our previous research in Somalia and Somaliland has shown this not to be the case, and that these groups are often excluded from programs' beneficiary selection and ignored in decision-making procedures. However, short of identifying the clan composition of the beneficiaries and then mapping out the community to view if some occupational minority groups are missing, we cannot deem one way or the other if any systemic exclusion of the occupational minority groups is taking place or not.

But let us now turn to assess the indicators against their targets, as set in the program logical framework.

SCHOOL MANAGEMENT AND GOVERNANCE: OVERALL FINDINGS

When looking at the program's goals as per the logical framework, the findings are mixed. One of the goals was to increase CEC monitoring of student attendance to 80 percent by ML2. The share of CECs doing that currently, according to the caregivers, is 68.2 percent. The findings are largely similar when looking at head teacher surveys where 65.6 percent report CEC monitoring of student attendance. The goal is thus not met. In terms of monitoring retention, only 15.9 percent of parents in intervention areas say that the CEC is engaged in this. At the same time, 37.5 percent of head teachers say the same.

Meanwhile, the program meets its objective of 30 percent of CECs taking action against corporal punishment and other child protection issues as such the share of intervention school head teachers reporting the CECs addressing child protection issues is 78.1 percent. Similarly, the head teachers saying that the CEC reinforces the use of non-violent discipline instead of corporal punishment in 68.8 percent of the intervention schools at Midline Round 2.

As per the final indicator relating to school management in the logical framework, the CECs' practices taking into consideration the needs of different vulnerable sub-groups, such as pastoralists, members of marginalised

[207] FGD CEC, Puntland, 201.
[208] FGD CEC, Galmudug, 301.

groups, IDPs and the disabled, the CEC members seem to be quite aware of the difficulties faced by these groups. However, most committee members viewed addressing barriers to their education as beyond their means due to the very severely limited financial resources at their disposal. Yet, some of the findings from the head teacher survey indicate at least partial improvements as 78.3 percent of school management plans in the intervention group now include plans to follow-up with dropouts.

Meanwhile, the qualitative data demonstrates that different respondents (CEC members, REOs and teachers) all value the importance of the CECs work and see them as key to managing the schools. The CECs are, moreover, viewed generally as very active. However, there is no evidence to say that this activity has increased significantly since the baseline. In addition, while the awareness of marginalisation and issues of retention and enrolment among some communities seems to be well understood and the CEC members and REOs and teachers alike seem conscious of these issues and are seeking ways to address them, respondents across the board do not view it as possible given the chronic lack of resources experienced by the actors. Most of the CEC activities, thus, tend to be things that either do not require much in terms of financing – such as monitoring the school or talking to parents or students. When it comes to more resource-intensive activities – such as giving scholarships, reaching out to more remote communities and paying teacher salaries – some communities have very little capacity to do this as there is little to extract in terms of financial support in the communities. This is particularly pressing in the rural communities where some of the other issues also likely concentrate, such as teacher absenteeism and high number of pastoralist children. This suggests, that successfully and comprehensively addressing the needs of the most marginalised groups in the future would require a more resource-intensive intervention.

Finally, it seems important to note that while some of the findings are not at the level that was desired for this milestone, the findings for arguably the most important variables – whether the school has a functioning CEC or not, or whether the school has a management plan – show positive results. Every intervention school currently has a functional CEC, according to the head teachers. And 72 percent of the intervention schools now have a management plan in place. Moreover, almost every indicator of CEC activity and performance is markedly higher in intervention schools than in comparison schools. While it has been difficult to show program impact as such due to the levels rising with a similar pace in comparison schools, the intervention schools most likely have a higher preparedness for sustainability as the levels indicating CEC activity are higher in these schools.

7.3 TEACHING QUALITY

Outcome	Indicator	BL	ML2 Target	ML2	Target achieved? (Y/N)	Target for next evaluation point	Will IO indicator be used for next evaluation point? (Y/N)
Teaching Quality	Percentage of teachers who apply improved teaching practices in literacy and numeracy	36.2%	30% increase in the proportion of teachers with self-reported use of formative assessments.	91.2%	Y	60% above baseline	Y
Main qualitative findings							
There was not direct qualitative support for the use of formative assessments during FGDs with teachers. There were no questions that asked teachers directly about their use of formative assessments. During questions with teachers about how their teaching practices had changed or how to improve the quality of instruction in school, teachers did not mention the use of formative assessments directly. Instead, there were some mentions of using tests (summative assessments) as a means of helping students learn.							
Teaching Quality	Shifts in teachers' awareness of quality education	N/A	Teachers acknowledge corporal punishment as a barrier to learning, teachers take proactive action to support students who are lagging behind in acquiring literacy and numeracy skills, including specific support to students with disabilities.	88.7% of treatment teachers acknowledged corporal punishment is a detriment to learning 92.3% of treatment teachers take proactive action to support students who are lagging behind.	Y	Teachers express awareness on the importance of child protection and gender-equitable practices and describe how they are using strategies to support struggling students, including children with disabilities.	Y
Main qualitative findings							

- During FGDs, teachers were asked about how they discipline students and the ways that they approach student discipline. Teachers regularly acknowledged that corporal punishment is not helpful for student learning and may lead to bad outcomes such as the student dropping out of school or not learning as well.
- During FGDs, teachers acknowledged the need to provide additional support to students who were struggling and tried to help students by offering extra classes and tutoring sessions or extra materials/assignments to take home. However, there was not much evidence of using differentiated instruction or altering teaching methods during instruction.
- Teachers express the need to support children who struggle with disabilities but seem unsure how to best do so. Teachers report taking very practical steps to help students with disabilities (such as having students with vision or hearing impairments sit at the front of the room), however teachers did not seem to be able to alter their instruction to support the needs of students with disabilities and often cited the need for further training to support learning for disabled students.

In addition to trying to remove barriers to enrolment and attendance in school, SOMGEP-T aimed to support school practices and conditions for girls once there. As such, teacher training programs aimed at improving pedagogical techniques and overall teaching quality were one of the major program interventions. Specifically, these teacher trainings focus on 1) improved delivery of literacy and English language, 2) improved delivery of numeracy in all 148 primary and 55 secondary schools, and 3) to provide structured remedial support to students at the primary and secondary level. Within these training goals, a particular emphasis was placed on the use of formative assessments[209] and a reduction of corporal punishment. By making strides in these areas, the hope is to create more positive and enriching learning environments that improve learning and student enrolment.

Teaching quality is an inherently difficult thing to assess as it can be evaluated along a number of different axes and involves a broad range of competencies. In this analysis, we have decided to focus on evaluating teaching quality in terms of teacher effort, teaching practices, and student and caregiver perceptions of teaching quality. Specifically, we examined the following five indicators of teaching quality:

- Caregiver perceptions of overall teaching quality, with a particular focus on changes over the previous 12-month period
- The use of learner-centred pedagogy in class and the use of formative assessments during instruction
- Classroom demeanour and safety: the use of corporal punishment in class, the extent to which teachers are respectful and welcoming toward students, and whether students feel safe and comfortable in the classroom.
- Gender equality in classroom interactions: do teachers interact differently with boys and girls in terms of the kinds or frequency of questions asked, the use of harsh language towards girls, the allocation of teacher attention or resources, … etc
- Teacher's effort level and preparedness for class, as indicated through the use of lesson plans and clear communication of learning goals at the start of the lesson.

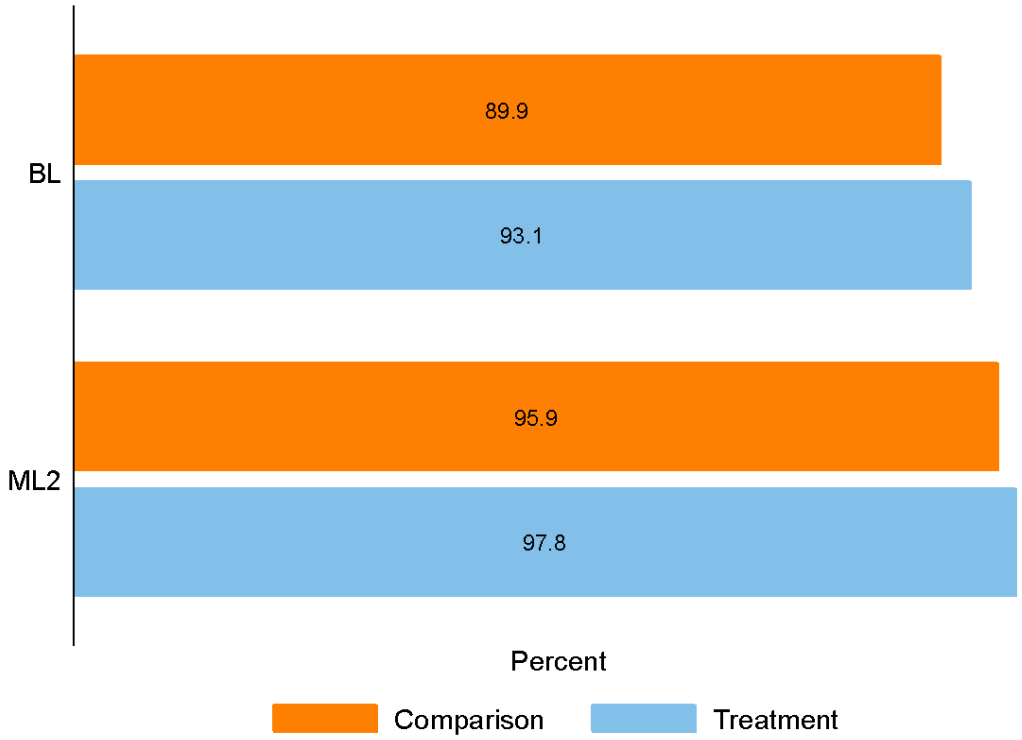
[209] The term "Formative Assessment" refers to a range of methods employed by teachers to elicit feedback from students, during learning, for the purpose of adjusting instruction to meet students where they are in the learning process. As opposed to formative assessments where the goal is to collect information at the end of some unit *of learning*, formative assessments are used to guide teacher and student actions in real time *for learning*. See: Black, P., and D. Wiliam. 2009. Developing the theory of formative assessment. Educational Assessment, Evaluation and Accountability 21.1: 5–31.

Data for these indicators was collected from three distinct data collection tools including a classroom observation tool, a girls' survey, and the caregivers' survey. The classroom observation tool was designed to capture information about what was going on the classroom and used to assess teacher effort, the pedagogical approach being used, and how the teacher was interacting with the students. In each of these observations, a researcher observed the classroom for three 15 minutes blocks and recorded notes about what kind of learning was taking place and how the teacher interacted with students. These observations were conducted in two classrooms per school at the baseline and a minimum of two were conducted at each school at the second midline. The girls' survey tool was incorporated into the household survey and was designed to gauge how comfortable girls felt in class and the potential for teachers to use corporal punishment when not being observed directly. The caregiver survey was administered as part of the household survey and was designed to capture caregiver perceptions on teaching quality.

Caregiver perceptions of teaching quality

Caregivers were asked to assess the quality of teaching in their girl's school and to indicate whether it has improved, stayed the same, or gotten worse in the last 12 months. Caregivers at both treatment and comparison schools have overwhelmingly positive perceptions of teaching quality at their girl's schools with greater than 95% of caregivers reporting good or very good teaching quality. As seen in the figure below, caregiver perceptions of teaching quality have improved since the baseline. At baseline, 89.9% of caregivers viewed teaching quality as good or very good. By the second midline, this number had increased to 95.9%. Caregivers of girls at treatment schools started with a higher perception of teacher quality at the baseline with 93.1% reporting teaching quality that was good or better and by the second midline this number had increased to 97.8%. At first glance, these results seem to suggest that the intervention is having little impact on caregiver perceptions of teacher quality, as perceptions at comparison schools increased by 6% since the baseline compared to 4.7% for the treatment schools. However, if we focus the analysis to only include caregivers who perceive teaching quality to be very good, we see a larger increase in the treatment schools. At baseline, 58.5% of caregivers at comparison schools viewed the teaching quality as very good compared to 61% of treatment schools. At the second midline the number of caregivers who perceived teaching quality to be very good rose 13.9% in comparison schools (72.4% in total) but in treatment schools this number had risen 18.5% (79.5% in total).

FIGURE 29: SHARE OF CAREGIVERS REPORTING GOOD OR VERY GOOD QUALITY TEACHING OVER PRIOR 12 MONTHS, BY ROUND



Caregivers were also asked about their perceptions of how teaching quality has changed within the last 12 months. While caregivers of girls at both comparison and treatment schools felt that teaching quality has improved, caregivers of girls at treatment schools were more likely to report that teaching quality had improved. 90.5 percent of caregivers at treatment schools report that teaching quality had improved while 85.1% of caregivers at comparison schools felt the same. Given that caregivers of girls at treatments schools were 5.4% more likely to report that teaching quality had increased and 5.6% more likely to report that the teaching quality was very good, it appears that the interventions at these schools are having a small, but discernible impact on teaching quality when measured by caregiver perceptions.

Focus group discussions (FGDs) with teachers suggest that teaching quality has improved. Many teachers cite the training they received and the increased guidance from school administrators and head teachers. As one teacher reports:

"If I try to answer this question, I think yes. The school has improved the quality of education. The school has changed a lot about the school system and teachers got so many trainings last year and this led many students to come to school more regularly than ever and the quality of the school is great now. Students are excellent and teachers are teaching with commitment and devotion. So, yeah, the quality of the school absolutely improved."[210]

[210] FGD with Teachers, Puntland.

The value of better school management, for driving better teaching practices – or at least caregiver perceptions of better teaching quality – is reflected in the quantitative data as well, at least tentatively. In schools where the head teacher reports that the CEC visited the school in the past year, and monitored teaching quality during their last visit, caregivers are more likely to view the quality of teaching their girl receives as "very good." In total, 82.9 percent of caregivers in communities where the CEC monitored teaching quality viewed teaching quality as very good, compared to 74.3 percent in communities where the CEC did not monitor teaching quality. Interestingly, this relationship is somewhat driven by other types of CEC monitoring visits – which may suggest simply that CEC monitoring either contributes to improved teaching or is perceived by parents as meaning that teaching quality is better.[211]

Despite the general sense of improvement among teachers, there were still some challenges that they faced with regard to teaching quality. Low salaries/lack of salaries, lack of necessary tools and resources, lack of training, and increased enrolment at schools (higher student/teacher ratios) are all challenges that teachers discussed with regards to teaching quality. These factors will be discussed in relation to each of the sections below to help gain a more complete understanding of the areas the project has best addressed, the remaining challenges, and the ways that the project can adapt moving forward in order to start addressing these challenges.

Learner-centred pedagogy

Learner-centred pedagogy broadly refers to the use of teaching methods that shift the focus of instruction from teacher to student. As opposed to a teacher-centred classroom where students are the passive recipients of knowledge, in classrooms that use a learner-centred approach students have a larger share of responsibility of the learning process and that process is shaped by the learners' needs, current understanding, and interests. While the concept of student-centred learning can encompass a wide range of activities, we used two metrics to assess the degree of learner-centredness: the degree to which classrooms were active and participatory and the use of formative assessments during instruction.

There are many ways for a classroom to be active and participatory and therefore numerous ways that this criterion can be assessed. In general, active and participatory classrooms involve students asking questions and generating discussion, have students engaged in active learning activities beyond copying from the board or taking notes during lecture. In our analysis, we looked at eight specific indicators of an active and participatory classroom. During classroom observations, researchers indicated "yes" or "no" if they saw one of these indicators was observed at any time during an observation block. A list of the eight indicators and the frequency of each of them being employed during observations during the baseline and second midline are presented in the table below. A difference in difference (DID) analysis was used in order to assess the effect of the project interventions on each of the indicators. Positive DID scores indicate that the intervention is having a positive effect on the outcome being analysed.

TABLE 90: INDICATORS OF LEARNER-CENTRED PEDAGOGY IN TREATMENT AND COMPARISON SCHOOLS FROM BL TO ML2

	Baseline	Midline 2	DID

[211] There does appear to be a tangible, if modest in magnitude, relationship between CEC monitoring activities and actual teaching quality, as observed by researchers during this round of data collection. Using the index of active teaching practices described later in this section, each additional visit by a CEC member over the previous year – as reported by the head teacher – is associated with a 0.6-point increase in the activity index calculated during our classroom observations. Again, this is a modest increase, as the index is scored on a 100-point scale, but the finding is robust to inclusion of control variables in a regression framework.

	Comparison	Intervention	Comparison	Intervention	
Did NOT spend most time copying from the board	44.8%	62.7%	50.8%	64.6%	-4.0%
Did NOT spend most time repeating words aloud	55.2%	72.9%	70.1%	80.0%	-7.9%
Use of student-centred games/activities	34.5%	40.3%	8.5%	13.8%	-0.4%
Students instructing each other	32. 8%	44.8%	30.5%	47.7%	5.2%
Teachers ask open-ended questions	43.1%	64.2%%	50.8%	60.0%	-11.9%
Teacher asks for student opinions	56.9%	64.2%%	49.2%	47.7%	-8.7%
Sought to involve student who was not participating	67.2%	62.69%	59.3%	60.0%%	5.2%
Students worked in groups	37.9%	32.8%	32.8%	16.9%	`13.5%

** None of the DID (difference in difference) estimates of project impact are statistically significant.

Broadly speaking, intervention classrooms were more likely to be observed using active teaching approaches during the second midline relative to comparison schools. Only two of the eight indicators (asking student opinions and group work) were more likely to be observed in comparison classrooms during this study. The most notable differences were observed in the frequency with which students spent most of their time copying from the board (64.6% in interventions schools compared to 50.8% in comparison schools), students instructing each other (47.7% vs. 30.5%), and students not spending most of the class repeating words aloud (80.0% vs. 70.1%). Conversely, student group work was almost twice as likely to be observed in comparison schools (32.8% in comparison schools compared to 16.9%).

The more frequent use of active teaching approaches among intervention schools is a positive for the effectiveness of the teacher training portion of the project interventions. However, when looking at the results of the first midline, there is some cause for concern. The table below compares the frequency with which these activities were observed in the first and second midline.

TABLE 91: INDICATORS OF LEARNER-CENTRED PEDAGOGY IN TREATMENT AND COMPARISON SCHOOLS FROM ML1 TO ML2

	Midline 1	Midline 2	Difference (midline 2 – midline 1)
Intervention schools			
Did NOT spend most time copying from the board	55.1%	64.6%	**9.5%**

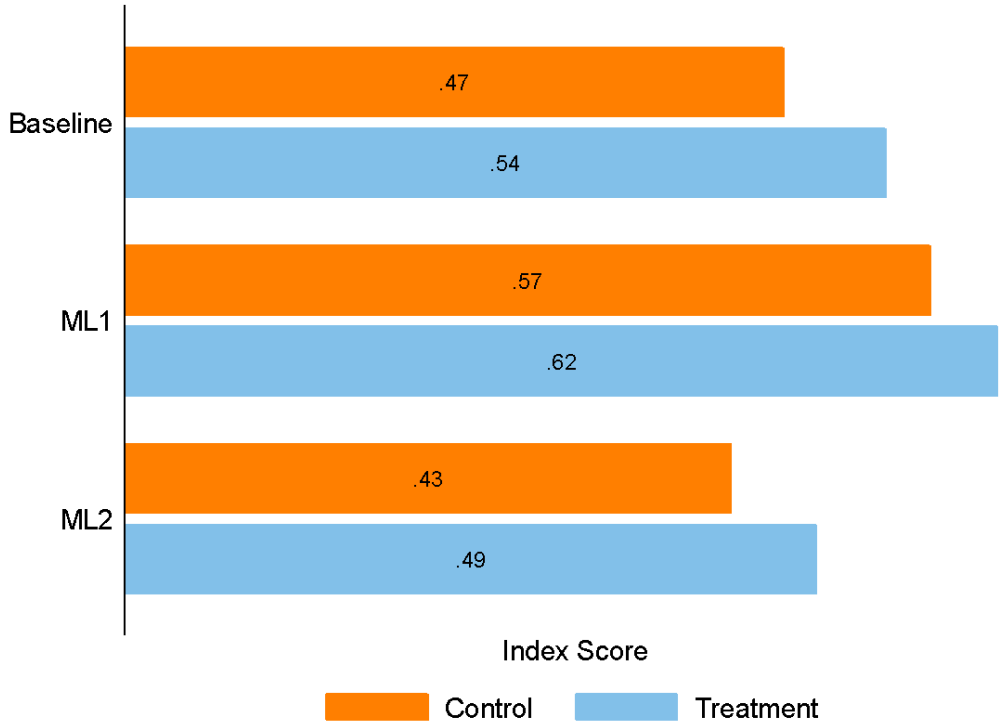
Did NOT spend most time repeating words aloud	82.8%	80.0%	**-2.8%**
Use of student-centered games/activities	44.8%	13.8%	**-31.0%**
Students instructing each other	55.2%	47.7%	**-7.5%**
Teachers ask open-ended questions	75.9%	60.0%	**-15.9%**
Teacher asks for student opinions	72.4%	47.7%	**-25.9%**
Sought to involve student who was not participating	79.3%	60%	**-19.3%**
Students worked in groups	27.6	16.9%	**-10.7%**

When comparing the frequency with which particular teaching methods were observed between the first second midline, the only indicators that did not see major decreases were time spent copying from the board and repeating words aloud. For the remaining six indicators, the frequency with which they were observed declined precipitously. Most notably, the use of student-centred games fell by 31% and the teacher asking for student opinions dropped by 25.9%. These results suggest that the gains observed in the first midline study may not be persistent. Incorporating the active strategies into lessons plans requires additional planning and teachers may require refresher trainings or periodic feedback from school administrators in order to ensure that the use of these active teaching methods continue throughout the school year and from year to year.

Each classroom is unique and there may be multiple routes to having a student-centered classroom, so in addition to looking at the use of individual pedagogical strategies, we also constructed an index to try to take a more holistic look at the how active and participatory the observed classrooms are. The index was constructed by converting the eight indicators into a single score ranging from zero to one.[212] The figure below shows the index of classroom activity and participation for each round and intervention status. Both control and treatment schools show signs of decreased classroom activity and participation relative to either the baseline or first midline. In treatment schools, the mean classroom activity index was 0.62, meaning that nearly 5 of the 8 indicators were being used on average during classroom observations on the first midline. By the second midline, the index score had dropped to 0.49, meaning that the average classroom employed approximately one fewer of the 8 strategies captured by the index.

[212] The index was created by taking the mean of each of eight indicators of classroom activity. The indicator ranges from 0 to 1 with zero meaning that none of the 8 indicators was observed and 1 being all eight indicators were observed. It isn't realistic to expect that all eight of the classroom strategies would be used in a single classroom and therefore the score shouldn't be interpreted in an absolute sense. Instead, the index captures the relative diversity of active approaches being employed in comparison and interventions schools.

FIGURE 30: INDEX OF ACTIVE AND PARTICIPATORY CLASSROOMS BY INTERVENTION STATUS AND ROUND



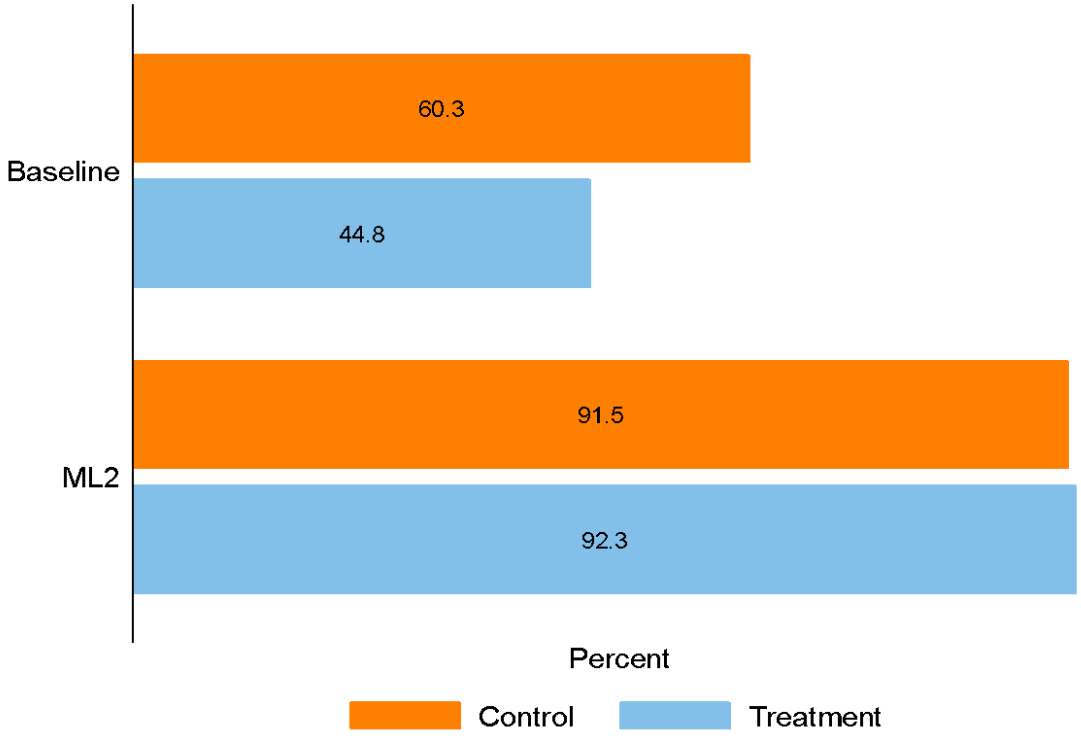
This finding is consistent with the interpretation that teachers may be less likely to employ these active teaching strategies over time without continued support and encouragement. However, there is a potential alternate explanation. Data for the second midline was collected from November 22nd – December 29th (approximately 3 weeks later than the first midline). During this time, many of the schools were nearing a break in instruction for the holidays. Many of the schools in Somaliland and Galmadug closed for the holiday soon after data collection while many of the schools in Puntland were getting ready for school-wide testing. It is possible that this lead up to the holidays or testing may have resulted in less-active classrooms than would be observed at another point in the school calendar year. This explanation is supported by the observed decrease of classroom activity in comparison schools relative to the baseline and first midline as well.

In addition to examining the use of participatory methods in class, we also examined the use of formative assessments. Either before or immediately after classroom observations began, teachers were asked whether they use formative assessments in their teaching. Teachers who indicated that they did use formative assessments were asked whether they had records or documentation of their use.[213]

[213] Teachers were not specifically asked to produce the documentation for review by the researchers. Since these observations were conducted during class time, we wanted to minimize the disruption of instruction by our researchers. Asking the teachers to produce the documentation may have caused a serious disruption and reduction in instruction time. Additionally, teachers were not informed in advance that they would need to furnish proof of their use of formative assessments and may not have been prepared to do so. So in effect, these results assume that the teacher's claims about the use of formative assessments and the possession of records are accurate.

In the figure below, we report the proportion of teachers who state that they use formative assessments. These results indicate that classrooms are becoming more learning-centered with respect to the use of formative assessments. While their self-reported use increased in both comparison and treatment schools, the gains exhibited by treatment schools is much larger with an increase of 47.5% in treatment schools (compared to 31.2% in comparison schools).

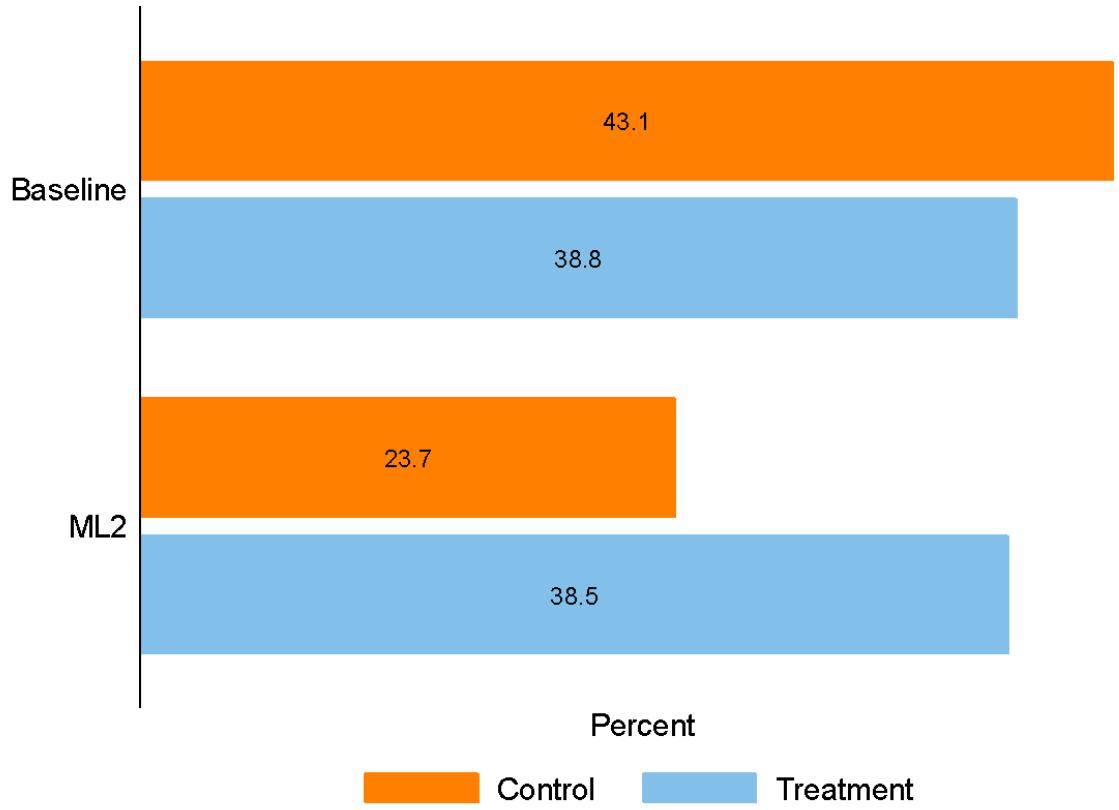
FIGURE 31: TEACHER SELF-REPORTED USE OF FORMATIVE ASSESSMENTS, BY INTERVENTION STATUS IN BL AND ML2



In terms of teachers documenting the use of formal assessments, there was little change in the intervention schools from baseline to the second midline (38.8% at baseline and 38.5% at second midline). Formative assessments can take on many different forms and it is not necessarily the case that teachers would have documentation of the use of formative assessments. Student questioning, checking for understanding verbally, the use of peer-to-peer feedback (like think-pair-share activities), and monitoring of students during class work can all be effective means of formative assessment that may not lend itself to documentation[214]

[214] Ideally, the use of these formative assessment strategies would be documented in a teacher's lesson plan, but given the nature of the question the teacher may not have considered their lesson plan as a form of documentation.

FIGURE 32: EVIDENCE OF USE OF FORMATIVE ASSESSMENTS, BY INTERVENTION STATUS IN BL AND ML2



In order to see if the fact that the schools sampled in each round of the evaluation changed, we also looked at the use of formative assessments in schools where classroom observation were conducted in all three rounds of the evaluation to see if changes in the sample could be driving the observed results. As can be seen in the two figures below, similar trends are seen when we compare only schools that were included in all three rounds of the evaluation (full panel sample). Reported use of formative assessments is almost identical when comparing the ML2 sample to the full panel sample. Evidence of the use of formative assessments declines slightly in the full panel sample (35.1%) when compared to the ML2 sample (38.5%), however the broad trends are the same. While the project appears to be having a large impact on the self-reported use of formative assessments, it does not seem to be having an impact on the teacher's ability to produce evidence of their use.

FIGURE 33: TEACHER SELF-REPORTED USE OF FORMATIVE ASSESSMENTS, BY INTERVENTION STATUS IN SCHOOLS OBSERVED IN THE BL, ML1, AND ML2

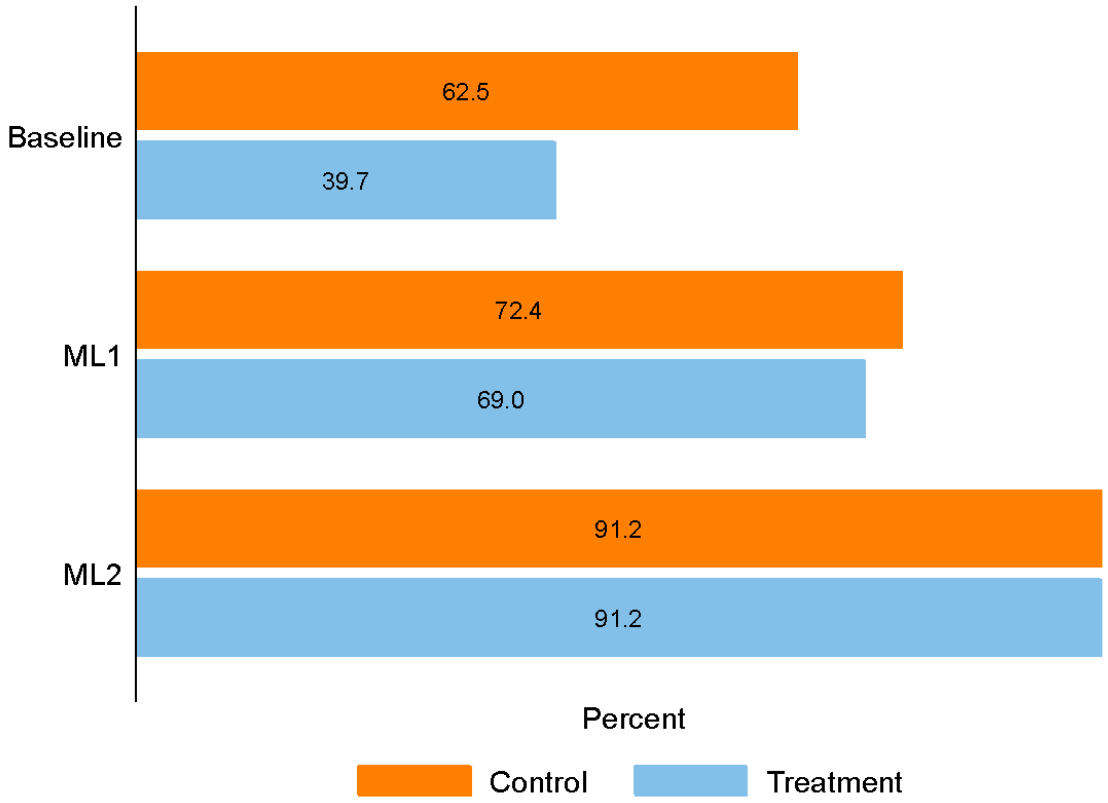
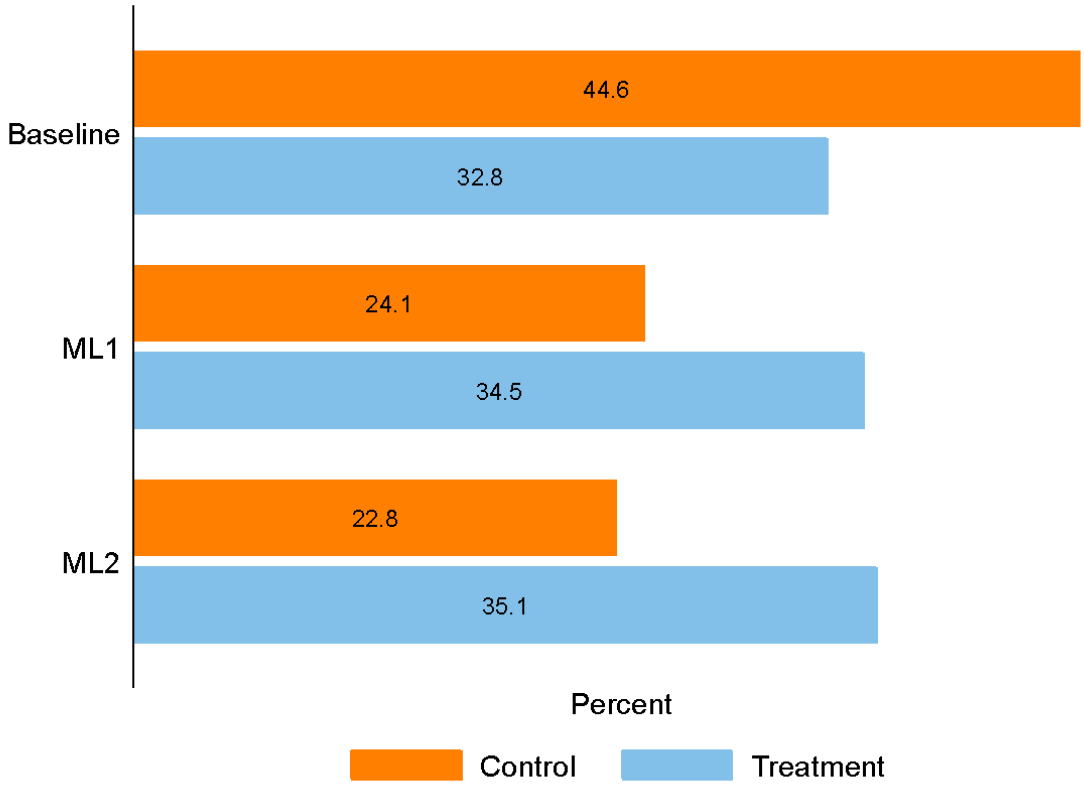


FIGURE 34: EVIDENCE OF USE OF FORMATIVE ASSESSMENTS, BY INTERVENTION STATUS IN SCHOOLS OBSERVED IN THE BL, ML1, AND ML2



While it is possible that this lack of documentation may relate to the specific formative assessment strategies being employed, the results of the previous analysis of the use of active classroom strategies does not support this interpretation. Classroom activities such as asking student opinions, the use of student-student instruction, student group work, and the use of student-centered games are both the kind of activities that align themselves with the use of formative assessments which may not be easily documented and are also the strategies observed with the lowest frequencies.[215]

While there is some doubt as to whether these results should be interpreted as an increase in the use of formative assessments by teachers or as an increased awareness of what formative assessments are, either interpretation represents a positive step toward improving teaching quality. Future assessments should try to gauge the degree to which teachers are eliciting student feedback during class and then acting on that feedback either in real time or in their lesson plans.

[215] At the core of formative assessment is the teacher eliciting feedback from students in order to adjust the lesson to suit the current understanding and needs of students. Activities such as group work and students instructing each other give the teacher opportunities to focus on gaining feedback on student's current understanding and to provide instruction based on that feedback. In this way, active classrooms and the opportunity for formative assessment to take place are linked.

During FGDs, teachers were asked questions about how they help students who are struggling and about how they approached teaching their subjects. During these interviews, there were no mentions of the use of formative assessments directly and few examples where the use of formative assessments were mentioned indirectly. When the use of formative assessments was mentioned, teachers seemed to generally use whole class questioning or one on one interactions to gauge student understanding and use that understanding to help students learn in real time. For example:

I'm a math teacher and I have been teaching for years. So, I have changed the way I used to teach math, because I teach different students and they understand things differently. So, sometimes, I go to the board and I write for them and explain. Also, I go to students face to face and I use their textbooks to give more personal help in order to make sure that everyone is on the same page. In the past, I used to stand in front of the board, and I used to explain orally and write it down on the board. But, I have changed this, because not a lot of students understood this. So, what I do now is that I sit with students and ask some questions using their textbooks and through this I make sure that enough students understand the concept. [216]

I always try to teach students with simple ways, and I work hard to make sure that all students understand maths. I often change my way of math teaching , because I do a daily reflection and when I see that many students are understanding the math problems with a different method, I change the way I teach, so yes I often change the way I teach.[217]

While there is some sign that some of the teachers have become more aware of formative assessment and some strategies for employing them in their classroom, future interviews and focus groups discussions will need to ask teachers more directly about their use of formative assessments to gain a better understanding of how teachers are understanding and using this strategy.

Classroom Environment

In addition to examining changes in instructional approaches, we also analysed classroom environment. In order to maximize learning and a desire to stay in school, it is important to create a positive learning environment where students feel comfortable and safe. In this analysis, the degree to which a classroom could be said to have a positive environment was a function of a teacher's demeanour and the disciplinary policies in place in the classroom. Specifically, the classroom environment was assessed using several indicators including: reported and observed use of corporal punishment, how respectfully the teachers interacted with students, and how welcome the students feel in class. These factors were measured through the use of classroom observations and student interviews.

Students who feel safe and secure in their classrooms are more likely to actively participate in classroom activities and discussions and are more likely to enjoy attending. As a result, it is likely that students who feel safe are more likely to continue attending and not drop out. Given that the majority of teachers are male[218] a teacher's demeanour and the classroom environment as a whole may be particularly important for reducing anxiety, increasing comfort, and increasing the feeling of belonging for girls.

[216] Teacher FGD Sanaag
[217] Teacher FGD Togdheer
[218] Of the 249 teachers observed, only 23 (9.2%) were female.

Corporal Punishment

In order to understand the use of corporal punishment, first, teachers were asked during classroom observations about their perceptions of the use of corporal punishment. 88.7% of teacher in the second midline indicated that they believe that corporal punishment slows down learning. researchers noted instances of corporal punishment during classroom observations and girls were asked about their teachers use of corporal punishment. Findings from classroom observations show a drastic decrease in the use of corporal punishment from baseline to the second midline. At the baseline, the observed use of corporal punishment on a student of any gender was 89.2% in comparison schools and 62.0% in intervention schools. However, by the second midline, the observed use of corporal punishment was 1.7% in comparison schools and 0% in intervention schools. The use of corporal punishment on girls specifically followed a similar trend with 85.7% of observations at comparison schools and 56.9% of observations at intervention schools saw evidence of corporal punishment at the baseline. During the second midline observation, only a single instance of corporal punishment on a girl was observed at a comparison school and no corporal punishment on a girl was observed at intervention schools.

While this is an overwhelming positive result, it should be noted that the observation of corporal punishment may be influenced by the effect of a classroom observer. Given that teachers know they are being observed and are aware that corporal punishment is considered unacceptable (particularly in intervention schools after having received training), it is possible that teachers may not have behaved normally during observations. As a result, girls were also asked about the use of corporal punishment in their classrooms in order to see if difference emerge in what is observed and what girls are experiencing when the teacher is the only adult present in the room.

The proportion of girls who reported that their teachers use corporal punishment in class or that punish students who get things wrong in a lesson slightly decreased from baseline to the second midline. Comparison school had similar reporting of the use of corporal punishment at baseline with 47.7% of comparison school girls and 42.0% of interventions school girls reported that their teachers used corporal punishment. By the second midline, that number had decreased in both comparison schools (31.7%) and intervention schools (40.2%). The trend of comparison schools having a greater decrease in the reported use of physical punishment is also reflected in the proportion of girls who report teachers physically punishing students for getting things wrong during a lesson with comparison schools dropping from 70.0% at baseline to 54.7% at midline two compared to intervention schools which only decreased from 70.0% to 65.0%. Girls were also asked if their teacher had used physical punishment in the last week. At comparison schools, the number of girls who reported that a teacher had used physical punishment on a girl in the last week decreased from 35.5% at baseline to 15.6% at midline 2. In intervention schools, this number increased from 15.7% to 19.2%.

TABLE 92: USE OF CORPORAL PUNISHMENT, BY INTERVENTION STATUS AND ROUND

	Baseline		Midline 1		Midline 2		DID baseline to ML2
	Comparison	Intervention	Comparison	Intervention	Comparison	Intervention	
Observed use of physical punishment (toward any gender)	89.2%	62.0%	20.7%	34.5%	1.7%	0%	25.2%*

Observed use of physical punishment (toward girls)	85.7%	56.9%	20.7%	31.0%	1.7%	0%	27.1%*
Student report – use of corporal punishment by their teacher in class	47.7%	42.0%	54.7%	53.1%	31.7%	40.2%	14.2%*
Student report – Teachers discipline or punish students who get things wrong in a lesson	70.0%	70.0%	76.5%	76.6%	54.7%	65.0%	10.3%*
Student report- Teacher used physical punishment on other students in last week	23.5%	15.7%	24.3%	24.1%	15.6%	19.2%	11.3%*

*Denotes the DID scores for change in comparison and intervention schools from baseline to ML 2 are significant.

Students from Somaliland were about half as likely (16.7%) as students from Puntland (31.1%) or Galmadug (27.3%) to report seeing instances of corporal punishment in the last week.

FGDs with teachers generally suggest that attitudes toward corporal punishment are negative with teachers generally opting to talk to students, offer advice and encouragement, and setting up conferences with administrators and/or parents instead. Furthermore, during FGDs, teachers generally expressed that they felt that corporal punishment had a negative impact on learning.

I think students shouldn't be punished. Beating students doesn't make students better students. Beating only makes the situation worse and it might not help students. So, I think students need to be given advice and not punishments[219]

I don't punish any student at all. This needs to be clear. When a student makes a mistake, regardless of their gender, I give advice and I tell them to not do it again. Then I give advice so that they can't do the mistake anymore[220]

While there were some examples like the quote below, these were generally few and far between. These qualitative findings are in line with classroom observations that suggest that the use of corporal punishment are declining.

[219] Teacher FGD Sanaag
[220] Teacher FGD Sanaag

No, Somali school culture is not usually a tradition of self-disclosure. But, I still think that the past behaviour is going to disappear now, since if you tell the student several times the student should feel scared, then he/she cannot take any discipline other than fear, so he/she must be disciplined to fear.[221]

Classroom Demeanour

In addition to the use of corporal punishment, we also assessed the classroom environment by the degree to which teachers made students feel welcomed and comfortable. When teachers do things like address the student by name, use respectful language, encourage participation by all students, and provide encouraging feedback, they can make their classroom environment more conducive to students feeling safe and comfortable to participate and take an active role in their learning. In order to assess this aspect of the classroom environment, researchers recorded whether teachers used respectful language toward students, addressed students by name, and avoided the use of harsh language when speaking with students.

The table below compares the indicators of classroom demeanour by intervention and round. There was a large increase in the proportion of girls who reported feeling welcome in both comparison and baseline schools from baseline to the second midline. In intervention schools, 79.0% of girls reported feeling welcome by teachers in the second midline compared to 50.6% at the baseline representing a 28.4% increase. While comparison schools saw a large increase as well of 24.0%, these results are still suggestive of improvement among intervention schools. The use of harsh language also decreased among comparison and treatment schools, however treatment schools saw a decrease in the use of harsh language that was 7.6% larger than comparison schools.

TABLE 93: TEACHER'S CLASSROOM DEMEANOR, BY INTERVENTION STATUS AND ROUND

	Baseline		Midline 2		DID
	Comparison	Intervention	Comparison	Intervention	
Harsh Language not observed	86.7%	84.4%	91.5%	96.9%	7.6%
Respectful language Observed			98.3%	98.4%	N/A
Teacher used student names			52.5%	47.7%	N/A
Students report feeling welcome by teachers	47.2%	50.6%	71.2%%	79.0%	4.4%

** ** None of the DID (difference-in-difference) estimates of project impact are statistically significant

[221] Teacher FGD Sool

While no baseline data is available for the use of respectful language, the high proportion of observations which report the use of respectful language at the second midline is encouraging. 98.4% of intervention observations and 98.3% of comparison observations noted the teacher using respectful language. This result is especially encouraging given that it represents a large increase compared to the first baseline analysis. In the first baseline, 65.6% of intervention observations and 76.7% of comparison observations noted respectful language being used. Given that this number has continued to increase to the point of being observed in nearly all observations, this suggests that teachers may have internalized the importance of using respectful language more than some of the other training foci. Surprisingly, the proportion of teachers who were observed using student names was lower at intervention schools (47.7%) than comparison schools (52.5%). This result is similar to what was reported in the first midline where 56.3% of intervention observations noted the use of student names by teachers.

It is worth noting that there appear to be gendered differences in how teachers behave in the classroom. For instance, female teachers are more likely to use very respectful language than their male counterparts. While the vast majority of all teachers observed use language deemed very respectful during classroom observations, male teachers were more likely to use less respectful language ("somewhat respectful") than female teachers. Similarly, female teachers were less likely to use harsh language directed at a specific student – the average female teacher used harsh language just 0.16 times during a lesson, across all three rounds of the evaluation, while the average male teacher used harsh language 0.28 times.[222]

Although we do not explore differences across schools in much detail, owing to the small sample size of classroom observations available, and the risk of parsing the data into such small subsamples that the results are not meaningful, there are notable distinctions across subgroups of *girls* in terms of how they view the environment of their classrooms. Most notably, girls with disabilities appear to feel somewhat less welcomed by their teachers: 84.6 percent of girls with disabilities strongly agreed that they feel welcome in their classroom at ML2, compared to 88.6 percent of girls who do not have a disability. This gap is actually larger in intervention – rather than comparison – schools, and it holds in the much smaller sample of ALP girls surveyed at ML2.[223] While these findings are merely suggestive – they are not statistically significant – they also align with at least one other metric of how teachers interact with girls with disabilities: girls with disabilities are somewhat less likely to report that their teacher suggests ways they can continue studying on their own at school, which we view as an approximate measure of teacher engagement and interest in student learning.

Gender-Equity in classroom interactions

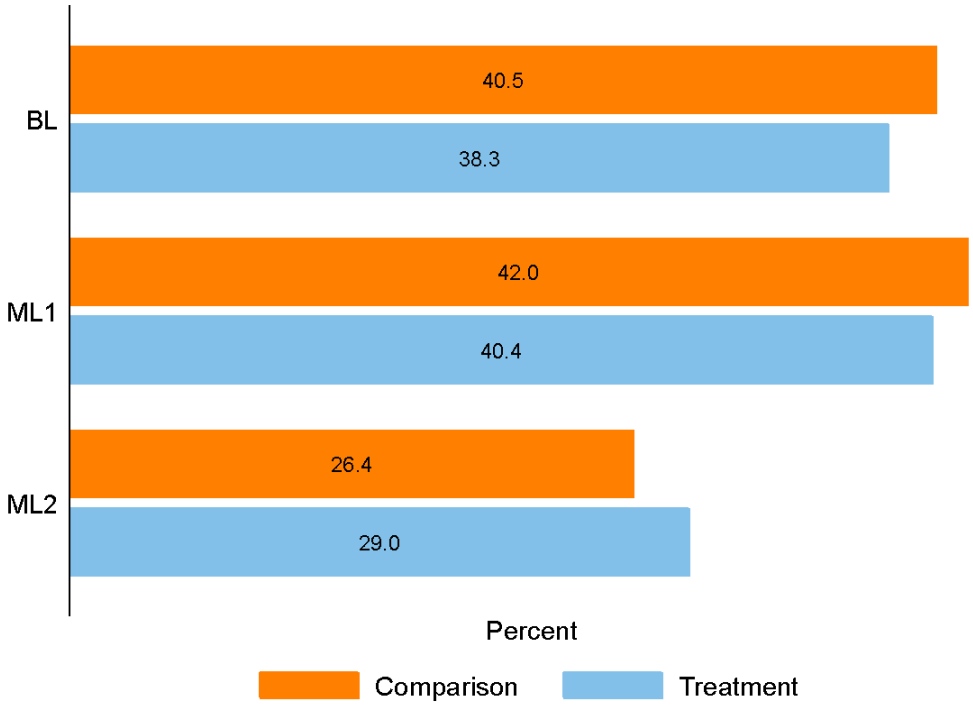
While it is important to assess the way in which teachers are interacting with the student body at-large, we also examined how teachers interact with students based on their gender. A teacher's attitude toward and perceptions of boy and girls has important implications for increasing girls' attendance and may impact their decision of whether to attend. A classroom environment that is conducive to girls' education is one in which their ability to participate in class, interact with teachers and their peers, receive feedback on their work, and have their opinions and thoughts heard. This section examines indicators of gender equity in classroom participation as well as how teachers interact with girls and boys in the classroom.

[222] We use all three rounds of data collection primarily due to the small number of female teachers observed in each separate round (just eight female teachers during ML2).

[223] Differential rates of feeling welcomed by their teacher are not statistically significant in any of the samples studied, owing partially to the small number of girls with disabilities, as well as to the relatively small differences identified.

Girls surveyed were asked if they felt that boys and girls were treated differently by their teachers. As the table below show, there has been a sharp decrease in the number of girls at both treatment and comparison schools who felt that teachers treated boys and girls differently at school. By the second midline, only 26.4% of girls at comparison schools and 29.0% of girls at treatment schools felt that they were treated differently than boys (down from 40.5% and 38.3% at baseline). That girls are slightly more likely to report being treated differently at treatment schools does not necessarily mean that greater degree of gender inequality is occurring at treatment schools, instead this could reflect teachers paying closer attention to the needs of girls in their class.

FIGURE 35: SHARE OF FEMALE STUDENTS WHO BELIEVE GIRLS AND BOYS ARE TREATED DIFFERENTLY AT SCHOOL



Girls in Galmudug were more likely to report being treated differently than boys with 39.0% of girls in Galmudug reporting that that teachers treat boys differently than girls. By comparison, only 25.8% of girls in Somaliland and 28.7% of girls in Puntland felt they were treated differently than boys by their teachers.

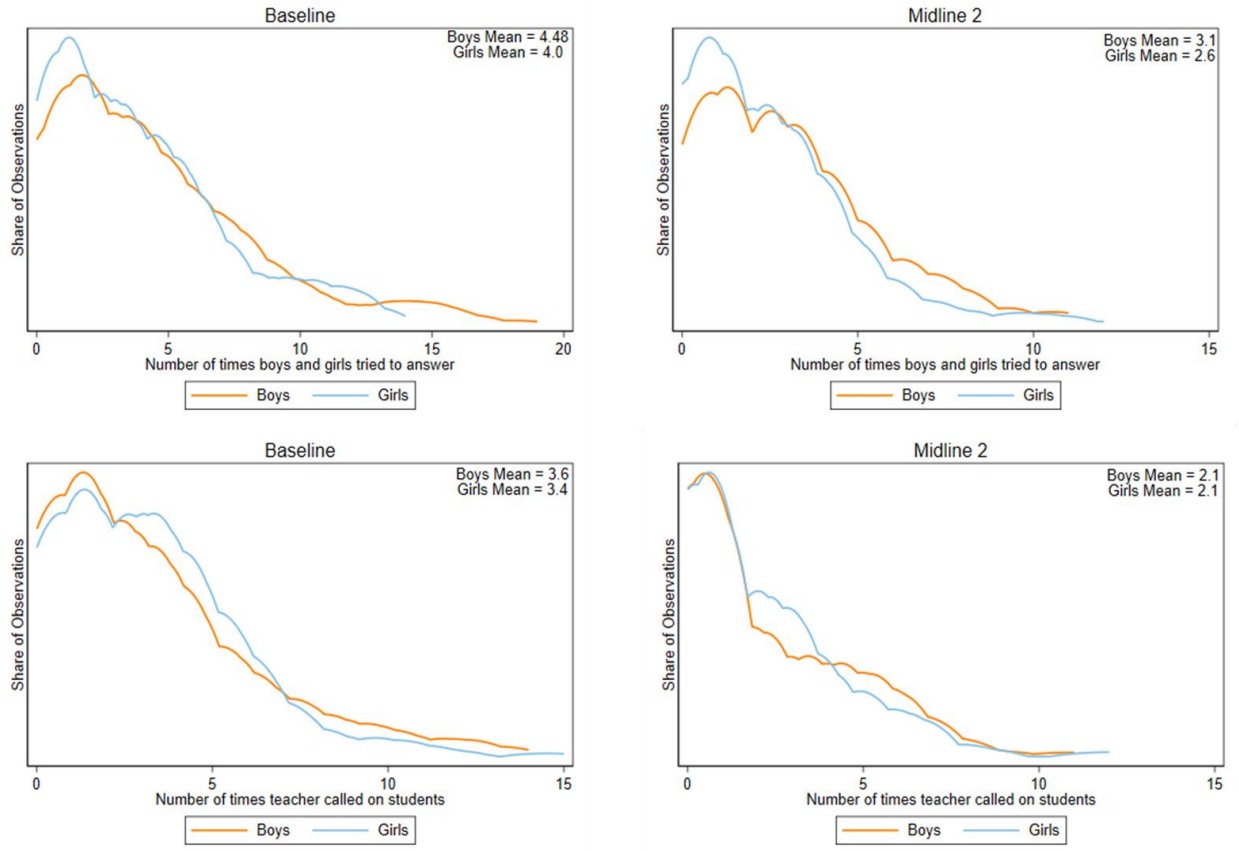
In order to more closely evaluate gender equality in the classroom, we reviewed additional indicators taken from classroom observations. Specifically, we looked at:

1. The number of times male and female students attempted to answer questions in class
2. The number of times teachers called on male and female students in class
3. The number of times the teacher provided positive feedback to male and female students
4. The number of times a teacher used a harsh tone to male and female students

During observations, enumerators recorded the number of times they observed one of these indicators occurring to boys and girls during three, 15-minute blocks. The table below compares the mean number of

times each of these was observed. The figure below plots the number of time boys and girls were called on by round and treatment. Results of the second midline indicate that boys and girls are called on at about the same rate (2.5 times on average). The number of times that teachers calls on students dropped from the baseline to second midline, but this result may be an indication of a change in teaching style or type of questions being asked (more open-ended questions may increase the amount of time spent questioning but reduce the number of questions asked) rather than teachers eliciting less student feedback. Likewise, the frequency by which boys and girls attempted to answer a question decreased from baseline to second midline, which is to be expected given the decrease in number of questions asked. While the observed number of times that girls attempted to participate (2.6) was lower than boys (3.1), the difference was small enough that it could not be distinguished from a null effect. These results suggest that there are not large differences in the extent to which teachers elicit participation from girls and boys in their classrooms.

FIGURE 36: FREQUENCY OF CLASSROOM PARTICIPATION BY BOYS AND GIRLS, BY ROUND



While the results based on girl's perceptions of how they are treated and classroom observations of participation are positive regarding gender equity, this evaluation also included information about the kind feedback boys and girls receive from their teachers. During classroom observations, enumerators were asked to record if encouraging feedback was offered to boys and girls. The use of encouraging feedback was much higher in treatment schools compared to comparison schools for both boys and girls. In intervention schools, boys were observed being given positive feedback 55.4% of the time while girls were observed being given positive feedback 60.0% of the time. These rates are similar to what was observed at the baseline (55.8% for boys and 58.4% for girls). Meanwhile, in comparison schools the use of encouraging feedback has dropped

significantly from baseline to midline for both boys and girls. It is not clear why the use of encouraging feedback may have dropped so precipitously over time, but the fact that intervention schools have been able to maintain a relative high level of the use of positive feedback for boys and girls is encouraging.

Surprisingly, encouraging feedback was less likely to be observed for boys or girls in Somaliland relative to observations in Puntland or Galmadug. Observations in schools in Somaliland did not indicate that boys were offered encouraging feedback more often than girls in the classrooms, but encouragement was only observed 43.7% of the time. By comparison, encouragement for boys was observed in 76.7% of observations and encouragement for girls was observed in 74.4% of observations in Puntland. In Galmadug, encouragement was observed for boys in 94.7% of classroom observations and for girls in all of the observations.[224]

TABLE 94: TEACHER'S USE OF ENCOURAGING FEEDBACK FOR BOYS AND GIRLS, BY INTERVENTION AND ROUND

	Baseline		Midline 2		DID
	Comparison	Intervention	Comparison	Intervention	
Teacher provides encouraging feedback to boys	69.3%	55.8%	49.15%	55.4%	19.7*
Teacher provides encouraging feedback to girls	68.0%	58.4%	40.68%	60.0%	28.8*

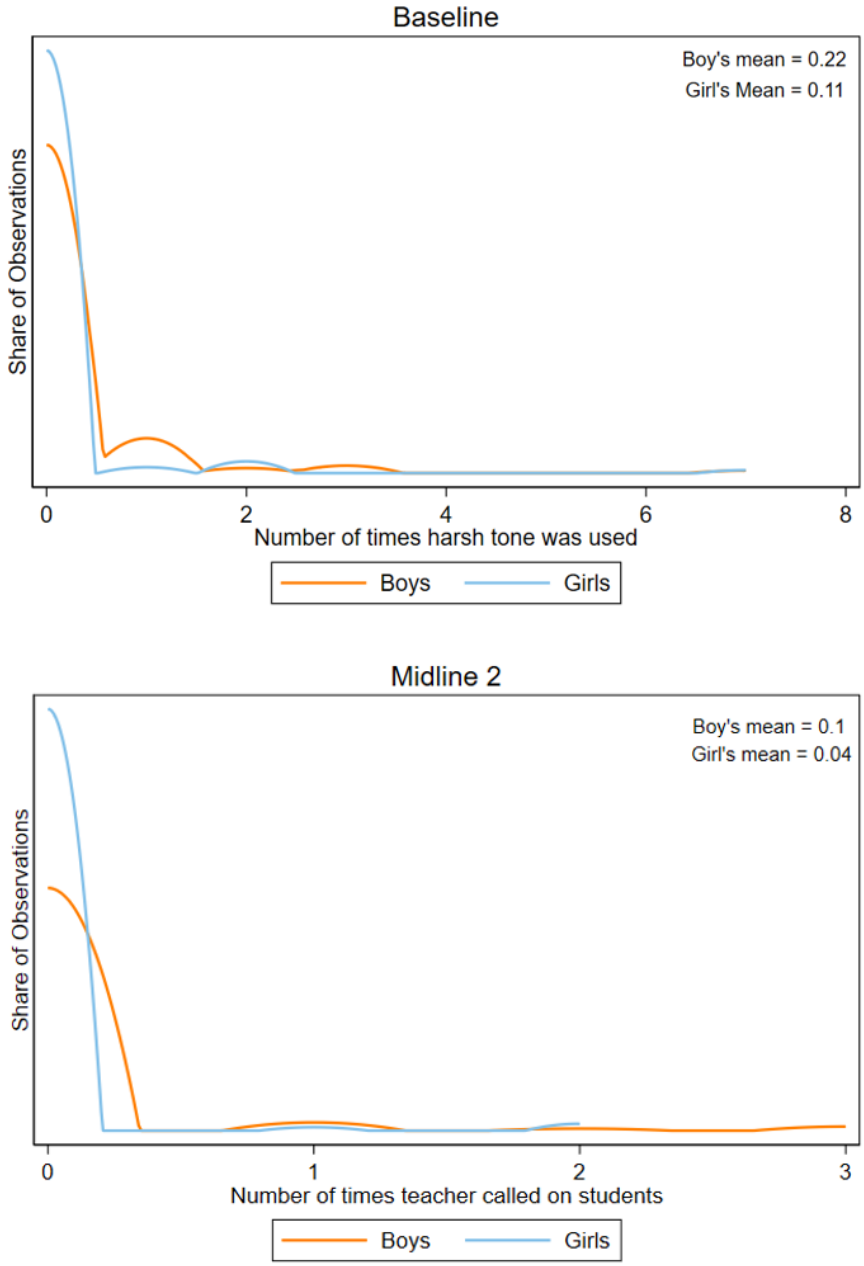
*Indicates the difference is significant at the p <0.05 significance level

During classroom observations, enumerators also counted the number of times that a harsh tone was used for boys and girls. The figure below summarizes these results by round and treatment status. At the baseline, the use of harsh tones was observed more frequently among boys (0.22 times per observations block on average) than girls (0.11). By the second midline, the frequency of the use of a harsh tone had decreased for both boys (0.1) and girls (0.04).

In all, the results of the gender equity analysis are quite encouraging. Girl's generally feel that they are treated the same as boys and the data regarding the number of times boys and girls are called on and participate in class support this. Furthermore, the interventions seem to be having a positive effect on how teachers treat their students in that the incidence of teachers being observed offering positive feedback has remained stable (despite large decreases in comparison schools) and that the use of harsh tones with students has decreased. Taken as a whole, these results show encouraging signs of teachers respecting girls in their classrooms and treating them the same way that they treat boys.

[224] There were only 19 classroom observations in Galmadug compared to 167 in Somaliland and 90 in Puntland.

VFIGURE 37: TEACHER'S USE OF A HARSH TONE WITH BOYS AND GIRLS, BY ROUND



In addition to monitoring teacher behaviour in class, teachers were also asked questions about their attitudes and the potential of boys and girls having different needs. The table below summarizes the response to three different questions about gender equity in a school setting. These questions were asked for the first time during this second midline, so we were unable to make comparisons to see changes over time. Almost all of the teachers surveyed agreed or strongly agreed with the statement that they have the same expectations for boys and girls in class, which lines up well with the previous findings suggesting that teachers are interacting with students equitably during class. Even though the observed teachers show a commitment to having equal

expectations for boys and girls, a large portion of teacher hold the view that boys and girls are better at different subjects. 79.7% of comparison school teachers and 73.8% of treatment school teachers believe that boys and girl are better at different subjects. However, this result may not be as straightforward as it seems. Teachers were asked during FGDs similar questions about the differences in performance between boys and girls. Overwhelmingly, teachers did not differentiate between boys and girls in their abilities (and when they did, it was usually to say that girls are the better students). Instead, these teachers focused more on the challenges that girls face in the classroom such as missing school for menstruation[225], feeling more self-aware and shy[226], and that they may face more barriers to attending school like greater household responsibilities or their family feeling that school isn't as important for their girl.[227]

I do not believe that there is a big difference between the students sitting in the classroom, as the teacher knows each one if their students, so there is no significant difference in their reactions, if any, to say the least. Sometimes the boys may also feel shy about some questions just as the girls do. The main challenges that girls face include shy, this is natural character most of the girls that Allah created them, we made motivation to inspire girls confidence and self-expression, this cause more girls to be more courage to be confident some of the girls are now more confident than boys.[228]

This year, they are improved, but the girls seem to have become more educated. The change has been accompanied by teacher training and girls training.[229]

There is a high recognition amongst teachers in the study of the need to design lessons in such a way that are sensitive to the different needs of boys and girls. 81.4% of comparison school teachers and 83.1% of treatment school teachers said that it is important to design lessons that are gender sensitive. While there is some evidence that biases about the importance of education and the academic abilities of girls may be different than boys, this analysis of gender-equity largely suggests that teachers are largely setting similar expectations for girls and boys academically and interacting with them as equals, while recognizing that each individual student may have specific needs.

TABLE 95: TEACHER ATTITUDES TOWARD GENDER EQUITY

Statement	Comparison	Treatment
I have the same expectations for boys and girls in class	94.9%	93.8%
Boys and girls are better at different subjects	79.7%	73.8%%
It's important to design lessons that are gender-sensitive	81.4%	83.1%

Teacher Effort

Teacher effort refers to the energy, time and resources spent by the teacher to prepare and deliver their lessons. High teacher effort is expected to result in more structured lessons and higher teaching quality overall

[225] Teacher FGD Sanaag
[226] Teacher FGD Togdheer
[227] Teacher FGD Sanaag
[228] Teacher FGD Sanaag
[229] Teacher FGD Sanaag

and improved learning outcomes. We used to indicators of teacher effort to try and gauge how much prepared teachers were to deliver their lessons. This information was not collected at the baseline, so in the following section the first and second midline will be compared.

Our first measure considers the extent to which teachers are prepared for class and appear to have a lesson plan ready when they arrive. Lesson plans are an important tool for teachers to set learning goals, determine the best teaching methods for reaching those goals, and set a standard set of expectations for the students in their classroom. In the first midline, 56.6% of comparison schools and 46.9% of treatment schools were observed to be very prepared by the enumerators. By the second midline, this number had dropped dramatically for comparison schools (37.3%), but intervention schools stayed the same (46.2%).

TABLE 96: TEACHER PREPAREDNESS AND EFFORT LEVEL

	Midline 1		Midline 2	
Indicator	Comparison	Treatment	Comparison	Treatment
Teacher seemed very prepared and had a plan for the observed lesson	56.6%	46.9%	37.3%	46.2%
Teacher clearly communicated the objective of the lesson	93.3%	78.1%	62.7%	73.8%

Our second measure of teacher preparedness and effort, reported in the table above, focuses on their communication of the lesson objectives. Clear articulation of the classroom objectives suggests that the teacher has a clear lesson plan or has at least thought enough about their lesson to describe precisely what it is they want to accomplish during the class period. This is also important for student learning in helping students connect to what they are learning and to be able to track their own progress. At the first midline, this behaviour was far les likely to be observed at treatment schools (78.1%) relative to comparison schools (93.3%). However, by the second midline, comparison schools (62.7%) had fallen below the treatment schools (73.8%). While teacher preparedness did not differ by zone, teachers in Somaliland were more likely to be observed communicating the objectives at the start of class compared to teachers in Puntland (805% compared to 55%).

Given that teacher preparedness and effort may be linked to their environment and the overall quality of the school in which they teach, we looked at these indicators of teacher effort in relation to indicators of school quality (access to water and electricity, cement vs. dirt floors, etc…). The tables below show how the likelihood of observing teachers who are well prepared or who communicate lesson objectives given these indicators of school quality.

TABLE 97: SCHOOL QUALITY INDICATORS

School Quality Indicator	Has feature	Teacher Prepared		Communicated Lesson Objectives	
		Comparison	Treatment	Comparison	Treatment
Have access to electricity sometimes or usually	Yes	100%	100%	64.2%	83.3%
	No	86.7%	80.9%	62.2%	70.2%
Has access to water	Yes	87.2%	85.7%	61.7%	73.5%

	No	100%	87.5%	66.7%	75.0%
Has cement floors	Yes	91.5%	85.2%	61.7%	73.8%
	No	83.3%	100%	66.7%	75.0%
Has functioning CEC	Yes	89.1%	86.2%	63.6%	73.8%
	No	100%	N/A	50.0%	N/A
Has CEC management plan	Yes	91.7%	91.3%	75.0%	71.7%
	No	88.6%	73.7%	54.3%	78.9%

We expected to see indicators of teacher effort increase in the presence of a higher quality school environment. Access to electricity did seem to increase the likelihood of observing a teacher who was prepared and shared the lesson objectives with the class, however factors like access to water and cement (as opposed to dirt) floors were not associated with increased teacher effort. Having a CEC management plan does seem to have a big impact of the likelihood of seeing lesson objectives communicated during the lesson in comparison schools, but overall these results do not suggest that the teaching environment is having a big impact on these measures of teaching effort.

The qualitative data suggests that a combination of lack of content training, resources (for planning interactive lessons, and teaching multiple courses/courses outside their primary subject area may be demotivating and make it difficult for teachers to plan high quality lessons.

The teachers interviewed most-often cited mathematics and language courses (Somali and English) as the most difficult to teach and many express a desire for more content-level training as they may have difficulties developing lesson plans without the help of a more experienced teacher. One teacher explains:

Mathematics is a good subject and I love it. It has complicated parts, and sometimes you would like to get help or ask for it. The place we are here is a village, so we need training and the math is difficult for us, We didn't receive any training since the curriculum was brought to us , especially I did math and did not get any training, I was a former Somali student so I was better at math, but we needed training and it was a challenge, and I was struggling with these challenges, When facing these challenges, I collaborate with people who I think are better than you, by asking the social media sites like WhatsApp.[230]

A lack of resources was also commonly cited as a challenge for these teachers. Without the materials needed to implement some of the curriculum (this seems to be a particular issue for science teachers) the teachers feel that they aren't capable of teaching the students properly. One teacher says:

The only subject that all students understand, and the teacher feel confident in, is Somali. On the other hand, science is very difficult to teach because when they are learning they need to see in practical. Therefore, we don't have the equipment to do so and that's challenge we face.[231]

[230] Teacher FGD Togdheer
[231] Teacher FGD Sanaag

Additionally, there is some suggestion that some of the teachers have too high a teaching burden, either teaching too many students, too many different subjects, and/or teaching courses they do not feel confident teaching in. This may be placing an undue burden on teachers and causing stress that may prevent them from being as engaged in developing their lessons than they might otherwise be. A teacher in Togdheer says:

All subjects need to be developed, whether they are trained in them or not. Seven subjects are taught in the school and the Somali is said to be the easiest. Yet from another perspective, it seems that the Somali language is the most difficult of all. Mathematics, as all teachers have already said, requires training. We also need to increase the number of teachers. Each teacher should only teach one subject because two subjects get stressful for them. Then either the Ministry or CARE should increase the teachers' salaries so that students can get enough teachers.[232]

Impact of CARE'S numeracy training

In addition to analysing changes in teaching quality from baseline to midline, this section also examines the impact of CARE's numeracy training on teaching quality. As a part of the project's first year implementation, CARE has conducted training in numeracy pedagogy with 834 teachers (at the time of this writing). We expect that the training programs to primarily have impacts on numeracy learning scores, but it is also possible that the training will improve pedagogy beyond the teaching of mathematics.

In order to assess the potential impact of the training program on teaching quality, we analysed five indicators representing the range of skills that go into teaching quality. Some of the indicators were chosen because we expected that they would be easy to change and may show a rapid shift in response to training (like the use of respectful language) while others (like the use of participatory methods) are fundamental aspects of pedagogy and ae likely to be emphasized in any training program (such as a class on numeracy-specific pedagogy).

Approximately 40% of the teachers observed during the second midline had received numeracy training (66 out of 174). The largest observed effect of the numeracy training is on the use of respectful language. In treatment school, those that had not received the training used respectful language in 74.4% of the observations. However, in 90.9% of the observation with teachers who had the numeracy training respectful language was used (p. = 0.086). While the numeracy training had a marginally significant impact on the use of student names, none of the remaining comparisons showed significant differences between teachers who participated in the numeracy training and those who did not. Using student names and reporting use of formative assessment was actually slightly higher in observations where the teacher hadn't received the training, while the opposite is true for observations where the teacher said they had records of their formative assessments. However, these differences are quite small (all less than 3% difference) and likely show that the numeracy training had little to no impact on these indicators. Likewise, teachers who received numeracy training had slightly more active classrooms (mean participatory index of 0.51 compared to 0.48 for comparison schools) but again, the difference is quite small and likely indicates the training is have little to no effect on the degree to which classrooms are active and participatory.

TABLE 98: EFFECT OF TEACHER TRAINING ON CLASSROOM DEMEANOUR, PARTICIPATION, AND USE OF FORMATIVE ASSESSMENTS

Indicator	Baseline	Midline 2

[232] Teacher FGD Togdheer

	Comparison	Treatment	Comparison	Treatment	
				Not Trained	Trained
Respectful language used (very respectful)			72.9%	74.4%	90.9%**
Teachers use student names			52.5%	48.8%	45.5%
Teachers use formative assessments (self-reported use)	60%	49.4%	91.5%	93.0%	90.9%
Teachers use formative assessments (reported keeping records)	46.7%	37.7%	23.7%	37.2%	40.9%
Average of index for participation	0.52	0.54	0.43	0.48	0.51

** Marginally significant difference (p. = 0.086) between teachers who received numeracy training and those that did not in treatment schools.

Much positive progress has been made in teaching quality particularly when it comes to caregiver perceptions of teaching quality, use of corporal punishment, reported use of formative assessments, classroom environment, and gender equity, the overall effectiveness of the teaching quality portion of the projects intervention are concerning. Indicators for an active and participatory classroom are all down relative to the second midline (and in many cases relative to the baseline). The gains in active and participatory classrooms observed during the first midline are in danger of being lost without further training for teachers on active and participatory methods. Furthermore, there is little supporting evidence to feel confident that the reported use of formative assessments reflects the reality of their use. Few teachers are keeping record of their use and an understanding of the role of formative assessments and there use in the classroom did not emerge from FGD discussions. In order to continue making progress on developing the skills and capacities of teachers and improving learning scores, more emphasis on pedagogy and its practical use should be a high priority moving forward.

Progress to Teaching Quality Target

Progress within the teaching quality component of the project is defined as improvement in the proportion of teachers who apply improved teaching practices in literacy and numeracy. This measure is formally defined as the proportion of teachers sampled who apply formative assessments among all the teachers sampled. The project's target during the midline evaluations is that the percentage of teachers who report using formative assessments will increase by 30 percentage points over what was reported in the baseline. In the baseline, 44.8% of teachers self-reported the use of formative assessments. By the second midline, this number had increased to 92.3% of teachers in intervention schools. This increase of 47.5% far surpasses the target for this midline assessment. Even when compared to the first midline, the use of formative assessments increased by 20.4%.

However, if we consider the proportion of teachers who have records showing that they use formative assessments, the results do not appear to be promising. The proportion of teachers who report having documentation of the use of formative assessments has stayed approximately the same (38.8% at baseline and 38.5% at the second midline). It is unclear which indicator best captures the use of formative assessments in practice in the classroom, however given the number of teachers who report using formative assessments is so high, it is safe to assume that awareness of formative assessment as a concept has increased. Future evaluations should try to more explicitly quantify the use of formative assessments during observations or by looking for evidence of their use in lesson plans or by asking teachers directly about how they use formative assessments.

When looking at the qualitative data, there was little evidence of an increased use of formative assessments among teachers. Although teachers where not asked explicitly about formative assessments, teachers were asked directly about how they improved their education or teaching methodology and how teaching practices have changed in mathematics and English and Somali literacy. There were no instances of teachers referring to formative assessments explicitly and scant mention of things that could be related to formative assessment (checking for understanding, adjusting lessons to meet student needs, or eliciting student feedback).When asked about discipline-specific teaching practices, there is some acknowledgment of students differing in their understanding or abilities, but not much indication that teachers are thinking about formative assessments as part of teaching best-practices.

In addition to increasing the percentage of teachers who apply improved teaching practices in literacy and numeracy, this project also seeks to create shifts in teachers' awareness of quality education. The project seeks to change teacher perceptions of good, quality education, incorporate child protection, gender equitable treatment in class, and strategies to support the worst performing students. Progress to this target in this second midline evaluation is specifically looking at evidence of 1) teachers acknowledging corporal punishment as a barrier to learning, 2) teachers taking proactive action to support students who are lagging behind in acquiring numeracy and literacy skills, and 3) including specific support to children with disabilities.

There is evidence to support that teachers are acknowledging that corporal punishment is a barrier to learning. 88.7% of teachers surveyed during classroom observations acknowledged that corporal punishment is a barrier to learning. This sentiment is supported form the qualitative findings as well as teachers routinely agreed that corporal punishment is a negative and that may cause students to drop out of school, harm relationships and trust between students and teachers, and does not benefit student learning. This attitude shift is also supported by the precipitous fall in observed instances of corporal punishment during classroom observations. During the baseline, 62.0% of intervention schools had at least one instance of corporal punishment observed. By the ML2, corporal punishment was only observed a single time during classroom observations, and not a single time in intervention schools.

Teachers also appear to be taking more proactive steps to support students who are lagging behind. 93.2% of teachers at treatment schools report that they take proactive action to support students who are falling behind. During FGDs, teachers regularly discussed offering extra classes or tutoring sessions for students who were falling behind and giving them extra materials and practice work to take home to try and improve their skills. For children with disabilities, teachers discussed moving children to the front of the room if they have hearing or vision impairments. However, there was very little evidence of teachers differentiating their instruction or making adjustments to lessons during instruction in order to meet the needs of these students.

7.4 LIFE SKILLS

Outcome	Indicator	BL	ML Target	ML	Target achieved? (Y/N)	Target for next evaluation point	Will IO indicator be used for next evaluation point? (Y/N)
Life Skills	Percentage difference in the average youth leadership index score at midline and endline, compared to the baseline	55.8%	10 percent above baseline	60.1%	No	20 percent above baseline	Yes
Life Skills	Girls describe examples of how they have engaged with others to express their needs and aspirations. Proxy: share who feel nervous speaking in front of an adult	Proxy: 64.4%	N/A	34.6%	N/A	Girls describe examples of engagement with others at school and household to express their needs and aspirations	No. There is a need to agree on indicator and measure as either a proxy or redesign qualitative tools to facilitate qualitative measurement

Main qualitative findings

- The primary takeaway from the qualitative data is the relationship between girls' self-confidence and participation in the Girls' Empowerment Forum and – more generally – the value of a support network for staying in school and keeping up with the material. The former finding is consistent with quantitative findings documenting the impact of GEF participation on life skills.
- Several quantitative proxies indicate that girls feel more self-confident speaking up in the presence of others and that adults take their opinions into account when making decisions. However, the qualitative data do not directly address the extent of empowerment or willingness to speak up at home or at school.

The fourth intermediate outcome targeted by SOMGEP-T consists of girls' life skills and self-confidence. The primary method by which the project seeks to improve life skills is through the Girls Empowerment Forums (GEFs) sponsored in intervention schools, which will give girls an opportunity to participate in the governance of their schools, complete training that encourages their leadership and engagement in decision-making, and offer financial literacy training, among other activities. The project's specific logframe indicator concerns girls' relative empowerment and self-confidence, captured through CARE's Youthe Leadership Index (YLI). Improvements in life skills are hypothesized to increase positive transition outcomes, as girls with a strong sense of empowerment are more likely to act in their own interests in household decision-making and more likely to *want* to stay in school, because they are more likely to recognize the value of schooling and their own

potential. Improvements should also positively impact learning outcomes, as girls become more willing to engage in classroom discussions, and more likely to attend school consistently.

In this section, we report on the impact of the SOMGEP-T program on girls' life skills over time, focusing especially on changes from baseline to ML2 among girls enrolled in school. We also perform additional analyses, studying change in girls' life skills as a function of their participation in the GEF, and perform a brief analysis of within-subgroup program impacts.

Our analysis concerns two different measures, following the reporting in previous rounds. The first is CARE's YLI score. The second is also an index, constructed from the "life skills" module of the GEC-T survey with girls. In supplementary analyses, we disaggregate the results of this life skills index further, first into three subcategories – developed by DFID – of life skills questions and, second, into a set of questions that all capture aspects of girls' self-confidence speaking in front of others or making their voices heard in class or their communities.

At the outset, we note that there is conceptual overlap between the Youth Leadership and Life Skills indices. The YLI score is meant to measure, per CARE's documentation, girls' "self-confidence, their decision-making, problem-solving and organizational skills, their sense of voice, and their ability to motivate others. The YLI also measures cooperation, diligence, independent thinking, personal responsibility, and leadership interest."[233] At the same time, many of the questions focused on life skills capture similar outcomes, especially self-confidence in school, the ability to form relationships with peers and adults, and relative empowerment over decisions that impact them. For instance, the life skills module includes these two representative questions, both of which are related to self-confidence: "I get nervous when I have to read in front of others" and "I feel confident answering questions in class." Throughout this section, we analyse the two indices separately, and refer to the YLI and Life Skills Index, respectively, in order to clarify the precise source of the data in each case.[234]

MAIN COHORT

Youth Leadership Index

The YLI is composed of 21 questions that, as noted above, capture aspects of self-confidence, ability to solve problems, organization, and ability to motivate peers, among other outcomes. The full set of questions is reproduced below. Beside each question, we note whether there are any restrictions on the type of respondents who were asked the question; specifically, four questions are asked only of in-school girls. More generally, YLI data was not collected at baseline among out-of-school girls. Our main focus in this section is on BL-to-ML2 comparisons among in-school girls, though we occasionally adjust this sample somewhat to check the robustness of the results. We also analyse changes among ALP girls, over the period ML1 to ML2, as girls enrolled in ALP are specifically targeted for life skills training. This analysis appears in a separate sub-section, below, on ALP girls.

Before turning to the key comparisons of interest, we briefly describe the construction of the index. Each of the questions in the box below is answered on a 4-point scale, with four response options: rarely, sometimes, most of the time, and almost always. In every case, responses associated with higher frequency (e.g., most of

[233] CARE. 2014. *Youth Leadership Index Toolkit*. Available at https://www.care.org/sites/default/files/documents/CARE-YLI-Toolkit-FINAL-WEB.pdf

[234] We do not report significant methodological analysis of the two indices in this section, partially because the YLI score has been used extensively by CARE in many contexts, and was the subject of an intensive development process. However, we do analyse the extent to which the different index items discriminate between low- and high-scoring girls

the time, almost always) denote more positive behaviours. The raw YLI score is calculated by scoring responses on a 1 (rarely) to 4 (almost always) scale and aggregating the scores for all 21 questions, resulting in a raw score ranging from 21 to 84. We transform the index to fit a 0-100 scale to ease interpretability. Our rescaling does not influence the relative ordering of respondents on the scale, or the spacing between them (there is perfect correlation between the two measures).

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

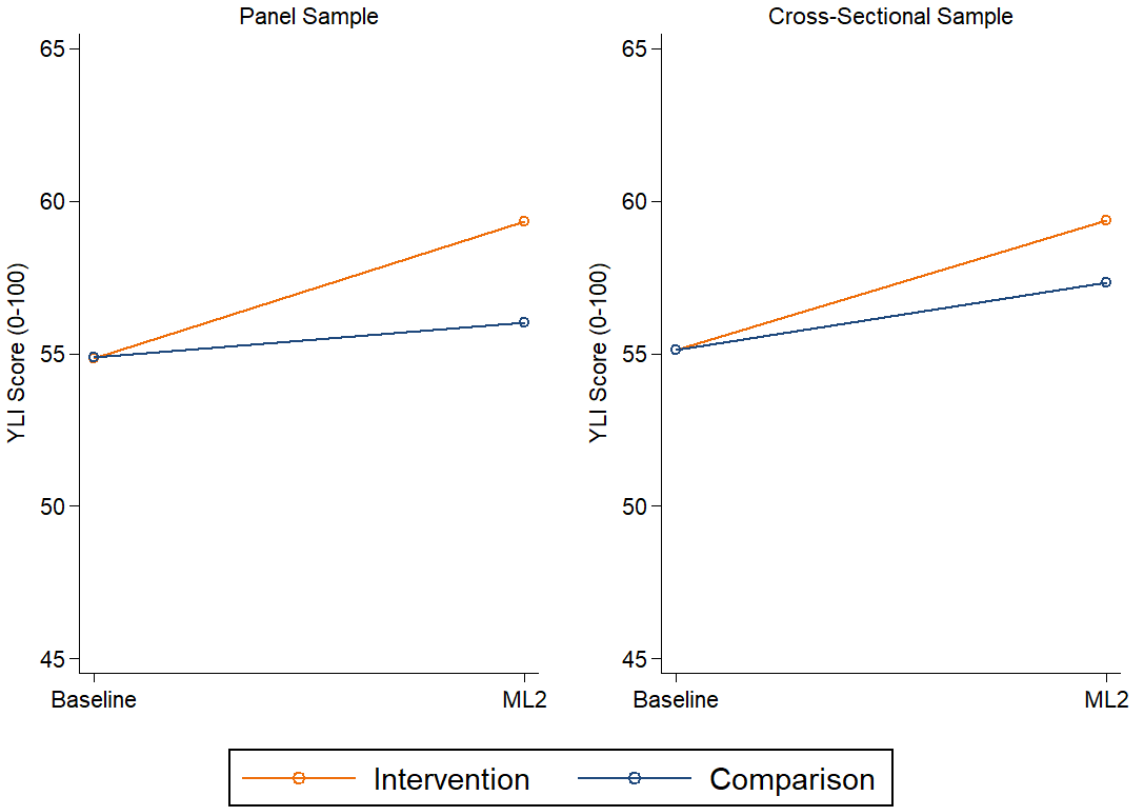
Across all cohorts interviewed during ML2, the mean YLI score was 55.2 on a 0-100 scale, with a standard deviation of 18.6 points. Scores are lowest among the OOS girl cohort (47.7 points), and highest among the ISG (58.2) and ALP girl (57.6) cohorts, respectively. The distribution of YLI scores in the sample approximates a normal distribution, with a slight rightward or positive skew, which is more pronounced in later rounds.

Our first analysis focuses on comparisons between baseline and ML2, as shown in the figure below. The left panel reports the evolution of YLI scores in intervention and comparison schools over time, restricting our attention to the "true panel" – the sample of girls who were successfully re-contacted and appear in both the baseline and ML2. This is the most comparable possible sample, as the change from baseline to ML2 cannot be attributed to sampling variation between the two time periods. The results strongly suggest an impact of the program on YLI scores: from a nearly identical starting point at baseline, girls in intervention schools gained 4.5 points, while girls in comparison schools improved by just 1.2 points.

The right panel broadens the sample to include girls who were replaced at ML2, meaning that the sample is not exclusively comprised of exactly the same girls from baseline to ML2. While the panel sample includes 708 girls in each round, the cross-sectional sample includes 838 girls in each round. The trade-off to this increase in sample size is a reduction in comparability from baseline to ML2. In this sample, girls in intervention schools still improved markedly, but girls in comparison schools improve somewhat more than

in the panel analysis, reducing the perceived impact of the program in this analysis. Nonetheless, intervention girls improve by 2.0 points *over and above* comparison girls over the same time period.

FIGURE 38: EVOLUTION OF YLI SCORES, BY INTERVENTION STATUS



In the table below, we expand this analysis in two ways. First, we estimate formal models of program impact – i.e. difference-in-differences – in a regression framework. Second, we analyse an additional sample of girls, restricting our attention to girls who remained enrolled between baseline and ML2.[235] The top panel of the table reports regression results for the true panel of in-school girls, who were successfully re-contacted at ML2; this sample is exactly equivalent to that reported in the left panel of the figure above. In the first regression, we do not incorporate additional control variables, reporting the raw difference-in-differences estimate; in the second regression we control for geographic zone and age, incorporating a dummy variable

[235] More precisely, this group of girls are those who were enrolled at baseline and were also enrolled at ML2, though we do not specifically assess whether they were also enrolled at ML1, because not all girls in the sample were contacted during ML1. In practice, the vast majority of girls enrolled in both time periods were also enrolled at ML1, but we do not impose this requirement, as it would limit the sample needlessly, given that some schools were not included in the ML1 sample.

for every zone and every age observed in the sample.[236] The third regression breaks down geography further, by controlling for region, in addition to age and grade.[237]

The results show a consistent positive trend in YLI scores among intervention girls, relative to comparison girls, but these changes are not statistically significant. It is also worth noting that the program impacts we estimate get smaller as we control for additional factors, such as region and grade. In the model that incorporates the most controls, the results suggest the program caused a 1.9 point improvement in YLI scores.

The middle panel of the table reports results for the set of girls who remained enrolled in both baseline and ML2. Here the results are more consistent across models: in all three regression models, we estimate the program improved YLI scores by between 2.6 and 2.9 points. None of the impacts we report are statistically significant at conventional levels, but our *best estimate* of program impact is that the program improved scores by approximately 2.6 points. The third panel expands the sample further, to include all in-school girls who were replaced, and their replacements, from baseline and ML2. In this sample, the results are less positive – while all of our analysis shows a positive impact of the program on average, none of the results are sufficiently large to distinguish them from a null effect.

TABLE 99: REGRESSION ESTIMATES OF PROGRAM IMPACT ON YLI SCORES

Regression Details	Impact Estimate (Regression Coefficient)	95% Confidence Interval	Sample Size
Girls Headcounts			
No controls	1.5	-5.8 - 8.9	832
Control for zone	1.7	-5.6 - 8.9	832
Control for zone and grade level	1.4	-5.9 - 8.7	832
Somaliland only, control for grade level	2.3	-7.7 - 12.3	537
Puntland only, control for grade level	-2.4	-11.7 - 6.9	265
Boys Headcounts			
No controls	5.9	-2.7 - 14.5	836
Control for zone	6.1	-2.3 - 14.4	836
Control for zone and grade level	6.1	-2.4 - 14.5	836
Somaliland only, control for grade level	14.0*	2.3 - 25.6	539
Puntland only, control for grade level	-8.4*	-15.9 - -0.9	266

[236] We typically employ age dummy variables, rather than a single variable capturing different values of age on a continuous scale, because the former approach is more flexible. Using dummy variables allows each age level to have a different impact on YLI scores, allowing a non-linear relationship between age and YLI scores, while using a single age variable forces the relationship between YLI scores and age to be linear (i.e. forces the effect of moving from age 12 to age 13 to be exactly the same as the effect of moving from age 17 to age 18).

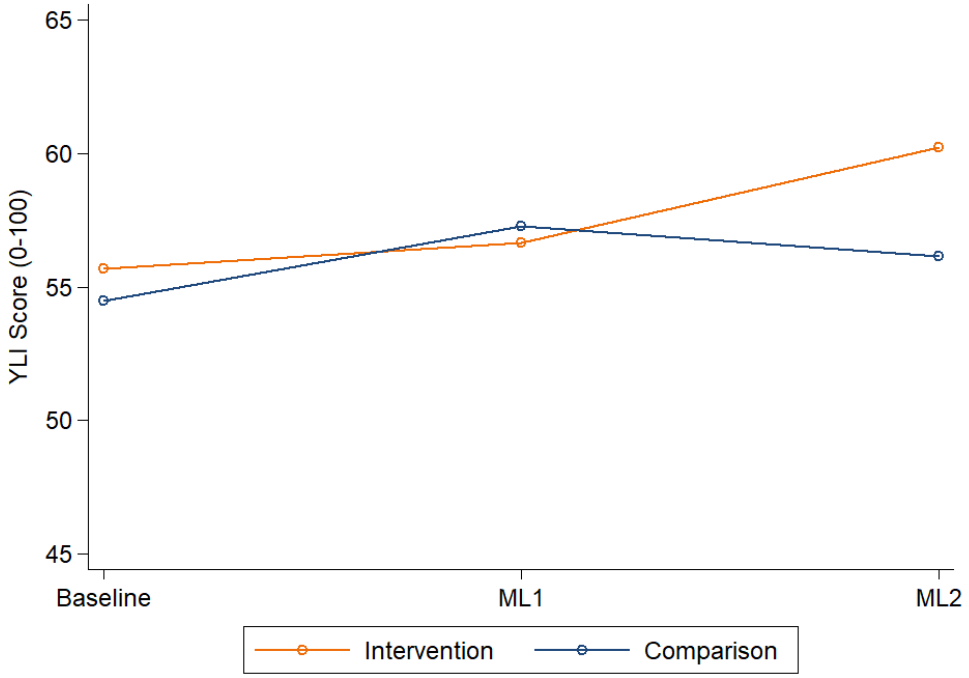
[237] Region in the Somali context refers to administrative regions within the broader constructs of Galmudug, Puntland, and Somaliland. There are nine distinct regions within the sample.

The main analysis thus far uses a sample of girls who were successfully tracked from baseline to ML2, regardless of whether they were re-contacted at ML1. While providing a larger set of girls to study – by not restricting the sample to those girls who appear in all three rounds, some of whom attend schools that were not visited at ML1 – this approach is limiting insofar as we are also interested in the evolution of YLI scores across all three rounds of the evaluation. To understand how YLI scores have shifted from baseline to ML1 and, eventually, ML2, we report the results for a smaller sample of girls who were successfully contacted at all three rounds of the evaluation (the "full panel", n = 522).

The figure below plots YLI scores for this sample. Interestingly, intervention and comparison communities appear to evolve opposite to one another: in the first year of program implementation, intervention girls experience a very minor improvement in overall YLI scores, but a clear relative loss when compared to girls in comparison schools. In the second year of program implementation, however, intervention girls' YLI scores improve dramatically, resulting in a moderate gap in scores at ML2. Among this sample, the net effect of the program on YLI scores from baseline to ML2 appears to be 2.9 points.

Based on past experiences related by CARE's Monitoring & Evaluation team, the trends shown in the figure appear to be consistent with trends observed in other girls' education and life skills-oriented programs. As they described, girls may enter a program targeting leadership skills overestimating their relative ability to lead, partially because they are not aware of what "leadership" truly entails, and because they may not be aware of topics like gender rights. After some degree of experience in the program, they begin to become "more aware of gender rights and stop accepting barriers to their realization as 'normal'." This produces a situation in which leadership scores initially drop as girls become more aware of the barriers they face, but rise again as barriers are overcome.

FIGURE 39: YLI SCORES IN THE "FULL PANEL" OF ISGS, BY INTERVENTION STATUS AND ROUND



Beyond the analysis reported above, we also tested the impact of the program on YLI scores that were adjusted to ensure the same set of questions were asked of all girls. Four questions in the index are only relevant for girls who are enrolled in school, because they specifically ask about a girl's comfort at schools or actions that can only be taken at school. For instance, one question asks girls how often they ask questions at school if they do not understand something. We calculated an alternative version of the YLI score that excludes these four questions, to ensure that differences in the components of the index between girls and over time are not driving our results.[238] While it is highly unlikely that this relatively small shift could influence our results in total, we verified by repeating our regression analysis. The results mirror those using the full YLI score, ranging from 2.4 to 2.6 points.

The YLI score is constructed to capture many related aspects of self-confidence, organization, and so forth. Across the 4,253 girls who have completed the YLI survey module since the SOMGEP-T baseline, the Cronbach's alpha is 0.92, which meets the highest standards for a scale's internal consistency.[239] Even in internally consistent scales, it is possible that the program has had an outsized impact on particular aspects of youth leadership, and smaller or negative impacts on other aspects. To assess this possibility, we studied the impact of the program on each individual item that comprises the YLI. The five items where we observed the largest positive program impacts were:

[238] At baseline, all girls in the sample analysed here were asked all 21 questions. At ML2, however, some of the girls in the sample were no longer enrolled in school; these girls were only asked the subset of 17 questions.

[239] Cronbach's alpha measures the covariance between individual pairs of questions in the scale. It is not a measure of how effectively the scale captures "leadership" or any other outcome, but high covariance between individual items in the index does imply that the scale is internally consistent, i.e. girls who score well on one item tend to score well on other items. The scale is broadly unidimensional, meaning that it is generally capturing one core outcome.

- My friends ask me for advice
- I am comfortable when my teacher calls on me to answer a question
- I contribute ideas to talks at home, even if they disagree with others
- I ask questions at school when I don't understand something
- The things I do set a good example for my peers

Four of these five items registered statistically significant improvements as a result of the program.[240] The program appears to drive positive impacts most on these five questions, though the coefficient of program impact is positive for 17 of the 21 scale items. This suggests that, in general, the program is having a broad-based impact on YLI scores. In general, 3 of the biggest effects are related to a girl's willingness to speak up, either at school or at home. It seems the program is having a particularly strong impact on a girl's self-confidence in settings where they might not normally have a voice.

Given the extent to which the evidence – in this and previous rounds – emphasises that girls do not speak up as often in school, due to "shyness", being uncomfortable speaking in public, or being uncomfortable speaking with a male teacher, this finding is particularly interesting. Girls' unwillingness to speak up in class is mentioned consistently across qualitative interviews with teachers, mothers, CEC members, and girls. For instance, one teacher indicated that girls and boys in their classroom are similar in all ways, "except that the girls are often shyer than boys."[241] The same finding has often emerged from classroom observations and, in previous rounds, surveys with teachers. For example, during the baseline, girls were less likely to *attempt* to answer questions than their male peers, during observation of classrooms.

Per the program's ToC and list of activities, we would expect much of the program's impact on self-confidence and life skills to be driven by the Girls Empowerment Forums. GEFs are specifically tasked with providing girls training in terms of life skills and encouraging them toward leadership within their schools and communities. The data seem to confirm this hypothesis, which is a key assumption of the ToC: girls who have participated in GEF activities experience greater growth in YLI scores over time than girls in intervention schools who *did not* participate in GEF activities.

The figure below shows the evolution of YLI scores among these three distinct groups of girls, all of whom were enrolled in school at the baseline. The comparison group consists of in-school girls enrolled in comparison schools. The orange line denotes girls in intervention schools who reported – in either round – that they had participated in a GEF activity or meeting at their schools. The black line denotes girls in intervention schools who reported that they had not participated in any GEF activities. In the analysis of changes from baseline to ML2 (left panel), all three groups of girls start from a similar starting point, but girls who participated in GEF activities outpace both other groups in YLI scores by the time of the midline. While comparison girls' YLI scores increased by 1.2 points and intervention girls who did not participate in GEF activities saw a gain of 3.1 points, intervention girls participating in GEF activities improved by 7.0 points over the same period. Although the results are not statistically significant, they do appear to confirm the idea that GEF participation is the main mechanism for improving YLI scores within intervention schools.[242]

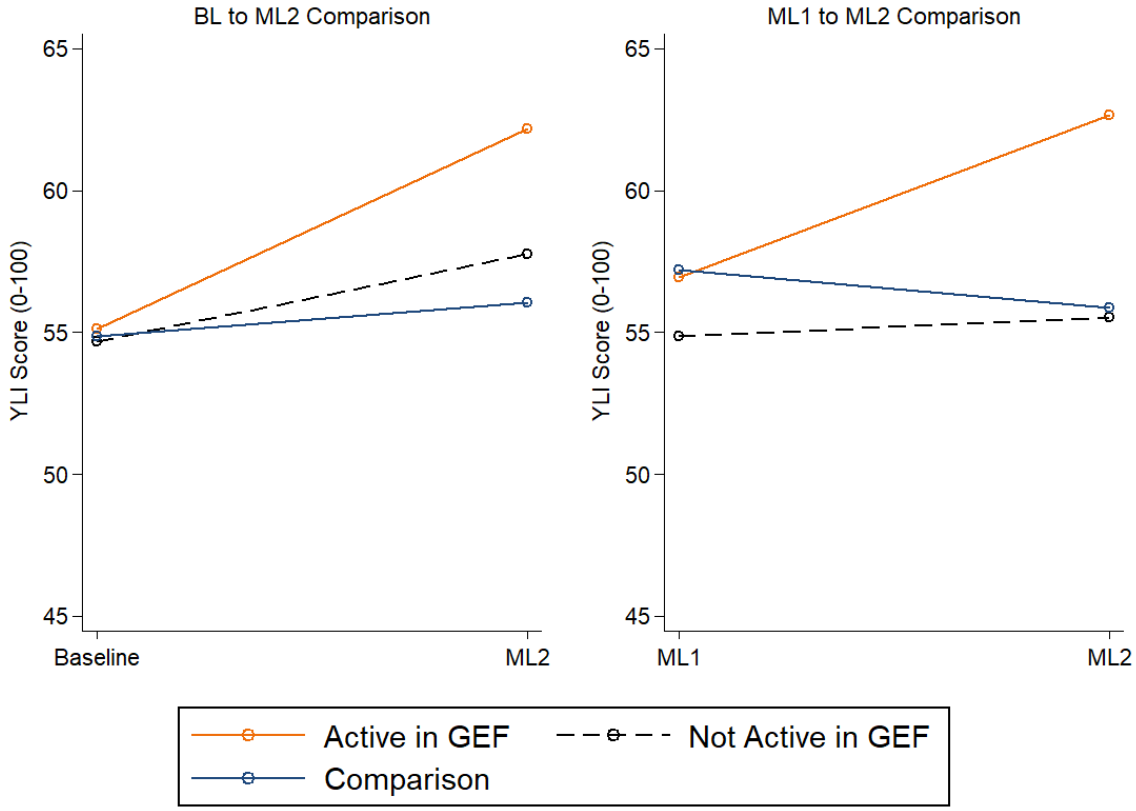
---

[240] At the 10 percent level. The third item did not meet this threshold ($p = 0.11$). However, we caution against over-interpreting statistical significance, as the effect sizes between variables with p-values that fall immediately on either side of whatever alpha cut-off is are not, themselves, distinguishable from one another.

[241] FGD with Teachers, Somaliland, Int. 131.

[242] In a variety of regression models, the impact of intervention and GEF participation combined is positive and substantively significant, but p-values range from 0.18 to 0.25, depending on the model specification and how the comparison group is constructed.

FIGURE 40: EVOLUTION OF YLI SCORES, BY INTERVENTION STATUS AND GEF PARTICIPATION



Two additional findings emerge from the figure above. The first is that the program, even in the absence of participation in GEF, seems to influence girls' YLI scores, relative to comparison girls. In both time periods considered, non-GEF intervention girls still gained ground relative to comparison girls, though the impact on YLI scores is much lower than among girls participating in GEF activities. The second finding is that starting points between these three distinct groups are very similar. Intervention girls who will eventually participate in GEF activities are indistinguishable – in terms of baseline YLI scores – from intervention girls who do not ever report participating in GEF activities. This finding is interesting, because we expected girls with the most confidence, empowerment or sense of identity to self-select into participating in GEF activities. This does not appear to be the case, implying that the GEF is having an impact on girls who are otherwise similar to their peers who did not choose to participate in the GEF.

Life Skills Index

Our discussion of girls' life skills continues by focusing on the Life Skill Index included in the household survey. As we described previously, the Life Skills Index captures aspects of a girls' personality and characteristics that overlap with those in the Youth Leadership Index. The life skills module applied to SOMGEP-T has been altered over time, which complicates comparisons made from baseline through ML2. In total, there are 40 distinct questions that have been used in the survey module, but many are not used in all three rounds, and others are only answered by a specific subset of girls (e.g., out-of-school girls under the age of 12). However, even where there is non-congruence across rounds or among all respondent types, many questions are very similar, with question wording adjusted to context. The table below lists the questions that were asked at baseline and ML2, which respondents received each question, and whether the

question is used in our construction of the Life Skills Index. Our interest is primarily in comparisons from baseline to ML2, which motivates our construction of the index to maximize similarity between the two rounds' indices.

TABLE 100: LIFE SKILLS INDEX QUESTIONS AND APPLICABLE SAMPLE AT BASELINE AND ML2

	Questions	Baseline Respondents	ML2 Respondents
Learning to learn	I am able to do things as well as my friends	All	None
	I can read as well as my friends	None	All
	I get nervous when I have to speak in front of an adult	OOS girls	OOS girls
	I get nervous when I have to speak in front of a group of people my age	OOS girls	OOS girls
	I get nervous when I have to read in front of others	ISGs	ISGs
	I get nervous when I have to do maths in front of others	ISGs	ISGs
	I feel confident answering questions in class	ISGs	ISGs
	I feel confident answering questions when I'm in a group of people	OOS girls	OOS girls
Learning for Life	I would like to continue studying/ attending school after this year	ISGs	ISGs
	I would like to continue learning by going back to school, learning a vocation or trade	OOS girls	OOS girls
	I recognise when choices I make today about my studies can affect my life in the future	ISGs over 12	OOS girls and ISGs over 12*
	I recognize when choices I make today can affect my life in the future	OOS girls over 12	None
	I can describe my thoughts to others when I speak	All	All
	I can work well in a group with other people	All	All
	When I have the opportunity, I can organize my peers or friends to do an activity	All	All
	I often feel lonely	OOS girls and ISGs over 12	OOS girls
	I often feel lonely at school	None	ISGs
	I ask an adult if I don't understand something	OOS girls	OOS girls
	I ask the teacher if I don't understand something	ISGs	ISGs
Agency	Who decides: Whether or not you will go to school	ISGs	ISGs
	Who decides: Whether or not you can go back to school or vocational training	OOS girls	OOS girls
	Who decides: Whether or not you will continue in school past this grade	ISGs	ISGs
	Who decides: If you will work after you finish your studies	ISGs	ISGs
	Who decides: How often you spend time with your friends	All	All
	Who decides: When/ at what age you will get married	All	All
	Who decides: Who decides what type of work you will do after you finish your studies	All	ISGs

	Who decides: What type of work you will do	None	OOS girls
	Who decides: How you spend your free time	OOS girls and ISGs over 12	OOS girls and ISGs over 12

In essence, there are three types of changes from baseline to ML2 in specific life skills questions. The first concerns cases in which a new question was added. We only utilize these questions when they are clearly meant to be equivalent to a question from the baseline.[243] The second type of change occurs when the set of girls to which a question is applied changed from round to round or when we restricted the set of girls for whom we use a question to ensure similarity across rounds.[244]

The two types of changes outlined above were readily handled during the data cleaning and analysis stage. The third type of change is different in this regard. In both rounds, the set of life skills questions applied to girls occasionally depended on their age -- girls under 12 years of age received a somewhat different set of questions. This same approach applied at ML2, but girls had aged two years in the meantime, meaning that -- for all girls who were 10-11 years of age at baseline -- the set of questions they received changed from baseline to ML2. We cannot adjust the index for this shift in questions. However, we do not believe it to be a concern for drawing conclusions about program impact on life skills, for the same reason that we are not overly concerned about very small changes in learning assessment difficulty from baseline to ML2.245

Ultimately, our construction of a life skills index relied on 28 distinct questions. However, when we account for overlap in questions that are intended to be equivalent, but are tailored slightly to different respondent populations, our index captures 16 individual items for OOS girls and 18 for in-school girls. The table above documents these questions, and breaks them into three categories of life skills, as defined by DFID. We reversed the scoring of questions, where necessary, such that higher values represented more positive outcomes. We standardized the score of each question onto a scale with mean zero and a standard deviation of one, in line with common practice when constructing indices from items with different raw scales (e.g., binary, 4-point, and 5-point items in the same index). Next, we aggregated the items into a raw score, with each item equally weighted in the calculation, and rescaled this measure to a 0-100 range to ease interpretation.

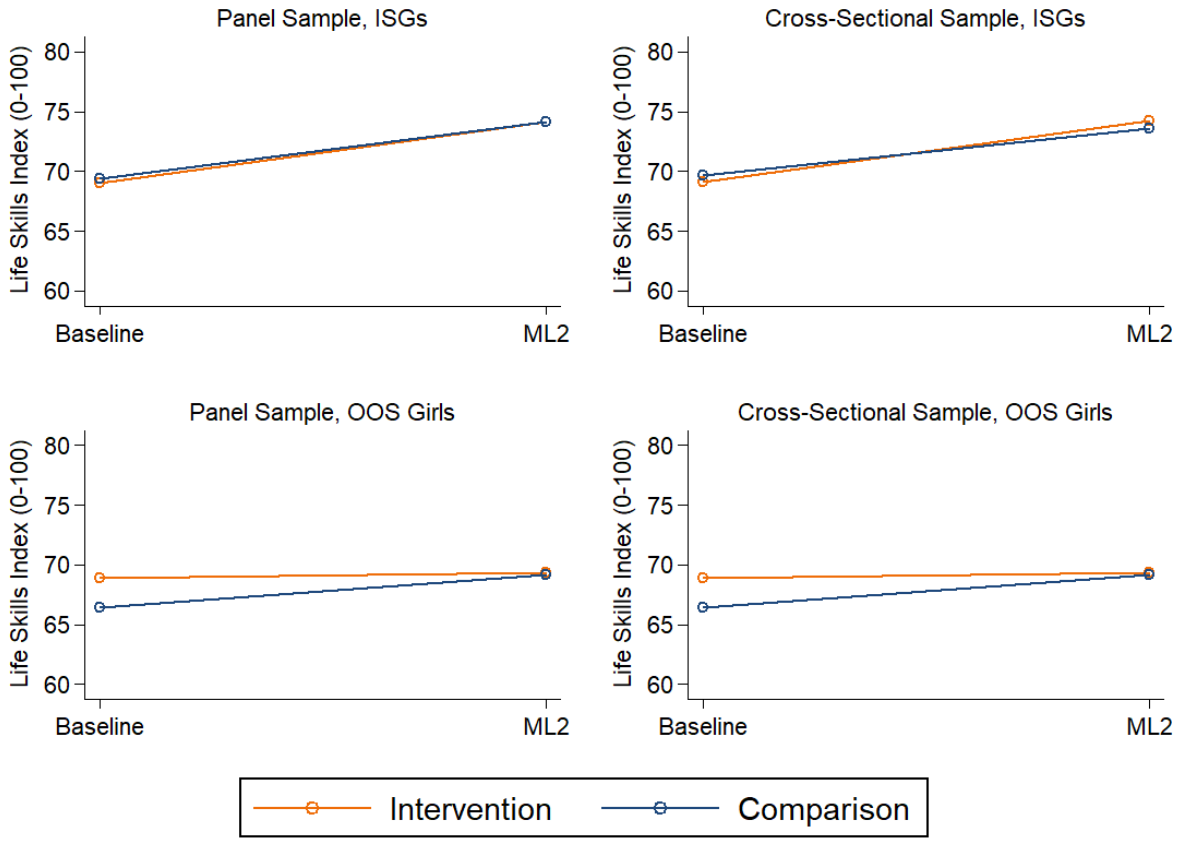
As in other sections of this report, our main analysis relies on comparisons between intervention and comparison girls over time, specifically from baseline to ML2. The figure below shows the evolution of life skills over time in two different groups of girls. The top two graphs report on in-school girls, who are expected to benefit directly from SOMGEP-T interventions -- such as sponsorship of GEFs -- meant to

[243] For instance, the question "Who decides what type of work you will do" was added after the baseline round. It is applied to OOS girls and is clearly intended to be a better-targeted version of, but equivalent to, the question "who decides what type of work you will do after you finish your studies." The latter question was applied to all girls at baseline and only in-school girls at ML2, whereas OOS girls received a question at ML2 that did not refer to school or studies in the question text. In contrast, a new question asking whether a girl has a trusted adult she can confide in was not included in the analysis, because no equivalent question was included at baseline.

[244] Typically, this happened when a new question was added at ML2 and applied to a subset of girls, while a similar question was kept from baseline also. In these cases, the original intention was for each question to apply to a subset of girls, but the evaluation team was occasionally overly cautious and applied the baseline question to all girls at ML2. In these cases, we corrected this issue during analysis, ensuring the indices only used one question -- from a pair of equivalent questions -- for each type of girl

[245] In line with our argument regarding learning assessments, the difference-in-differences design controls explicitly for small changes in measurement approach, as long as those changes are universally applied to all girls in the ML2 sample. Because the intervention and comparison groups include an approximately equal number of girls who were under 12 years of age at baseline, this shift at ML2 should not bias the analysis in this section.

increase life skills. The bottom two graphs show scores for OOS girls, who do not have access to such training, in general. In both cases, we prefer the panel sample, from a methodological perspective, because it restricts the sample to precisely the same girls over time. This means that changes in sample composition cannot drive changes in life skills over time, and any shift in life skills over time can be attributed more plausibly to the program's impact.



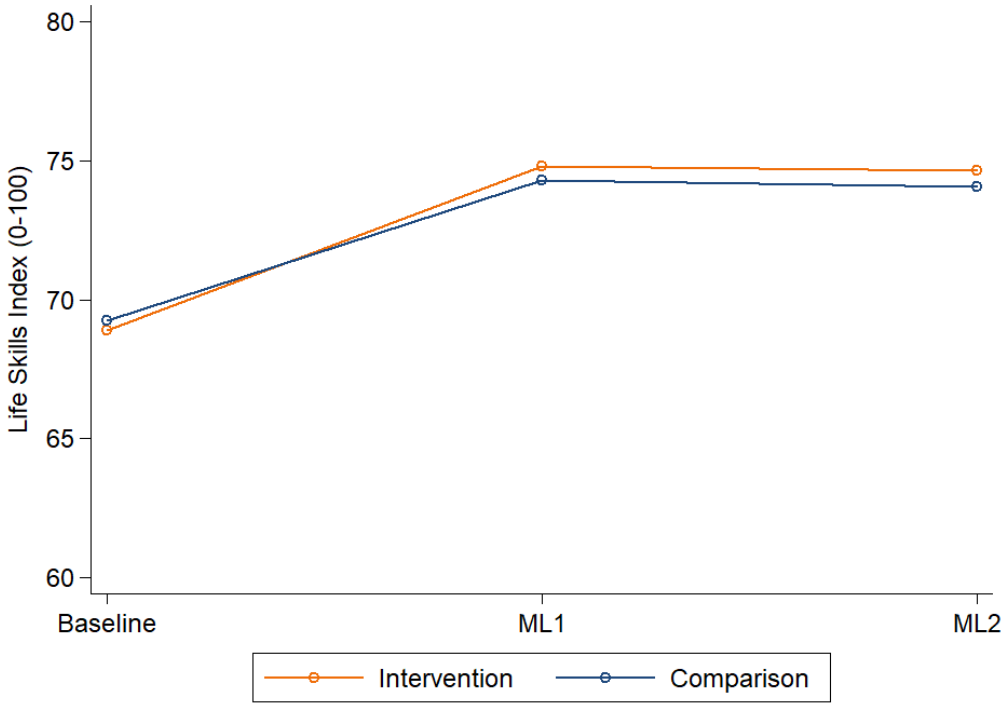
In contrast to our findings regarding the YLI, above, there is very little evidence that the program has improved life skills, on average. Among in-school intervention girls who were successfully re-contacted at ML2, there is an improvement of 5.1 points from baseline in terms of life skills. However, girls in comparison schools experience a similar increase, improving by 4.7 points over the same period. Therefore, our estimate of program impact is just 0.4 points, a result that is not statistically, nor substantively, significant. The results among the cross-sectional sample is only marginally more positive, with an estimate of 1.2 percentage points for intervention girls, over and above comparison girls.

The bottom panel shows that life skills actually decreased in intervention communities, among OOS girls, relative to OOS girls in comparison schools. It is unclear why this would be the case: while we did not expect to find a tangible increase in life skills among OOS girls at all, there is no reason to expect performance among OOS girls in intervention communities should actively decline, relative to their counterparts in comparison communities. In one sense, this decline invites the question of whether something structural has changed in intervention communities that would account for such a decline. If this is the case, it might also explain the lack of impact on life skills among in-school girls -- the program may have had an impact in the absence of an exogenous shock to intervention communities. However, this argument is entirely speculative, as we have

not identified a specific shock and, if such a shock occurred, it did not have as significant an impact on YLI scores as it did on life skills.

The same finding is supported, broadly, when we analyze life skills scores across all three evaluation waves at once. Using the "full panel" described in the YLI section above (n = 523), the figure below reports the life skills scores of the same set of girls from baseline to ML1 and ML2. Whereas YLI scores in intervention communities showed an initial relative decline and then a sharper increase, we do not observe this trend in life skills, with both intervention and comparison communities evolving very similarly. One explanation for the discrepancy between YLI and life skills results, as noted elsewhere in this section, is that YLI represents a set of skills and knowledge very specifically targeted by SOMGEP-T, while life skills are a broader construct designed for GEC-T programming in general, and not specifically tied to SOMGEP-T's activities.

FIGURE 41: LIFE SKILLS INDEX SCORES, BY INTERVENTION STATUS AND ROUND



We investigated life skills in more detail in a regression framework, in a similar structure as the YLI analysis above. Our approach uses a linear regression model, with an interaction term between treatment and round to identify the impact of the program on the life skills index. We used three different samples in this analysis:

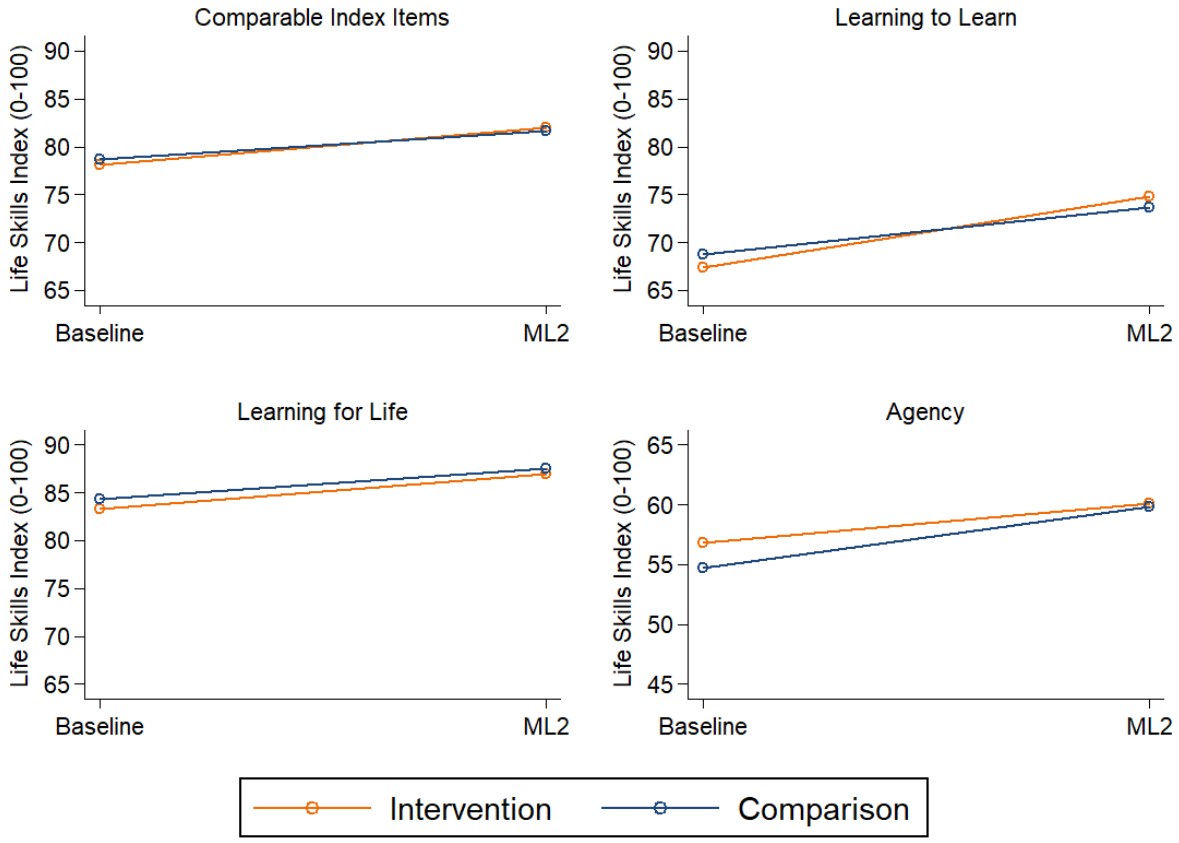
- The true sample of in-school girls (n = 708 per round)
- The true sample of in-school girls who remained enrolled at ML2 (n = 635 per round)
- The cross-sectional sample of in-school girls (n = 838 per round)

We restrict our attention to in-school girls in this analysis, as they are the girls most likely to be affected by the program's activities. In each case, we estimated program impact after accounting for a variety of factors that could also influence life skills, such as geographic zone, age, and grade. Across a number of model

specifications, we estimated a very small but consistently positive impact of the program; however, these findings never approached statistical or substantive significance, ranging from -0.1 points to 1.2 points on a 100-point scale. Even at its largest, this 1.2-point estimate represents just 0.1 standard deviations on the life skills index ($p = 0.43$).

Given our earlier discussion of comparability across rounds, it is reasonable to wonder whether the construction of the life skills index accounts for our null result. This is especially true, given that the program's impact on YLI was more substantial. To guard against this possibility, we constructed an alternative life skills index, using the index items that are precisely comparable from baseline to ML2. We also disaggregated the index into three thematic areas, as defined by DFID, and as outlined in the table of index items presented earlier in this section. It is possible that the program had an impact in one thematic area, which was offset by declines in performance in another area -- a possibility that might emerge if the program's activities focused on self-confidence but did not directly address barriers to girls' participation in household decision-making. Improving girls' comfort speaking up in class is primarily a function of improving the self-confidence of girls and training teachers to facilitate girls' participation, whereas changing the nature of household decision-making might require more activist interventions into community attitudes. Therefore, it may not be surprising if the program influences the former aspect of life skills but not the latter.

The figure below reports findings for these alternative constructions of the life skills index. We find no evidence that focusing on the exactly identical questions from baseline to ML2 alters our conclusions regarding program impact on life skills.



More importantly, the results seem to confirm our hypothesis regarding the area of most likely impact from GEF training. We observe a small but meaningful change over time in "learning to learn" outcomes, which focus on self-confidence, and nervousness or confidence in school and in front of adults. In contrast, we observe no effect on the other two thematic areas, reported in the bottom two panels of the figure. With respect to the "learning to learn" index items -- of which there are 8, but only 4 for in-school girls -- we find that in-school girls in intervention communities improved by 2.6 points relative to girls in comparison communities. While this result is not statistically significant, the results are strongly suggestive of program impact. First, a focus on this particular set of outcomes, as opposed to the entire life skills index, is theoretically motivated, given the nature of SOMGEP-T interventions. Second, the impacts are even greater when we limit the sample to girls who remained enrolled at the time of ML2, which is consistent with the idea that longer exposure time to SOMGEP-T interventions yield greater impact.246 Third, as we show below, the results are even more pronounced among girls participating in GEF, as we would expect if the program's impact is emerging via this specific channel.

Our final results concern the impact of participation in GEF activities on life skills. As with our YLI analysis, we define a set of girls who have reported participating in GEF activities at either baseline or ML2, a set of girls who attend intervention schools but do not report participating in GEF activities in either round, and a set of girls who are enrolled in comparison schools. We study the differential impact of the program, depending on participation in GEF activities.

Looking first at the full life skills index, there is strong evidence that GEF activities have improved scores. Among intervention girls who do not participate in a GEF, the impact of the program, vis-a-vis comparison communities, is actually negative (-0.9 points). On the other hand, among intervention girls who participate in a GEF, the program has generated a 2.9 point increase in life skills ($p = 0.10$). This is very similar to our findings for YLI scores, where girls participating in GEF activities saw outsized impacts. As with that analysis, it is conceivable that girls with higher levels of intrinsic motivation or self-confidence self-select into participation in GEF activities. However, our analysis controls for baseline life skills and YLI scores and studies the relative change over time. In order to undermine these findings, we would have to believe that girls who participate in GEF activities have become more motivated or self-confident between baseline and ML2 as a result of some process that is not attributable to the program, which seems unlikely.

Here, again, the data make it clear that the program has its biggest impact on the aspect of the life skills index that focuses on self-confidence. When we limit our analysis to the "learning to learn" sub-index, girls participating in GEF activities see a 6.1-point increase in their score from baseline to ML2, over and above the comparison group. This result is statistically significant at the 1 percent level, and is meaningful in a substantive sense, representing a 0.4 standard deviation improvement in this index. Meanwhile, intervention girls who do not participate in GEF activities see no tangible improvement in this sub-index.

Self-Confidence

Building on our analysis above, we delve more deeply into outcomes related to self-confidence in this section. The "learning to learn" index consists of several measures that capture aspects of self-confidence; however, following the approach taken in the previous midline evaluation report, we analyse a set of self-confidence questions from this sub-index separately. The six questions that center on self-confidence are:

- I get nervous when I have to speak in front of an adult

[246] Among this latter sample, girls in intervention communities have improved by 3.4 points as a result of the program, a result that is marginally significant ($p = 0.09$).

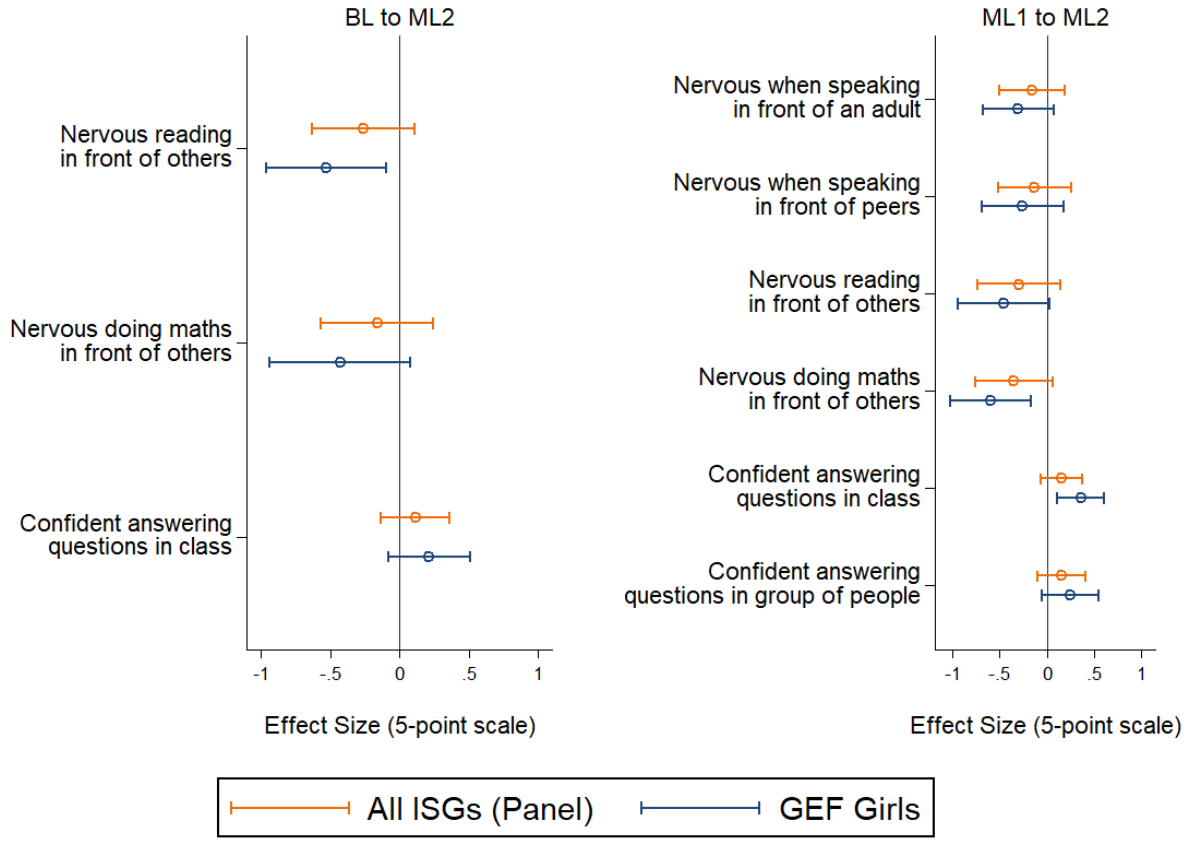
- I get nervous when I have to speak in front of a group of people my age
- I feel confident answering questions when I'm in a group of people
- I get nervous when I have to read in front of others
- I get nervous when I have to do math in front of others
- I feel confident answering questions in class

As noted, these questions are constituent parts of the "learning to learn" sub-index discussed above. Below, we analyse these items individually, to provide a more fine-grained analysis of the program's impact on self-confidence.

We first use difference-in-differences to estimate the program's impact on each of three self-confidence measures from baseline to ML2, among the true sample of in-school girls. Note that each measure uses a 5-point likert scale, from strongly agree (5) to strongly disagree (1).247 Therefore, the direction of "positive" impact varies, depending on the specifics of the measure.

The left panel of the figure below reports findings for the three self-confidence questions that were applied to a large sample of in-school girls at both baseline and ML2. In total, 571 girls are included in this analysis, resulting in a sample of 1,142 observations across the two rounds. Results in orange represent the effect of the intervention on each self-confidence measure -- these results are derived from difference-in-differences, so they can be interpreted as the extent of a change in self-confidence caused by the program from baseline to ML2 among in-school girls. The dot represents the regression coefficient or point estimate -- our best estimate of the program's impact -- while the bar that surrounds it represents the 95 percent confidence interval for that estimate. Meanwhile, the blue dots and bars represent the program's impact among intervention girls who self-report participation in GEF activities, relative to in-school girls in comparison communities. Again, this can be interpreted as the program's impact *among this specific subgroup* of girls.

[247] Throughout this section, we utilize linear regression models. We also checked our finding using ordinal logistic models, which account for the categorical but ordered nature of the outcome variables, but the results are substantively unchanged. We prefer the linear models reported here because they are more directly interpretable.



The findings in the left panel are not entirely surprising, given our earlier findings regarding the "learning to learn" sub-index. Nonetheless, these results are instructive, because they make clear the substantive effect size and the differential impact of GEF participation. The program reduces the mean score for girls expressing nervousness while reading in front of others by 0.26 points (on a 5-point scale), among in-school girls. Among GEF girls, this reduction is starker, at 0.53 points, and statistically significant ($p = 0.02$; in the full sample, the p-value is 0.18).

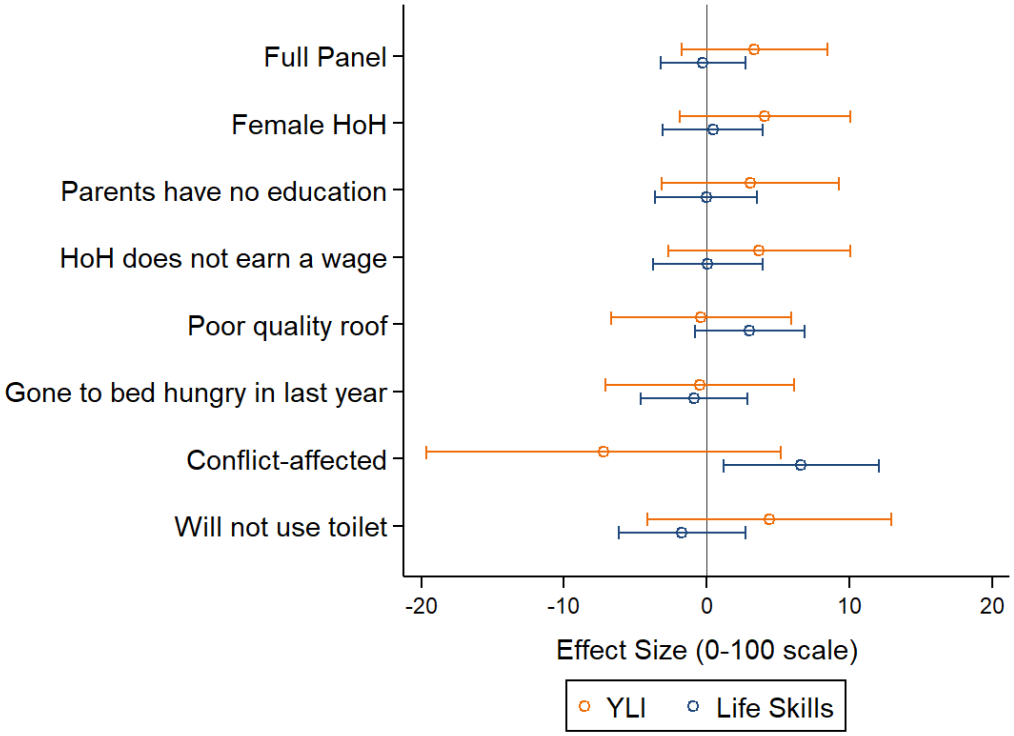
Subgroup Analysis of Life Skills

Thus far, the analysis of life skills has largely focused on aggregate impacts of the program on varying measures of life skills, and different methods of parsing life skills into specific categories (e.g., self-confidence, agency, and so forth). The only exception to this focus on aggregate impact is our analysis of GEFs and their apparent role in promoting self-confidence. In this section, we disaggregate program impact on life skills into relevant subgroups, as we have done throughout this report. We do not report findings for every possible measure of life skills, focusing instead on the YLI and "life skills index" measures described above.

A number of individual and household characteristics – as well as exogenous events – can theoretically impact self-confidence and leadership skills. However, our interest in this section is primarily on subgroup-specific impacts of the program. If the program has had outsized effects on one or a few subgroups, this can help shape programming going forward or suggest methods for distributing impact more evenly.

The figure below reports our core results. For each subgroup specified along the y-axis, we estimated a difference-in-differences model analogous to our primary models of program impact, but limiting the sample

to girls – in both intervention and comparison schools – who fall into that subgroup. The results plotted are the impact estimates, with their 95 percent confidence intervals, for the two main measures of life skills. It is important to note that some of the subgroups studied have relatively small sample sizes; however, we have excluded theoretically-relevant subgroups that have too few observations for fruitful analysis.



The results in the figure make clear that the program's impact on YLI scores has been only somewhat broad-based. While parental attributes – such as the occupation of the head of household – do not appear to change the overall impact we estimate in the full sample (i.e. a substantively meaningful, but statistically insignificant effect on YLI), the same cannot be said for groups that are economically marginalized. Among households with a poor-quality roof or who have gone to bed hungry several or more nights in the past year, the program's impact on YLI scores is almost exactly zero. This suggests that the program's impact has been concentrated among girls from relatively advantaged – by local standards – backgrounds. When we analyse the inverse groups: those with a higher-quality roof and with no experience of hunger in the last year, we find that the impacts of the program on YLI scores are large, positive, and marginally significant.

A group excluded from the figure above are girls with disabilities. We did not include them in the models reported above, partially because the sample size of girls with disabilities is fairly low. Using the same subsample as the regression models reported above, there are just 40 girls with reported disabilities at baseline. Despite this small sample size, there is a clear gap in YLI scores between girls with and without disabilities in the baseline data, with the former group scoring about 12.7 points lower on the index than the latter (equivalent to 0.73 standard deviations, a sizable difference). At ML2, this gap had declined markedly, to just 2.6 points, with the vast majority of the change caused by an increase in YLI scores among girls with disabilities, and only a small decrease (3.5 points) in the mean YLI score of girls without disabilities. Again, while the sample of girls who were tracked from baseline to ML2 for life skills assessment includes just 40 girls with reported disabilities at baseline, this shift is substantively meaningful in any sample size. Note,

however, that the program does not appear to have had an outsized impact among this subgroup – rather, girls with disabilities have seen large-magnitude improvements in their YLI scores relative to girls without disabilities across *both* intervention and comparison schools, though parsing the sample this extensively risks drawing tenuous conclusions from too little data.

Beyond these substantive subgroups, it is also possible that program impacts are concentrated exclusively among groups with the lowest starting points of life skills, or among a particular age group. To study the former, we estimated regression models after limiting the sample to girls in different age brackets at baseline (all girls are part of a panel, so the age distribution stayed the same from baseline to ML2, but shifted two years upward). For instance, we estimated program impact among girls aged 10-11 years at baseline, and – likewise – girls aged 12-13 years. No obvious pattern emerged from this analysis – program impacts were largest among the 12-13 and 16+ year old groups, but were null among the 10-11 and 14-15 year cohorts. If age was a determining factor in shaping program impact, we would expect a monotonic relationship between age and impact estimates, but this is not what we find. Rather, the differences across age groups are likely driven by random noise.

Finally, we were also interested in whether the program primarily improved life skills among girls who were already relatively confident, or whether the program was generating "catch up effects", in which girls who are at the bottom of the scale at baseline are brought up to be closer to their peers in terms of life skills. To answer this question, we split the sample into quartiles based on girls' YLI scores at the baseline. The lowest quartile of girls had an average standardized YLI score of 33.8 points, and the highest quartile averaged 78.5 points.

Based on this analysis, it appears that the program has had slightly more impact among girls who were on the upper end of the YLI spectrum to start. However, the differences are relatively minor: among girls with a baseline YLI score of 53.7 points (and an average of 41.1 points), the program drove a 2.1 point increase in YLI scores in intervention communities, over and above any gains made in comparison communities. Among the higher two quartiles – with an average baseline YLI score of 69.1 points – the program drove an improvement of 4.3 points. In general, this seems to indicate that the program had its greatest impact among girls who were already in the top half of the life skills or leadership spectrum. However, even among girls with a lower starting point at baseline, the program still improved their YLI score, just less markedly so.

7.5 Community Attitudes

Positive attitudes toward education are foundational to much of what SOMGEP-T is seeking to achieve in terms of improving enrolment rates, attendance, and enhancing girls' learning. Support from the community is critical to many different avenues of change. Within households, parents who do not support girls' education are unlikely to make the hard financial choices necessary to educate their daughters, nor will they prioritize schoolwork over the completion of chores. Girls whose parents value education but who live in a community where it is not valued may not feel it is worthwhile to continue their schooling, or may feel discouraged by the lack of support – or active discouragement – they receive from adults in the village.

But community attitudes are more critical to the proposed ToC than even this implies. Communities that are pro-education make the work of CECs easier, by contributing their time to engage in awareness-raising, by contributing funds to support students through the CEC, and by contributing materials for improving the school. They make efforts to recruit better teachers through the use of financial incentives. In reality, pro-education views in the community are, in many ways, a prerequisite for other aspects of program impact.

Our assessment of community attitudes draws from a range of both quantitative and qualitative respondents, with the goal of shedding light – first – on whether the program has improved community attitudes, and – second – on how attitudinal barriers to girls' education manifest in practice. In terms of quantitative indicators, we draw from responses given by caregivers, who are taken to represent the community more broadly; head teachers, whose opinions of community attitudes are indirect and subjective but, nonetheless, informative; and girls themselves, who report the extent to which they feel supported by their parents.

Following the previous midline report, we first report attitudes among caregivers. As part of the household survey, caregivers were asked a short set of questions designed to understand their attitudes toward girls' education. First, they were asked whether girls' education is a worthwhile investment, even in a context of limited financial resources. Second, they were asked whether girls are just as likely to use their education as boys, a measure designed to determine whether parents see equal extrinsic value in educating their daughters and sons. Third, they were asked how decisions around their girls' education are made, and whether they consider the opinions of the girl in making those decisions.

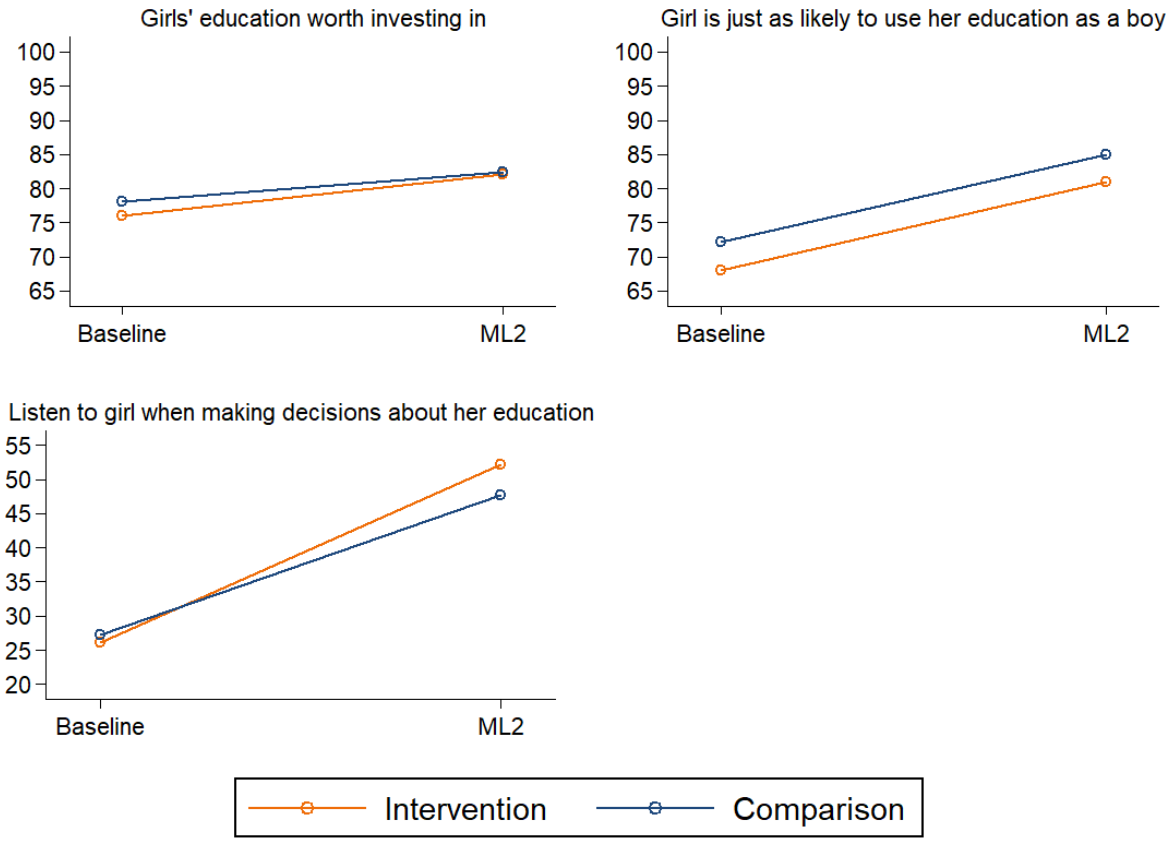
As with our analysis throughout this report, we employ a difference-in-differences model to analyse program impact, focusing on the baseline and ML2 data. We lean heavily on the "true panel" of caregivers, based on the same logic we have discussed previously – by focusing on the exact same set of respondents over time, we avoid introducing bias due to the replacement process or differential attrition.

The figure below reports the trends in caregiver answers over time, disaggregated by intervention and comparison groups; this figure is a visual representation of the difference-in-differences models reported later in this section. The upper-left panel of the figure reports the share of caregivers who believe a girls' education is worth the investment. In many ways, this question addresses the crux of educational attainment in this region: while it is easy to value girls' education in the abstract or in principle, real-world parents face hard trade-offs between paying for their daughter's education, paying for their son's education, feeding their families, maintaining their livestock herd or other source of livelihood, paying medical bills, and so forth.[248] As we discuss briefly below, financial hardship is a central refrain from qualitative interviewees, who suggest that financial constraints are the most important, or among the most important, barrier to enrolment.

As the figure shows, there has been a small increase since baseline in the share of caregivers who believe girls' education to be a worthwhile investment, and this overall increase is matched by a small increase in support among intervention communities, relative to comparison communities.[249] The upper-right panel also shows a general trend toward greater valuation of girls' education, but there is no evidence that the program has contributed to this increase, given the intervention and comparison communities have evolved similarly. Finally, we also document a relatively small but positive impact of the program on the share of caregivers who report factoring their daughter's preferences into decision-making around her education.

[248] The question is also related to the economic view of education as a long-run investment that does not receive a higher level of investment because the returns are significantly delayed, are uncertain (the child's education may not improve their life prospects or those of their parents), and – in some cases – not shared with the individuals (parents or caregivers) making the investment.
[249] In general, there are large improvements from baseline to ML2 in both intervention and comparison communities across all three metrics. In the most extreme case, the share of caregivers indicating their daughter has input into decision-making around her own education has risen from 26.6 percent at baseline to 50.1 percent at ML2 among the exact same set of caregivers.

FIGURE 42: EVOLUTION OF CAREGIVER ATTITUDES TOWARD GIRLS' EDUCATION, BY INTERVENTION STATUS



The results in the graph encouraging for two reasons. The first is that attitudes toward girls' education appear to be improving across all communities, not just those which have received benefits from SOMGEP-T programming. If this is part of a wider shift in attitudes, it will promise significant improvements for girls across the region, not strictly in communities targeted for educational interventions. The results also show that the program has had an outsized impact, over and above that observed in comparison communities.

There is reason to be cautious regarding the results of these measures, however. Elsewhere in this report, we have noted the issues of social desirability bias in attitudinal questions of this kind; we also note that attitudes do not translate directly into changes in behaviour, especially in a context of resource scarcity. More importantly, the measures do not address differential attitudes toward high- and low-performing girls. Caregivers may believe that education is a worthwhile investment for girls, as a general rule, but wish to remove their daughter from school when she performs poorly – in which case, the justification might be that investing in this *specific* girl's education is not a good use of resources. This attitude is not explicitly expressed in any of the qualitative interviews. It is present, though, when CEC members, mothers, or teachers are asked about the performance of girls versus boys, and their answers fixate exclusively on the fact that the school's

top students are often girls.250 This positive statement about high-achievers does not reflect the reality for most girls and may overstate the gains made in girls' education. Along the same lines, focusing on the benefits that may accrue to a girl or her community when she is educated (e.g., she will succeed in life, or bring benefits to the community) is certainly a positive outcome, but is less desirable – as an end goal – than a community that intrinsically values education.

The quantitative data lend mild support to the idea that caregivers of higher-achieving girls are more likely to – in principle, when asked whether it is worthwhile – support investing in girls' education. Simply to illustrate this relationship, among girls who scored below 20 percent, combined, in Somali literacy and numeracy, 77.7 percent of caregivers strongly agreed that it was worth investing in girls' education, despite limited funds. Among girls who scored 80 percent or higher, the share of caregivers who strongly agreed rose to 86.4 percent. The relationship between girls' performance is caregiver attitudes is extremely sensitive, however, to the round and sample employed.[251] Nonetheless, combined with insights from the qualitative data, it is clear that some disconnect exists between investment in a specific girls' education and investment in girls' education as an abstract principle.

In order to better assess the magnitude of program impacts, we estimated a series of regression models that provide a difference-in-differences estimate for each outcome. As the results in the table below show in the context of our aggregate models, the program has produced small but positive impacts in two of the three indicators, though none of the effects can be distinguished from a null result.

In addition to aggregate findings, we also report findings for subgroup analyses – the impact of the program on caregiver attitudes among the in-school and OOS girl cohorts (i.e. caregivers of girls who were in school or out-of-school at baseline). Interestingly, we find that, for the two outcomes in which aggregate impacts were highest, the impact is more positive among the cohort of OOS girls. This suggests that the program has had its most substantial impact on caregiver attitudes among caregivers who daughters were previously out of school, likely the best target for attitudinal change, as the goal of the program is to increase enrolment and positive transition outcomes.

Unfortunately, none of the results reported in the table are statistically significant, despite effects sizes – at least in the case of the final outcome, girls' input into decision-making – that are substantial. Although the sample size available for analysis is relatively large, the outcome itself and the effect of the program is simply too noisy to statistically identify the impact. In light of this issue, we also estimated results using the full cross-sectional sample from baseline to ML2, expanding the sample beyond the true panel to include girls who were replaced and their replacements. In this sample, the results are more consistent, ranging from effect sizes of 2.4 to 4.0 points (on a 100-point scale) for each outcome. However, none of these results obtained statistical significance either.

[250] For instance, when asked whether their daughter was doing well in school, one respondents discussed the fact that girls' education in Somalia is improving, citing female university graduates, news reporters, and government ministers (FGD with Mothers, Galmudug, Int. 311).

[251] Of course, the data do not support the claim that higher achievement by girls changes their caregiver's attitudes, which this analysis implies, because we would expect girls with more supportive caregivers to perform better in school partially as a result of that support. In short, it is very difficult to disentangle causation in this case, and this may be a point worth further targeted investigation as part of the endline or as part of other GEC-T evaluations.

TABLE 101: REGRESSION ANALYSIS OF PROGRAM IMPACT ON CAREGIVER ATTITUDES

Regression Model Description	Impact Estimate (Regression Coefficient)	Standard Error	Sample Size
Strongly agrees girls' education is a worthwhile investment, even when funds are limited			
Aggregate sample	1.9	5.4	2008
In-school girls only	1.9	5.4	1278
OOS girls only	5	8.1	730
Strongly agrees girls are as likely to use their education as boys			
Aggregate sample	0.2	5.6	2008
In-school girls only	1.6	6.3	1278
OOS girls only	-2	8.5	730
Girls' views are considered when making decisions regarding her education			
Aggregate sample	5.5	6.5	2008
In-school girls only	1.4	6.8	1278
OOS girls only	9.2	9.1	730

It is worth noting that, insofar as the analysis above demonstrates small positive impacts of the program, they seem to be driven by changes between baseline and ML1, with limited impact from ML1 to ML2. For instance, when we limit our analysis to ML1 and ML2 only, we actually find that the program has had a *negative* impact on caregiver attitudes. Likewise, when analyse the full sample, across three rounds, we see a sharp improvement in caregiver attitudes in intervention communities, vis-à-vis comparison communities, from BL to ML1, and no discernible improvement from ML1 to ML2. This is not a criticism of the project – the large gains made from BL to ML1 are still valuable, and we view the program as having a net positive impact on caregiver attitudes, even if relatively small, since baseline.

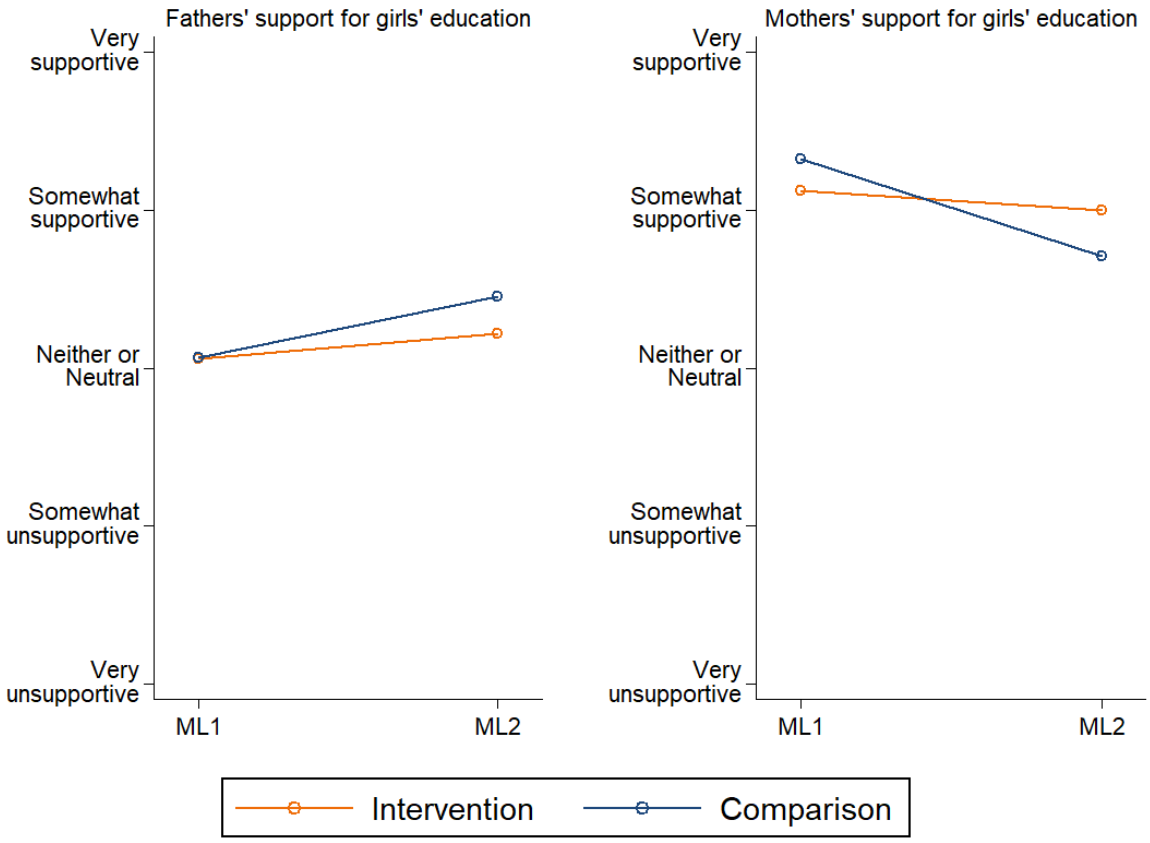
These mixed and inconclusive results continue when we consider the viewpoints of head teachers. We asked head teachers a number of questions gauging their perception of attitudes among the community in which they serve. It is important to note the limitations of this approach, as it relies entirely on the subjective perceptions of head teachers, rather than the actual attitudes of community members. However, we consider head teachers well-informed observers of their communities, especially in relation to attitudes toward education. They are decidedly not objective observers – they likely have a viewpoint that education is unambiguously good – but their views of their communities (and, especially, changes in their views over time) are useful for triangulating community attitudes.

To start, we asked head teachers to rate the support for girls' education shown by fathers and mothers, respectively, in their communities. Head teachers were not faced with this question at baseline, so our analysis focuses exclusively on changes from ML1 to ML2, as shown in the figure below. Head teachers rated support levels on a 1-5 scale, ranging from very unsupportive to very supportive. For the sake of this analysis, we

treat this scale as a continuous, rather than ordinal, scale and analyse the mean support levels reported by head teachers.[252]

As the figure shows, head teachers in intervention schools viewed fathers in their communities as slightly more supportive of girls' education at ML2 than at ML1. However, this small improvement was surpassed by the shifts seen in comparison communities. On the other hand, reports from head teachers imply a positive program impact on support for girls' education among mothers: both groups of communities show a decline from ML1 to ML2, but the decline was less steep in intervention communities.[253]

FIGURE 43: HEAD TEACHERS' PERCEPTIONS OF COMMUNITY SUPPORT FOR GIRLS' EDUCATION, OVER TIME AND BY INTERVENTION STATUS



[252] As with our other analysis, we study only the head teachers who were contacted at both ML1 and ML2, limiting the sample size to 63 head teachers or schools in each of the two rounds. It is worth noting that head teachers are not static positions, and some head teachers may have been newly hired since the ML1 data collection round. This sample is consistent across rounds with respect to the schools visited, but not necessarily the respondents to which field teams spoke.

[253] One should not place too much weight on the fact that perceived maternal support for girls' education has declined. As noted, head teacher reports are subjective, and the set of head teachers changed over time. It is possible that newer head teachers have more stringent standards for considering mothers in their community "supportive." While this would *seem* to imply that the same trend should occur among fathers, it is also possible that head teachers hold mothers to a higher standard of support than they do fathers. In any case, our focus is on differential trends that represent program impact, not the direction of the trend itself.

While the results for mothers are not statistically significant, this is primarily a function of the limited sample size, and not an indictment of the effect size itself. We estimate that the program has generated a 0.49 point change in perceived maternal supportiveness, on a 5-point scale. This represents a change of approximately 0.29 standard deviations, typically seen as a substantively meaningful effect size.[254]

Finally, we also presented head teachers with a hypothetical scenario describing a boy and a girl, in separate vignettes, who has achieved high marks in school and has been accepted to university. The child's family is presented as unable to pay the cost of university; head teachers were asked the likelihood that the community would rally to support the child and raise the money necessary for them to continue their education. The fundamental idea of this question is that it moves beyond attitudes alone, and asks head teachers to assess *which members* of the community other community members would be willing to sacrifice to support.

The results for this set of questions are reported in the table below. Each head teacher was asked to rate the likelihood of community support in the context of a girl and a boy. As the table shows, head teachers increasingly believe that communities would be able to raise funds to support either a boy or girl to attend university. Gains are seen most clearly in the context of girls – the share who believe that their community is very likely to support a girl in this way has risen from 23.8 percent to 38.1 percent since ML1. Of course, these measures do not imply communities actually *could* provide such support – we expect most head teachers overestimate the likelihood of such support – however, the trends are meaningful, regardless.

TABLE 102: PERCEIVED COMMUNITY SUPPORT FOR BOYS' AND GIRLS' EDUCATION, ACCORDING TO HEAD TEACHERS

Likelihood of Raising Funds to Support Child's Education	ML1		ML2	
	Girls	Boys	Girls	Boys
Very Likely (4)	23.8%	57.1%	38.1%	63.5%
Somewhat Likely (3)	27.0%	22.2%	20.6%	12.7%
Somewhat Unlikely (2)	20.6%	1.6%	17.5%	3.2%
Very Unlikely (1)	28.6%	19.1%	23.8%	20.6%
Mean (1-4 Scale)	**1.46**	**2.17**	**1.73**	**2.19**

Unfortunately, the program itself does not appear to have driven these changes, at least insofar as the difference-in-differences is able to isolate program impacts. We calculated a measure of perceived "preference toward boys" – the extent to which head teachers believed their community favoured boys over girls – and

[254] On the other hand, the negative result we report for fathers may be taken as offsetting the improvement among maternal attitudes. Of course, this is only a single measure of paternal and maternal attitudes. We also analysed data in which head teachers were asked whether mothers, fathers, or both parents would attend a meeting with their daughter's teacher, if one was called. Head teachers were asked to imagine a hypothetical set of parents and a hypothetical meeting, in order to gauge whether they think a typical father would make a tangible effort – in the form of sacrificing their time – to support their daughter's education. Very few head teachers (just 7.6 percent across both evaluation rounds) believed that a father would attend such a meeting. However, the share who believe they would has increased in intervention communities, relative to comparison communities, suggesting a subtle and small shift in perceived support for girls' education among fathers. Again, this finding contradicts that reported above, suggesting that no major shifts in fathers' support – or head teachers' perception of it – have occurred.

studied the impact of the program on this metric. Ultimately, the program appeared to have a small negative effect, where a negative effect refers to an increase in net preferential treatment toward boys.

Our final quantitative metric of community attitudes toward girls' education come from an entirely different source: the girls themselves. As with head teachers, girls are not entirely objective barometers of public opinion. Moreover, they are less astute observers of public opinion than head teachers. On the other hand, to some degree girls' perceptions of community attitudes is quite proximate to that which actually interests us. Specifically, if girls feel their community supports their education, that is meaningful in and of itself; even if it does not accurately reflect community attitudes, perceived positive attitudes likely have benefits for girls' self-esteem and interest in continuing their education. And, in the end, we expect girls to have some insight into true public attitudes, at least among the set of community members they encounter and with whom they interact.

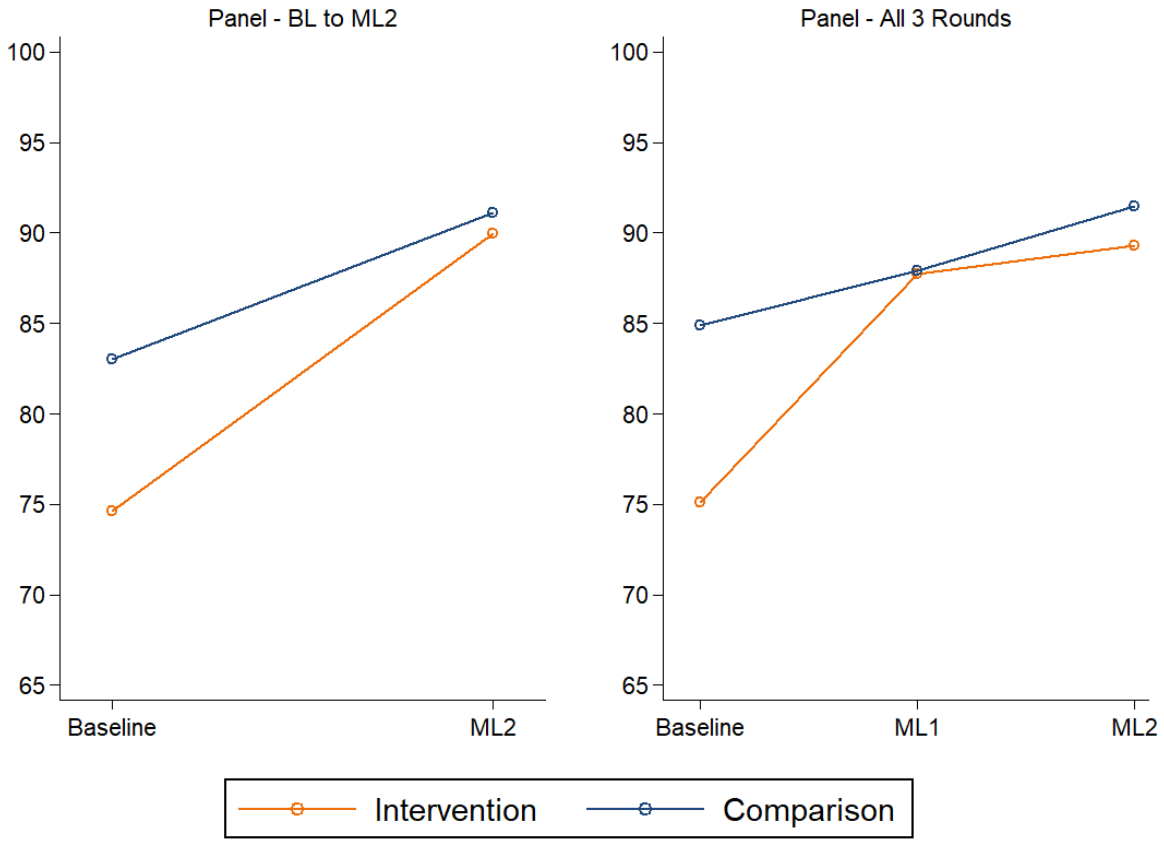
The indicator we study with regard to girls is the extent to which they feel they receive the support they need from their parents to continue to attend school and perform well. Girls were given four response options, ranging from "agree a lot" to "disagree a lot", with no neutral option provided ("don't know" was an option, though few girls availed themselves of it).

The results of our analysis are shown in the figure below. This question was originally targeted exclusively at girls who were enrolled in school, as the premise assumes the girl is enrolled, so our sample consists of in-school girls who were successfully contacted at both baseline and ML2 (left panel) or baseline, ML1, and also ML2 (right panel). That is, the right panel includes only girls who appeared in all three rounds of data collection. The figure reports the share of girls who strongly agree that they receive the support they need from their parents.

The figure shows a marked increase in the support girls feel they receive from their parents from baseline to ML2, an increase which is especially pronounced in intervention communities. In comparison communities, the share of girls who strongly agree that they receive sufficient support from their families increased from 83.0 to 91.1 percent, an increase of 8.1 points. In intervention communities, over the same time period, the share of girls who strongly agree rose from 74.6 percent to 90.0 percent, or a 15.4 point increase. In total, we estimate the program has increased the share of girls who feel supported by their parents in this way by 7.2 points, over and above the improvements in comparison communities.[255]

[255] This finding is not statistically significant, but it is one of the larger effect sizes we document in this report, and its p-value is lower ($p = 0.18$) than those of other marginal effects.

FIGURE 44: SUPPORT GIRLS PERCEIVE FROM THEIR FAMILIES, BY INTERVENTION STATUS



Beyond this aggregate impact, we were interested in whether girls whose households were particularly disadvantaged financially saw similar gains. The logic of this analysis was that girls in disadvantaged households were less likely to feel supported, because their continued schooling may seem tenuous, dependent on their parent's ability to pay their school fees or otherwise continue to sacrifice to keep them enrolled. Our hypothesis was that the program might have outsized impacts specifically among this group, because it might have eased the financial burden of marginalized families, resulting in increased perceptions, among girls, of support from their families.

We studied subgroups of girls whose caregivers reported that they had reduced their food expenditure sometime in the last three months, gone without food an entire day in the last 30 days, or reduced the number of meals they ate sometime in the last 30 days. In the end, we classified a family as economically marginalized if they had experienced any of these three metrics of hardship (n = 238 when we consider only girls in the panel who were asked all the necessary questions). When we analysed the program's impact on perceptions of support among girls whose families had experienced hardship, we found a 9.6 point impact; when we perform the same analysis among the sample of girls whose families had not experienced such hardship (n = 343), the estimated impact of the program declines to 5.1 points. While the difference between these estimates is not statistically distinguishable from zero, the gap is suggestive: tentatively, the program may have increased the extent of support girls believe they receive from their families especially strongly in families who face more severe resource limitations.

The available qualitative data reinforce many of the themes that emerged in our analysis above. However, they also paint a picture of communities in which attitudes have shifted somewhat, but where much improvement is still needed.

The first theme that emerges from the qualitative interviews is the importance girls place on support for and demand for education. Girls participating in participatory story-telling exercises emphasized repeatedly that a girl's future was heavily dependent on her parent's desire for her to pursue education. In the story-telling exercises, the moderator described a hypothetical girl, her family, her educational background, and a problem or scenario that the girl faced, such as wanting to continue her schooling but being simultaneously responsible for household chores. Participants tended to be optimistic about a girl's prospects, especially if her parents supported her. For instance, one participant insisted that a girl could continue until university "because her mother wants her to continue to study."[256] Girls also tended to focus on the role that teachers and friends can play in encouraging girls to remain in school, which capture a slightly different form of pro-education attitudes.[257] Girls' belief that pro-education attitudes matter also extended to their own attitudes. That is, girls seem to place a lot of weight on girls *wanting* to stay in school. For instance, some participants cited the hypothetical girl's love of education as a reason why she will remain enrolled[258]; on the other side, participants believed a girl who was not particularly interested in education was much more likely to drop out.[259]

With respect to attitudes of the broader community, the qualitative evidence is decidedly mixed. In some FGDs with CEC members, the members reported that members of their community actively opposed education for their children.[260] One CEC indicated that parents actually resent the actions their committee was taking to raise awareness and encourage enrolment, accusing CEC members of being self-interested, as parents suspected the CEC members were compensated based on the number of students they enrolled.[261] While this is, admittedly, an extreme case, other qualitative interviewees also reported that some members of their communities do not appreciate the value of education. One mother described her community: "There are people that don't want the school and are against studying. They say that the school is bad and spread untruths around for people."[262] Others report a lack of active support or a lack of understanding of the importance of education, including among nomadic families[263]; in contrast to the egregious opposition to education cited above, these reports are slightly less concerning, but still indicate that community attitudes are far from universally pro-education.[264]

Of course, many participants cited positive attitudes within their communities.[265] Girls occasionally reiterated this finding: one set of girls in Puntland were neutral about a hypothetical girl's prospects for staying in school; however, when they were asked how the girl would fare if she lived in their village, their responses became much more positive. One girl said "[Barwaaqo, the hypothetical girl] will get support from the community and continue her education."[266] Our interpretation of the qualitative evidence is that a majority of participants viewed their community as broadly supporting girls' education. However, our view is that qualitative

[256] Vignettes FGD with girls, Puntland, Int. 241.
[257] Vignettes FGD with girls, Puntland, Int. 242; Vignettes FGD with girls, Somaliland, Int. 141.
[258] Vignettes FGD with girls, Puntland, Int. 242.
[259] Vignettes FGD with girls, Somaliland, Int. 142.
[260] FGD with CEC members, Somaliland, Int. 104.
[261] FGD with CEC members, Galmudug, Int. 301.
[262] FGD with mothers, Galmudug, Int. 311.
[263] FGD with teachers, Puntland, Int. 231.
[264] FGD with CEC members, Somaliland, Int. 104.
[265] FGD with CEC members, Puntland, Int. 201; FGD with CEC members, Somaliland, Int. 102; FGD with teachers, Galmudug, Int. 331; FGD with mothers, Galmudug, Int. 311; FGD with teachers, Puntland, Int. 231.
[266] Vignettes FGD with girls, Puntland, Int. 242.

interviewees tend to underreport negative attitudes or actions. If a few qualitative interviewees cite problems such as lack of support for education, we interpret this to reflect a relatively broad trend – not in the sense that *most* community members oppose girls' education, but that opposition exists and is probably more widespread than many participants will admit or more common than they are aware.

A similar result obtains when we consider religious leaders. Our qualitative interviewees – especially CEC members – were asked whether religious leaders in their community support education. In general, most respondents indicated that religious leaders were supportive, sometimes extremely supportive. One CEC member reported that their local religious scholars "give speeches to girls about the benefit of girls' education."[267] Another CEC member described the composition of their committee, which included Quranic teachers and religious scholars, their dual role implying their support for girls' education.[268] Other CECs reported that religious leaders make scriptural arguments in favour of girls' education.[269] Unfortunately, there was at least one instance in which religious leaders were not viewed as supportive of girls' education. Specifically, CEC members in Galmudug reported that "religious elders believe if students regularly focus on academic achievement, they might also forget to follow their Islamic culture."[270] This finding was not widespread, but may hint at pockets of opposition to girls' education among religious scholars in some areas.

In addition to the relatively mixed evidence above, a broader problem may be the general lack of an accurate reference point for what constitutes "pro-education attitudes" or actions that support girls' education. By this we mean that girls, their parents, and community members may perceive themselves as supportive of girls' education without fully understanding the manner in which they obstruct girls' education, reifying existing barriers. For instance, one mother reported that girls had begun attending school more often because they saw how "the community wants to educate boys", so girls began attending school of their own volition.[271] While the girls who enrolled under these circumstances should be commended, girls' motivation should not be derived from experiencing gender bias. Some parents made efforts to spare their daughters the extra burden of completing household chores.[272] At the same time, others claimed to support girls' education while describing the chores they asked their daughter to complete before going to school:

> *Yes, I wake her up early in the morning and she makes us breakfast and I tell her to be fast when she is getting dressed and that's how I help her.*[273]

The disconnect between abstract support for girls' education and tangible actions to support girls' education is also supported by the quantitative data. At the time of ML2, 66.4 percent of caregivers – across all cohorts of girls – reported that their girl spends either a half day (49.5 percent) or the entire day (16.9 percent) completing various household chores, helping with agricultural work, and so forth. We might expect caregivers that express the strongest support for their girl's education would seek to reduce this chore burden; however, this is not the case, in general. Among caregivers who strongly agree that a girl's education is a

[267] FGD with CEC members, Somaliland, Int. 102.
[268] FGD with CEC members, Somaliland, Int. 102; FGD with CEC members, Puntland, Int. 201.
[269] FGD with CEC members, Puntland, Int. 201; FGD with CEC members, Somaliland, Int. 107; FGD with CEC members, Somaliland, Int. 105.
[270] FGD with CEC members, Galmudug, Int. 301.
[271] FGD with mothers, Galmudug, Int. 311.
[272] FGD with mothers, Somaliland, Int. 116.
[273] FGD with mothers, Somaliland, Int. 118.

worthwhile investment, even when funds are limited, 69.1 percent report their girls complete a whole or half day of chores, a rate that is higher than that when caregivers express less emphatic support for girls' education. A similar result obtains when we investigate chore burden among girls whose caregivers say that girls are just as likely to use their education as boys. Overall, the self-reported value that caregivers place on girls' education does not seem to translate into fewer chores, which may reflect a lack of understanding about how chores impede girls' education or a lack of willingness to alter household dynamics and the household economy to concretely support girls' education.

We do not wish to minimize the seriousness of the trade-offs families face and the decisions they must make. In practice, households have to make choices that will influence their survival and economic security. The lack of financial resources is consistently raised as a key barrier to girls' education, because it shifts those trade-offs against girls' education. Mothers and girls alike recognised the importance of financial resources.[274] One girl identified the precise decision – which children to educate – many parents have to make, and indicated that an elder sister would elect for her brothers to be enrolled in school, rather than herself.[275]

Money is never far from the surface when discussing barriers girls face. One girl put it bluntly, when considering a hypothetical story about Barwaaqo and her family:

> *I think that her future will be destroyed since she is poor.*[276]

Other circumstances, such as a nomadic or pastoralist lifestyle, impose similarly difficult decisions for parents.[277] As one teacher described it, "Most of the nomadic girls love education , however when their families move to different areas, the girls will drop out of school because they won't get people who will help them get textbooks , pens ,school uniforms and housing. They therefore drop out and give up on education."[278]

The economic barriers households face are not surprising. The critical issue is that many of these barriers force choices that are often decided in favour of boys in the household, to the detriment of girls. The girls themselves recognize this, and may have internalized it, as with the participant cited above, who stated that many girls drop out so that their younger brothers can enrol in school. Parents and others occasionally recognise it outright; in other cases, parents believe that they are being fair to their daughters, while citing household chores they must complete before going to school.[279] This relates to the analysis above, which showed that head teachers still perceive significant bias in the community toward boys' over girls' education.

The conclusions that should be drawn from this analysis are not immediately obvious, but reveal a consistent set of findings with further review. Across a number of indicators and quantitative data sources, we documented a small improvement in community attitudes from either baseline or ML1 to ML2. This result was not universal, but was consistent across most metrics, including measures of caregiver attitudes and the

[274] FGD with mothers, Galmudug, Int. 311; Vignette FGD with girls, Somaliland, Int. 141.
[275] Vignette FGD with girls, Puntland, Int. 242.
[276] Vignette FGD with girls, Somaliland, Int. 141.
[277] FGD with teachers, Somaliland, Int. 134; FGD with teachers, Galmudug, Int. 331.
[278] FGD with teachers, Somaliland, Int. 134.
[279] FGD with mothers, Somaliland, Int. 118.

perception of girls themselves. None of your analysis documented substantively meaningful *negative* effects on community attitudes.

The qualitative data shed only limited light on changes in community attitudes over time. But it is clear from the data that most community members support girls' education, while a minority of community members are either ambivalent or vehemently opposed to girls' education. Whether this latter group is sufficiently large to influence girls, their self-esteem, and their dedication to education is an open question. But the communities studied are clearly not universal bastions of pro-girls' education attitudes. Further effort needs to be made to emphasise that girls' and boys' education are equally important, and that, when faced with real-world trade-offs, the default outcome should not be that boys attend school while girls drop out.

7.6 SCHOOL-RELATED GENDER-BASED VIOLENCE

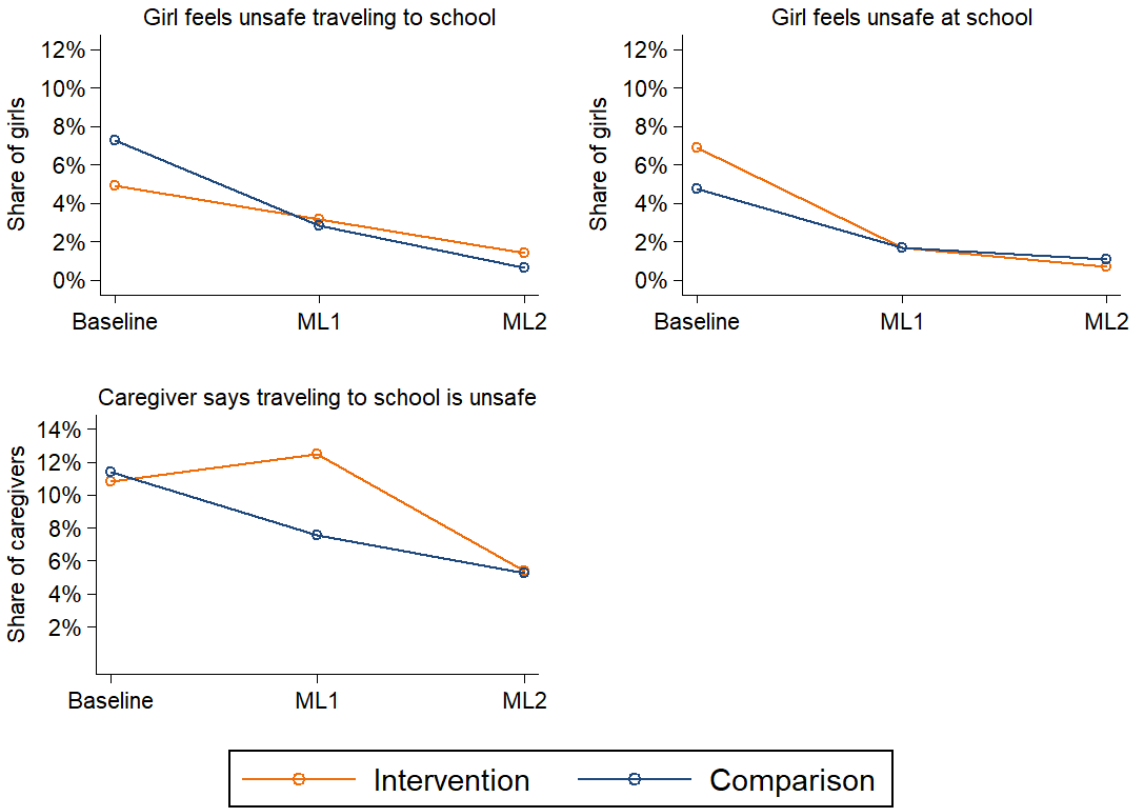
The final intermediate outcome we consider is gender-based violence. Safety and security at school and on the journey to school is an essential component of improving attendance and retention, because – as we will discuss in greater detail below – safety is a major motivating factor in girls remaining out-of-school.

Our focus in this section is on safety and security, somewhat broadly construed. In this, we mirror the previous midline evaluation, which was primarily concerned with gender-based violence, including harassment and other forms of violence, but which was not limited exclusively to violence that occurred in the school. We present results concerning general safety concerns, even if the violence is not explicitly gendered. Given the context in which SOMGEP-T is being implemented, a broader view of violence and security is useful for highlighting the types of problems girls face while trying to complete their education. We also cast a fairly wide net because it is difficult to ask girls direct questions regarding many forms of violence, especially sexual violence, and because indirect questions often elicit somewhat indirect answers. Using these responses requires some degree of contextual knowledge and interpretation; we tend to be cautious in our interpretation, but this means that our discussion is often around topics such as harassment or general feelings of insecurity, rather than very specific acts of violence or potential violence.

Our analysis is oriented heavily toward qualitative data, but we attempt to triangulate the information collected against quantitative data wherever possible. Often, this triangulation requires use of quantitative data in slightly indirect ways.

At the broadest level, there is a trend toward increased perceptions of safety at school, and on the journey to school, from baseline through the ML1 and ML2 rounds of data collection. Both girls and caregivers were asked questions regarding safety at school and on the journey to school, which we report in the figure below. Girls were asked a dichotomous question regarding whether they feel safe traveling to and from school (upper-left panel) and a similar yes/no question regarding whether they feel safe at school (upper-right panel). These graphs below report the share of girls who stated that they *do not* feel safe in those two situations. The results show a consistent and substantial decrease in the share of girls who feel unsafe at school or on the way to school, from baseline through the present. As the graphs also make clear, we cannot attribute this impact to the program itself, because we observe similar downward trends in both intervention and comparison schools. However, the downward trend is positive regardless of its source.

FIGURE 45: PERCEIVED SAFETY OF ATTENDING AND TRAVELING TO SCHOOL, OVER TIME

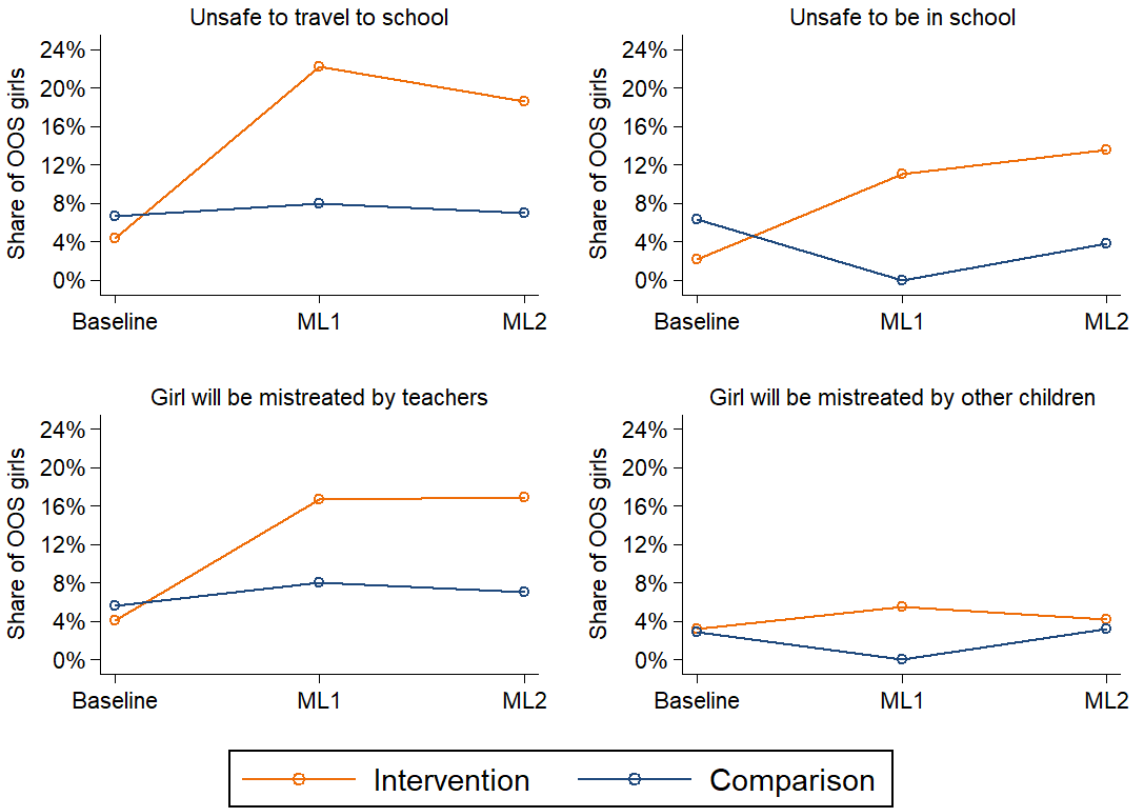


Caregivers were asked to rate the relative risk girls face when traveling to school, which we report in the lower-left panel. We interpret any response other than "very safe" as an implicit statement that it is not entirely safe for a girl to travel to schools in the area, and we report that share of caregivers who imply it is not entirely safe in the figure. As with reports from girls, the general trend is toward increased perceptions of safety over time, a trend that is shared across intervention and comparison groups.

The analysis above is based on a large sample of respondents, as all girls enrolled in school were asked these questions, and all caregivers – regardless of their girls' enrolment status – was asked about safety of traveling to schools in their area. However, we also analysed data from caregivers of out-of-school girls, and the reasons they gave that their girls were not enrolled in school. This analysis produced dramatically different results, as shown in the set of graphs below. Caregivers could give multiple reasons why their girl was not enrolled; we plot the share of caregivers who selected each option, focusing exclusively on responses that express concern about safety, including harassment or mistreatment. The figure below plots the share of OOS girls' caregivers who selected each response.[280]

[280] A significant caveat applies to this analysis. While it is true that the sample size of OOS girls is substantial (199 at ML1, on the low side, up to over 700 at baseline), the sample size shifts dramatically over time. By implication, the sample's composition is changing over time, which could be the reason why reasons given for non-enrolment change so much from round to round.

FIGURE 46: SAFETY-ORIENTED REASONS OOS GIRLS ARE NOT ENROLLED, OVER TIME



In three of the four cases, the share of caregivers who cite a given reason is increasing over time, and also increasing in intervention vis-à-vis comparison schools. For instance, the share of OOS girls whose caregivers cite potential mistreatment by teachers has risen from 4.1 percent at baseline to 16.9 percent at ML2. Similar changes are observed in the share of caregivers citing a lack of safety on the trip to school and at school. In most ways, the results of this analysis contradict those of caregivers and girls reporting their perceptions of safety, which have tended to improve over time.

It is possible that the type of girls who remain out-of-school has changed over time, especially in response to the program's interventions encouraging enrolment. For instance, if the program has prompted increased enrolment, the girls who remain out of school could be those in more conflict-affected or less secure areas. Similarly, the program may have impacted the "low-hanging fruit" of OOS girls, leaving primarily those who are concerned about safety as the core of the remaining OOS girls.[281] In short, if the program has reduced barriers to enrolment that are primarily *not* safety-oriented, it would explain a shift toward greater numbers of intervention-area caregivers citing safety as one of the main reasons their girl is not enrolled.

Our interest in this section is not exclusively in perceptions of safety among OOS girls. The qualitative data provides a wider view of safety concerns and gender-based violence that impacts key project outcomes, such as learning, attendance, retention, and so forth. The main qualitative tool that captures views on safety is a

[281] This should not be considered a negative comment on the program itself, as the "low hanging fruit" of OOS girls are still generally marginalized, and this would still represent positive impact relative to comparison schools.

participatory risk mapping exercise, which CARE developed during the previous midline (ML1) evaluation round. In the ML1 evaluation report, results from this risk mapping figured prominently in the analysis of gender-based violence, and this continues in this round. Of course, we also supplement this discussion with information gleaned from other qualitative interviews, observations of our team leaders in the field, and quantitative data wherever possible.

As in the previous round, one of the most common themes that emerged from the risk mapping was fear of the roads that lead girls to and from school. Participants in all four risk mapping FGDs cited the risk posed by crossing streets, walking along the roads, etc. mainly in terms of the potential for being struck by a vehicle.[282] As one girl described it, "I feel very unsure while passing the road because they drive fast."[283]

Fear of roads is not limited to the potential of being struck by a vehicle, however. Roads can also be open, desolate areas – especially in the types of rural areas where SOMGEP-T is being implemented – which pose the potential for violence. This finding fits in with broader themes from the qualitative data that open spaces or places devoid of people were generally risky.[284] Roads can be risky because they pass through empty spaces, and because they expose girls to harassment or violence from a wide range of people. Girls are sufficiently concerned about travel along the roads that at least one girl suggested teachers and the CEC could improve safety at their school by accompanying them when they leave the school.[285]

A key consideration to keep in mind is the extent to which girls in the communities visited often live among a low, but omnipresent, level of latent violence. Our team leaders reported several stories, which arose from discussions with head teachers and others, about criminal and inter-personal violence in their communities.[286] One girl in the cohort sample had been killed in the previous two years, in the context of a dispute between her father and another man.[287] Another team leader described an incident in which a girl was accidentally hit by a stone during a fight among the boys in a school. Her father brought a gun to the school and threatened the teachers and students until community elders and school staff were able to resolve the situation.[288]

But there is also a sharp distinction in girls' experiences, because broad conflicts are quite localized, in specific regions or even specific villages. Moreover, most villages have areas that are widely viewed as safe and areas that are seen as potentially or definitely unsafe, implying that girls will have different experiences, depending on where they live and attend school. This distinction emerges quite sharply in both the qualitative and quantitative data.

At the overall community level, some communities experienced conflict in the past year that was sufficient to affect schools for weeks or more. In total, 4.6 percent of respondents during ML2 data collection stated that their village had experienced conflict in the past 12 months. But respondents do not all agree on whether their village was affected by conflict; if we consider a village conflict-affected if at least two respondents reported it as such, 14.5 percent of girls in ML2 lived in communities that had been affected by conflict. In

[282] Risk Mapping Exercise with Girls, Puntland, Int. 221; Risk Mapping Exercise with Girls, Somaliland, Int. 121; Risk Mapping Exercise with Girls, Somaliland, Int. 122; Risk Mapping Exercise with Girls, Somaliland, Int. 123
[283] Risk Mapping Exercise with Girls, Somaliland, Int. 123.
[284] Risk Mapping Exercise with Girls, Somaliland, Int. 122.
[285] Risk Mapping Exercise with Girls, Somaliland, Int. 122.
[286] Interview with Fieldwork Team Leader, Somaliland; Interview with Fieldwork Team Leader, Galmudug.
[287] Interview with Fieldwork Team Leader, Somaliland.
[288] Interview with Fieldwork Team Leader, Galmudug.

total, 12 schools had conflicts where at least one respondent stated the school was closed for at least one day as a result. Most of these schools were in the Sool and Sanaag regions. One of the schools most heavily impacted (Dharkeyn Primary) was already discussed in the methodology section above, as it was removed from the ML1 round of data collection due to ongoing clan conflict.

The qualitative data tells a similar story about localization. For instance, some girls perceive their community as quite safe, reporting that "the whole village's safety is good."[289] In contrast, girls in one village described their recent lives as a state of outright inter-communal conflict between two clans that occupy their area.

> *Yes, there was a conflict between two tribes and it did affect us physically and mentally because we lost our education and a lot of family members.*[290]

Girls in this group stated that their fathers and brothers had left the community due to the fighting, leaving women and children in the village alone. One girl stated that "I believe each one of us lost someone in that conflict."[291] Although this conflict has since ended, the insecurity has not entirely left: girls described a military encampment near the school where a man had been accidentally shot, in front of some of the students:

> *There was a gunshot in front of us. There was a man who got hurt. They were playing with the gun and it accidentally fired. One man got shot on his hand, it was very scary.*[292]

The team leader conducting data collection in this village clarified that the military encampment was not *near* the school, but was physically located *inside* the school. Soldiers from the Somaliland military – who were sent to keep peace between the two warring clans – were occupying two rooms in the school, which meant that girls interacted with the soldiers and went to school in the shadow of the conflict.[293]

Even within communities that are not wracked by such widespread conflict, there are areas where girls do not feel safe, where they face harassment or potential violence. As noted above, open spaces are considered risky, though the specifics vary from community to community. One example concerned an area - the telecommunications tower in town, which is mostly empty, and is a place the girls appear to fear.[294] But other areas of concern are specific to the village: some girls report that they feel safe in some parts of the local market, but not in others, and one girl reported a specific shop or set of shops they feared.[295]

---

[289] Risk Mapping Exercise with Girls, Puntland, Int. 221.
[290] Risk Mapping Exercise with Girls, Somaliland, Int. 123.
[291] Risk Mapping Exercise with Girls, Somaliland, Int. 123.
[292] Risk Mapping Exercise with Girls, Somaliland, Int. 123.
[293] Interview with Fieldwork Team Leader, Somaliland. The presence of soldiers can be interpreted in either of two ways: on one hand, the school is likely very safe from inter-communal conflict as a result of the soldiers being stationed there; on the other hand, incidents like that in which a man was accidentally shot expose girls to a different type of violence and insecurity.
[294] Risk Mapping Exercise with Girls, Somaliland, Int. 121.
[295] Risk Mapping Exercise with Girls, Puntland, Int. 221; Risk Mapping Exercise with Girls, Somaliland, Int. 121.

Upon first asking girls to describe where in their communities they feel safe, most girls immediately indicated they felt safe at school. However, digging deeper into their responses, this is far from universally true. One set of girls highlighted the improvements in their security that had been made since their school hired a watchman and constructed a fence around the school, implying that the school was not physically secure previously.[296] Of course, the added security should be viewed as a positive sign. Girls in school also express concern about using the toilets, due to concern about being spied on.[297] Between this concern, the need for a fence and a watchman, and the stationing of soldiers in one school, it is clear that schools are not universally safe areas for students.

Before discussing issues of safety in schools in more detail, it is worth noting the gendered nature of many complaints girls raise about their communities. Outside of roads and cars, and the threat of military or inter-communal conflict, most of the girls' concerns about safety in their community at large center on harassment or abuse by boys and men. For instance, one girl said that she does not like to visit the football pitch where boys play, because they harass girls.[298] The aforementioned shop that girls fear is scary *because of the men there*, who the girl specifically fears. Whether she is concerned simply because they are men who she does not know, or because they frighten her for specific, tangible reasons is unclear, but her concern is specifically about the men in the shop.[299] Another example concerns an area girls in one school pass on their way to school, where men will harass, verbally abuse, or make sexually suggestive comments toward the girls. One girl described her feelings about the area as follows:

> *I don't like where they sell khat because these men harass you and I hate crossing that street because they will look at you.*[300]

Within schools, concerns around safety change – at least insofar as girls are forthright in their answers – from concerns about verbal harassment and outright violence to safety concerns. Infrastructure features heavily here, as girls in two of the four risk mapping exercises described in detail their fear of the wells where they gather water. Girls in both groups indicated that the well has an open surface that is large enough for them to fall inside, and that they fear doing so.[301] As noted previously, concern about the privacy available at the toilets is also common -- while one participant stated explicitly that they fear boys are spying on them, another girl couched a potentially similar concern in terms of her own "shyness."[302] Other girls highlighted the dirty state of the toilets, an issue which was also raised by one of our team leaders in their post-visit school reports and in post-fieldwork debriefing sessions.[303] Given that some schools still lack a girls' toilet with privacy walls, this concern seems to be well-founded, especially since even privacy walls do not prevent purposeful spying

[296] Risk Mapping Exercise with Girls, Puntland, Int. 221.
[297] Risk Mapping Exercise with Girls, Somaliland, Int. 123.
[298] Risk Mapping Exercise with Girls, Puntland, Int. 221. This comment from a single girl in an FGD may be part of the explanation why so many girls report that they do not use the common areas at school where children play and socialize. The share of girls who do not use these areas has declined over time – though this may be due to cohort girls getting older and more confident or less fearful – but is still substantial: 36.4 percent of cohort girls interviewed at ML2 stated that they do not use such areas, though the reason for their choice is not known.
[299] Risk Mapping Exercise with Girls, Somaliland, Int. 121.
[300] Risk Mapping Exercise with Girls, Somaliland, Int. 123.
[301] Risk Mapping Exercise with Girls, Somaliland, Int. 122; Risk Mapping Exercise with Girls, Somaliland, Int. 123.
[302] Risk Mapping Exercise with Girls, Somaliland, Int. 121.
[303] Risk Mapping Exercise with Girls, Somaliland, Int. 121; Interview with Fieldwork Team Leader, Somaliland.

by other students.[304] Other infrastructure-oriented safety concerns centered on the school building itself, as one team leader reported at least two schools in which the ceilings had begun to crumble or fall and represented a risk to students.[305]

With regard to their teachers and other school staff, girls generally do not report any issues. This is actually somewhat surprising, given the rate of corporal punishment that still exists in the sampled schools, and the fact that harassment by teachers is a documented issue in other schools in Somalia. Girls in the risk mapping exercises occasionally expressed fear of their teachers, but mostly in the context of being fearful of punishment for doing something wrong, rather than something more nefarious. This evidence is not dispositive, however, as girls may hesitate to report incidents such as this to a stranger who will soon leave their school. Teachers participating in several FGDs also reported that harassment – in general, not specifically by teachers – has not occurred in their school in the recent past, though it is also not clear the extent to which this evidence is credible and reliable.[306] Corporal punishment, though not necessarily related to harassment by teachers, has gone down in intervention schools, as documented in Section 3 above. The lack of evidence *for* a hypothesis of harassment, the declining rate of corporal punishment, and the fact that girls increasingly feel safe at school seems to suggest the development of a more positive school environment, from the perspective of safety from violence.

Finally, it is worth noting the particular challenges faced by girls with disabilities (GWDs) and the extent to which these may have improved somewhat since last year. Data collection included 13 KIIs conducted with GWDs, who were identified from the baseline and ML1 samples. In the vast majority of these interviews, girls indicated that they were broadly supported by their families to attend school, by their teachers at school, and by their classmates and friends throughout.[307] As one girl described it, her teachers treat her well – "with respect, grace, and dignity"; her family encourages her schooling; and her neighbors inspire and motivate her education.[308]

From a practical and methodological perspective, interviewing GWDs regarding their disability is problematic, because girls with disabilities may not perceive themselves as having a disability, or they may not place as much importance on it as the interview seems to require. For instance, several girls were fairly insistent that they had no physical limitations, and an interviewer who suggests or implies otherwise may be committing a serious ethical violation. Where interviewers may normally push respondents to interrogate their answers more deeply, this is not advisable in this context. Therefore, some girls spoke relatively little about their disability, despite this being the main topic of the interview. On one hand, this reduces the quality of the data that is collected; on the other hand, the fact that many GWDs do not define themselves by their disability is, itself, a finding of consequence.

Despite this generally positive appearance, there is some evidence that GWDs still face discrimination and the potential for harassment. We do not focus, in this section, on the barriers to attendance and enrolment that they face, as this is discussed under the sections relevant to those outcomes. Rather, we focus on the

[304] In the most recent round of data collection, 17.4 percent of schools lacked a dedicated girls' toilet with privacy walls. This figure was not different between intervention and comparison schools.
[305] Interview with Fieldwork Team Leader, Somaliland.
[306] FGD with Teachers, Galmudug, Int. 331; FGD with Teachers, Puntland, Int. 231; FGD with Teachers, Somaliland, Int. 136.
[307] KII with GWD, Somaliland, Int. 158; KII with GWD, Somaliland, Int. 155; KII with GWD, Puntland, Int. 253; KII with GWD, Puntland, Int. 251; KII with GWD, Puntland, Int. 252.
[308] KII with GWD, Somaliland, Int. 154.

threat of harassment or violence GWDs face. In one group of mothers we interviewed, a woman stated rather bluntly that other children occasionally "insult, harass, and segregate" children with disabilities.[309] In an FGD with CEC members, they described the extent to which they have raised awareness of this issue, and their efforts to improve educational prospects for GWDs; they noted that they "make sure that students do not harass or insult them."[310] However, the very fact that they must pursue this objective actively implies that it is an underlying issue, and will be a problem in other schools or, indeed, in their own school.[311] A Regional Education Officer interviewed made a fairly broad claim about the extent to which mistreatment of individuals with disabilities can be, as he stated "The Somali people have a habit of insulting the person with a disability."[312] One GWD interviewee reported that students insult her openly and that she has been in fights -- though it is not entirely clear whether these fights were physical – with them as a result.[313]

The diversity of issues and sources of evidence used in this section makes it difficult to draw firm conclusions. To what extent has gender-based violence – or violence and insecurity more broadly – changed since baseline? In terms of safety at school and on the way to school, there are still significant risks, based on qualitative interviews with girls. But the quantitative evidence suggests that the perceptions of risk have declined among both caregivers and girls. Our tentative viewpoint is that localized safety around and in schools has likely increased somewhat, but that this is probably mostly attributable to changes in communities and the broader political, social, and conflict environment, rather than specific program impacts.

Much of the risk girls face is driven by local conflict dynamics. While the regions where SOMGEP-T is being implemented are not as directly impacted by the al Shabaab insurgency as areas of southern and south-central Somalia, girls' education is taking place in an environment of low-level violence and occasional bursts of highly localized, intense conflict. This fact further complicates conclusions around changes in overall violence levels and renders causal attribution nearly impossible.

Putting aside the issue of program impact, this section has illuminated the varied and multiple risks that girls face. Girls face real danger of harassment and abuse on their journeys to school and in their community. Our concern is that this evidence is based largely on four FGDs with girls, which do not reflect the broad range of experiences of Somali girls.

In an effort to triangulate the results of this section slightly more, the table below reports the type of risks that caregivers perceive for their children on the journey to school. Caregivers who indicated that the journey to school was unsafe for either boys or girls were asked to indicate the reasons why they felt the journey was risky. The results confirm many of the themes discussed above, as do many of the quantitative asides interspersed throughout this section. Concerns about safety on the road to school are paramount. Notably, caregivers are not solely concerned about traffic – though a significant number of caregivers did cite this as an issue – but they are also concerned about the sheer distance to school. We interpret this concern as partial confirmation of the fact that girls are often scared of the road to school because it passes through empty, ungoverned or unpopulated places where risks to their safety are higher.

[309] FGD with Mothers, Somaliland, Int. 113.
[310] FGD with CEC Members, Somaliland, Int. 102.
[311] FGD with CEC Members, Somaliland, Int. 102.
[312] KII with Regional Education Officer, Puntland, Int. 263.
[313] KII with GWD, Puntland, Int. 254.

TABLE 103: REASONS JOURNEY TO SCHOOL IS UNSAFE, AT ML2, ACCORDING TO CAREGIVERS

Source of Risk	Share of caregivers reporting risk on journey to school citing this specific risk
Long distance	75.6%
Verbal, physical or sexual abuse by other children or adults	46.3%
Traffic or Poor Roads	20.7%
Heat or rain	14.6%
Wild animals	13.4%
Conflict or open fighting	6.1%
Environmental disasters (fires, floods, etc.)	3.7%
Kidnappings	3.7%

The results in the table also seem to confirm the extent to which girls experience harassment or abuse on the way to school. We have grouped many such categories in the quantitative survey together, and find that 46.3 percent of caregivers citing risk on the journey to school (n = 82) are concerned, in part, about abuse or harassment. Tellingly, the most common concerns center on sexual abuse – by other children, young people, or adults – while physical and verbal abuse is a prominent, but less common concern.[314] Other risks also exist, including the threat of wild animals, and open conflict, though these risks are also likely to be quite localized.

The analysis in this section is only suggestive when it comes to broad trends in safety from baseline to ML2. As stated above, our tentative viewpoint is that violence levels have likely decreased over time, and that schools are likely somewhat safer than they were previously. Meanwhile, this section has also documented the sheer number and variety of risks that girls can and do face in their pursuit of education.

[314] This finding might reflect the actual probability of risk or, more likely, the perception of the severity of such abuse. Girls seem to indicate that verbal harassment is more common than sexual abuse, but caregivers may focus on sexual abuse because it is a more severe offense. At the same time, girls may focus on verbal harassment in the FGDs because they are not willing to talk about or bring up the topic of sexual abuse to a stranger conducting an interview.

8. ABE GIRLS BASELINE

The 2018 baseline report of SOMGEP-T indicated that approximately 13,400 more girls were out of school in the target areas than was originally anticipated.[315] As a result, Alternative Basic Education (ABE) programming was implemented in order to support acquisition of foundational skills and competencies for these out of school girls and to provide an additional point of transition into formal school settings for out of school girls in remote and rural communities. Given that the ABE program was instituted after the first midline study, this section will serve to provide baseline data of ABE girls that can be used in future assessments examining the effect of participation in ABEs on learning scores, life skills, and successful transition into a formal school setting.

The ABE program specifically targeted girls from 10-15 years of age who have never been enrolled in any form of primary school and who were from displaced, pastoralist families.[316] ABE classes were held in primary schools or NFE centres when regular programs were not in session and focused on Arabic, English Literacy, Life Skills, Numeracy, Science, and Somali Literacy. Girls participating in the program spend one year completing level 1 and 2 of the learning program with the goal of joining a formal primary school for grade 3. Alternatively, girls may opt to spend an additional year in the ABE program completing levels 3 and 4 before transitioning into a formal school for grade 5.

Surveys of ABE girls were conducted concurrently with data-collection of ISG, OOS, and ALP girls. The ABE sample was determined by identifying the ABE sites that overlapped with the existing formal school and ALP samples (this list of overlapping ABE sites was provided by CARE). These sites were included in the sample for logistical and budgetary reasons (in agreement with CARE). These ABE school made up just over half of the total ABE sample. The remaining ABE sample was determined by random sampling of additional ABE sites, trying to recover zone-level distribution ABE girls in the population. Once the sample of ABE centres was selected, girls within those centres were selected randomly.[317]

8.1 PROFILE OF ABE GIRLS

Given that the ABE program targeted girls who have not been previously enrolled in primary school and who are from displaced, pastoralist families, we would expect them to differ from other girls in the sample. Most significantly, we expect that ABE girls would face greater barriers to educational attainment due to the increased vulnerabilities as a result of displacement. In order to understand the unique challenges that the ABE girls face, we examined household characteristics, economic status, attitudes toward education (of both the girls and caregivers), and the life skills and confidence of each of the cohort, OOS, ALP, and ABE girls surveyed at the midline.

Before examining potential differences in economic, demographic, and attitudinal differences between these groups of girls and exploring their potential to serve as barriers or facilitators of girls' education, we first examined the reasons, cited by their caregivers, for their girls not being enrolled in a formal primary school. It should be noted that these responses only capture caregivers who said that their daughter was not enrolled in school. Relatively few caregivers of ABE girls (n=46) said their girl is not enrolled indicating that many caregivers of ABE girls considered their girl's attendance in the ABE program as being enrolled in formal

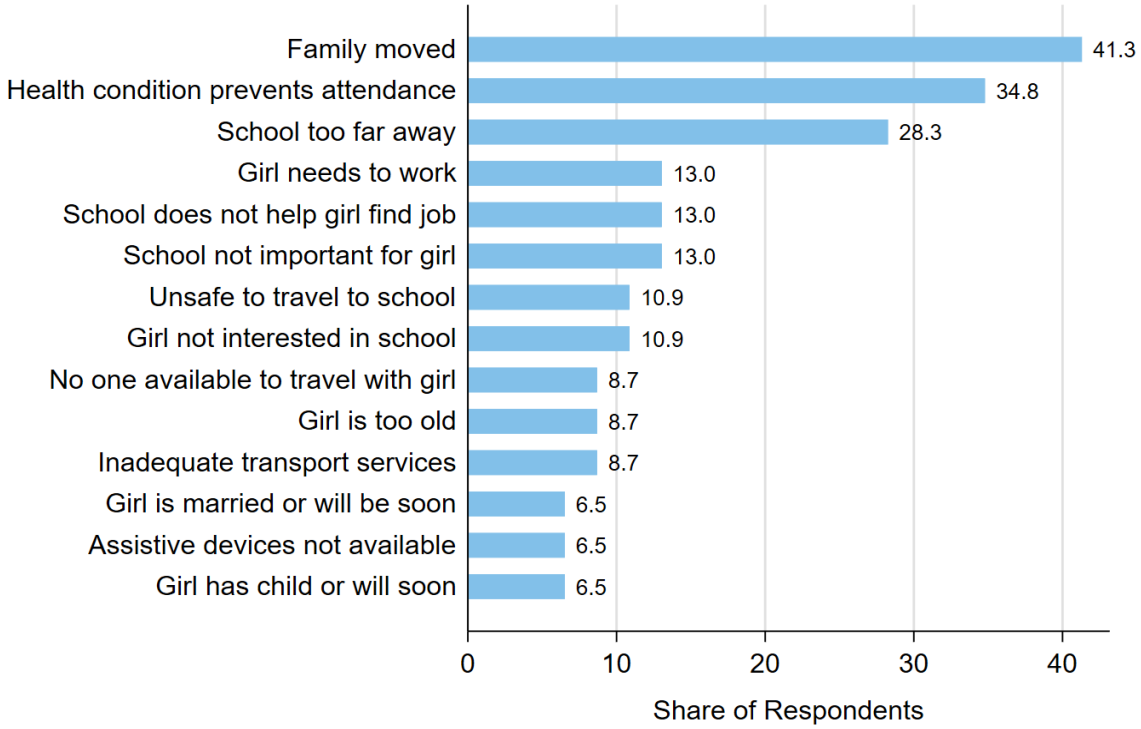
[315] CARE SOMGEP-T Baseline 1

[316] Most of the targeted girls are from pastoralist families who were displaced during the 2016-2017 drought.

[317] See the "Sampling ABE Girls" subsection of the methodology section for more details.

school. Despite the small sample, his analysis will help to shed initial light onto the barriers that ABE girls face and to help provide context for some of the later findings. Figure 1 below shows the most common challenges ABE girls face to attending school. Given that the ABE program targeted girls from displaced households, it is not surprising that the family having moved (41.3 percent) was the major cause given by caregivers for an ABE girl not being enrolled in a formal school. Health concerns were also very common with 34.8 percent stating that the girl has a health condition that prevents them from attending and another 6.5 percent cite a lack of available assistive devices.[318] Given the low number of caregivers in the overall sample for this question, one should be careful as interpreting this result to mean that there is a large number of girls with a health problem that prevents their attendance.[319] Instead, this result shows that for ABE girls with health problems, they may be lacking the support they need to be able to attend school. Being unable to travel to the school due to distance (28.3 percent), safety (10.9 percent), lack of someone able to travel with the girl (8.7 percent), or insufficient transport services are also a major barrier for girl's attendance. This travel barrier may be especially problematic for girls with health issues. Girl's with health issues were twice as likely (60.0 percent) as girls without health issues (30.4 percent) to say that the school was too far away for them to attend.

FIGURE 47: REASONS FOR ABE GIRLS NOT BEING ENROLLEDIN FORMAL SCHOOL



Attitudinal barriers were not uncommonly given as a reason for their girl not being in school. 13 percent of respondents felt that the girl needs to work instead of attending school and/or that school is not important for a girl. Additionally, 10.9 percent of caregivers cited a lack of interest on the girl's part in attending school. Surprisingly, relatively few caregivers cited marriage or having children as a major barrier (6.5 percent in each case). These results suggest that while a few barriers such as migration and health concerns serve as the

[318]

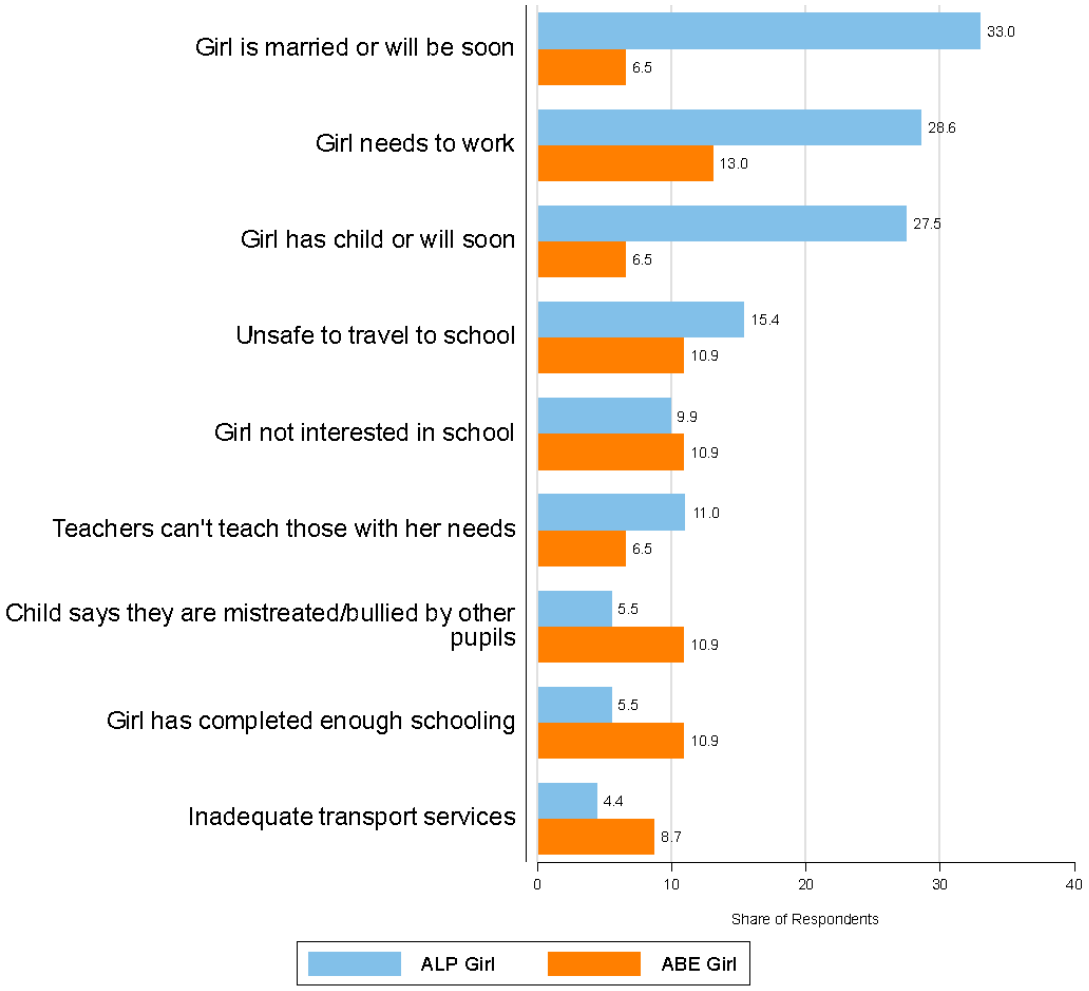
[319] In the household characteristics section, we explore the frequency of disability at the household level.

most common barriers, there are a multitude of barriers from factors within and outside the home that make it difficult for ABE girls to receive a formal education.

These results taken as a whole suggest that the reasons that ABE girls are not enrolled in school are multi-faceted and that they face many barriers (often more than one at once) that may keep them from attending a formal education program. While it may be difficult to draw conclusions about the specific nature of the barriers households that have moved face, it is clear that these households are vulnerable to lacking the requisite support needed to help their child go to school, even when the desire is there. For many households, having a school that is close by, with a means to get their girls their safely, and that has the assistive devices required to meet their needs is accessible to them. Furthermore, access to appropriate healthcare services is serving as a major impediment to their attendance.

When these trends are compared to other girls who are also not enrolled in traditional school settings, some interesting differences emerge between ABE and ALP and OOS girls. We first compared ABE to ALP girls to see if they face different challenges or barriers to their attending formal school. The categories with the largest differences between ABE and ALP girls are listed in the figure below.

FIGURE 48: REASONS GIRLS IS NOT ENROLLED IN FORMAL SCHOOL – ABE VS ALP

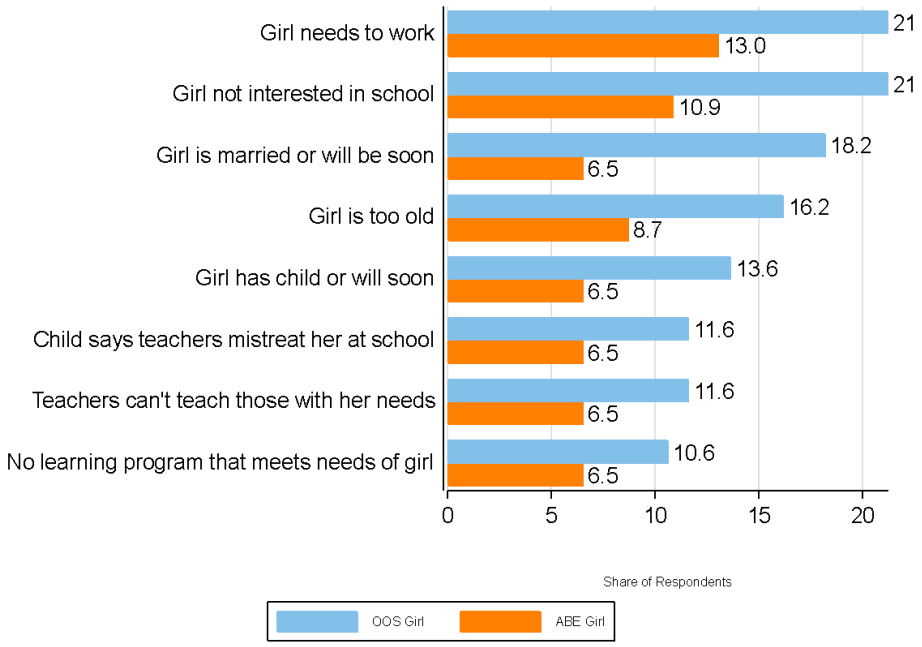


Three key categories emerge as important differences between ABE and ALP girls in terms of the reasons that they are not currently enrolled in school. Pressure to marry and have children is often thought of as a primary reason for a girl not being enrolled in school. ALP girls are far more likely to not be in formal school settings due to marriage or the expectation of marriage. In fact, a full third of ALP caregivers cited marriage as a reason for their girl not being in school compared to only 6.5% of ABE girls. Similarly, having children or the expectation of having a child soon was the second largest difference between the groups with 28.8% of ALP girls compared to 6.5% of ABE girls citing this cause. A need to work was also a major difference between the two groups as 28.6% of ALP caregivers (compared to only 13% of ABE caregivers) cited the girl needing to work as a cause to not be enrolled in formal school. These discrepancies are likely due in large part to differences in age between ALP and ABE girls. girls are on average 13.7 years old while the average ALP girl is 21.0 meaning that ABE girls are likely too young to be facing the same pressures to marry, have children, and work.

While ABE girls do not appear to be facing the same pressures to start a family and begin caring for their family when compared to ALP girls, they do appear to be more likely to face other familial expectations that prevent them from attending formal school. ABE girls were twice as likely (10.9% for ABE compared to 5.5% for ALP) to have a caregiver say that the girl has already received enough schooling than ALP girls. Additionally, ABE girls report barriers to having transportation to school (8.7% to 4.4%) and mistreatment once they are there (10.9% compared to 5.5%) than ALP girls.

The reasons ABE girls were not attending school were also compared to the reasons OOS girls did not attend. Figure 3 below summarizes the categories with the largest differences between ABE and OOS girls. We found that for every large difference, OOS girls were much more likely to report that reason for being out of school than ABE girls. This result does suggest that ABE girls are much less vulnerable to not receiving an education as OOS girls appear to face a greater variety of barriers to their attendance.

FIGURE 49: REASONS GIRL IS NOT ENROLLED IN FORMAL SCHOOL – ABE VS OOS



Household characteristics for each of the four types of girls were examined in order to understand how home life may be impacting educational outcomes. The most surprising result from this analysis is that ABE girls were no more likely to have a head of household who is a pastoralist (and are less likely than OOS girls). This is surprising given that the ABE program sought to recruit girls from displaced, pastoralist households. However, part of the explanation for this result may lie with ABE girls being more likely than any other type of girl to come from a home where the head of household does not have employment (56.7 percent of ABE households have a head of household who is unemployed). It is difficult to draw any conclusions about why this may be the case, but it would not be unlikely that many of these heads of household, who are categorizing themselves as unemployed, may in fact be former pastoralists who have not been able to pursue this trade as result of a loss of livestock or other difficulties. ABE girls are more like to have a head of household who is illiterate and a caregiver who has not received a formal education. The higher rate of unemployment by the head of household and low educational opportunities afforded to caretakers within the household may put ABE girls at a stark disadvantage in terms of their ability to receive academic support from inside the home.

Despite not being enrolled in a formal school, ABE girls are at least as likely as other girls to report having access to a primary school within a 15 minute walk from their home and are far more likely to report having access to secondary school. Additionally, ABE girls are far more likely to have their mother with them in the household than either OOS or ALP girls. This result is surprising because we would expect that these factors may be similar for other types of girls who are not enrolled in formal schools (OOS and ALP), however ABE are far more similar to ISG girls in this regard. It should be noted that while access to schools is at least as good as it is for other types of girls, there are still a large number of households who do not have access to schools which are reasonably close.

TABLE 104: HOUSEHOLD CHARACTERISTICS OF ABE GIRLS

	Group Mean				Difference in Means (t-test)		
Indicator	ABE	ISG	OOS Girls	ALP Girls	ABE vs. ISG	ABE vs. OOS	ABE vs. ALP
Household Size	5.4	5.5	5.2	5.0	-0.1	-0.2	0.4*
Female HoH	43.1	47.2	26.0	45.5	-4.1	17.1*	-2.4
HoH has no occupation	56.7	44.9	50.5	47.0	11.8*	6.1*	10.1*
HoH is a pastoralist	5.6	7.1	10.7	6.9	-1.5	-5.0*	-1.2
HoH has no formal education	81.5	73.2	74.9	68.1	8.3*	6.6*	13.4*
Caregiver is illiterate	79.4	65.8	74.6	61.2	13.6*	4.8*	18.2*
Primary school 15+ minutes away	42.8	40.1	37.3	30.3	2.7	5.5	12.5*
Secondary school available	66.0	41.8	21.4	31.8	24.2*	44.7*	34.3*
Secondary school 30+ min away	31.9	35.6	34.5	44.2	3.6	2.6	-12.3*
Mother not in HH	12.9	13.0	18.9	29.6	-0.1	-5.9*	-16.6*
Has disability	17.7	17.9	17.5	17.9	-.001	-.1	-0.2

Economic factors were also considered when trying to understand the unique challenges that ABE girls may be facing. Table 14 below summarizes these results. There were stark differences between ABE girls and other girls along some key economic indicators (summarized in table 3 below). ABE girls were far more likely to have gone to bed hungry and lacked access to clean water in the past year. ABE girls were also far more likely to be living in a home with a poor-quality roof.[320] The large gap between ABE girls and other girls along these economic indicators suggest that they face more day-to-day challenges to their well-being and this is likely constituting a major barrier to their attendance in a formal school setting.

While ABE girls are lagging behind in the ability to meet their most basic needs, we found that ABE girls were generally more likely to own property that might be more typically associated with relative wealth. ABE girls were significantly more likely to own a cellular phone (84.9 percent) than any of the other girls surveyed at midline. Similarly, they were more likely to solely own land (64.6 percent) than ISG or OOS girls (59.1 percent and 35.8 percent respectively). Generally, livestock ownership did not vary much between girls surveyed, with exception of OOS girls were far less likely to own medium livestock. These findings may not be as surprising as they seem at first glance. The ABE program targeted girls from households who had been displaced as a result of the 2016-2017 flood. The disparity between their relative inability to secure their day-to-day needs and high ownership of land, livestock, and cellular phones is likely more reflective of their previous economic condition rather than the currently economic realities they are facing.

TABLE 105: ECONOMIC STATUS OF ABE, ALP, ISG, AND OOS GIRLS

	Group Mean				Difference in Means (t-test)		
Indicator	ABE	ISG	OOS Girls	ALP Girls	ABE vs. ISG	ABE vs. OOS	ABE vs. ALP
Poor quality roof	42.0	23.9	15.9	27.1	18.2*	26.2*	14.9*
Own large livestock	10.1	11.7	9.5	14.0	-1.5	0.6	-3.8*
Own medium livestock	59.0	60.0	33.2	60.0	-0.1	25.8*	-0.9
Own mobile phone	84.9	78.5	45.6	77.8	6.4*	39.3*	7.1*
Solely own land	64.6	59.1	35.8	63.8	5.4*	28.8*	0.7
Went to bed hungry 1+ nights in the last year	44.2	26.2	20.2	29.3	18.0*	24.0*	14.9*
Lacked clean water for HH use 1+ days last year	71.2	50.0	35.6	58.9	21.2*	35.8*	12.2*
Reduced food expenditures in the last three months	38.5	30.4	20.5	30.1	8.1*	18.1*	8.4*

[320] A roof was defined as being poor quality if it is made of mud, thatch, cardboard, or tarp/plastic.

Went without food for an entire day in the last month	22.6	16.4	12.1	17.8	6.1*	10.5*	4.8
Reduced the number of meals eaten in the last month	34.8	25.1	18.2	28.5	9.7*	16.6*	6.3*
Did not have a source of protein in their diet in the last 24 hours	38.5	18.1	11.4	20.5	20.4*	27.1*	17.9*
Mean Dietary Diversity Score	3.8	4.6	2.5	4.5	-0.82*	1.3*	-0.75*

In addition to examining the demographic and economic characteristics of these girls, we were also interested in gaining a better understanding of how household attitudes toward education may be impacting ABE girls' educational opportunities. How households and caregivers view the importance of education, relative to other concerns (such as marriage or household chores) may be a key factor in determining if a girl is able to take advantage of educational opportunities. In table 4 below, we examine some indicators of caregiver attitudes toward girls' education as well as girls' views of their ability to make decisions that will affect their opportunities to attend school. Caregivers of ABE girls in this survey held similar views about the worthiness of girl's education compared to ISG and ALP caregivers with only 16.5% stating that a girls' education is not worthwhile. ABE caregivers were far more likely to express a desire for their child to attend university than either OOS or ALP caregivers.

Girls were asked if they felt that had either sole or joint input into marital decisions and about their schooling. ABE girls were far more likely than OOS girls to feel that they had input into these decisions and more autonomy than ISG girls when it comes to school decisions. However, ALP reported a far greater degree of input into both of these decisions than ABE girls. This large divide between the degree of input that girls have in making decisions that will affect their education has potentially interesting implications. Given the questions that girls were asked, it is possible that these answers reflect, to some extent, girls' views of their ability to influence their educational outcomes. Girls enrolled in either ALP or ABE programming were significantly more likely to report a sense of input into these large decisions and it is not clear if having decision making power increases the likelihood of participating in some sort of educational program, or if participation is driving a feeling of empowerment.

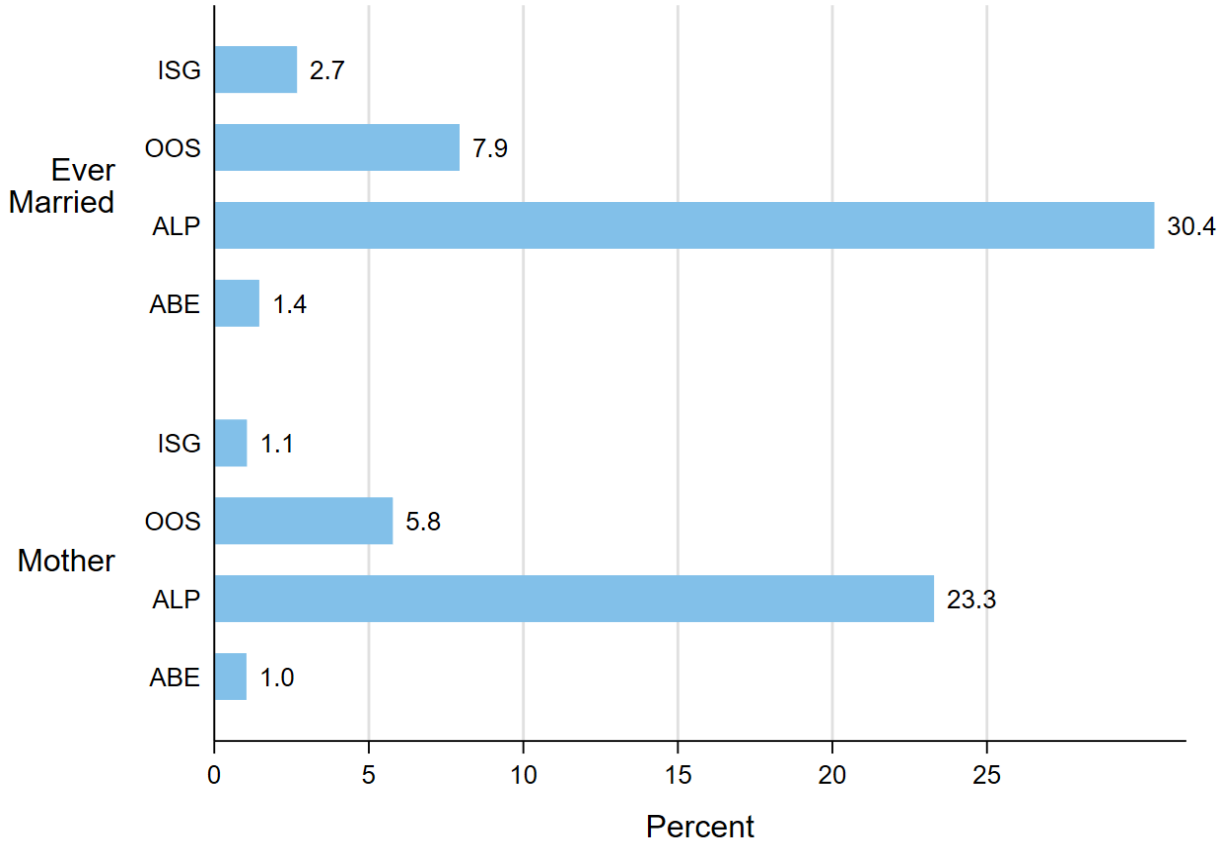
TABLE 106: VIEWS OF GIRLS' EDUCATION AND DECISION-MAKING POWER

	Group Mean				Difference in Means (t-test)		
Indicator	ABE	ISG	OOS Girls	ALP Girls	ABE vs. ISG	ABE vs. OOS	ABE vs. ALP
Caregiver aspires for girl to attend university	77.2	75.0	36.7	66.8	2.3	40.6*	10.4*
Adults make schooling decisions (adult self-report)	63.6	41.6	30.6	41.4	21.9*	33.0*	22.2*
Caregiver says girls' education is not worthwhile	16.5	16.6	21.0	14.0	-.1	-4.6*	2.4
Girl has sole/joint input into schooling decisions	69.2	61.3	40.3	82.2	7.8*	28.9*	-13.0*

Girl has sole/joint input into marital decisions	38.5	36.6	27.0	61.9	1.9	11.5*	-23.4*

ABE girls were very similar to ISG in their rates of marriage and having children. Among these group marriage rates were low (1.4% of ABE girls and 2.7% of ISG girls had ever been married) and the proportion of girls had children (1.0 percent of ABE girls and 1.1 percent of ISG girls have had children) is very low. Conversely, other girls who were not in formal school settings were much more likely to be married with 30.4 percent of ALP girls and 7.9 percent of OOS girls having been married at some point. Having children was also much more common among these groups with 23.3 percent of ALP girls and 5.8% of OOS girls having had children.

FIGURE 50: RATES OF MARRIAGE AND MOTHERHOOD, AMONG COHORT GROUPS



As discussed previously, ABE girls generally report a higher degree of decision - making power than ALP, OOS, or ISG girls. However, ABE girls report a significantly higher degree of nervousness when being asked to speak in front of adults or peers and to practice skills like reading and math in front of others. That ABE and ALP girls are more likely to be nervous than either ISG and OOS when speaking to adults or peers ABE girls display nervousness when speaking front of adults and peers at a similar rate to ALP girls, but OOS and ISG are far less likely to claim nervousness when doing these tasks. While it is not clear why, this similarity between ABE and ALP girls may stem from their relative lack of experience in doing these activities compared with girls who are enrolled in formal school, but are confronted with having to speak in front of others more often relative to OOS girls. A similar relationship emerges when looking at feelings of nervousness around

doing math and reading in front of others. While ALP girls reported being less nervous than ABE girls, OOS girls were the least likely to report feeling nervous about these activities.

Despite their nerves when asked to speak in front of adults or peers and to display their reading or math abilities, ABE girls report being generally confident answering questions in class or group settings. In fact, ABE girls are more confident than any of the other girls in these settings.

TABLE 107: LIFE SKILLS AND SELF-CONFIDENCE OF ABE, ALP, OOS, AND ISG

	Group Mean			Difference in Means (t-test)			
Indicator	ABE	ISG	OOS Girls	ALP Girls	ABE vs. ISG	ABE vs. OOS	ABE vs. ALP
Nervous when speaking in front of an adult	38.3	29.5	20.9	36.2	8.8*	17.4*	2.1
Nervous when speaking in front of peers	30.0	21.5	17.3	27.9	8.5*	12.7*	2.1
Nervous when reading in front of others	37.1	28.0	9.8	27.4	9.1*	27.2*	9.7*
Nervous doing maths in front of others	33.5	24.4	9.4	27.4	9.2*	24.2*	6.1*
Confident answering questions in class	79.1	70.3	20.6	64.1	8.8*	58.5*	15.0*
Confident answering questions in group of people	81.4	72.5	35.9	77.5	8.9*	45.4*	3.8

Based on the indicators examined in this analysis, it appears that ABE girls tend have stark disadvantages relative to other girls in terms of being able to meet their most basic day-to-day needs, but that they benefit from a higher degree of ownership over decision making processes and favourable attitudes toward school coming from within their homes.[321]

8.2 ABE GIRLS' LEARNING OUTCOMES

This section summarizes the learning outcomes for ABE girls in Somali literacy, numeracy, and financial literacy and compares their mean scores to ISG and ALP girls interviewed during this midline. Given that this is the first time that data has been collected from ABE girls, this section of the report will establish baseline learning outcomes that can be used for establishing learning gaps that can be used for targeted learning-based interventions in the future.

Large differences in the learning outcomes for ISG, ABE, and ALP girls are expected. The ABE program is designed to reach girls who have never attended school or who dropped out early in primary. This means that ABE girls survey for the first time in this second midline report are starting significantly behind both ISG girls and ALP girls (who have generally completed lower primary school). This discrepancy in the amount of

[321] ABE girls were not given English literacy assessments since they had not yet received English literacy courses at the time of this second midline.

schooling that each type of girl had going into the baseline data collection is important to keep in mind when comparing learning outcomes.

Somali Literacy

The table below summarises the mean Somali literacy scores for ABE, ISG, and ALP girls. The scores have been binned by age since neither ALP or ABE girls are assigned to traditional grade levels. The baseline scores of the ABE and Baseline girls are then compares to both ISG and ALP girls scores in the same age band who were also interviewed at the second midline. As can be seen in Table 18, the mean literacy score for ABE girls was 13, which is a full 49 percentage points than ISG girls and 35 points lower than ALP girls.

TABLE 108: ABE SOMALI LITERACY (EGRA/SEGRA)

Age	ABE Mean	ISG Mean	Difference (ABE-ISG)	ALP Mean	Difference (ABE-ALP)	Standard Deviation for ALP
12-13	11	53	-42	56	-45	35
14-15	13	68	-54	37	-24	36
16-17	13	64	-51	49	-36	36
18-19	35	67	-33	50	-16	34
20-22	53	42	11	49	4	33
Overall	13	62	-49	48	-35	36

As might be expected, the mean literacy score for ISG girls does increase as function of age, with 18-19 year olds seeing a large increase in literacy even when compared to 16-17 year olds.[322] While literacy scores do appear to take a large jump as ABE girls enter into adulthood, there are still significant gaps between ABE and ALP girls at every stage from 12 to 19.

While mean literacy scores remain quite low among ABE girls, variation of scores within each age band remains quite high. This significantly lower score among ABE girls is predominantly driven by scores of zero on the literacy assessment (396 girls scored zero on this part of the assessment). In a separate analysis, all girls with a score of 0 were dropped in order to see how girls performed who were able to complete even one portion of the assessment. The overall mean of these ABE girls' literacy scores is 29 and the distribution of girls in the sample is approximately uniform. These results suggest that among girls who can complete even one section of the report, thy still lag behind their peers, albeit to a lesser degree. However, the main takeaway from this result is that many ABE girls are especially vulnerable compared to other types of girls not as a matter of degree, but as a matter of even having the most basic literacy skills. This pattern persisted when we examined the literacy scores within age bins as well, suggesting that opportunities for literacy acquisition outside of school will be limited. It will be interesting in subsequent assessments of the ABE program if after completing the program girls are able to significantly close the gap among their peers, or these large deficits in literacy skills make it even more difficult to close the gap between other girls.

[322] While the ABE girls in the 20-22 age bin do exhibit a larger mean literacy score than either the ISG or ALP girls, the number of girls in these age bins (ABE = 4 and ISG = 8) is quite low and drawing inferences from these sub-samples is not advisable

Numeracy

The table below summarizes the average ALP numeracy scores by age-bin in the same fashion as the literacy scores summarized above. The mean numeracy score among all ABE girls in the sample was 30. ABE girls score far lower than the ISG girls (difference of -30) and ALP girls (difference of -23).

Numeracy scores increase as a function of age across all girl types. When compared to ISG girls, ABE girls do not significantly close the gap. At ages 12-13, the difference between ABE and ISG girls is -27 and has only shrunk to -25 by the time the girls are 18-19. While still lagging behind ALP girls, ABE girls do significantly close the gap by the time they are 18-19 (-22 at age 12-13 and -13 at age 18-19). These results suggest that there are massive skill gasps in numeracy between ABE and the ISG and ALP girls.

It is interesting to note that the gap between ABE and ALP girls is much smaller for numeracy than for literacy. Furthermore, nearly every girl displays and ability to complete the most basic numeracy assessment (whereas most ABE girls are not able to do the same for the literacy assessment). A closer look at the specific skills assessed within the literacy and numeracy assessments may help to shed further light on these key differences between ABE girl performance on literacy vs. numeracy assessments and will be explored further in the Foundation Skill Gaps and Subgroup Analysis sections.

TABLE 109: ABE NUMERACY (EGMA/SEGMA)

Age	ABE Mean	ISG Mean	Difference (ABE-ISG)	ALP Mean	Difference (ABE-ALP)	Standard Deviation for ALP
12-13	26	54	-27	48	-22	25
14-15	33	62	-30	42	-10	25
16-17	34	64	-30	52	-18	26
18-19	42	67	-25	55	-13	23
20-22	55	58	-3	56	-1	21
Overall	30	60	-30	53	-23	26

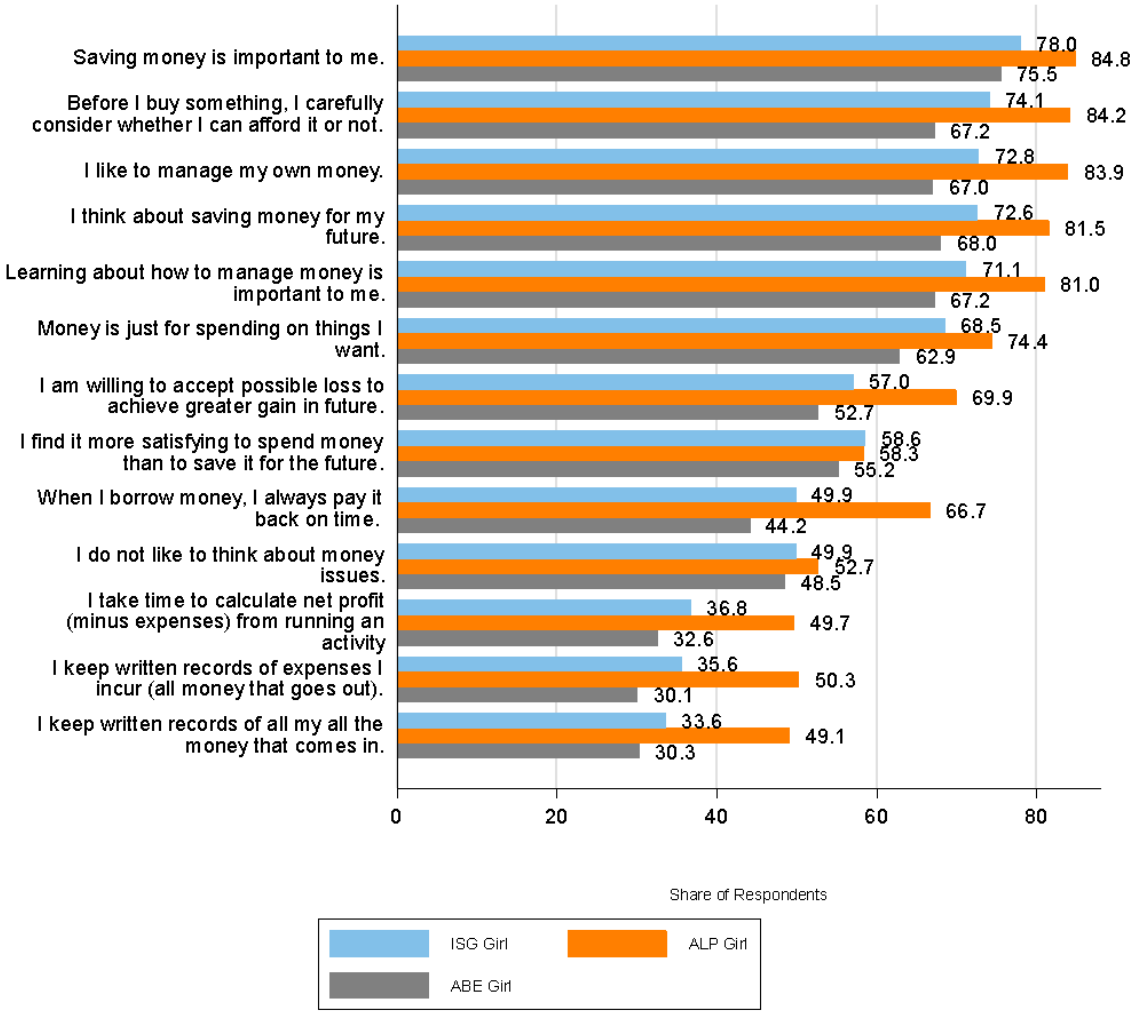
Financial Literacy

The financial literacy assessments aimed to help gain an understanding of girls' perceptions about savings and the use of money and their ability to apply their numeracy and literacy skills to solving word problems that involved utilizing basic financial knowledge.

Figure 17 below shows ABE, ISG, and ALP girl's attitudes towards savings and use of money. Across all 14 surveyed attitudes about savings and use of money, ALP girls were the most likely to agree with each statement, followed by ISG girls, and ABE girls were the least likely to agree with the statement about savings and use of money. When asked questions about the importance of savings, thinking about purchases before making them, and expressing a desire to learn about managing money, the differences between girls were the smallest, however quite substantial. This finding is consistent with the findings in the ABE baseline that show that ABE girls are more likely to not be having their basic needs met. Having positive attitudes towards saving may prove more difficult if purchases or more closely attuned with basic needs. Concerning questions about formally and systematically managing money (For example, keeping records of expenses incurred), only about 30% of ABE girls responded taking such measures. Given the large differences between ABE girls and

both ISG and ALP girls in literacy and numeracy, it may not be surprising that their tendency to engage in practices that rely on those skills would be lower. While this is true when compared to ALP girls, this expectation is supported. However, ISG girls answered very similarly to ABE girls (within a few percentage points) and scored higher than either ABE or ISG girls. Given this observation, it is likely that ABE girls attitudes are generally less positive than ALP and ISG girls due to their and their households relationship to money and their ability to meet everyday expenses and needs.

FIGURE 51: ATTITUDES ABOUT SAVINGS AND USE OF MONEY FOR ABE, ISG, AND ALP GIRLS



In part two of the financial literacy assessment, girls were asked to solve a series of word problems that correspond with their ability to manage money in personal spending and business contexts. Given the large gaps between ABE and both ISG and ALP girls in the literacy and numeracy assessments, we expected to find that ABE girls would score significantly lower on the financial literacy assessment. These findings were largely true with ABE girls averaging a 7 on the financial literacy assessment compared to a 23 and 22 for the ISG and ALP girls respectively.

The financial literacy scores did not show a clear relationship with age for any of the girl types as was the case for literacy and numeracy assessments. Among ABE girls, 16-17 year olds were actually scored worse (mean = 5) than 12-13 year olds (mean = 6) or 14-15 year olds (mean = 7). However, by the time these girls had reached young adulthood (19-19) their scores had risen dramatically. Like the analysis of girl's attitudes toward savings and use of money, this analysis (when conducted by age group) does not suggest that literacy and numeracy are driving differences in financial literacy. Instead, it is likely more strongly related to other factors. The subgroup analysis below will explore some of these factors more closely in order to see if some purported drivers of financial literacy can be identified among ABE girls.

TABLE 110: ABE FINANCIAL LITERACY

Age	ABE Mean	ISG Mean	Difference (ABE-ISG)	ALP Mean	Difference (ABE-ALP)	Standard Deviation for ALP
12-13	6	18	-12	25	-18	23
14-15	7	25	-18	16	-9	27
16-17	5	29	-24	20	-15	30
18-19	15	28	-13	20	-5	27
20-22	16	20	-5	29	-13	33
Overall	7	23	-17	22	-15	27

8.3 ABE GIRLS' SKILL GAPS

The analysis of ABE girls and their learning outcomes thus far has focused on their aggregate performance and demographic, household, and personal barriers – among others – to that performance. In this section, we dig more deeply into specific foundational skills that girls need in order to be considered fully numerate or literate. Our goal in this section is to understand the skills that ABE girls have at baseline, so that programming can be adjusted to match their existing skills and skill gaps. We also wish to understand divisions among girls, to help shape efforts at remedial education or tailored approaches to ABE pedagogy.

We do not provide a full discussion of the foundational skills gap analysis, or how we approach it. Readers interested in more details are referred to the learning outcome section for non-ABE cohort girls above, and to the SOMGEP-T baseline report. Briefly, we score ABE girls' performance on each subtask and place them into four groups (non-learner, emergent learner, established learner, proficient learner) based on their performance on each subtask, with score ranges of 0 percent, 1-40 percent, 41-80 percent, and 81-100 percent. The results are intended to reveal where girls' learning stops, where girls become divided into low- and high-performers, and other trends that can guide programming.

The table below reports the share of girls in each of the four categories for the 11 subtasks that appeared on the numeracy learning assessment. The first finding that stands out, from subtask 1, is the fact that ABE girls have poor facility recognizing basic number patterns. The typical girl – both the median and modal girl – was able to identify just three of ten missing numbers successfully.

By comparison, girls are much better at performing simple, 1-digit arithmetic operations, captured in subtasks 2 and 3. The majority of girls have achieved proficiency in both 1-digit addition and subtraction. Noteworthy,

however, is the sharp increase in the number of girls who are entirely unable to perform subtraction versus addition – there are more than double the share of non-learners of 1-digit subtraction than there are for 1-digit addition. While girls enrolled in formal school also exhibited poorer performance in 1-digit subtraction than 1-digit addition, the gap in performance between the two tasks was not nearly as dramatic as is the case for ABE girls.

TABLE 111: FOUNDATIONAL NUMERACY SKILLS AND SUBTASKS AMONG ABE GIRLS

Subtask	1	2	3	4	5	6	7	8	9	10	11
Skill Assessed	Missing number	Addition (Level 1)	Subtraction (Level 2)	Addition (Level 2)	Subtraction (Level 2)	Word problems (add/sub)	Multiplic. (Level 1)	Multiplic. (Level 2)	Division (Level 1)	Division (Level 2)	World problems (mult/div)
Non-Learner (0%)	10.4%	15.1%	31.5%	50.5%	64%	48.2%	82.2%	96.1%	91.1%	97.7%	92.3%
Emergent Learner (1-40%)	46.6%	3.5%	5.2%	8.9%	3.9%	7.9%	2.5%	0.8%	2.3%	0.6%	0%
Established Learner (41-80%)	30.8%	19.5%	11.4%	16.1%	11.4%	23.4%	5.8%	1.7%	4.1%	0.8%	3.9%
Proficient Learner (81-100%)	12%	61.7%	51.8%	24.2%	20.5%	20.3%	9.3%	1.2%	2.3%	0.6%	3.5%

Similar to formal school girls, there is a sharp break between subtasks 2 and 3 and subtasks 4 and 5, the latter of which assess 2-digit addition and subtraction, respectively. ABE girls start from a lower base level of competency in 1-digit arithmetic than their peers in formal schools, but the drop from 1-digit to 2-digit operations is also steeper: where the share of formal girls who achieved proficiency in addition or subtraction dropped 25-30 percentage points when moving from 1-digit to 2-digit test items, the share of ABE girls who achieved proficiency dropped around 30-40 points. Starting from a lower base of proficiency on 1-digit tasks means that only 24.2 percent of ABE girls are proficient in 2-digit addition, compared to 59.5 percent of girls enrolled in formal schools.

A further distinction between ABE and formal school girls is what occurs following this decline. Formal school girls demonstrated remarkably steady performance across subtasks 4-7, which move from 2-digit addition and subtraction to word problems based on these same operations, to 1-digit multiplication. In contrast, ABE girls' performance on subtask 7 shows another sharp decline, down to a proficiency rate of just 9.3 percent, and a non-learner rate of 82.2 percent. Formal school girls, on the other hand, had a proficiency rate of 61.0 percent and a non-learner rate of 20.0 percent on this same subtask. ABE girls' performance never truly experiences an upswing during the numeracy assessment, and shows a general, steep decline from subtask 2 onward.[323]

Turning to literacy, we consider ABE girls performance on the nine subtasks of the Somali literacy assessment. As noted above, ABE girls did not complete an English literacy assessment, because ABE programming does not specifically target literacy in English as a core goal.

The starkest findings in the table below actually require comparisons to non-ABE girls. While our earlier analysis of ABE girls' learning performance showed that ABE girls significantly lag behind girls enrolled in school – a finding that is, naturally, consistent with the fact that ABE girls have not been enrolled in school – what is surprising is the extent to which ABE girls do not just have marginal or weak literacy, but are mostly non-literate in a deeper sense. The results for subtask 1 make this clear, as only 17.8 percent of ABE girls achieve proficiency in identifying and reading a set of common Somali words. The median girl was able to identify just 13 of 60 words. Insofar as identifying common words is a natural precursor to reading simple sentences, this suggests that ABE girls have significant steps needed to achieve functional literacy. We view this finding as surprising because, while we expected ABE girls to lag dramatically behind in-school girls in terms of reading comprehension, and even reading simple sentences, we expected them to have higher facility in terms of identifying words.

[323] It is important to note that low performance on the latest subtasks may overstate the share of girls who should be classified as non-learners, because enumerators were able to skip later subtasks if girls had received a score of zero on three consecutive subtasks. Non-learner classification requires a girl to score exactly 0 percent on a subtask, and some of the girls who are thus classified on subtasks 8-11 may have been able to answer some test items correctly – thereby moving into the emergent learner category – if they had been given the opportunity to do so. The choice to skip subtasks is consistent with administration during past rounds of SOMGEP-T, and was made on ethical grounds, as insisting that a girl continue to attempt questions that she cannot do has the potential to harm her self-esteem or upset her. This decision does make it possible to overestimate the number of non-learners, however, as noted above.

TABLE 112: FOUNDATIONAL SOMALI LITERACY SKILLS AND SUBTASKS AMONG ABE GIRLS

Subtask		1	2	3	4	5	6	7	8	9
Skill Assessed		Reading Words	Reading Comp (easy)	Reading Comp (medium)	Reading Fluency	Reading Comp (difficult)	Writing (fill blank)	Writing (negative form)	Writing (future tense)	Sentence completion
Intervention Schools										
Non-Learner (0%)		8%	13.8%	19.8%	12.2%	22.4%	22.1%	32.5%	39.8%	50%
Emergent Learner (1-40%)		7.6%	4.4%	6.7%	12.4%	9.9%	10.4%	4.6%	4.6%	6.9%
Established Learner (41-80%)		15.5%	26.1%	30.9%	30.6%	39.2%	19.3%	7.8%	8.3%	13.1%
Proficient Learner (81-100%)		68.9%	55.7%	42.6%	44.9%	28.4%	48.2%	55.1%	47.3%	30%

The results for later subtasks are simultaneously more concerning and a reason for small amounts of optimism about the starting point for ABE girls. On the negative side, the number of girls who can read a simple story and answer questions based on the content is very low – 66.3 percent of girls are unable to answer any questions about the story at all. However, the share of girls achieving "established learner" and "proficient learner" status provides a reason for optimism. In total, 16.6 percent of girls answered either three or four out of four comprehension questions correctly. Although the majority of girls were not able to answer any questions correctly, there is a small set of girls who have already gained meaningful reading skills. These girls can be targeted for a more accelerated program, while the majority of girls who are beginning from a lower starting point can have their program tailored to their specific needs.

AVERAGE SKILL PERFORMANCE

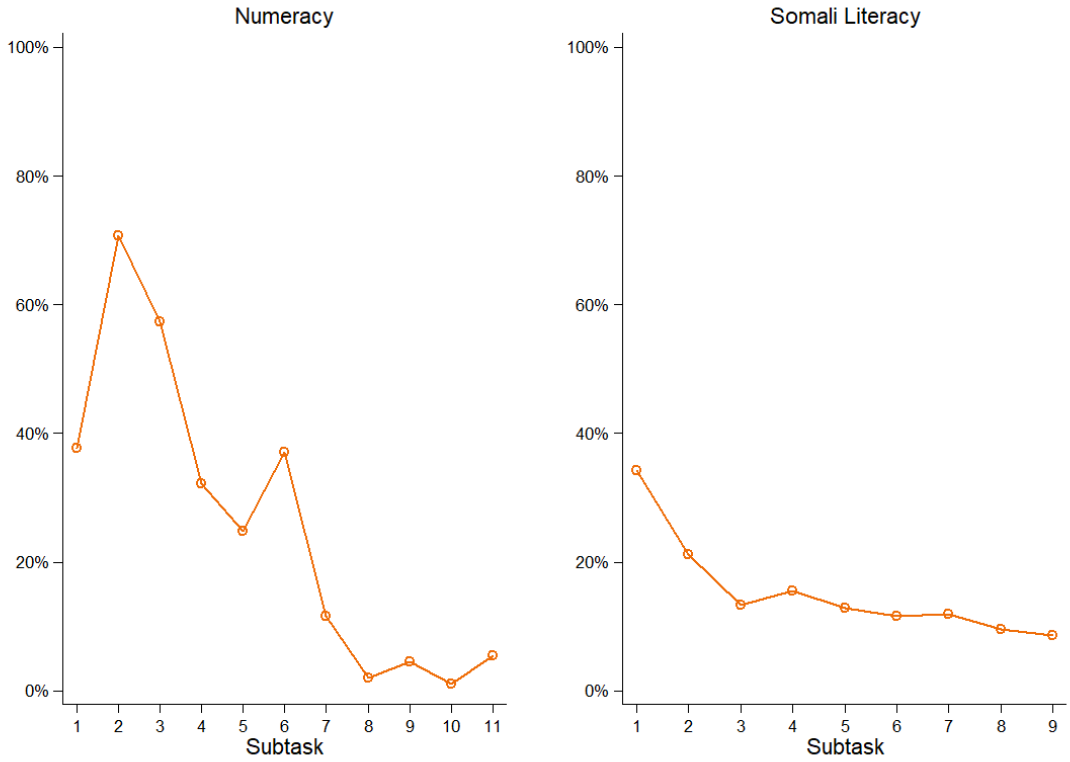
The foundational skills gap analysis, which is a staple of GEC-T evaluations, highlights a number of key trends in ABE girls' learning across the different numeracy and Somali literacy subtasks. However, because the categorization of girls into four groups can occasionally obscure trends – typically because many of the subtasks consist of just a few test items, which forces girls into particular categories when they may be at the margins of two categories – we also report subtask-specific mean scores for ABE girls in the figure below.

Numeracy scores, provided in the left panel, largely mirror the results of both OOS girls discussed earlier in this report and the foundational skills gap analysis for ABE girls reported just above. As we noted in that discussion, girls experience a sharp decline from 1- to 2-digit addition and subtraction, and almost no ABE girls are able to reliably perform multiplication or division operations, even when limited to 1-digit problems.

Where the focus on means is useful is in illustrating the performance of ABE girls on subtask 6, which is a set of word problems utilizing addition and subtraction. When we reported results from the foundational skills gap analysis, we stated that ABE girls' performance on this subtask was not markedly different from their performance on subtask 5; in short, that they did not "recover" from poor performance on subtasks 4 and 5 by improving their performance on subtask 6, as did in-school girls. However, the figure below suggests that this was not entirely true. In-school girls, including both intervention and comparison communities, saw a 10.9 point increase in performance from subtask 5 to subtask 6, and a nearly equivalent decline on subtask 7. As a result, mean scores on subtask 6 are seen as a spike relative to its neighbouring tasks. While we reported that ABE girls did not see such a spike in performance, based on the foundational skills gap analysis, their mean scores reported below suggest otherwise. In fact, ABE girls see a *larger* spike in performance on subtask 6 – a larger gain vis-à-vis subtask 5 *and* a larger decline in subtask 7 – than in-school girls. What is different between the two groups is the relative score at which this occurred, and its placement relative to the thresholds for proficiency versus other categories specified in the foundational skills gap analysis. In short, the finding reported above regarding subtask 6, in the context of ABE girls, does not appear to hold, as ABE girls also recover from their poor performance on earlier subtasks.

The numeracy results also highlight the almost universal lack of facility in multiplication. Unsurprisingly, ABE girls are not able to complete more complex multiplication and division problems reliably. However, they are also generally not able to complete 1-digit multiplication, with a mean score of just 11.7 percent on this task. This score falls below even girls enrolled in grades 2 or below, which is surprising, as girls in these lower grades are presumably not exposed to extensive teaching in multiplication.

FIGURE 52: MEAN SUBTASK SCORES AMONG ABE GIRLS



In the case of Somali literacy, ABE girls are broadly similar to OOS girls and girls in lower grade levels, in that they achieve low-moderate performance in terms of identifying Somali words, but are largely unable to read simple stories. Girls who do moderately well in terms of identifying Somali words still face the relatively common barrier that they are unable to transition into reading full sentences.

8.4 ABE GIRLS – SUBGROUP ANALYSIS

This section provides an analysis of the learning outcomes by key subgroups of the population of ABE girls. The table below summarizes the mean literacy and numeracy scores based on those subgroups that were previously identified as being particularly important during the baseline study of cohort girls. Some sub-categories (such as mothers under the age of 16) are not included because none of the sampled girls belonged to this category. Other categories, particularly within the disability category, have few girls in them so our ability to draw strong inferences about learning in those subgroups is limited. Nonetheless, a few subgroups have emerged as strong predictors of learning outcomes among ABE girls.

TABLE 113: ABE LEARNING SCORES BY KEY SUBGROUPS

	Mean ABE Somali literacy score	Mean ABE numeracy score	Number of observations for subgroup
Characteristics:			

All ABE girls	13	30	482
Somaliland	10	27*	257
Puntland	15	32	195
Galmadug	25	40	30
Disability			
Vision impairment	2*	18	7
Mobility impairment	1*	11*	4
Hearing impairment	2*	14	2
Mental health impairment	1*	11*	3
Anxious	8	36	18
Depressed	7	33	15
Problems with self-care	1*	13*	4
Problems communicating	2*	14	2
Any disability	6*	26	29
HOH and Carer Characteristic			
Living without both parents	12	30	92
HOH no wage-earning occupation	13	29	272
HOH no education	13	29	391
HOH female	16	32	208
HOH Pastoralist	12	26	27
Carer no education	13	30	482
Household Assets			
Owns camels	10	26	52
Owns medium-sized livestock	10*	28	286
Owns small livestock	13	30	482
Owns mobile phone	12*	29*	410
Access to water reservoir/storage	11	30	229
Owns land	13	30	312
Poverty			
House is informal/temporary structure	15	31	80
Gone to sleep hungry most days	14	28	18

Gone without enough clean water most days	**10**	**26**	**52**
Gone without medicines or medical intervention most days	**7***	**25***	**99**
Gone without cash income most days	**10**	**29**	**90**
Migration			
Displaced or moved in past 12 months	**13**	**21**	**6**
Other			
High chore burden (whole or half day spent on chores)	**21**	**35**	**45**
Married	**38***	**59***	**4**
Mother, under 16	**20**	**46**	**5**

*Note, an asterisk indicates results that are statistically significant at the 95% confidence level in a regression with cluster-robust standard errors.

When comparing literacy and numeracy scores at the regional level, ABE girls in Somaliland appear to score lower than ABE girls from Putland or Galmadug. Somaliland numeracy scores were the only comparison that emerged significant at the 95% confidence threshold, however it should be noted that the observed lower literacy scores for Somaliland girls could be considered marginally significant ($p = .011$). This trend is consistent with previous reported baseline scores for cohort and ALP girls.

The overall sample size of girls with various kinds of disabilities is extremely low, making it difficult to draw strong conclusions about the relative effects of different kinds of disabilities on learning outcomes. In fact, reported disability among girls in the sample overall was low (less than 10% of all girls in the sample report any disability). On average, ABE girls with disabilities scored lower on both Somali literacy and numeracy than the other ABE girls in the sample. And, despite the small sample size, their performance on Somali literacy (6 percent average score) was significantly lower than those of other ABE girls (13 percent average score). These girls are likely to face particularly high barriers to catching up to their peers, or transitioning into other educational opportunities.

As we have briefly noted elsewhere in this report, we believe the disability prevalence rate described here is lower than the true rate of disability among the sample as a whole and the ABE sample specifically. While good estimates of the overall disability rate in Somalia are lacking, it is thought that the disability rate in Somalia is likely higher than the global average of 15% reported by the World Health Organisation (WHO), given the country's history of conflict, widespread poverty, and lack of access to healthcare.[324] One explanation for the lower rate could be that our sample population consists of adolescents, who experience disability at lower rates than adults and, therefore, the general population that the WHO study considered.

[324] WHO World- World Bank World Report on Disability 2011

However, even relative to an expected lower rate of adolescent disability, our reported rates are low, which we expect may be attributable to methodological and cultural explanations.Under-reporting of disability could arise from cultural factors and social desirability bias, as individuals – girls, caregivers, or both – wish to avoid being characterized as disabled. Anecdotally, the stigmatization of disability in Somalia is widespread. Additionally, there may be a "self-esteem protection mechanism" at play in under-reporting, as reporting disability can be seen as admitting "weakness" and harm one's self-confidence – again, anecdotal experience suggests that admissions that can be construed as "weakness" may be under-reported in Somalia.

Additionally, the questions used to identify disability are unlikely to capture all but the most severe or obvious disabilities. For assessing mental health, caregivers were asked "How often does [girl] seem very anxious, nervous, or worried" and "How often does [girl] seem sad or depressed". Interpretation of terms like anxiety and depression may vary among caregivers and are not necessarily in line with clinical definitions.[325] For physical disabilities, girls were asked questions about their ability to perform certain tasks (such as "Do you have difficulty raising your arms"). Girls were not categorized as disabled if they responded that they had "no difficulty" or "some difficulty". As a result, this survey may have had a difficult time identifying some mental disabilities or limiting, but not debilitating, physical disabilities.

Note from the project: Given that the Washington Group questions ask about the degree of difficulty in performing tasks rather than about disability itself, the likelihood of a girl self-underreporting or a caregiver underreporting the case is lower. The word 'disability' is not mentioned at all. The questions that ask directly about a 'weakness' – how often the girl is anxious or depressed – actually had the highest reporting rates, countering the hypothesis above. In addition to this, it should be noted that the project conducted an exhaustive canvassing of the communities to identify CwDs for support, preceded by extensive sensitisation and community mobilisation. The results do not show a higher rate of disability – and similar results were found by Consilient in Southern Somalia, under the AGES baseline study.

There is a simpler and more logical reason for the relatively low prevalence of disability – survival rates. Children with severe disabilities resulting from micronutrient deficiency at pregnancy, genetic variations or infections such as meningitis / polio often face serious heart, liver and respiratory system conditions that require extensive treatment and/or corrective surgery, in addition to physical disabilities. In a context where medical care is nearly non-existent, there is a high likelihood that these children may not survive, particularly when exposed to extreme conditions such as malnutrition and displacement. In the case of children who face a disability as a result of violence or accidents, the lack of medical care severely limits their ability to survive, particularly in a situation where minimal hygiene conditions cannot be met during conflict. These are, unfortunately, terrible realities that should be considered in an extreme setting such as Somalia, particularly for an ultra-marginalised group such as ABE girls.

Household characteristics and poverty measures were generally not strong predictors of ABE girls learning outcomes. The strongest predictor among these subgroups were those who do not have access to medicine or medical intervention. Approximately 20% of the ABE households surveyed reported not having access to medicine or medical devices regularly and 35% of households indicated that their girl had a medical condition

[325] The Washington Group questions inclusion of anxiety and depression is a positive improvement over previous approaches that focused exclusively on physical disabilities. However, the still-narrow focus, with regard to mental health disabilities, may contribute to lower prevalence rates in the sample, as the data do not capture some disabilities – such as post-traumatic stress disorder – that we would expect to be more common in a conflictual and post-conflict environment like Somalia.

that prevent their attendance in a formal school. The low access to and high demand for medical care make ABE girls particularly vulnerable to low performance and dropout.

While few of the ABE households surveyed directly identify as pastoralists, subgroups which align with pastoralist lifestyles (such as owning camels or medium-sized livestock) trend toward lower scores for both literacy and numeracy (lower literacy scores among households than own medium livestock were statistically significant).

ABE girls appear to be at much higher risk of underperforming relative to any other type of girl in this sample. ABE girls at all age groups score lower than their ISG or ALP counterparts in both Somali literacy and numeracy. Like the other girls in the sample, ABE girls face many challenges and barriers to education, however it is difficult to pinpoint the specific underlying causes. The high number of households reporting girls with medical conditions that prevent their attendance in formal schools and the low access to medicine and medical intervention appears to be a major driver of this discrepancy among ABE and other cohort girls. Finding additional single predictors of ABE success (or lack thereof) proved difficult, the high underperformance among ABE girls is likely a function of ABE girls facing a multitude of interacting barriers. Households of ABE girls are likely to have a head of household who has not received and education themselves and who is unemployed (many of which go without cash income frequently), are likely to have a caregiver who is illiterate, and which often struggle to meet basic needs. While the effects of any single one of these factors may not individually predict lower performance, taken as a whole, these factors are likely to create a very challenging educational environment for these girls.

9. CONCLUSIONS AND RECOMMENDATIONS

9.1 CONCLUSIONS

Learning

Three general observations emerge from the aggregate learning analysis. Firstly, the findings derived from the analysis of different panels of data suggest that the program has had an impact on financial literacy. When using the pure longitudinal panel of all the individuals who overlap between Baseline and Midline Round 2, the impact is a substantive 8.4 percentage in favour of intervention schools. The impact is nearly of the same magnitude when looking only at the girls who have been enrolled since the baseline, the group most exposed to the program's interventions, including better-trained teachers.[326]

Secondly, while aggregated numeracy outcomes are yet to be statistically significant in terms of the difference in improvement vis-à-vis the comparison group, increases in numeracy outcomes are systematically higher than in the comparison group. Assuming that the impact of the program in this regard may take some time, it is particularly encouraging that the outcomes have markedly improved since Midline Round 1. When looking at the true panel of the respondents who are the same across the three waves of data collection, the gap in learning between intervention and comparisons schools has almost entirely occurred between the two midline evaluation rounds. This bodes well for the endline because it appears that the program is starting to produce the results that were intended. Moreover, the panel consisting of girls who have been enrolled since baseline, and as such most engaged with the teaching, shows that the intervention cohort has improved their results on average by 4.6 points more than the comparison group since the baseline.[327] Thus, we can cautiously expect the gap between intervention and comparison schools to grow as we approach the endline and as girls' exposure to the program increases in duration.

Thirdly, while the program has shown impact in terms of financial literacy, and potentially will do the same for numeracy moving forward, it has struggled to make improvements in literacy outcomes. In fact, depending on the panel used, the comparison schools have often shown more marked improvement since the baseline. It is not entirely surprising that the program has had an impact on financial literacy, over and above comparison schools, as financial literacy is not included in the curricula of comparison schools at all. No similar progress has been observed in terms of literacy outcomes.

With respect to the results of subgroup learning, they generally reflect and confirm those of the aggregate analysis. For instance, numeracy results which are consistently positive – even if not statistically significant – imply that the numeracy gains observed are generally broad-based and not exclusive to a particular group of girls. Similarly, the results for literacy are not being dragged down by specific subgroups; the progress tends to be broadly null across all subgroups, with only minor variations. The same logic applies equally to financial literacy, where program impacts were greatest in the aggregate: in both groups where you would expect to see learning difficulties, as well as among groups where you would expect the students to learn with relative ease, the program has improved financial literacy in the intervention versus the comparison group.

[326] Despite a much smaller sample size, the difference of differences estimate of 8.1 is marginally statistically significant.
[327] This finding is one of the closest to statistical significance, with the p value of 0.126.

Transition

Transition rates have increased sharply since the baseline in both intervention and comparison communities. The program has been reasonably effective in improving transition rates since the baseline, with successful transition increasing 3.2 points in intervention communities, over and above, comparison communities. The impact is not limited to one particular type of girl or one particular transition pathway, either; for instance, it would be of concern if the program's entire impact on transition rates focused solely on the number of girls advancing grades, and failed to improve enrolment rates or address dropout rates. The program, however, has increased within-school grade progression, reduced dropout rates, and increased enrolment into non-formal education, when judged relative to comparison communities.

ABE Girls Baseline

ABE girls face many challenges in terms of attending or staying in school. Girls in the ABE sample cited displacement, health conditions, and not having a school close enough to attend as the primary barriers to their attendance. ABE girls are particularly disadvantaged, even in comparison to OOS and ALP girls, in terms of household environment with respect to the potential for HoH/caregiver support and the economic hardships they face. In most cases, ABE girls' caregivers are illiterate (79.4 percent), unemployed (56.7 percent) or did not have a formal education (81.5 percent). All these factors contribute to the challenges of ABE girls, who often lack support at home and may, thus, find it rather difficult to attend and stay in school. ABE girls also struggle financially; a large proportion of these girls (42 percent) come from a home with a poor-quality roof (an indicator of low economic prosperity). Nearly half of girls reported that they went to bed hungry at least once in the previous year and nearly three-quarters said that they did not have access to clean water at least once in the previous year (71.2 percent). Living under these harsh conditions may make it difficult for ABE girls and their families to make education their priority.

Despite the hardships endured in many ABE girl's homes, we found that these girls and their caretakers generally have attitudes toward education that are positive. Caregivers of ABE girls were more likely than OOS or ALP caregivers to express a desire for their daughter to attend university (77.2 percent) and a relatively low proportion of ABE girls' caregivers expressed that girls' education was not worthwhile (16.5 percent). While ABE girls did not report a high degree of autonomy over input into marriage decisions compared to other types of girls (38.5 percent reported having sole or joint decision over marital decisions), they were found to be much less likely to be wed or become mothers compared to OOS or ALP girls (their marriage and motherhood rates were on par with those of ISG girls at around 1 percent).

Given the unique challenges that ABE girls face and that the ABE program was their first opportunity to attend formal school, the learning results were generally poor (in absolute or relative terms). The gaps between ABE girls and their ISG or ALP counterparts are large in Somali literacy (-49 percent compared to ISG girls and -35 percent compared to ALP girls), numeracy (-30 and -23 respectively), and financial literacy (-17 and -15 respectively). Furthermore, these huge differences did not seem to be very sensitive to age as the gap did not shrink much by the time a girl was 18-19 relative to younger girls.

Perhaps the most surprising finding in the ABE baseline came from analysis of the Somali literacy sub-task scores. While the existence of a large literacy gap between ABE girls and their peers was expected, we did not expect the gap to manifest itself as the most basic levels of literacy. Only 17.8 percent of ABE girls achieved proficiency in identifying and reading a set of common Somali words with the median girls only being able to read 13 of the 60 words they were presented with. This result suggests that ABE girls need the most basic of literacy training and face a steep learning curve if they are to close the gap with their peers. The results of the numeracy sub-task analysis yielded results more in line with expectations in that the majority of ABE girls are considered to be established learners in the most basic numeracy sub-tasks (identifying a missing

number, basic addition and subtraction) and start to fall behind as the sub-tasks become more challenging. By the time girls are asked to do multiplication, most of the them (82.2 percent) are categorised as non-learners, scoring 0 percent.

Attendance

The program appears to be having a positive impact on attendance rates, in general, though these effects are noisy in a statistical sense. The attendance analysis used three different quantitative data sources—derived from entirely different data collection tools—to triangulate attendance rates at ML2 and in previous rounds. The findings across these sources were consistent: attendance rates in intervention schools had increased vis-à-vis comparison schools in small but meaningful ways. Our best estimate of the program's impact on attendance is that girls' attendance has increased 1.4 percentage points from baseline to midline. From a starting point at which attendance rates were between 80 and 85 percent, an improvement of 1.4 points is substantively meaningful, though smaller than might be desired after two years of program implementation.

There is no evidence that the program is having particularly large or minimal effects among specific subgroups of girls or schools, aside from the fact that the program's impact is significantly higher in Somaliland than in Puntland. On the other hand, intervention-specific effects seem to confirm the impact of the program, as girls participating in GEF activities have improved their attendance at a higher clip than either girls in comparison schools or girls in intervention schools who do not participate in GEF activities. Among girls active in their local GEF, the program has improved attendance rates by 6.1 percentage points.

Teaching Quality

When looking at the indicators of teaching quality, there is some reason to be concerned about the overall effect of the intervention on teaching quality. While perception of teaching quality remains high (greater than 95 percent of caregivers rate teaching quality to be good or very good in both comparison and intervention schools), indicators of student-centred pedagogy do not support this optimism. In intervention schools, improvements to learning centred pedagogy have been found in a reduction of students spending most time repeating words aloud (-7.1 percent from baseline to midline 2), a reduction in spending most of the time copying from the board (-1.9 percent) and an increase in the frequency of students instructing one another (+2.9 percent). However, observed use of the remaining five learner-centred approaches was actually lower than during the baseline. When all indicators were aggregated into an index of classroom activity and participation, schools scored higher in the baseline than they did during the second midline.

Self-reported use of formative assessments increased dramatically from baseline to the second midline with greater than 90 percent of intervention and comparison schools reporting their use. However, the proportion of teachers who reported having documentation of their use remained unchanged. Given the low proportion of teachers who reported being able to provide documentation and the decrease in the use of active teaching strategies observed, there is reason to be cautious about widespread use of formative assessments among intervention classrooms.

Similarly, observed use of physical punishment in the classroom has gone from 'common' at the baseline to 'unobserved' during ML2. However, students reported that the use of corporal punishment in the classroom actually increased from the baseline to Midline Round 2. Intervention girls also reported that the use of corporal punishment by teachers in the classroom rose from 23.2 percent at baseline to 30.9 percent at ML2,

whereas, students at interventions schools who reported that their teacher would punish them for getting things wrong in a lesson[328] increased from 40.8 percent to 54.7 percent.

These results are discouraging given the general increase in teaching quality observed during the first midline. They suggest that without persistent reinforcement of teaching best-practices, long-term sustainability of program interventions may be diminished.

School Management and Governance

For school management, the most important variables – i.e. whether the school has a functioning CEC or not, or whether the school has a management plan – has improved since the baseline. According to the head teacher surveys, all intervention schools now have an active CEC, and nearly 72 percent of intervention schools have developed their management plans at ML2 while at ML1 only 53.1 percent had. The same increase is not witnessed in comparison schools where the schools that have a management plan remains nearly unchanged, at 41.9 percent.

Secondly, intervention schools tend to have a much higher baseline for most of the indicators, and in fact report a higher share for all but one indicator at ML2.[329] The higher levels of baseline activity in intervention schools have two important consequences. Namely, it is possible that we are witnessing something like diminishing returns in the results. In other words, it is possible that seeing an increase from 20 percent to 40 percent is possible with relatively limited efforts compared to an increase from 60 to 80 percent. Indeed, comparison schools had a baseline level of less than 30 percent for half of the indicators, and that of less than or equal to 50 percent for all of them. Meanwhile, intervention showed a level of more than or equal to 50 percent for five of the eight indicators. In addition, when considering why these variables are measured here, the question of the program's impact on these indicators becomes less important. To elaborate, while it may not be possible to attribute higher CEC activity to the program per se, this point is not entirely relevant. For sustainability of the program may, in fact, be better that the CEC activity is not dependent upon the program. The more independent the CEC, the more sustainable the program. As such, the fact that the intervention schools continue to have a much higher level of CEC activity as reflected in the responses to the questions analysed above, bodes well for the program.

When looking at the program's goals as per the logical framework, the findings are mixed. One of the goals was to increase CEC monitoring of student attendance to 80 percent by ML2. The share of CECs doing that currently, according to the caregivers interviewed, is only 68.2 percent. The findings are largely similar when looking at head teacher surveys where 65.6 percent report CEC monitoring of student attendance. In terms of monitoring retention, only 15.9 percent of parents in intervention areas say that the CEC is engaged in this and 37.5 percent of head teachers share the same opinion.

Meanwhile, the program has met its objective of 30 percent of CECs taking action against corporal punishment and other child protection issues. The share of intervention school head teachers reporting the CECs addressing child protection issues is 78.1 percent. Similarly, the head teachers stated that the CEC

[328] Note from project: The actual question is indirect and does not refer to the student herself. The actual wording is *"Do your teachers discipline or punish students who get things wrong in a lesson?"*

[329] We cannot say with certainty whether this difference in baseline outcomes relates to program activities already implemented at the time of the baseline, random sampling variation, or the impact of SOMGEP phase I activities. The latter explanation is the most straightforward and likely, in our view, given the length of time SOMGEP has been working in some of the targeted communities.

reinforces the use of non-violent discipline as opposed to corporal punishment in 68.8 percent of the intervention schools at ML2.

As per the final indicator relating to school management in the logical framework, the CECs' practices taking into consideration the needs of marginalised sub-groups, such as pastoralists, members of marginalised groups, IDPs and the disabled, the CEC members seem to be quite aware of the difficulties faced by these groups. However, most committee members viewed addressing barriers to their education as beyond their means due to the very severely limited financial resources at their disposal. Although findings from the head teacher survey indicate at least partial improvements – as 78.3 percent of school management plans in the intervention group now include plans to follow-up with dropouts – this, nonetheless, calls into question the feasibility of some of the factors required for the long-term sustainability of the project. In particular, while the awareness of barriers relating to enrolment and retention as it pertains to marginalised groups (pastoralists, disabled girls, members of minority groups) is high, targeting the most marginalised is largely seen as beyond the means (in terms of financial or other resources) of the CECs. As such, it is most likely that members of these groups will continue to fall through the cracks.

Life Skills

As with several other outcomes (e.g., attendance, learning) studied in this evaluation, the program's impact on life skills appears to be positive, meaningful in a substantive sense, but too small to distinguish it, statistically, from a null result. Using CARE's Youth Leadership Index (YLI) as a metric for problem-solving abilities, self-confidence, organisation, and ability to motivate their peers, the program has increased girls' scores by around 1.9 points on a 100-point scale. While this effect is small, it is fairly clear – cohort girls in intervention and comparison schools had nearly identical scores at baseline, and girls in intervention schools have improved markedly since that time, while scores for girls in comparison schools have remained flat. Notably, we interpret these results unequivocally as the program's *causal* impact on YLI scores: thanks to the quasi-experimental design, the findings are intended to represent the program's true impact; moreover, our analysis suggests that girls with greater exposure to programming—those who remained enrolled across rounds—saw larger improvements in self-confidence, consistent with the program itself being the mechanism of change. As with attendance rates, we find that participation in the GEF was associated with even larger gains in self-confidence and leadership skills.

Our results are less clear-cut when it comes to the life skills index, which differs from the YLI in the specific questions asked, but often has conceptual overlap with the YLI. On this second index, the program does not show any meaningful impact over time. However, when we split the life skills index into its constituent parts, the evidence suggests that the program improved the self-confidence aspects of the life skills index. The conclusion that emerges is that the program—and the GEF intervention in particular—has had a positive impact on girls' self-confidence and confidence-adjacent outcomes, such as willingness to speak up at home and in school. But the program has had less, or no, impact on other aspects of "life skills", such as feelings of loneliness, agency over life decisions, and desire to stay in school.

Community Attitudes

In the context of the SOMGEP-T evaluation, community attitudes are expected to have effects on transition rates by encouraging re-enrolment of OOS girls and continued enrolment for girls already enrolled. Community attitudes have a more varied, but similarly important impact on learning outcomes: parents who visibly value girls' education will encourage higher attendance, increase the importance girls place on their own education, and assist them with their schoolwork, where possible.

The value that caregivers place on girls' education has improved as a result of the program, though the improvements have not been dramatic. In general, improvements have been largest from the perspective of girls: the program has increased the share of girls who feel they receive support from their family to stay in school and perform well by about 7.2 percentage points since baseline, over and above the improvements seen in comparison schools. Caregivers also report an increased role of girls in making decisions that influence their education, and an increased belief that a girl's education is a worthwhile investment, even if funds are limited.

However, the impacts among caregivers are of small or moderate size—none are statistically significant, though some approach that standard—and head teachers' impressions of attitudes in their community have not improved as a result of the program. Specifically, the share of head teachers who perceive positive attitudes among mothers and fathers in their community has not changed as a result of the program.

School-Related Gender-Based Violence

Girls attending SOMGEP-T schools are not monolithic—their experiences vary dramatically based on the specific contexts in which they live. On the whole, perceptions of safety at and during the journey to school, among both caregivers and girls, seem to be improving over time. However, our quantitative metrics of safety are blunt, and the qualitative interviews demonstrated that girls often report feeling safe but, under additional questioning, readily reveal significant safety concerns. Triangulating from the reports of fieldwork researchers working in these communities also reveal specific instances of violence that confirm girls' reports, and highlight the variation in girls' experiences depending on the specifics of their communities. Some girls go to school with soldiers stationed in their schoolyard; others have experienced school closures of a week or longer due to conflict in their communities; still others have seen a parent come to school with a gun.

Outside of these more extreme cases, girls generally face multiple distinct risks in their communities and at their schools, though they tend to be unified by their gendered nature. They report being harassed by boys, spied upon while using the toilet (or, equally worrying, they avoid using the toilet out of fear they will be spied upon), and verbally harassed or gawked at by men in specific parts of their village and along the open roads they often travel to get to school. The theme here is one of harassment by men, whether their peers, older children, or adults. Given the program's emphasis on girls' self-esteem, leadership skills, and empowerment—and the concerns raised elsewhere in this report about girls' willingness to speak up in class (especially girls' "shyness" when interacting with male teachers)—safety concerns of this kind are especially problematic. They likely undermine the willingness of parents to continue sending their daughters to school, and reinforce the shyness that inhibits girls' participation in class and their learning outcomes.

Sustainability

When looking at CEC financial support as a measure of community level sustainability, the findings are somewhat bifurcated. The head teacher surveys indicate that while CEC share of paying for teacher salaries increased between Baseline and ML1, it has since declined. Moreover, the level of scholarship support across schools is currently more than 20 percent lower than at ML1. The CECs' share of the scholarships has also declined very slightly. Yet, it is possible that this is due to increases in other sources of funding for the schools. Both qualitative interviews and household survey data in fact suggest that CECs have been more active in raising funds and in other forms of financial support since the baseline. Nevertheless, fewer community members viewed their communities as having a functional CEC. Similarly, when looking at CEC functionality and level of activity, the different data sources paint opposing pictures. According to head teachers, there are more functional CECs now, school management plans are more common, and monitoring visits by the CEC have increased across the sample. Conversely, community members reported fewer active CECs while

reporting some increased frequency in the communications of existing CECs. Given the discrepancy in the different data sources, the findings regarding sustainability at the community level are inconclusive.

At the school level, similarly, the evidence does not conclusively support that the use of formative assessments would have become more prevalent since ML1. While the reported use of the assessments had increased, this might be a result of social desirability bias rather than anything else – as teachers know they *should* use the assessments in their work, and report accordingly. The share of teachers who can show records of using them has not changed significantly since the last round of data collection. Yet, the differences are markedly more positive in intervention schools, which would indicate that the sustainability in them is increasing.

At the system level, talk of ALP specifically is very limited in interviews with MoE respondents. Many respondents mention efforts for including the nomadic population and disabled children. But most respondents do not view these as sufficient at this time, due to financial constraints and other operational challenges. When looking at MoE activities, the findings there are somewhat positive. Some evidence suggests that the gender units in the ministries are more active than at Midline round 1. MoE activities seem to be more focused on improving teaching quality and addressing issues of retention and transition. The ministries are increasing their monitoring, seeking funding for teachers and training more female teachers. Specific activities have also been designed for pastoralists and disabled children. The emphasis on training more female teachers seems to have increased across the three states. Overall, the qualitative information for the system level changes at the MoE are somewhat encouraging but inconclusive to show consistent improvements across the board. When assessing CEC activity in the communities as a proxy for the MoEs' efforts, this seems to have increased visibly across the board. Schools are much more likely to have a management plan, and CECs are much more likely to conduct monitoring visits. These changes have taken place in both intervention and comparison schools which would suggest that the MoEs' work can explain the positive changes, at least partially. Thus, it can be cautiously interpreted that some of the system level changes required for sustainability are coming into place.

Yet, the financial sustainability of these activities could not be established, and this ought to be a concern as the respondents continue to list many challenges relating to funding. At endline, it will be important to establish the source of funding for different elements of the MoEs' work that relate to the sustainability of this project.

Overall, thus, it seems that some gains in way of increasing the sustainability have been made when looking at the system and school level. However, the measures on sustainability at the community level are more inconclusive. Nevertheless, much work remains in order to ensure the sustainability of the program.

9.2 Recommendations – Monitoring, Evaluation, and Learning

Measurement of Transition

A key difficulty in studying the impact of the program on transition rates is the extent of attrition among the sample of OOS girls. In the absence of a logical mechanism for selecting new OOS girls into the sample, the share of OOS girls in the primary cohort has declined precipitously from baseline (n = 754) to ML2 (n = 368). Given the program's emphasis on enrolment of OOS girls into alternative learning outlets, such as ABE centres, a dropping number of OOS girls in the primary cohort sample is arguably unfair to the program, as it reduces the weight placed on a group within which program impact may be especially significant. Of course, large-scale attrition is also methodologically unsound, presenting inferential problems across many aspects of

our analysis. At the endline, an explicit component of the evaluation should be seeking to re-contact any and all girls who fell out of the sample since baseline, even if by telephone. A similar tactic undertaken in this round applied only to in-school girls, as no OOS girls had fallen out of the sample prior to this round; however, this approach should be expanded to all girls who fell out of the sample at any previous round. Moreover, the procedures should be expanded – for girls who cannot be located, their information should be passed to office-based researchers, who can call the given contact numbers for households periodically; in addition, head teachers should be pressed for additional information about the girls and their families, so that they can be contacted. This effort will take additional time. However, girls will not be replaced at endline, freeing up fieldwork time that would otherwise be spent interviewing replacement girls. By maximising the number of girls who can be tracked—even if only for the purposes of assessing their transition status—the rigor of the endline evaluation will be improved.

A limitation in the transition analysis also arises when girls leave and re-enter the sample, because they cannot be included in the full analysis across multiple rounds of data collection. In the current evaluation, this applied most prominently – but not exclusively – to OOS girls, who were not contacted at ML1. As a result, transition analysis for OOS girls only considered their status at baseline and ML2. However, this problem will become more noticeable at endline, as girls who are successfully brought back into the sample from baseline or ML1 lack one or more rounds of data on transition status.[330] At the endline, a set of transition questions should be added to the household survey to retrospectively capture enrolment status and grade level for each year since the baseline. Capturing this status for each prior round – and cross-checking it against the data collected in the past – will facilitate analysis of transition across all four rounds for a relatively larger sample of cohort girls.

Measurement of Formative Assessment Use

Given that an increase in the use of formative assessments is one of the primary goals under the teaching quality component of this project, future evaluations should develop more effective means of assessing their use. While it may be onerous, asking teachers to produce documentation of the formative assessments (for example asking to see exit tickets) or looking at teacher lesson plans to see if teachers are building in opportunities to elicit and incorporate student feedback may be necessary in order to properly evaluate the use of formative assessments during instruction. Additionally, training enumerators to better understand formative assessments and how to look for teachers using feedback from students during instruction may allow for a more reliable way of assessing the use of formative assessments in classrooms. Lastly, incorporating questions during teacher FGDs about the strategies they use to formatively assess students and how they approach incorporating student feedback into their lessons and lesson plans may help shine a light on if teachers are using formative assessments in their classroom. For instance, teachers should be asked what proactive steps they take to help students who are falling behind when planning their lessons, during class, and outside of class. They should also be asked how they are incorporating formative assessments into their instruction and how, specifically, the results are being used to guide instruction and lesson plans. Alternatively, some of this information could be derived from alternative sources, including surveys with head teachers.

[330] For instance, consider the 359 OOS girls who fell out of the sample between baseline and ML2, because they could not be located in the latter round. If 120 of these girls can be re-located during the endline evaluation and all 368 successfully re-contacted OOS girls at ML2 are also re-contacted at endline, it would increase the sample size of OOS girls tracked from baseline from 368 to 488. However, 368 of the girls would have data for baseline, ML2 and endline, while the remainder would have data only for the baseline and endline. Both samples could be fruitfully analyzed, but the analysis will become needlessly disjointed and nuances in trends across rounds would be lost. If retrospective data were collected for all OOS girls, the full sample of 488 girls could be studied across all four rounds, even if this approach is only used as a robustness check.

Measurement of Community Attitudes

As it stands, the evaluation of SOMGEP-T has sought to use the best possible measures of community attitudes, and triangulate these findings across multiple types of respondents and multiple question types. Measurement of community attitudes has become more nuanced and extensive since the baseline. However, there are still shortcomings in how community attitudes are assessed. Currently, questionnaire design does not make explicit distinctions between high- and low-performing students when asking caregivers about the value of girls' education. For instance, when a caregiver is asked "Is girls' education worth investing in, even when funds are limited?" they can simultaneously agree strongly in principle but advocate removing their own daughter from school because she is performing poorly and they perceive *their* girls' education to be fruitless. The available evidence suggests that community members and caregivers may be making this distinction implicitly when asked about girls' education – focusing on the highest-achieving girls and emphasising the value of girls' education in those terms.

At the endline, we recommend attempting to distinguish between attitudes toward high-performing girls and toward those who have not performed well or who are not expected to continue into the workforce. Specifically, we suggest adding nuance to existing questions regarding whether investing in girls' education is worthwhile and whether girls are just as likely as boys to use their education, by framing the questions with reference to whether the girl is performing well or poorly in school. We also suggest more targeted probing questions be added to the qualitative tools – alongside training to emphasise how to use these probes – to assess whether broad stated support for girls' education extends to girls who have fallen behind for their age, or are performing poorly. Finally, further analysis of the relationship between girls' performance and community attitudes may be worth considering, including how community attitudes change in response to high performance for just one or two girls, and how caregiver attitudes change in response to positive improvement by their own girls. Efforts of this kind will provide insight into the dynamics of attitudinal change. They may also provide insights that can be useful for a future iteration of GEC or LNGB programming in Somalia.

Improved Use and Collection of Qualitative Data

At the outset of the endline evaluation, specific effort should be dedicated to rewriting the qualitative tools with new purpose. At baseline, the qualitative tools were designed largely to understand the existing situation in communities, identify barriers to girls' education, and understand subgroups that faced the most significant barriers to enrolment and learning. This goal is reflected throughout the qualitative tools. However, as the program has progressed, the goal of the evaluation has shifted from understanding the existing context within communities and guiding targeting of the program toward assessing what changes have taken place since baseline. Before the endline, the qualitative tools should be revised to emphasise this goal, asking interviewees what has changed in their community since baseline, how attitudes toward girls' education have changed and among which groups in the community, whether girls are more likely to enrol and stay in school now than they were previously, and whether this progress is uniform across all girls. Beyond the different goals of baseline, midline, and endline evaluations, the qualitative tools should also be revised simply because conducting qualitative interviews over three rounds has given us more information about what types of questions work and which do not – learning that should be more actively incorporated into tool design in the next round.

Qualitative analysis in this evaluation round focused predominantly on differences across subgroups, i.e. cross-sectional differences in learning or transition, many of which stem from structural or systematic gaps that existed before the program began and are likely to continue – even if in muted form – at endline. However, by reorienting the qualitative tools toward changes in communities, the qualitative analysis at

endline can shed light on program impact, differential impacts across groups, and the barriers that have either continued to inhibit progress or have been overcome. In short, the qualitative analysis will be better aligned with the retrospective, evaluative goals of the endline.

Approach to Assessing Impact

As we have noted repeatedly in this and previous report rounds, the difference-in-differences approach to evaluating SOMGEP-T allows for a rigorous assessment of program impact, requiring comparatively few assumptions to draw valid inferences. However, the nature of interventions like SOMGEP-T is that they are multi-faceted, in the sense that they bring together many different activities, and are geographically clustered. At the same time, many of our most interesting analyses were those that studied the impact of a specific intervention (e.g., GEF participation) on outcomes, because the expected impact of that activity is very closely tied to specific outcomes.

In future GEC-style programs, a more targeted approach to assessing impact may be worthwhile. For instance, if we expect community outreach campaigns to alter local attitudes, a randomised-rollout, dose-response, or other approach to assessing the impact specifically of the campaign on attitudes may be preferable to annual or biannual evaluations, where impacts are obscured by noise and, in some cases, insufficient statistical power. While such an approach would not capture cross-fertilisation between different activities and their interactive effects, the design may make them more tractable and, in most cases, less expensive to implement.

Refinement of Logframe Indicators

Several of the logframe indicators present difficulties in operationalization. First, there is a lack of clarity in the definition of the learning indicator, specified as "number of girls sampled who demonstrate increase in their SEGRA/SEGMA scores." This definition is unclear because nearly every girl improves over time with regard to these scores, due to maturation effects; moreover, if a comparison were to be made to a specific girl in the comparison group, each girl should have been assigned into a matched-pair design. We recommend reframing the indicator with explicit reference to the comparison group, by making the indicator the difference (in percentage points) in SEGRA/SEGMA scores between intervention and comparison groups. For each evaluation round, the relative gap between intervention and comparison groups can be set with specific targets.

More problematic, in many ways, are the qualitative-oriented indicators of attendance, teaching quality, life skills, and school management. In general, these indicators were not explicitly addressed in the baseline report, and it does not seem that the tools were generally designed to target these qualitative indicators. At times, proxy indicators have been used to fill the logframe (e.g., a quantitative proxy indicator for girls' feeling empowered to speak up). However, agreement should be made going forward on whether these proxy indicators will constitute the actual logframe indicator, or whether qualitative evidence will be used to assess if the target has been reached. If the qualitative indicators are to be used at endline, we recommend a longer-than-normal tool development period, with either CARE or the external evaluator developing qualitative questionnaires that more concretely address these indicators, and with additional review by the other parties and the FM. Whereas the qualitative tools currently employed are extremely useful for understanding barriers to girls' learning and transition, they are less suited to measuring changes in the IOs over time, and this will need to be remedied prior to the endline evaluation.

9.3 RECOMMENDATIONS – PROGRAMMING

Given that this report covers the last evaluation round prior to the endline, and there will be limited opportunities to implement large-scale changes prior to the close of the program, our recommendations related to programming are of two types. The first type are actions that could feasibly be taken between now and the planned winding down of SOMGEP-T interventions in late 2020 or early 2021. The second type are those that are potentially more ambitious or less feasible in that timeframe, but which may be useful for informing future educational programming in Somalia, whether implemented as part of the GEC or not.

- Both analysis of aggregate learning outcomes as well as the sub-group analysis of learning outcomes seem to suggest that the program will be able to show positive impact—at the endline—in terms of financial literacy. The program has also made significant progress in the field of numeracy, as much improvement in learning scores has taken place since ML1. As it pertains to literacy outcomes, however, no impact can be detected at this stage. Prior to the endline, we recommend that the program undertake a final push of activities designed to boost literacy outcomes, alongside additional, targeted pedagogical training for teachers focused on literacy. This recommendation relates to those, below, regarding teaching quality and strategies for addressing foundational gaps in literacy, such as letter recognition.
- A second recommendation regarding literacy concerns future investments in girls' education programming in Somalia. We recommend investigating *why* the program seems to not have achieved the same level of impact in the field of literacy without any evidence of neglecting it in the programming. This might require a more comprehensive profiling of the comparison schools to identify whether and what kind of other education investments are taking place in those areas.
- New data in this evaluation round revealed that there is a correlation between the cognitive function – in terms of working memory – of children and their dietary diversity, and consistency of food consumption.[331] While a full-scale school feeding intervention is well beyond the scope of the program, targeted supplementation with protein sources may be worth considering, either over the next year or in any future stages of GEC-T implementation in Somalia. Supplementation of diet with a daily protein source would be cost-effective and potentially yield gains in cognitive development.
- ABE girls report experiencing indicators of poverty and an inability to meet basic needs (low quality roof, going to bed hungry, and lacking clean water), have the highest likelihood of having an unemployed head of household, and are more likely to report having a medical condition that prevents them from attending school. If future program interventions and programs target economic well-being or health/nutrition, ABE girls may be the group that benefits most from these interventions.
- Many interviewees continue to call for training of the CECs. The lack of CEC member capacity was identified as a source of contention between the CECs and the head teachers during the ML1 evaluation round. In this round, some community members, teachers and CEC members themselves continue to lament the lack of committee member awareness and capacity as a barrier for their work being more effective. The program could thus conduct another round of training of committee members before the end of the project.

[331] This finding is consistent with at least some of the existing literature. See, e.g.: Whaley, et al. 2003. "The Impact of Dietary Intervention on Cognitive Development of Kenyan School Children." *The Journal of Nutrition* 133 (11): 3965S-3971S.

- As it pertains to addressing broader barriers in education, we suggest exploring the current adoption of best practices among MoE respondents, and promoting their broader adoption into institutional mandates. Namely, in some areas in Somaliland, the REOs are ensuring that when a pastoralist child migrates, the school records of the child are sent to the new school in order for the child to continue from where they had left off. While sustainable and systematic change might require a more comprehensive elimination of the educational barriers faced by pastoralists and children with disabilities, such efforts require more resource-intensive programming, while the efforts noted above by MoE employees can be implemented on a smaller scale and with a smaller footprint.
- In each round of the SOMGEP-T evaluation, the most common disability indicated by girls and their caregivers relate to frequent anxiety and depression, with anxiety reported by a slightly higher share of caregivers than depression. Although we are not able to provide strong empirical evidence that experiences of conflict or specific instances of violence are correlated with anxiety rates in this sample, that relationship is both theoretically justified and supported in the broader literature.[332] While providing mental health counselling to students in all program schools is likely cost-prohibitive, some aspects of mental health counselling could be integrated into GEF activities, and efforts could be made to target more specific counselling services to students in communities known to have been impacted by violence (e.g., the ongoing clan conflict in Ceerigaabo). Similarly, girls in schools known to have been impacted by widespread hunger—which can be identified partially based on data collected during this evaluation—could be targeted for mental health support designed to ease the anxiety that arises from such long-run uncertainty and stress.
- Large skill gaps in both Somali literacy and numeracy were observed between ABE girls and their ISG and ALP counterparts. However, the skills gap in Somali literacy is especially troubling given that a high proportion of ABE girls struggle with the most basic skills needed to recognise words (a precondition for reading comprehension or writing). Aggressive approaches to teaching or remediating basic literacy skills may be necessary for ABE girls to be able to achieve significant progress and close skill gaps (especially given the low rates of formal education and literacy among ABE HoH/Caregivers). There is a large body of research on the link between developing foundational literacy skills at a young age and future academic success.[333] Future interventions with administrators and teachers may focus on training educators in teaching and reinforcing foundation literacy skills such as phonological awareness and developing curriculum that addresses the specific needs of girls struggling with literacy.[334]
- Educators have become increasingly aware of the importance of teaching literacy in all subjects, not just language arts. For example, when common core standards were adopted in the United States, literacy requirements were included into all subject area teaching standards (for a closer look at K-12 literacy standards under common core in the United States, see: http://www.corestandards.org/ELA-Literacy/). Given the lack of foundational knowledge in the most basic reading skills, working with school administrators and teachers to help teachers of all

[332] Ayazi, Touraj, et al. 2014. "Association between exposure to traumatic events and anxiety disorders in a post-conflict setting: a cross-sectional community study in South Sudan." *BMC Psychiatry* 14; Bronstein, Israel, and Paul Montgomery. 2011. "Psychological distress in refugee children: a systematic review." *Clinical Child and Family Psychology Review* 14: 44-56.
[333] Catts, H.W., Fey, M.E., Zhang, X., & Toblin, J.B. (1999). Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. Scientific Studies of Reading, 3,331-361.
[334] Most people who struggle with recognising printed words struggle with phonological awareness (decoding and encoding) written words. See: Moats, L.C. 2001, *When Older Students Can't Read,* Educational Leadership: Journal of the Department of Supervision and Curriculum Development, N.E.A 58(6) 36-40

subjects to incorporate lessons which teach and reinforce basic literacy skills may help narrow the gap.

- The index of active and participatory classrooms decreased from the first to the second midline rounds in intervention schools. While certain indicators of an active classroom remain high such as not spending most time copying from the board (64.6 percent), teachers asking open-ended questions (60 percent), and not spending most time repeating words aloud (80 percent) other indicators of the student-centred approach like students working in groups (16.9 percent), teachers asking for student opinions (47.7 percent), students instructing one another (47.7 percent), and the use of student-centred games/activities (13.8 percent) remains low. Refresher trainings and providing teachers with example curriculum/lessons that incorporate a more student-centred approach may be necessary to ensure that teachers do not fall back into the teaching strategies they may be most comfortable with.
- Reported use of formative assessments increased drastically in both comparison and intervention schools from baseline to midline (more than 90 percent of teachers in both schools reported using formative assessments). However, documentation of formative assessments remains low (only 38.5 percent of teachers at intervention schools report that they would be able to provide documentation of formative assessments). Given the infrequent use of student-centred activities that would lend themselves to teachers eliciting feedback from students and assess their knowledge during learning, there is reason to be cautious about the use of formative assessments to guide instruction. Additional training on the importance and utilisation of formative assessments as well as strategies for incorporating formative assessments into instruction may be necessary.
- Girls Empowerment Forums stand out as one of the most effective SOMGEP-T activities. While it is difficult to attribute improvements in life skills and attendance directly to girls' participation in GEF activities, as girls can self-select into participation or not, it is clear that girls participating in GEFs have gained greater self-confidence than other girls. Particular effort should be made to ensure the sustainability of GEFs after the end of SOMGEP-T programming should be encouraged. If possible, their expansion through recruitment of additional girls should also be encouraged.
- For schools in rural areas with open spaces or dangerous roads connecting the school to its respective community, or in which girls travel long distances, organizing a way for them to travel to school in safety is needed. There are many mechanisms by which this could be achieved. Teachers could walk with students, though this asks teachers to put in additional time and effort that they may resist. Parents could also be organized, via the CEC or the Girls Empowerment Forums, to walk with students who live in their part of the community, perhaps on a rotating basis. This could serve to reinforce the value of the GEF as an institution and give it longevity after the conclusion of SOMGEP-T's intervention. It is also a genuine exercise in empowerment, fitting with the mission of GEFs. The ideal approach might be to have the BEFs and GEFs at the same school coordinate this effort, to improve boys' understanding of the risks girls face and promote cross-gender camaraderie.
- While it is not part of the school management and governance target, CECs are also responsible for safeguarding children and ensuring child protection policies are upheld within their schools. Currently, a majority of CECs reinforce the use of nonviolent discipline in schools and address child protection issues. However, the trend since ML1 has only been marginally positive. Given the CECs' role as the link between parents and the school, their participation in child protection issues should be emphasized going forward. The evidence suggests that active CECs are able to influence outcomes in the areas where they focus (e.g., attendance) and their active participation is appreciated by parents, based on qualitative interviews. A child protection group could be formed

in each school, bringing together the head teacher, 2-3 CEC members, and one of the traditional elders in the community, which would allow them to resolve conflicts that occur in school or spill over from the community into the school. They could also raise awareness among the community on child protection issues and the importance of reporting issues that occur in the school.

- Efforts to improve attendance should focus on the role of domestic work in reducing girls' available time and energy for schoolwork and their ability to arrive on time and stay in the classroom consistently. Community awareness around the importance of girls' education appears to have improved; at this stage, it is important that adults recognize that sacrifices beyond the financial may need to be made to ensure their girls stay in school; namely, household chores can be divided more evenly. If communities shift toward this way of thinking – emphasizing greater equity in the distribution of household work among children and between children and adults – it could generate social pressure to avoid saddling adolescent girls with too many responsibilities.
- The four biggest pathways in which transition rates have improved as a result of the program are enrolment of OOS girls into school, movement of OOS girls into informal education, reducing the number of in-school girls who drop out and move into informal education, and increasing the share of girls who remain enrolled and progress a grade. At this stage, additional attention should be placed on increasing grade-to-grade progression for girls who remain in school. As learning scores improve, this should naturally be tied to higher pass rates; increased progression can also motivate girls to continue their education, and assuage their family that progress is being made. While it is important that standards are not reduced to accommodate this, increasing grade progression can also reduce dropouts and may have positive downstream effects in other areas of programming.
- At the school level, there is a small but meaningful relationship between attendance and availability of gender-segregated toilets. Access to separate toilets is substantially more impactful on girls' attendance than it is on boys' attendance. This trend is supported by qualitative data, in which respondents cite a deficiency in the availability of facilities required for girls during menstruation. While infrastructural inputs such as the provision of latrines across all targeted schools may remain beyond the scope of Care activities, highlighting the importance of gender-segregated facilities within CEC's and MOE departments may encourage local resource mobilization. Should CEC members (and other local proponents of girls education) better understand and articulate the significance of access to female-only bathrooms on girls attendance (and in turn education outcomes), this particular challenge may be addressed through locally driven action.

ANNEXES

ANNEX 1: INTERVENTION ROLL-OUT DATES

TABLE 114: INTERVENTION ACTIVITIES AND IMPLEMENTATION TIMELINES

Activities	Start	End
Output 1: Improving access to post-primary options		
Meetings with MoEs, specialists and other stakeholders to develop ALP model	October, 2017	December, 2017
MoE subject specialist workshop to develop ALP modules	January, 2018	February, 2018
Validation, translation, production and distribution of ALP modules	July, 2018	September, 2018
Roll out of ALP classes (ALP implementation)	September, 2018	Ongoing - end date October, 2021
Develop girls' life skills in upper primary through ALP, including leadership skills, financial literacy and business selection and management of income generation activities; participation in Girls' Empowerment For a	July, 2018	Ongoing - end date October, 2021
Expand ALP to enrol 2,345 OOSG within an additional 34 villages, increasing ALP coverage from 76 villages to 110 villages	June,2019	Ongoing - end date October, 2021
Provide 2 years of Alternative Basic Education (ABE) classes for 2,029 marginalized girls (older girls who have never attended school, particularly girls from displaced families) and link them with existing schools to join formal education depending on learning achievement	August 2019	December 2020
Training CECs across 199 villages in identification of different type of disabilities and support to girls/boys with disabilities	April 2019	July 2019
Work with CECs to liaise with parents of displaced OOS girls and girls with disabilities, provide	April 2019	Ongoing - end date October 2021

targeted social support and track their attendance		
Assessment of girls with disabilities for placement in regular schools or referrals to special needs facilities	February 2020	Ongoing - end date October 2021
Training of CECs to improve retention and transition (33 additional secondary schools)	February, 2018	July, 2018
CEC Coaching on improving retention and transition (199 schools)	March, 2018	Ongoing – end date September, 2020
Provide partial grants to girls from poor families	November, 2017	Ongoing – end date March, 2020
Equip and enrol girls into boarding schools	Not started	N/A
Output 2: Supportive school practices and conditions for marginalised girls		
Train teachers on improved delivery of literacy and English language, supported by digital content in all 148 primary and 55 secondary schools	February, 2019	Ongoing – end date May, 2019
Recruitment of consultant to develop manual and train teachers on improved delivery of numeracy	October, 2017	December, 2017
Refresher and advance numeracy TOT training	January, 2018	March, 2018
Train teachers on improved delivery of numeracy in all 148 primary and 55 secondary schools (cluster training)	July, 2018	September, 2018
Train teachers to provide structured remedial support to students at primary and secondary level	July 2018	Ongoing - end date October 2021
Train and coach teachers to deliver the ALP curriculum	July 2018	Ongoing - end date October 2021
Recruit additional ALP teachers to mitigate the effect of teacher's attrition and retraining of new ALP teachers	June,2019	December , 2020
Train and coach teachers to deliver the ABE curriculum	August 2019	Ongoing –December 2010
Train and coach teachers to deliver ABE in villages with large	August 2019	Ongoing –December 2010

populations of displaced pastoralists		
Training of teachers and MoE staff to identify and support girls with disabilities	October 2019	November 2019
Train teachers on inclusive and special needs education. The training will include basic special education; identification and basic assessment of girls with disabilities; building inclusive classroom environments; guidance and counselling. The training will include residential training and follow up on-site sessions.	October 2019	November 2019
Work with CECs and teachers to address corporal punishment, particularly against overage/ displaced adolescents and those who are struggling to learn, and promote community-managed self-monitoring of community efforts in addressing corporal punishment	August 2019	Ongoing - end date October 2021
Strengthening Student and Teachers Attendance tracking and monitoring	March 2019	Ongoing - end date October 2021
Increase reading time by establishing and supporting CEC managed reading clubs associated to GEFs/BEFs and promote the use of culturally appropriate local learning materials	October 2019	Ongoing – end date October 2021
Construct additional classrooms in remote primary schools; build water facilities in new secondary schools; and provide solar chargers for mobile devices/tablets and sanitary pads to schools	April, 2018	March, 2019
Incorporate life skills and financial literacy training into GEFs and BEFs	April 2019	Ongoing – end date October 2021
Provide career guidance in schools	November 2018	Ongoing - end date October 2021
Output 3: Positive shifts on gender and social norms at community and individual girl level		
Engage community-level stakeholders including religious	February, 2017	September, 2018

leaders, women's groups, men and boys		
Expand and strengthen GEFs and create BEFs to develop leadership and mentorship skills	September, 2018	Ongoing – end date: March, 2019.
Provide adult literacy and financial literacy classes for mothers	May, 2018	March, 2019
Support the financial empowerment of mothers through savings groups (VSLA), business selection, and business coaching and mentoring	February, 2018	Ongoing – end June, 2019
Incorporate sessions on identification and support for Girls with Disabilities in stakeholder forums and others NFE and VSL groups.	Jan 2020	Ongoing - end date October 2021
Provide support to VSLA groups to start business upon completion of the VSLA cycle through competitive selection of most viable business ideas	Oct 2019	March 2020
Encourage girls and boys from pastoralists families to participate in empowerment forums to enhance their confidence and address negative stereotypes associated with their itinerant lifestyle	April 2020	Ongoing - end date October 2021
Social mobilization campaigns to encourage pastoralists to bring their children to school and actively participate in their education.	Jan 2020	Ongoing - end date October 2021
Output 4: Enhanced MoEs' capacity to deliver quality and relevant formal and informal education		
Strengthen Gender Departments' capacity to improve girls' education outcomes through trainings, development of action planning and provision of incentives to retain the gender focal points especially in rural areas	December, 2017	Ongoing - end date Dec. 2019

Support quality assurance and standards (QAS) functions at all MoE levels	September, 2018	Ongoing – end date June 2020.
Provide support to Regional Education Officers (REOs) and District Education Officers (DEOs) to mainstream improved teaching practices and address retention/ transition	January, 2018	Ongoing – end date June, 2019
Work closely with MoE on NFE for mothers and entrepreneurships skills for girls	April, 2018	December, 2018
Development of project IEC materials in conjunction with MoE for use at stakeholder advocacy and promotion events	July, 2018	Ongoing – end date March, 2021
Strengthen the MoEs' Special Education Unit through tailored trainings on advocacy and support for special and inclusive education	October 2019	Ongoing – end date March, 2021
Training of education officials at national, regional, district levels on basic special education concepts, support required by children with disabilities and including special needs education data in EMIS	October 2019	Ongoing – end date March, 2021
Introduce components of inclusive education/special education in pre-service teacher training curriculum for both primary and secondary school level.	October 2019	Ongoing – December, 2020

ANNEX 2: EVALUATION APPROACH AND METHODOLOGY

PRIOR TO DATA COLLECTION

Training

Training for the SOMGEP-T midline evaluation took place in Hargeisa from November 17-21, 2019. Training was led by Consilient's technical evaluation team, alongside four of Consilient's permanent Somali staff members, all of whom have been fieldwork team leaders on GEC and GEC-T projects for multiple years. Several CARE staff members were also in attendance and led specific sessions on child protection and an introduction to the SOMGEP-T program, in addition to providing clarification on aspect of program and evaluation design, and otherwise assisting to make sure training ran smoothly. In total, 11 teams were

included in the training. Almost all teams consisted of a team leader and four enumerators, and the data collection team consisted of 54 individuals in total.

Often during the training, team leaders and enumerators were split into separate sessions. The logic of this decision was that team leaders are solely responsible for a number of tasks during fieldwork – conducting classroom observations, attendance headcounts, a survey with head teachers, most qualitative interviews, and monitoring and managing their team members – that require specialized training. Enumerators, on the other hand, need to fully understand the household survey and learning assessments, and time spent training them on the responsibilities that fall to team leaders detracts from time that could be better spent practicing the learning assessments.

The first four days consisted of both classroom-style training and practice. During this time, all team members participated in an overview of the project and a discussion of the methodology, including the types (i.e. respondent type, whether enrolled in ALP, formal school, and so forth) of girls that we would be interviewing, re-contact procedures, what to do in case a replacement girl was needed, research ethics, child protection, and how to use the paper tracking sheet system.[335] Enumerators were trained on:

- The program, its activities, and the overall evaluation methodology
- Child protection and research ethics
- Learning assessment administration
- Household survey administration
- Re-contact procedures
- Specific scenarios they might face during re-contact and during the learning assessment

They also spent significant time practicing the household survey and learning assessments in one-on-one sessions.

Child protection training and research ethics were incorporated into the curriculum. Child protection was covered during a dedicated session with all evaluation team members – enumerators, fieldwork team leaders, the quality assurance officer, fieldwork manager, and team leader – CARE staff, and the MOE staff that were tasked with accompanying some of the field teams during their work. The training was led by CARE's specialist in child protection, and included theoretical precepts of child protection as well as interactive discussion, group work, and a discussion of specific scenarios. It also included a discussion of the process for reporting child protection issues that may arise during fieldwork or which researchers may become aware of during data collection. Part of this discussion emphasized the multiple available pathways for reporting child protection issues, including reporting directly to the Research Officer overseeing the project, or referring cases to Consilient Research's Human Resources focal point; additionally, CARE staff could be approached directly – the main purpose of having multiple channels for reporting is to facilitate reporting for all individuals. All researchers signed both Consilient's internal child protection policy and CARE's more specific child protection protocol.

Team leaders also participated in several group-wide sessions, but spent considerable time in separate training. These breakout sessions focused on administration of the attendance headcount, classroom observation, head teacher survey, and qualitative tools. A participatory approach to training was

[335] All enumerators already have experience using ODK Collect for mobile data collection and interviewing. The SOMGEP-T evaluation also relies on a backup system of paper tracking sheets, however, which requires training on its and emphasis on its importance.

emphasized. For example, team leaders conducted a mock classroom observation while observing one of Consilient's staff members training enumerators on the household survey, assessing the extent to which he encouraged participation, asked open-ended questions, explained concepts in a different way, etc. Newer team leaders led mock focus group discussions in which other team leaders played the role of mothers or teachers in a focus group, confronting the moderator with problems they had personally experienced on past evaluations.

Training was capped by a pilot test on the fifth day. The pilot was held in three schools in Hargeisa, with each team leader and enumerator completing at least one full household survey and learning assessment, including assessments of Somali literacy, English literacy, numeracy, working memory, and financial intelligence. Where time allowed, enumerators completed additional financial intelligence assessments with girls and boys in the pilot schools, as this assessment was viewed as the most difficult to administer well, and enumerators had the least prior experience with it.[336] During the pilot, three female enumerators were selected for a separate training – they did not complete the pilot – on the participatory qualitative exercises CARE designed, and which were completed during the first midline round. These participatory exercises – risk mapping and vignette-completion exercises described in detail elsewhere in this report – were completed with girls and covered sensitive topics; therefore, teams selected for these exercises who also had male team leaders needed to incorporate qualitative training for a female enumerator, who was tasked with completing these interviews with girls on behalf of their team leader.

Sample Design and Sampling

A full description of the sample design for SOMGEP-T is provided in the main body of the report and in this annex, combined. In general, issues related to overall sample design (selection of schools), new samples of cohort girls (i.e. ABE girls) or discussion of re-contact and attrition are discussed in Sections 2.1 through 2.4 of the report. In contrast, sampling for tools other than the household survey and learning assessments – including qualitative tools, classroom headcounts, classroom observations, and head teachers surveys – are described in the following two sections.

Qualitative Sampling

Qualitative data collection occurred alongside quantitative data collection, over the same time period and in the same overall set of schools. As noted above, there were seven distinct qualitative tools across six groups of interviewees: mothers, teachers, CEC members, Ministry of Education (MOE) officials, girls with disabilities, and girls. The latter group was recruited into two different types of participatory exercises, as described previously. In this section we briefly review the sampling approach for qualitative tools and the achieved sample.

The first three types of interviews were FGDs with mothers, teachers, and CEC members. Schools were targeted for these interviews randomly, stratifying the share of interviews allocated to each zone proportionally to the share of sample schools in that zone. For instance, schools in Somaliland comprise 62.5 percent of the overall sample, it was assigned six of the ten FGDs with mothers. A similar logic was applied

[336] The financial intelligence assessment was included as part of the SOMGEP-T baseline, but not the ML1 round completed in late 2018, and it was not a component of other GEC-T evaluations completed by Forcier/Consilient in the past. Therefore, at most, enumerators had one round of prior experience with the financial literacy assessment, compared to multiple rounds (up to 5, in some cases) of experience completing the literacy and numeracy assessments.

to each of the three FGD respondent types. Within zones, communities were selected randomly for participation.

MOE officials and GWDs were selected purposively for inclusion in the qualitative sample. For MOE officials, CARE provided a list of desired interviewees, and facilitated their contact. For GWDs, girls were selected from the baseline and ML1 cohorts – inclusive of all cohorts, including ALP girls – for a qualitative interview centred on their experiences at school and in their community. We oversampled this group dramatically, targeting around 20 girls in total, with the understanding that many girls would not be available or willing to participate in an interview, and others might deny or not view themselves as having a disability, making it difficult to conduct the interview in a manner that is ethically responsible.

Finally, the participatory exercises with girls were targeted to communities with Girls Empowerment Forums (GEFs) that were known to be active. Schools were selected into this sample if they had an active GEF, and if at least 15 percent of girls in the ML1 cohort – among those enrolled in school at the time of ML1 data collection – had participated in activities at their school led by the GEF. This approach differs slightly from the previous ML1 round, in that our standard for what constitutes an active GEF has increased, in line with the idea that the program itself has made GEFs more active. In the last round, schools were eligible if a sufficient number of girls were aware of the GEF; in this round, we require GEFs to be sufficiently active that a considerable number of girls have actually participated in a GEF activity. Out of the eight participatory exercises, five were assigned to Somaliland, two were assigned to Puntland, and one was assigned to Galmudug.

In practice, the sample deviated somewhat from this planned distribution. The table below describes the achieved sample of qualitative interviews, by type of interview and zone. In the case of mothers, the sample is skewed toward Somaliland-area schools, due to the gender composition of our team leaders, who were responsible for qualitative data collection. Of the three teams that visited Puntland, only one had a female team leader – because participatory exercises with girls, and KIIs with girls with disabilities required a female interviewer, and FGDs with mothers were preferably conducted by a female interviewer, we were compelled to prioritize the team leader's time in conducting qualitative interviews. We opted to have her focus on the participatory exercises and interviews with GWDs, at the expense of completing only one FGD with mothers. The greater number of female team leaders in Somaliland allowed these teams to handle more FGDs with mothers, as well as GWDs.

In addition, the team working in Galmudug did not include an experienced qualitative interviewer who was female. For this reason, we opted not to conduct any risk mapping or vignette exercises in Galmudug, due to the trade-off we assumed would result between stronger geographic coverage and higher-quality data. Given the few participatory exercises to be conducted, we did not wish to risk one or more poor-quality interviews, for the sake of wider coverage. Therefore, the Galmudug team focused on interviews that could be conducted by their experienced team leader, who was male.

Finally, note that the number of interviewed GWDs exceeded the targeted sample of ten, due to the purposive oversampling mentioned above. As multiple teams were conducting KIIs with GWDs on consecutive days – occasionally in areas without good mobile network coverage – more KIIs were completed than originally planned. However, given the relative difficulty of completing an effective and informative KII in this case – due to the complexities of asking girls about their disabilities – extra interviews is useful.

TABLE 115: ACHIEVED SAMPLE OF QUALITATIVE INTERVIEWS, BY ZONE

	Somaliland	Puntland	Galmudug	Total
FGD Mothers	8	1	1	10
FGD Teachers	6	3	1	10
FGD CEC Members	7	2	1	10
KII MOE Officials	6	3	1	10
Girls with Disabilites (GWDs)	9	4	0	13
Risk Mapping	4	0	0	4
Vignette Exercises	2	2	0	4

Sampling for Other Tools

In addition to cohort girls re-contacted from previous rounds and ABE girls recruited during ML2 data collection, three other quantitative samples were recruited in this round. In all three cases, the first stage of sample selection was the selection of schools or learning centers and our sampling approach mirrored that taken in previous rounds.[337]

Attendance headcounts were conducted in each formal primary school. Team leaders completed the headcounts, as they require coordination with the head teacher and teachers in the school, access to the enrolment records, and particular training to conduct properly. In each school, team leaders completed one headcount per grade level from 1 to 8; in schools with only 6 or 7 grade levels represented, team leaders completed 6 or 7 headcounts, respectively.[338]

Classroom observations were also conducted in each formal primary school, by team leaders. Two classrooms were targeted for observation in each school, with two stipulations: classes from grade 3 to 8 were targeted, and classes where Somali, English, or mathematics were being taught. In practice, these two requirements occasionally conflicted, especially given the short timeframe (2 days) teams were in each location and the short length of school-days in some locations. In those cases where a targeted subject was being taught to children in grades 1-2 and where a class with older children (grades 3-8) was available but not in a targeted subject, team leaders were instructed to complete the classroom observation with the class learning the targeted subject. That is, subject matter took priority over the specific grade, *in those cases where it was not possible to complete a classroom observation that met both criteria*.[339] Of course, some classrooms included children of lower and upper grades mixed together, so many grade 1-2 students were captured even in classrooms with grades 3-8 students. In those rare instances where a team leader had their option of selecting more than two eligible classrooms, they selected them randomly.

[337] We do not refer to the head teacher survey as a sample in the same sense, because no second-stage sampling occurred below the school level. Once schools were chosen, all head teachers were recruited into the head teacher sample, as each school has just one head teacher (or equivalent top-level administrator/principal).

[338] In practice, few schools in the SOMGEP-T sample have multiple classes for a given grade level. In the rare cases where they do, team leaders selected a classroom within the grade randomly.

[339] The logic of this decision is that many schools have only one or two teachers who teach a given subject (e.g., one mathematics teacher for all grade levels). Therefore, observing that teacher in front of grade 2 students was seen as more informative than observing a different teacher, teaching science or social studies, to the appropriate grade level. Given SOMGEP-T's emphasis on literacy and numeracy, subject matter was prioritized.

Boys were recruited into the evaluation through the household survey, in the same manner as the baseline and ML1. In all three rounds, girls participating in the evaluation, in any cohort, were asked to complete a household survey with their caregiver and head of household. At the time of this interview, the head of household was asked whether there were any boys in the household, in the targeted age range, available for an interview. In each round, this age range was equivalent to that applied to cohort girls – 10-19 years at baseline, moving up to 12-22 years at ML2. If multiple boys in a household were eligible, a boy was selected randomly using a pre-programmed random number generator in Consilient's survey software (ODK Collect). Boys recruited into the sample were presented with a few demographic questions and completed the full battery of learning assessments including financial intelligence and working memory. At baseline, the achieved boys sample was 510 boys; however, when we limit the school-level sample to the schools that were visited at both baseline and ML2, the baseline sample was 466 boys, compared to 416 boys at ML2.

The table below provides the geographic breakdown of the achieved sample at ML2. Note that the table does not report figures for cohort girls of any kind, as we discuss their achieved sample in more detail below and in later sections of this report. Moving back to the table, it is important to note that no specific sampling targets were set for attendance headcounts, as the number of headcounts to be completed is a function of the number of distinct classrooms operating in each school. On the other hand, targets were set for head teacher surveys, but this target – 1 per school, implying a total of 69 surveys – was met. An upper target of 138 classroom observations were sought, with a goal of two per school; as the table shows, this target was not achieved.

TABLE 116: GEOGRAPHIC DISTRIBUTION OF ACHIEVED SAMPLE

Data Collection Tool	Somaliland	Puntland	Galmudug	Total
Headcounts	294	159	18	471
Head Teacher Surveys	44	21	4	69
Classroom Observations	82	38	4	124
Boys Learning Assessments	235	163	18	416

Note that at least one classroom observation was completed in 63 of the 69 sampled formal primary schools. Two schools were completing examinations or revisions for examinations during scheduled fieldwork, and standard classes were not available for observation.[340] A further two schools were closed entirely during fieldwork. The remaining two schools did not have eligible classes for observation, because the subject matter was not English, Somali, or mathematics. The same schools that were closed impacted the total sample of headcounts completed, as headcounts could not be completed in schools that were closed or actively completing examinations. The only other school-based data collection tool – the head teacher or school survey – was not impacted by school closures or examinations, as head teachers were willing to work with our data collection teams even if the students were not present at the school.

[340] Team leaders were instructed to only observe classrooms if they were hosting instruction, not completing tests or preparing for examinations, as these processes do not reflect normal teaching practices.

Given the lack of targets and the complexities of sampling at schools that are in operation at the time of data collection, perhaps a more informative metric is to compare the number of completed surveys relative to the baseline. In the table below, we limit the sample to schools that were contacted in both rounds, to ensure a fair comparison between the two rounds (as the ML2 sample is smaller than the baseline by 7 schools). As the table shows, the ML2 evaluation captured a larger number of headcounts, an equal number (the maximum possible) of head teacher surveys, and nearly as many classroom observations, as were captured from the same set of schools during the baseline.

TABLE 117: BASELINE AND ML2 ACHIEVED SAMPLES, WITHIN EQUIVALENT SET OF SCHOOLS

Data Collection Tool	Completed at Baseline	Completed at ML2
Headcounts	434	465
Head Teacher Surveys	69	69
Classroom Observations	125	124

TOOL DETAILS

Tool (used for which outcome and IO indicator)	Beneficiary group	Actual sample size intervention and (control group) - if appropriate	Remarks: 1) Attrition rate from baseline to midline 2) Re-contacted sample vs replaced sample 3) Major changes to tools or differences between anticipated and actual sample sizes
SeGMA and SeGRA (for learning outcome); financial literacy assessment; working memory test	In-school and out-of-school cohort girls (aged 10-19 years at baseline)	922 in-school girls (517 intervention; 405 comparison) 368 OOS girls (172 intervention; 196 comparison) 482 ABE girls (no comparison) 336 ALP girls (no comparison) 416 boys	Targeted all girls who had previously been contacted at either baseline or ML1, regardless of whether they had been successfully re-contacted from baseline at ML1. Identified 2,588 girls who fell into the sample across all sample groups. The majority of attrition from the target sample size occurred among OOS girls, as they were not replaced if they could not be located. Boys were not re-contacted from previous rounds. They were recruited from households visited for the household survey with girls of all respondent types. **Outcomes Measured:** Learning of all types, including working memory and financial literacy
Household survey	In-school and out-of-school cohort girls and their households	922 in-school girls (517 intervention; 405 comparison) 368 OOS girls (172 intervention; 196 comparison) 482 ABE girls (no comparison) 336 ALP girls (no comparison)	Same sampling notes as SeGMA and SeGRA, above (same sample composition and replacement rates) **Outcomes Measured:** Transition; community attitudes; life skills; attendance; teaching quality
Head Teacher (School) Survey (for school management and other outcomes)	All head teachers at sampled schools	69 (37 intervention, 32 comparison)	**Outcomes Measured:** School management and governance
Headcount Tool (for attendance outcome)	All classrooms at sampled schools	471 (248 intervention, 223 comparison)	Census of classrooms in grade 1-8 at sampled schools. Note that combined classrooms – I.e. those with grade 7 and 8 students – were recorded separately as two observations with distinct headcounts and attendance rates. **Outcomes Measured:** Attendance

Classroom observation (for teaching quality outcome)	2 classrooms per school	124 (65 intervention; 59 comparison)	• Six schools in the sample at ML2 lack classroom observations, in most cases because normal classes were not operating at the time of fieldwork in their specific schools (typically in preparation for exams). • Targeted classes were teaching Somali, English or mathematics • Added short module of questions directed at teachers to measure their attitudes toward pedagogical practices (due to removal of teacher survey since baseline) **Outcomes Measured:** Teaching quality

DURING DATA COLLECTION

Data collection took place during a single wave, without staggering for different tools or modes of data collection. Data collection began on November 25, 2020, and continued through December 29. In total, 11 teams participated in data collection; all teams except one consisted of a team leader and four enumerators, with a total of 43 enumerators, and 11 team leaders. Qualitative and quantitative data collection occurred simultaneously for logistical reasons. Given the remote areas in which SOMGEP-T schools are located, visiting communities twice, in separate waves, was not considered feasible. Team leaders were responsible for most qualitative data collection except – as noted above – where the team leader was male and the interview was being conducted with school-age girls.

Prior to the start of data collection, CARE's Monitoring & Evaluation team and Consilient's Fieldwork Manager both reviewed the list of target schools for accessibility and security concerns. As we discuss in the methodology section of the report, two schools were removed from the sample at this stage: one school was removed due to security concerns, and the second school was removed because it had been destroyed during a conflict that also prevented its inclusion at ML1. Consilient's Fieldwork Manager and CARE staff were in contact with teams throughout fieldwork, to discuss security concerns and contingency plans if necessary. CARE staff or members of the MOE also accompanied most of the teams during fieldwork. In practice, the primary contingencies faced were travel delays and adjustments to the schedule due to flooding in some areas of Somaliland and Puntland.

For a full discussion of sampling, achieved sampled sizes, re-contact procedures, the process for selecting replacement girls, and an analysis of replacement rates, please see Sections 2.2 (sampling and achieved samples) and 2.4 (re-contact procedures and attrition rates) of the primary report. We view these issues as sufficiently important to warrant extensive discussion before considering the primary substantive findings from the evaluation.

During data collection, Consilient's Fieldwork Manager and the technical evaluation team provided extensive quality control of the incoming data. However, quality assurance began in earnest prior to fieldwork, as the survey scripts were reviewed for mistakes, incorrect or unclear translations, errors in skip or filter logic, and general areas of concern. The quantitative tools remained mostly the same from the previous round, with adjustments for the sampling and re-contact process, the fact that OOS girls were included in the sample, and addition of a few new questions. During the inception phase at both ML1 and ML2, the evaluation team sought to script in purposeful checks of data quality. For instance, each girl's information – and that of her household – is scripted into the survey, such that it is shown to the enumerator during the re-contact process. We also added confirmatory steps, asking the enumerator to confirm the outcome of the re-contact procedures and what steps they followed after failing to locate a girl. As an example, if a girl was replaced and the replacement girl's grade level did not match the replacement procedures established during training, enumerators were required to provide a reason why they were interviewing a replacement who did not meet the formal requirements. Finally, we imposed constraints on questions where appropriate, and generally sought to minimize measurement error through careful survey design.

Once data collection began, the evaluation team conducted quality control testing using a pre-designed Stata .do file, which was designed specifically for this project. The process included checking for known error "hotspots":

- Any learning assessment question that involved counting the number of letters known or words read
- Physical headcounts and enrolment numbers from school records
- Later subtasks in learning assessments, where enumerators can skip questions if the girl was completely unable to answer earlier subtasks
- New dietary diversity questions, due primarily to the fact that they were newly added this year
- Working memory assessment, because enumerators had not completed such an assessment before, and consistency in administration is critical

In addition to these areas of complexity where we wished to monitor the data for errors, we also monitored re-contact, replacement, and attrition rates, as well as the number of households in which an eligible boy was located. We also monitored GPS coordinates and survey duration to guard against data fabrication, and checked for "enumerator effects" in learning assessment scores to check whether any enumerators appeared to be administering the assessment in fundamentally different ways. Some of these results, during ML1, were included in the report, and they are available upon request for this evaluation round.

When errors were discovered through the quality control process, the information was relayed to the Fieldwork Manager and Team Leader, so that corrective action could take place. Where corrections to the data were needed, the evaluation team made these changes immediately once feedback was received from the team leader in the field. All data cleaning that occurred concurrently with fieldwork was recorded in a Stata .do file, with justifications for all cleaning decisions, to ensure replicability and transparency.

Qualitative data was checked in a similar, though obviously less structured, fashion. Team leaders submitted audio recordings of their qualitative interviews to Consilient's office-based researchers and the Fieldwork Manager. The first audio files submitted by each team leader was reviewed by the Fieldwork Manager for quality of the interview, and they provided feedback to the team leader to improve their interviewing techniques.

With regard to data transmission and storage, survey responses were transmitted as soon as network coverage allowed from enumerator's phones to secure servers hosted by Ona, which runs a server for ODK Collect submissions. The data is encrypted end-to-end from the phone to Ona's servers, preventing data loss and ensuring confidentiality of the data. Once receipt of the data was confirmed by the Fieldwork Manager and technical team, surveys were deleted from enumerator phones as, again, a method of ensuring anonymity. Qualitative interviewers were audio recorded. Audio recordings were transmitted to Consilient Research's Google Drive folder dedicated to the project and with access restricted to the technical evaluation team, and then deleted from enumerator phones and audio devices. All data storage took place on Consilient Research's Google Drive account. When data was transmitted to CARE's Monitoring & Evaluation team, data was password-protected in a .zip archive and transmitted via email; the password was provided separately to CARE's focal point via Skype.

POST-DATA COLLECTION

During and after the completion of fieldwork, the technical evaluation team performed additional checks of the data for consistency and cleaned the data. Our strategy was to flag interviews or observations which had some form of inconsistency, and follow up with the team leader in question for clarification or additional information. In some cases, the Fieldwork Manager called girls' households or head teachers to clarify information provided. Some of the most common cases which we flagged were:

- Girls whose given age or grade was not consistent with prior years
- Physical headcounts and enrolment numbers that deviated significantly from previous years (e.g., school in which grade 3 went from having 10 students to 35 students).
- Reading scores which were inconsistent across subtasks (e.g., very strong word identification scores but very low scores on reading fluency, or vice versa)
- Very high or very low scores on the working memory test

We emphasise that these cases were not dropped from the sample or cleaned automatically. Rather, they were reviewed in light of additional information elsewhere in the survey, from prior rounds, and after discussions with the team leader or enumerator who conducted the interview. Much of this quality control occurred during fieldwork, to maximize the likelihood that enumerators would recall the interview in question.

As noted above, qualitative data was also subject to quality control. In addition, we conducted post-fieldwork debriefing sessions with several of the team leaders, with the goal of improving data collection for future rounds, and to understand subtleties of the qualitative data – or issues that were revealed outside of qualitative interviews. At times, these post-fieldwork interviews with team leaders are cited in this report.

Qualitative data was transcribed verbatim in Somali and then translated into English, resulting in complete English-language transcripts of all FGDs, KIIs, and participatory exercises. Qualitative analysis was the responsibility of all evaluation team members, rather than assigning a dedicated qualitative analyst. The reason for this decision was to maximize the insights gained from the qualitative data, as multiple individuals reading the interviews notice different aspects of participants' answers. This approach also ensured that there was not a sharp disjuncture between quantitative and qualitative analysis, as the goal was for the two sources of data to "speak to each other".

In terms of process, the evaluation team prioritised reading the complete qualitative interviews before turning to in-depth quantitative analysis. This first reading produced a number of specific findings – many of which were not explicitly addressed in the quantitative tools – which were noted for incorporation into a given section of the report. This process ensured that themes could emerge organically from the qualitative data, rather than being tied too tightly to explaining or providing nuance to quantitative results. Next, the team performed the core quantitative analysis; again, the benefit of fully reading the qualitative transcripts in advance was that insights from the qualitative interviews often suggested adjustments that should be made to the quantitative analysis.[341] Following the quantitative analysis, team members went back to the qualitative data in a more purposeful way, seeking to understand specific findings from, provide nuance to, and find evidence that either confirmed or contradicted, the quantitative findings. While our approach to qualitative analysis was not as systematic – in the sense of establishing a formal and rigorous coding scheme – as some forms of qualitative data analysis, our preferred approach ensured that we maximized insights from the qualitative data and that findings from the one data source influenced analysis from the other in productive and symbiotic ways. We employed a similar approach, where appropriate, to the incorporation of multiple

[341] For instance, the risk mapping exercises revealed outright violent conflict in some of the communities visited for this evaluation. These findings motivated a review of the available quantitative data on localized conflict, and prompted us to re-interview some of the team leaders, with the goal of specifically understanding the latent level of violence or insecurity in the communities they visited.

quantitative data sources, attempting to triangulate across multiple data sources in a manner that maximized the quality of our inferences and the value of the analysis.

The ML2 evaluation constitutes the third round of contact between the evaluation team and its fieldwork researchers, on one hand, and the SOMGEP-T intervention and comparison communities, on the other. In actual practice, some of the communities were part of the SOMGEP phase I evaluations, which were completed by many of the same fieldwork researchers. For the purposes of facilitating effective fieldwork, we made an effort during ML2 to dispatch field teams back to sites they had visited previously. This approach enhances their ability to work productively with the school administrators, teachers, and the community; it also increases the likelihood that they will successfully re-contact girls, because they are familiar with the community and may actually remember some of the girls in question.

Tracking and re-contact procedures from ML2 to endline will mirror those used in the previous three rounds. Information about each girl was recorded in the household survey, on paper tracking sheets, and in a school data collection report completed by the team leader at the conclusion of fieldwork in each village. This information includes her name, age, grade, and geographic details (school, village name). It also includes information about her household, including her caregiver's name, name of the head of household, and two contact phone numbers for the household in most cases. Given multiple rounds of data collection, we now have up to six contact phone numbers for some households. Finally, GPS data was recorded as part of the household survey. It is important to note that we generally do not rely heavily on GPS data for tracking respondents. Instead, our teams rely on their experience in and knowledge of the village, and the assistance of the head teacher and community members to find the girls in question. GPS data is available as a backup, as well as a data quality control check.

In general, the procedures for re-contact at endline should remain the same. However, as we note in the recommendations section of this report, we suggest specifically emphasising partial re-contacts of *all* girls recruited since baseline, even if they cannot be located to complete the full interview. During ML2, when teams encountered girls who could not be located or who were unavailable for an interview, they were asked to find someone who knew the girl or her family – a close family member, a neighbour, friend, or teacher at school. We asked these individuals whether they knew where the girl was, and – if this did not make it possible to interview her (e.g., she lives in a distant village or in another region) – we asked whether she was enrolled in school. This information allowed us to understand the nature of panel attrition and – we hoped – would be useful for analysis of transition outcomes among a broader sample than those who were successfully re-contacted.

We suggest a more robust approach to this goal be employed at endline. In lieu of replacing girls who fall out of the sample at endline – which is not currently planned and would be of limited value inferentially – the time and resources saved should be redirected to making repeated efforts to re-contact girls via phone or other means. First, any girl who has been contacted in the SOMGEP-T evaluation should be re-contacted at endline, even girls who fell out of the sample at an earlier stage. Girls residing in communities that cannot be visited for accessibility reasons should be contacted by phone. For girls who cannot be located to complete the full interview, fieldwork teams should collect as much detail about the girl's location and contact details as possible. They should then be sought by phone, not to complete the full interview, but to complete a

special "transition-only" module of the household survey. Formal procedures mimicking the re-contact procedures – which specify the number of contact attempts that should be made and how they should be spaced temporally – should be designed for contact attempts by telephone, to ensure the highest number of girls, who would otherwise not appear in the endline sample at all, complete the transition-only module. Finally, the evaluation team should consider dedicating one team leader in each region to working for an extra week to revisit locations with high attrition rates, seeking to re-contact girls a second time. Given the number of girls who have fallen out of the sample due to short-term unavailability, this approach should contribute to lower attrition rates at endline.

Data analysis was conducted primarily in Stata, using .do files for all analysis to enhance replicability both internally and externally. The evaluation lead reviewed the cleaning .do files that had been constructed during fieldwork to correct day-to-day quality control issues; they then completed additional cleaning and merged the data with the baseline and ML1 datasets, and combined datasets where needed for the analysis (e.g., combining the head teacher survey, which provides information about school characteristics, with the household survey). The cleaning and data compilation process was documented via .do files as well.

Qualitative data was analysed in Excel. The English transcripts were completed in Excel, and the design of the transcript form allowed the evaluation team to filter questions by topical area and respondent type, in order to read many related responses at once. Rather than code qualitative interviews in a formal manner, we made an effort to read transcript relevant to a given section verbatim, as critical information is often hidden in responses to questions that one would not expect. The goal of qualitative analysis was to allow natural themes to emerge, which we accomplished by reading the qualitative transcripts, as much as possible, prior to starting the quantitative analysis. Following the quantitative analysis, we revisited the qualitative data with new questions and themes in mind, seeking to understand the trends and anomalies in the quantitative data.

COHORT DEMOGRAPHICS AND COMPOSITION

The following tables break down the Midline Round 2 (ML2) evaluation sample by cohort, geographic zone, grade, age, and disability status, disaggregated by intervention and comparison schools. The first table specifically outlines attrition and re-contact rates. Note that we draw a distinction between attrition that results in replacement of a girl from the sample ("replaced") and attrition that results in the girl being lost from the sample and not replaced ("attrition"). While both outcomes have consequences for the sample, attrition produces a smaller sample, while replacement results in the same size sample, but a shift in precisely how comparable the sample is across rounds. This table provides a relatively brief look at sample composition; however, readers interested in the evolution of the cohort's characteristics over time, and its representativeness, should review Annex 3.

TABLE 118: ATTRITION AND RE-CONTACT RATES, BY COHORT

Cohort group	Midline sample (treatment)	Recontacted (treatment)	Attrition (treatment)	Midline sample (control)	Recontacted (control)	Attrition (control)
In-School Girls	**517**	**439**	**71 attrition 78 replaced**	**405**	**338**	**54 attrition 67 replaced**

Out-of-School Girls	172	172	177 attrition 0 replaced	196	196	148 attrition
ALP Girls	365	257	29 attrition 79 replaced	N/A	N/A	N/A
ABE Girls	482	N/A	N/A	N/A	N/A	N/A

TABLE 119: REGIONAL BREAKDOWN OF SAMPLE

	Intervention (recontacted)	Control (recontacted)
Sample breakdown (Girls)		
Somaliland (62.6%)	412 (59.7%)	397 (66.1%)
Puntland (32.0%)	226 (32.8%)	187 (31.1%)
Galmudug (5.4%)	52 (7.6%)	17 (2.8%)
Girls (1,290)	689	601
Sample breakdown (Boys)		
Somaliland (56.5%)	165 (53.6%)	70 (64.8%)
Puntland (39.2%)	130 (42.2%)	33 (30.6%)
Galmudug (4.3%)	13 (4.2%)	5 (4.6%)
Boys (416)	308	108

TABLE 120: GRADE BREAKDOWN OF SAMPLE

	Intervention (recontacted)	Control (recontacted)
Sample breakdown (Girls)		
Grade 1 (2.9%)	2.0%	3.2%
Grade 2 (2.7%)	2.8%	2.7%
Grade 3 (5.0%)	4.6%	5.5%
Grade 4 (7.1%)	6.5%	7.8%
Grade 5 (15.0%)	16.7%	13.1%
Grade 6 (14.5%)	16.7%	12.0%
Grade 7 (15.9%)	15.2%	16.6%
Grade 8 (12.3%)	14.5%	9.8%
Form 1 (1.6%)	1.3%	1.8%
Form 2 (0.1%)	0.0%	0.2%
Form 3 (0.0%)	0.0%	0.0%
Form 4 (0.1%)	0.2%	0.0%
OOS girls (21.4%)	17.9%	25.5%
Alternative education (1.2%)	1.3%	1.0%
Girls (1290)	N = 689	N = 601
Sample breakdown (Boys)		
Grade 1 (5.3%)	6.5%	4.9%
Grade 2 (3.9%)	4.6%	3.6%
Grade 3 (11.1%)	13%	10.4%

Grade 4 (9.1%)	0.9%	12%
Grade 5 (12.3%)	19.4%	9.7%
Grade 6 (8.9%)	11.1%	8.1%
Grade 7 (7.9%)	7.4%	8.1%
Grade 8 (7.0%)	7.4%	6.8%
Form 1 (3.6%)	0.9%	4.6%
Form 2 (1.4%)	0.9%	1.6%
Form 3 (2.4%)	0.9%	2.9%
Form 4 (2.4%)	3.7%	2%
OOS girls (20.7%)	23.2%	19.8%
Alternative education (3.6%)	0%	4.9%
Boys (416)	N = 308	N = 108

TABLE 121: AGE BREAKDOWN OF SAMPLE

	Intervention (recontacted)	Control (recontacted)	Total
Sample breakdown (Girls)			
12-13 years	32.5%	34.6%	33.5%
14-15 years	36.0%	31.3%	33.8%
16-17 years	20.5%	20.6%	20.5%
18-19 years	8.1%	9.5%	8.8%
20-22 years	2.9%	4.0%	3.4%
Total	N = 689	N = 601	N = 1,290
Sample breakdown (Boys)			
12-13 years	39.9%	47.2%	41.8%
14-15 years	28.9%	26.9%	28.4%
16-17 years	15.9%	11.1%	14.7%
18-19 years	11.7%	6.5%	10.3%
20-22 years	3.6%	8.3%	4.8%
Total	308	108	416

TABLE 122: DISABILITY PREVALENCE IN SAMPLE, BY INTERVENTION STATUS

Sample breakdown (Girls)	Intervention (recontacted)	Control (recontacted)	Household Survey and Girls School survey – Washington Group and child functioning questions
Girls with disability	9.6%	8.0%	

Provide data per domain of difficulty			
Difficulty seeing	0.7%	0.7%	
Difficulty hearing	0.4%	0.2%	
Difficulty walking or climbing steps	0.4%	0.2%	
Difficulty remembering or concentrating	0.2%	0.2%	
Difficulty with self-care	0.0%	0.0%	
Difficulty communicating	0.3%	0.2%	

SAMPLE COMPOSITION, REPRESENTATIVENESS, AND ATTRITION BIAS

In the previous section, Table 115 documented the replacement and attrition rate in the cohort sample across three evaluation rounds. A key point made in reference to that table is the extent to which both full attrition (total loss of a girl, without replacement) and replacement can influence the composition of the sample. Here, we also refer readers to Section 2.2 of the report (especially Tables 7 and 9) and Section 2.3, which performs methodological analysis of re-contact and replacement as an outcome.

Section 2.2 highlights the tradeoffs made depending on the sample selected for analysis. Specifically, there is a tension between analysing the maximum available sample size by including as many interviewed girls from each round as possible in the sample to be analysed, and analysing the most comparable possible sample, by limiting the sample to those who were successfully re-contacted and never replaced. Similarly, there is a tension between capturing change over as many periods (baseline, ML1, ML2) as possible, and maintaining the largest possible comparable sample, because the set of girls who were successfully re-contacted in *all three* rounds is significantly smaller than the set of girls who were successfully contacted at the baseline and ML2 only (for in-school girls, this difference is 529 girls versus 716).

Section 2.3 of the report approaches this issue in a slightly different way, analysing whether girls in the cohort are "missing at random" during ML1 and ML2. To put it differently, we analyse whether dropping out of the sample – whether being replaced or falling out of the sample altogether – is correlated with a girl's characteristics. If panel attrition is correlated with a girl's characteristics, it means the sample may suffer from attrition bias. To summarize, the data shows that in-school girls do not fall out of the sample randomly; instead, attrition is highly correlated with older age and somewhat correlated with geographic zone. Importantly, demonstrating attrition bias also requires studying whether the replacement process *always* replaces girls successfully, and whether they replace girls in a like-for-like fashion. Section 2.3 also analyses this question. Unfortunately, the number of girls who have fallen out of the sample altogether – without replacement – suggests that, even if the replacement process were perfect, there may still be attrition bias.

Further, the analysis in Section 2.3 (especially Tables 15-16 and Figures 3-4) shows that replacement girls are different from the girls they replaced in several fundamental ways – most problematically, their learning scores are different from what we would expect from the girls they replaced. Learning scores are dramatically higher among replacement girls than the girls they replaced. Interestingly, however, our analysis suggests that

learning scores among replacement girls are actually *lower* than they should be, if replacement girls are to be considered a fair substitute for the girls they replaced. In fact, the replacement process is almost certainly dragging down the estimated changes in learning scores from baseline to later rounds, at least compared to the larger gains made among re-contacted girls.

In this section, we expand on the analysis presented in Sections 2.2 and 2.3 in two main ways. First, we provide two detailed tables showing the differences between:

- Replaced and replacement girls from baseline to ML2 (disaggregated by intervention status)
- Re-contacted ("true panel") girls from baseline to ML2 (disaggregated by intervention status)
- Replaced and replacement girls from ML1 to ML2 (disaggregated by intervention status)
- Re-contacted ("true panel") girls from ML1 to ML2 (disaggregated by intervention status)
- The table reports values for the in-school girl cohort only.

The concern motivating this analysis and discussion is this: replaced and replacement girls are somewhat different from one another in terms of their household and personal characteristics. However, some of these differences could be explained by changes in the sampled communities over time (for instance, improvements in food security). If this is the case, we should also observe these changes for "panel girls" who were re-contacted from BL to ML2, because they live in the same communities.

If there are shifts in characteristics/barriers over time for replaced/replacement girls and those shifts are not mirrored among re-contacted girls, it indicates that replaced and replacement girls are not comparable to the re-contacted girls, and are likely to produce bias in our estimation of learning scores. If household and personal characteristics evolve differently among replaced/replacement girls than among re-contacted girls, bias is almost certain: the only context in which bias would not occur is if personal and household characteristics are uncorrelated with learning outcomes, which is extremely unlikely.

Second, we provide a table documenting the changing sample composition across all three rounds, if we include replacement girls and the girls they replaced in the analysis (i.e. the "cross-sectional" sample). While the first set of tables show that replacement and replaced girls are different from one another, it is also the case that the aggregate impact on sample *representativeness* is low. Replacements make up a small overall share of the sample, and therefore exert little influence on sample demographics. To summarize our conclusions in advance of the analysis: we are concerned about the inclusion of replacement and replaced girls in the analysis because doing so can shift our estimates of program outcomes, such as learning – as we discuss below – but we are not concerned about their inclusion from the perspective of making the sample more or less representative of the underlying population of SOMGEP-T girls.

Baseline to ML2 Sample

The first table below documents the changes in the sample from BL to ML2. The simplest way to interpret this table is to look at the four columns labelled "Diff", which is the difference from BL to ML2. To illustrate, consider the share of girls who are living without their parents (i.e. they live in households that do not include their mother or father). In intervention schools, the share of replaced/replacement girls in this situation increased by 10.3 points, while the share declined by 2.8 points among the same replaced/replacement girls in comparison schools. On its own, this is not problematic, because there could be broad differential trends between intervention and comparison schools. This is where a comparison to the group of panel girls is useful: in this group, the share of girls in both intervention and comparison schools who live without their parents increased slightly.

The problem is that this trend is different between replaced/replacement girls and re-contacted girls: the replacement process is increasing the number of girls living without their parents in intervention schools, relative to their re-contacted peers (the upward trend is steeper among replaced/replacement girls than among re-contacted girls). But in comparison schools, the replacement process is *decreasing* the share of girls living without their parents, relative to their re-contacted peers (the trend is down among replaced/replacement girls, while it is upward among re-contacted girls).

Of course, we would expect small-to-medium deviations between trends in the two groups, due to sampling variation. However, there are more discrepancies in this table than we can easily explain as a result of sampling variation, and they *tend* to occur in a consistent direction. Replaced/replacement girls in comparison schools seem to face fewer household barriers/disadvantages than re-contacted girls in the same schools. The replacement process causes a decline in the share of girls in comparison schools who have these characteristics:

- Female-headed household
- Living without both parents
- Head of household has no formal education
- Caregiver has no formal education
- Girls with disabilities (of any kind)
- Household went hungry at least 1-2 days in last year
- Household went without clean water at least 1-2 days in last year

Other shifts were less pronounced, so we have focused on these variables. The issue, in our view, is that these changes are all in a consistent direction. The inclusion of replacements in the sample makes comparison girls *less marginalized* at ML2 than they were at BL, while the same is not true of intervention girls. Consider household hunger: replaced/replacement girls in intervention schools mirrored their re-contacted peers almost exactly in terms of their evolution – on this characteristic – from BL to ML2. Among both groups in intervention schools, experiences of household hunger declined very slightly. The evolution of household hunger in comparison schools was similar *for re-contacted girls*. But it was dramatically different for replaced/replacement girls in comparison schools – replacement in comparison schools results in a sample that is significantly less marginalized than it would be if using the panel or girls were not replaced.

TABLE 123: CHARACTERISTICS OF REPLACED/REPLACEMENT GIRLS & RE-CONTACTED ("PANEL") GIRLS IN INTERVENTION AND COMPARISON SCHOOLS (BL TO ML2)

Girl Characteristics	Replaced/Replacements – Intervention			Replaced/Replacements - Comparison			Recontacted – Intervention			Recontacted - Comparison		
	BL	ML	Diff	BL	ML	Diff	BL	ML	Diff	BL	ML	Diff
Living without both parents	4.4	14.7	**10.3**	10.8	7.9	**-2.8**	6.7	7.8	**1.1**	8	12	**4**
Orphan	0	0	**0**	0	0	**0**	.2	.5	**.3**	.6	.6	**0**
HoH is female	47.1	58.8	**11.8**	60	55.6	**-4.4**	45.7	51.3	**5.6**	42.8	54.7	**11.9**
HoH has no education	80.6	58.7	**-21.9**	75.8	35.2	**-40.6**	66.5	44.4	**-22.1**	61	42.4	**-18.6**
HoH has no formal education	80.6	82.5	**1.9**	75.8	63	**-12.8**	66.5	70.2	**3.7**	61	71.4	**10.4**
Primary caregiver has no education	79.4	58.8	**-20.6**	83.1	42.9	**-40.2**	69.2	51.8	**-17.5**	70.3	50.2	**-20.1**
Primary caregiver has no formal education	86.8	89.7	**2.9**	87.7	73	**-14.7**	80.1	78.9	**-1.2**	77.6	81.6	**3.9**
Currently married	1.5	0	**-1.5**	0	0	**0**	1.5	2.8	**1.3**	1.6	3.9	**2.3**
Has ever been married	1.5	0	**-1.5**	0	0	**0**	1.5	3	**1.5**	1.6	4.9	**3.3**
Mother, under 16	0	0	**0**	0	0	**0**	.5	.5	**0**	.6	2.9	**2.3**
Mother, over 16	0	0	**0**	0	0	**0**	.5	0	**-.5**	.3	0	**-.3**
Girls with any disability	3	5.9	**2.9**	10.8	0	**-10.8**	7.4	9.2	**1.8**	3.8	8.6	**4.8**

Household Characteristics	Replaced/Replacements – Intervention			Replaced/Replacements - Comparison			Recontacted – Intervention			Recontacted - Comparison		
	BL	ML	Diff	BL	ML	Diff	BL	ML	Diff	BL	ML	Diff
HH is pastoralist	10.3	8.8	**-1.5**	4.6	7.9	**3.3**	9.7	6.3	**-3.4**	7.3	7.1	**-.2**
HoH has no wage-earning occupation	42.6	41.2	**-1.5**	35.4	36.5	**1.1**	37.7	44.8	**7**	38	48.5	**10.5**
Moved in the past 12 months	0	2.9	**2.9**	1.5	1.6	**0**	.2	2.3	**2**	1	1.3	**.3**
Seasonal migration	5.9	4.4	**-1.5**	4.6	4.8	**.1**	3	3.8	**.8**	3.8	3.6	**-.3**
Owns camels	10.3	11.8	**1.5**	7.7	20.6	**12.9**	7.7	8	**.3**	7	19.7	**12.7**
Owns medium-sized livestock	60.3	58.8	**-1.5**	40.6	60.3	**19.7**	58.3	67.9	**9.7**	64.9	72.2	**7.3**
Owns land	81.3	73.1	**-8.1**	85.9	69.5	**-16.4**	79.6	72.2	**-7.4**	82.4	72.3	**-10.1**
House is informal / temporary structure	3	3.2	**.2**	4.7	8.5	**3.8**	3.5	4	**.5**	6.5	7.2	**.7**
Poor quality roof	32.4	29.4	**-2.9**	30.8	33.3	**2.6**	28.2	23.2	**-5**	35.5	32.9	**-2.6**
Mobile	98.5	92.6	**-5.9**	95.4	88.9	**-6.5**	96.3	89.7	**-6.6**	95.2	90.3	**-4.9**
Household Deprivation	**Replaced/Replacements – Intervention**			**Replaced/Replacements - Comparison**			**Recontacted – Intervention**			**Recontacted - Comparison**		
	BL	ML	Diff	BL	ML	Diff	BL	ML	Diff	BL	ML	Diff
Goes to sleep hungry many/most days	10.4	5.9	**-4.6**	9.2	6.8	**-2.5**	8.4	12.4	**3.9**	7.7	12.1	**4.3**
Goes to sleep hungry some days or more often	29.9	29.4	**-.4**	36.9	25.4	**-11.5**	31.8	31.1	**-.7**	30.2	31.3	**1**
Gone without clean water many/most days	32.4	25.4	**-7**	16.9	32.8	**15.9**	26.8	29.3	**2.5**	28.1	29.9	**1.8**

Gone without clean water some days or more often	69.1	53.7	**-15.4**	60	47.5	**-12.5**	69	58.8	**-10.1**	62.6	62	**-.6**
School Characteristics	**Replaced/Replacements – Intervention**			**Replaced/Replacements - Comparison**			**Recontacted – Intervention**			**Recontacted - Comparison**		
	BL	**ML**	**Diff**	**BL**	**ML**	**Diff**	**BL**	**ML**	**Diff**	**BL**	**ML**	**Diff**
Language of instruction not Somali	13.2	13.2	**0**	3.1	3.1	**0**	10.2	10.2	**0**	4.5	4.5	**0**
School Barriers	**Replaced/Replacements – Intervention**			**Replaced/Replacements - Comparison**			**Recontacted – Intervention**			**Recontacted - Comparison**		
	BL	**ML**	**Diff**	**BL**	**ML**	**Diff**	**BL**	**ML**	**Diff**	**BL**	**ML**	**Diff**
Caregiver believes travel to school is unsafe for girls	5.9	3.1	**-2.8**	6.2	1.6	**-4.5**	5.2	1.9	**-3.4**	6.2	1	**-5.1**
Girl feels unsafe travelling to school	0	3	**3**	14.5	0	**-14.5**	4.9	.8	**-4.1**	8.5	.4	**-8.1**
Doesn't feel safe at school	5.1	4.5	**-.6**	9.1	0	**-9.1**	6.3	.3	**-6**	5.6	1.1	**-4.4**
Difficult to move around school	15.3	6	**-9.3**	18.2	18.5	**.3**	18.2	13.4	**-4.8**	20.2	13.8	**-6.5**
Doesn't use drinking water facilities	18.6	3	**-15.7**	43.6	18.5	**-25.2**	21.5	11	**-10.5**	31.7	12.6	**-19.1**
Doesn't use toilet at school	15.3	28.4	**13.1**	47.3	41.5	**-5.7**	25.3	26.9	**1.6**	30.6	21.2	**-9.4**
Doesn't use areas where children play/socialize	37.3	26.9	**-10.4**	63.6	20	**-43.6**	41	26.9	**-14.2**	45.8	35.4	**-10.3**
Disagrees teachers make them feel welcome	1.7	3	**1.3**	3.6	1.6	**-2.1**	5.7	2.2	**-3.6**	8.9	3	**-5.9**

Agrees teachers treat boys and girls differently in the classroom	49.2	31.8	**-17.3**	51.9	28.6	**-23.3**	40	31.2	**-8.8**	39.7	32.3	**-7.4**
Agrees teachers often absent from class	34.5	23.1	**-11.4**	50.9	34.4	**-16.5**	34.3	22.4	**-12**	36.5	20.2	**-16.4**
Afraid of teacher	64.4	67.2	**2.8**	64.8	68.3	**3.4**	56.8	64.9	**8**	55.4	71.9	**16.6**
Does not feel comfortable asking teacher questions	6.8	0	**-6.8**	9.1	4.7	**-4.4**	4.3	1.6	**-2.7**	5.5	1.5	**-4**
Teacher punishes students who get things wrong in a lesson	79.7	59.7	**-20**	76.4	76.6	**.2**	76.9	69.9	**-7**	75.7	65.1	**-10.7**
No computers at school	86.2	88.1	**1.9**	96.3	96.9	**.6**	88.3	88.1	**-.1**	95.6	92.9	**-2.7**
Cannot use books or other learning materials at school	16.9	10.4	**-6.5**	20	16.9	**-3.1**	20.1	15.1	**-5**	29	17.8	**-11.2**
Not enough seats for all students	18.6	0	**-18.6**	40	1.5	**-38.5**	23.3	1.6	**-21.7**	28	2.6	**-25.4**
Caregiver says principal is of poor quality	9.4	3.1	**-6.2**	8.3	3.2	**-5.2**	4.1	1.1	**-3**	5.6	1.6	**-4.1**
Caregiver says teaching is of poor quality	1.6	1.5	**0**	5	3.3	**-1.7**	3.1	1.1	**-2**	5.6	.8	**-4.9**
Girls says teacher does not asks girls and boys an equal amount of questions	8.5	11.9	**3.5**	12.7	7.8	**-4.9**	8.7	5.6	**-3.1**	11.8	5.7	**-6.1**
Girl says teacher does not ask girls and boys	10.3	10.4	**.1**	10.9	9.4	**-1.5**	11.7	8.4	**-3.4**	11.1	8.7	**-2.4**

questions of equal difficulty												
Household Barriers	**Replaced/Replacements – Intervention**			**Replaced/Replacements - Comparison**			**Recontacted – Intervention**			**Recontacted - Comparison**		
	BL	**ML**	**Diff**	**BL**	**ML**	**Diff**	**BL**	**ML**	**Diff**	**BL**	**ML**	**Diff**
High chore burden	10.3	11.8	**1.5**	13.8	4.8	**-9.1**	10.7	9	**-1.6**	10.5	9.7	**-.8**
Doesn't get support from family to stay in school	3.4	4.4	**1**	1.8	1.6	**-.3**	4.9	1.7	**-3.2**	1.5	2.2	**.8**
Girl feels that she has no choice whether to attend or stay in school	81.8	89.7	**7.9**	94	89.1	**-4.9**	89	86	**-3**	84.6	83.7	**-.9**
Girl feels family makes schooling decisions for her	25.8	29.4	**3.7**	21.5	41.5	**20**	23.8	33.5	**9.7**	29.6	32.6	**2.9**
Distance to school is greater than 30 minutes	1.5	0	**-1.5**	0	4.8	**4.8**	1	1.9	**.9**	0	2.6	**2.6**
PCD and family members are not involved in CEC	12	20.9	**8.9**	17.6	20.6	**2.9**	16.7	20.7	**4.1**	18.5	29.9	**11.4**

ML1 to ML2 Sample

Expanding on the analysis in the previous section, we now turn to changes in the composition of the sample between ML1 and ML2 that could arise from the inclusion of replacement and replaced girls. The structure of the table mirrors that presented for BL-to-ML2 comparisons, above. The findings are also similar to those for the BL-to-ML2 comparison:

- The trend among replaced/replacement girls is for the comparison group to become *less marginalized* from ML1 to ML2, relative to the intervention group.
- This trend is fundamentally different from the trend among the panel sample

Combined, these two facts suggest that the selection of replacement girls changes the sample composition in a way that produces bias *against* finding program impacts, because the comparison group becomes less marginalized as a result of the replacement process.

Again, the results for Somali literacy and numeracy (last two groups in this section) are consistent with this idea. Among the panel sample, the program has no discernible impact on either learning outcome – the trends over time from ML1 to ML2 are nearly parallel. However, among the replaced/replacement sample, the comparison group has far outpaced the intervention group, which is what we would expect if replacements selected in comparison schools were more privileged – relative to the girls they replaced – when compared to those selected in intervention areas.

TABLE 124: CHARACTERISTICS OF REPLACED/REPLACEMENT GIRLS & RE-CONTACTED ("PANEL") GIRLS IN INTERVENTION AND COMPARISON SCHOOLS (ML1 TO ML2)

Girl Characteristics	Replaced/Replacements – Intervention			Replaced/Replacements - Comparison			Recontacted – Intervention			Recontacted - Comparison		
	ML1	ML2	Diff	ML1	ML2	Diff	ML1	ML2	Diff	ML1	ML2	Diff
Living without both parents	10	15.7	**5.7**	9	7.7	**-1.3**	9.5	9.3	**-.2**	9.3	10.7	**1.5**
Orphan	0	0	**0**	0	0	**0**	0	.6	**.6**	.3	1	**.7**
HoH is female	51.4	58.6	**7.1**	55.2	55.4	**.2**	44	51	**7**	45.4	55.7	**10.3**
HoH has no education	49.3	60	**10.7**	51.5	35.7	**-15.8**	48.3	47.5	**-.8**	49	42.4	**-6.6**
HoH has no formal education	61.2	83.1	**21.9**	75.8	62.5	**-13.3**	70.4	71.5	**1.1**	69.2	71.4	**2.2**
Primary caregiver has no education	58.6	58.6	**0**	62.7	44.6	**-18.1**	51	54.1	**3.1**	59.3	50.7	**-8.6**
Primary caregiver has no formal education	78.6	88.6	**10**	88.1	73.8	**-14.2**	80.1	78.9	**-1.2**	85.1	82.6	**-2.5**
Currently married	4.3	0	**-4.3**	3	0	**-3**	.3	2.5	**2.2**	1.7	3.4	**1.7**
Has ever been married	5.7	0	**-5.7**	3	0	**-3**	.3	2.8	**2.5**	1.7	3.7	**2**
Mother, under 16	1.4	0	**-1.4**	1.5	0	**-1.5**	0	.3	**.3**	2	2	**0**
Mother, over 16	0	0	**0**	0	0	**0**	0	0	**0**	1	0	**-1**
Girls with any disability	8.6	5.7	**-2.9**	20.9	0	**-20.9**	14.5	8.9	**-5.5**	13.9	8.9	**-5**
Household Characteristics	**Replaced/Replacements – Intervention**			**Replaced/Replacements - Comparison**			**Recontacted – Intervention**			**Recontacted - Comparison**		
	ML1	**ML2**	**Diff**	**ML1**	**ML2**	**Diff**	**ML1**	**ML2**	**Diff**	**ML1**	**ML2**	**Diff**

HH is pastoralist	2.9	8.6	**5.7**	3	7.7	**4.7**	6.4	6.2	**-.2**	10.3	7.4	**-2.9**
HoH has no wage-earning occupation	45.7	40	**-5.7**	44.8	36.9	**-7.9**	48.5	45.4	**-3.1**	45.7	50	**4.3**
Moved in the past 12 months	1.4	2.9	**1.4**	0	1.5	**1.5**	.3	1.4	**1.1**	.7	1.3	**.7**
Seasonal migration	1.4	4.3	**2.9**	0	4.6	**4.6**	.8	3.6	**2.8**	2	3.4	**1.4**
Owns camels	8.6	12.9	**4.3**	10.4	23.1	**12.6**	7.8	9.8	**2**	14.6	20.1	**5.5**
Owns medium-sized livestock	62.9	58.6	**-4.3**	61.2	61.5	**.3**	66.3	68.5	**2.2**	72.5	72.8	**.3**
Owns land	78.6	73.9	**-4.7**	87.5	70.5	**-17**	73.8	72.4	**-1.4**	62.5	74.3	**11.8**
House is informal / temporary structure	8.8	3.1	**-5.7**	4.5	8.2	**3.7**	6.6	3.9	**-2.7**	6.8	7.5	**.7**
Poor quality roof	28.6	30	**1.4**	19.4	33.8	**14.4**	26.7	24.3	**-2.4**	34.6	33.3	**-1.2**
Mobile	90	92.9	**2.9**	77.6	89.2	**11.6**	90.8	91	**.2**	83.8	89.9	**6.2**
Household Deprivation	**Replaced/Replacements – Intervention**			**Replaced/Replacements - Comparison**			**Recontacted – Intervention**			**Recontacted - Comparison**		
	ML1	**ML2**	**Diff**	**ML1**	**ML2**	**Diff**	**ML1**	**ML2**	**Diff**	**ML1**	**ML2**	**Diff**
Goes to sleep hungry many/most days	4.5	5.7	**1.2**	12.1	6.6	**-5.6**	10.3	12.6	**2.3**	11.4	12.2	**.8**
Goes to sleep hungry some days or more often	28.8	28.6	**-.2**	28.8	24.6	**-4.2**	29.3	32	**2.7**	28.9	33.4	**4.6**
Gone without clean water many/most days	27.5	26.1	**-1.4**	23.9	32.3	**8.4**	32.9	29	**-3.9**	28.7	31	**2.3**
Gone without clean water some days or more often	66.7	53.6	**-13**	58.2	48.4	**-9.8**	73	60.3	**-12.7**	67	64	**-3**

School Characteristics	Replaced/Replacements – Intervention			Replaced/Replacements - Comparison			Recontacted – Intervention			Recontacted - Comparison		
	ML1	ML2	Diff	ML1	ML2	Diff	ML1	ML2	Diff	ML1	ML2	Diff
Language of instruction not Somali	12.9	12.9	**0**	3	3	**0**	7.2	7.2	**0**	4.6	4.6	**0**
School Barriers	**Replaced/Replacements – Intervention**			**Replaced/Replacements - Comparison**			**Recontacted – Intervention**			**Recontacted - Comparison**		
	ML1	ML2	Diff	ML1	ML2	Diff	ML1	ML2	Diff	ML1	ML2	Diff
Caregiver believes travel to school is unsafe for girls	15.4	3	**-12.4**	0	1.6	**1.6**	11.9	2.7	**-9.2**	2.3	1.1	**-1.2**
Girl feels unsafe travelling to school	7.9	2.9	**-5**	0	0	**0**	2.3	.6	**-1.7**	3.5	.7	**-2.8**
Doesn't feel safe at school	3.2	4.3	**1.2**	1.7	0	**-1.7**	1.4	0	**-1.4**	1.8	1.1	**-.6**
Difficult to move around school	11.1	7.2	**-3.9**	27.1	17.9	**-9.2**	17.6	14.7	**-2.8**	23.9	14.2	**-9.7**
Doesn't use drinking water facilities	7.9	2.9	**-5**	27.1	17.9	**-9.2**	10.4	10.9	**.5**	24.6	13.1	**-11.5**
Doesn't use toilet at school	15.9	27.5	**11.7**	32.2	40.3	**8.1**	21.3	25.7	**4.4**	29	22.5	**-6.5**
Doesn't use areas where children play/socialize	31.7	26.1	**-5.7**	55.9	19.4	**-36.5**	35.4	27.8	**-7.6**	52.8	35.3	**-17.5**
Disagrees teachers make them feel welcome	4.8	2.9	**-1.9**	5.1	1.5	**-3.6**	3.5	2.7	**-.8**	5.7	3	**-2.7**
Agrees teachers treat boys and girls differently in the classroom	41	30.9	**-10.1**	51.7	29.2	**-22.5**	43.6	33.1	**-10.4**	45.4	31.5	**-13.9**
Agrees teachers often absent from class	10.3	22.4	**12**	35.1	33.3	**-1.8**	17.9	22.6	**4.7**	28.5	20.8	**-7.7**

Afraid of teacher	73	66.7	**-6.3**	84.7	67.7	**-17.1**	75.8	65.7	**-10.1**	79.9	70.6	**-9.3**
Does not feel comfortable asking teacher questions	1.6	0	**-1.6**	1.7	4.5	**2.9**	1.4	2.1	**.6**	3.5	1.1	**-2.4**
Teacher punishes students who get things wrong in a lesson	84.1	58	**-26.2**	94.9	75.8	**-19.2**	80.4	67.5	**-12.9**	84.5	64.4	**-20.1**
No computers at school	100	88.4	**-11.6**	89.8	97	**7.2**	95.1	91.4	**-3.7**	94.7	93.2	**-1.5**
Cannot use books or other learning materials at school	15.9	10.1	**-5.7**	33.9	19.4	**-14.5**	20.5	15.1	**-5.4**	29.2	16.9	**-12.4**
Not enough seats for all students	15.9	0	**-15.9**	20.3	1.5	**-18.8**	13.5	2.1	**-11.5**	15.1	2.6	**-12.5**
Caregiver says principal is of poor quality	6.7	3	**-3.6**	3.4	3.1	**-.3**	3.5	1.8	**-1.7**	1.5	1.6	**.1**
Caregiver says teaching is of poor quality	3.4	1.5	**-1.9**	1.7	3.2	**1.5**	2.3	1.2	**-1.1**	2.1	.8	**-1.4**
Girls says teacher does not asks girls and boys an equal amount of questions	14.3	13	**-1.2**	10.2	7.6	**-2.6**	8.6	5.6	**-3**	9.2	5.3	**-3.9**
Girl says teacher does not ask girls and boys questions of equal difficulty	15.9	13	**-2.8**	16.9	9.1	**-7.9**	14.4	8.3	**-6.1**	10.6	8.3	**-2.2**
Household Barriers	**Replaced/Replacements – Intervention**			**Replaced/Replacements - Comparison**			**Recontacted – Intervention**			**Recontacted - Comparison**		
	ML1	**ML2**	**Diff**	**ML1**	**ML2**	**Diff**	**ML1**	**ML2**	**Diff**	**ML1**	**ML2**	**Diff**

Doesn't get support from family to stay in school	1.6	4.3	**2.7**	0	1.5	**1.5**	2	1.7	**-.3**	3.9	2.3	**-1.6**
Girl feels that she has no choice whether to attend or stay in school	85.7	88.6	**2.9**	98.3	89.4	**-8.9**	85.9	85.8	**-.1**	93	84.1	**-8.9**
Girl feels family makes schooling decisions for her	25.7	31.4	**5.7**	22.4	40.3	**17.9**	34	31.5	**-2.5**	35.8	34.8	**-1**
Distance to school is greater than 30 minutes	1.6	0	**-1.6**	0	4.6	**4.6**	0	1.8	**1.8**	1.1	3	**1.9**
PCD and family members are not involved in CEC	16.3	20	**3.7**	8.8	20	**11.2**	16.6	21.8	**5.2**	19.3	28.4	**9.1**

Replacements in Full Sample

The analysis in the previous two sections showed that replacement girls were fundamentally different from the girls they replaced in many cases. More problematic, however, is the finding that these differences vary between intervention and comparison communities. The inclusion of replacement girls in the sample essentially has contributed to a reduction in marginalization among the comparison sample vis-à-vis the intervention sample. The result is that the intervention and comparison samples – where they include replacement and replaced girls – have not just become *less similar* to one another, but that the underlying trend lines between them have diverged. Recall that the assumption underlying difference-in-differences analysis is that, in the absence of an intervention, the intervention and comparison groups would evolve along parallel trends. Including replacement and replaced girls in the sample alters those underlying trends, reducing our confidence in the parallel trends assumption. This fact is the primary motivation behind our heavy reliance, throughout this report, on the "true panel" sample, which includes only those girls who were successfully re-contacted in all of the rounds under consideration, thereby excluding replacement and replaced girls.

In the tables below, we provide an additional view of these trends. The first table documents the characteristics of the "replacement" sample – all those girls who were replacements or replaced, excluding any true panel girls – across rounds, disaggregated by intervention and comparison communities. We calculate the change in characteristics between baseline and ML2 for this replacement sample, and then include the change over the same time period in the "re-contacted sample" (the true panel) for comparison. For instance, in intervention communities, the share of girls in the replacement sample who live apart from their parents has shifted by 5.7 points from baseline to ML2, reflecting both changes in the same girls' status, and the changing composition of the sample as girls were replaced. In contrast, the share of girls living without their parents – in the same communities – in the re-contacted/panel sample has shifted by 3.2 points. The gap between these two shifts is an estimate of how much the replacement process might be changing the nature of the sample, as the gap is a function of either differences between replacement and replaced girls (over and above any changes that occurred among girls in the community), which alters the sample composition, or changes in the re-contacted/panel girls' responses over time.[342]

The second table below provides describes the true panel and cross-sectional samples over time, in terms of the characteristics of girls and their households. This table simply provides an alternative way of thinking about the differences between the replacement and the re-contacted/panel sample. The "true panel" includes 529 girls who were successfully contacted in each round of the evaluation. The cross-sectional sample includes all of those girls, plus the girls who were replaced, and their replacements. The only in-school girls excluded from the cross-sectional sample are those who fell out of the sample and were not replaced. Therefore, the cross-sectional sample is the union of the re-contacted/true panel sample, and the replacement sample, and consists of 798 girls in each of the three rounds of data collection.

The important conclusion that emerges from the second table is that the differences in trends between the true panel and the cross-sectional samples over all three evaluation rounds are not particularly different.

[342] The latter point, while concerning, is almost certainly true in some cases. Research in survey methodology has shown that respondents with fixed characteristics often report them differently across waves of a survey, or even within the same survey. Therefore, we cannot rule out the possibility that some re-contacted/panel girls have changed their answers without their underlying demography changing. Nonetheless, the critical takeaway point is that the replacement and re-contacted samples have not evolved in the same way, generally, in intervention and comparison communities, and that this underscores the value of using the true panel sample.

There is variation, of course, but the trends generally align with one another and the BL-to-ML2 differences calculated for the two samples are similar. This is not at all surprising – 66.3 percent of the girls in the cross-sectional sample are "true panel" girls, so trends among this group tend to outweigh other changes over time among the smaller set of replaced/replacement girls that make up the remainder of the cross-sectional sample. That is, there are too few replaced/replacement girls to fundamentally shift the cross-sectional sample relative to the true panel.

We emphasize that we are making an important distinction here regarding our reason for relying on the true panel heavily throughout the report. Our concern is *not* about the representativeness of the sample, which might be better served by using the cross-sectional sample, which its larger number of observations and less aggressive exclusion of girls from the analysis. We are not overly concerned about representativeness because the basic characteristics of the sample are not altered dramatically by relying on the true panel instead of the cross-sectional sample – as the second table below shows. At the same time, we do have significant, and empirically-validated concerns regarding the similarity between replacement girls and the girls they replaced, and differential levels of similarity between intervention and comparison schools. This was demonstrated conclusively in the previous section for BL-to-ML2 comparisons, ML1-to-ML2 comparisons and again below (in the first table below). The replacement sample tends to produce a less marginalized comparison group over time, vis-à-vis the intervention group, contributing to potential bias in our estimates of program impact. These two findings, combined, motivate our reliance on the true panel sample throughout most of this report.

TABLE 125: REPLACED/REPLACEMENT GIRLS BY ROUND, DISAGGREGATED BY INTERVENTION STATUS

Girl Characteristics	Replaced/Replacements – Intervention				Re-Contacted – Intervention	Replaced/Replacements – Comparison				Re-Contacted – Comparison
	BL	ML1	ML2	BL-ML2 Diff	BL-ML2 Diff	BL	ML1	ML2	BL-ML2 Diff	BL-ML2 Diff
Living without both parents	5.6	10.5	11.3	5.7	3.2	11.9	11.1	9.8	-2.1	4.2
Orphan	0	0	0	0	0.4	1.6	0.8	0.8	-0.8	0.8
HoH is female	47.6	47.6	54.5	7	9.4	52.4	56.3	56.1	3.7	13
HoH has no education	77.6	47.5	55.9	-21.7	-18.6	66.7	46.7	41.1	-25.5	-20.9
HoH has no formal education	77.6	64	77.2	-0.4	6.6	66.7	72.1	69.2	2.5	8
Primary caregiver has no education	81.8	56.6	58	-23.8	-12.6	71.4	61.1	48.8	-22.6	-23.7
Primary caregiver has no formal education	86.7	79.7	85.2	-1.5	0.2	79.4	87.3	78.9	-0.5	2.2
Currently married	2.8	2.1	0	-2.8	2.1	1.6	2.4	2.4	0.9	1.7
Has ever been married	2.8	2.8	0	-2.8	2.5	1.6	2.4	2.4	0.9	2.1
Mother, under 16	0.7	0.7	0	-0.7	0	0.8	1.6	1.6	0.8	1.3
Mother, over 16	0.7	0	0	-0.7	-0.3	0.8	0	0	-0.8	0
Girls with any disability	3.5	9.8	8.5	4.9	-0.3	7.1	16.7	3.2	-4	5.3
Household Characteristics	**Replaced/Replacements – Intervention**				**Re-Contacted – Intervention**	**Replaced/Replacements – Comparison**				**Re-Contacted – Comparison**
	BL	**ML1**	**ML2**	**BL-ML2 Diff**	**BL-ML2 Diff**	**BL**	**ML1**	**ML2**	**BL-ML2 Diff**	**BL-ML2 Diff**
HH is pastoralist	13.3	5.6	10.5	-2.8	-4.2	5.6	5.6	9.8	4.2	-1.2
HoH has no wage-earning occupation	32.9	52.4	41.3	8.4	4.5	34.9	46.8	43.1	8.2	11.7

Moved in the past 12 months	0.7	0.7	2.1	1.4	1.4	0.8	1.6	1.6	0.8	0.4
Seasonal migration	4.9	1.4	4.2	-0.7	0.4	2.4	1.6	4.1	1.7	-1.6
Owns camels	13.3	7.7	14.8	1.5	2.5	6.3	15.1	22.8	16.4	12.2
Owns medium-sized livestock	63.6	69.9	66.2	2.6	10.7	47.2	61.9	67.5	20.3	5.4
Owns land	79	74.7	73.9	-5.1	-7.2	83.1	66.1	73.9	-9.1	-9.6
House is informal / temporary structure	3.5	9.3	4.5	1	0.6	4	4	6.9	2.9	0.5
Poor quality roof	32.9	30.8	31.2	-1.7	-6.6	29.5	27	32.5	3	-1.8
Mobile	97.9	90.9	93.7	-4.2	-6	94.4	82.5	88.6	-5.8	-5.1
Household Deprivation	**Replaced/Replacements – Intervention**				**Re-Contacted – Intervention**	**Replaced/Replacements – Comparison**				**Re-Contacted – Comparison**
	BL	**ML1**	**ML2**	**BL-ML2 Diff**	**BL-ML2 Diff**	**BL**	**ML1**	**ML2**	**BL-ML2 Diff**	**BL-ML2 Diff**
Goes to sleep hungry many/most days	9.9	8.6	7.7	-2.1	6.4	7.9	8.9	9.2	1.3	3.9
Goes to sleep hungry some days or more often	33.8	30.2	29.6	-4.2	3	28.6	28.5	28.6	0	0
Gone without clean water many/most days	32.2	28.2	29.1	-3.1	1.7	18.3	29.6	28.3	10.1	1.8
Gone without clean water some days or more often	74.8	74.6	53.2	-21.6	-7.7	60.3	67.2	55	-5.3	1.1
School Characteristics	**Replaced/Replacements – Intervention**				**Re-Contacted – Intervention**	**Replaced/Replacements – Comparison**				**Re-Contacted – Comparison**
	BL	**ML1**	**ML2**	**BL-ML2 Diff**	**BL-ML2 Diff**	**BL**	**ML1**	**ML2**	**BL-ML2 Diff**	**BL-ML2 Diff**
Language of instruction not Somali	9.8	9.8	9.8	0	0	2.4	2.4	2.4	0	0

School Barriers	Replaced/Replacements – Intervention				Re-Contacted – Intervention	Replaced/Replacements – Comparison				Re-Contacted – Comparison
	BL	ML1	ML2	BL-ML2 Diff	BL-ML2 Diff	BL	ML1	ML2	BL-ML2 Diff	BL-ML2 Diff
Caregiver believes travel to school is unsafe for girls	9.1	12.5	2.9	-6.1	-1.6	4.8	1.9	0.8	-4	-5.4
Girl feels unsafe travelling to school	2.4	3.7	1.4	-0.9	-3.1	11.2	2.5	0.8	-10.4	-7.5
Doesn't feel safe at school	6.3	1.5	2.2	-4.1	-4.6	6.6	2.5	0.8	-5.8	-4.7
Difficult to move around school	13.4	16.2	14.5	1.1	-6.5	20.6	21.2	16	-4.6	-5.7
Doesn't use drinking water facilities	23.6	8.8	9.4	-14.2	-10.3	43	28.8	17.6	-25.3	-18
Doesn't use toilet at school	22.8	24.3	25.4	2.5	2.6	46.7	32.2	34.5	-12.3	-6.3
Doesn't use areas where children play/socialize	44.9	39	26.8	-18.1	-11.2	51.4	57.6	26.9	-24.5	-12.4
Disagrees teachers make them feel welcome	4	3	4.3	0.4	-3.9	6.5	5.1	0.8	-5.7	-4.7
Agrees teachers treat boys and girls differently in the classroom	37.8	42.1	33.1	-4.7	-8	44.3	50.4	29.2	-15.1	-8.1
Agrees teachers often absent from class	31.7	13.8	20.9	-10.9	-11.9	36.8	33.3	27.9	-8.9	-17.9
Afraid of teacher	58.3	74.3	71.7	13.5	6.5	57.1	78.8	64.7	7.5	16.4
Does not feel comfortable asking teacher questions	4.7	1.5	1.4	-3.3	-3.1	8.4	4.2	2.5	-5.9	-2.8

Teacher punishes students who get things wrong in a lesson	76.2	83.1	58	-18.2	-6.4	69.2	93.2	75.4	6.3	-17.6
No computers at school	89.7	97.8	89.1	-0.6	1.1	97.2	92.3	94.1	-3.1	-0.9
Cannot use books or other learning materials at school	19.7	19.1	14.5	-5.2	-6.5	28	33.9	17.6	-10.4	-9.4
Not enough seats for all students	19.7	14.7	2.2	-17.5	-20.3	32.7	21.2	2.5	-30.2	-26.3
Caregiver says principal is of poor quality	5.8	4.7	3	-2.8	-2.8	6	4.3	3.5	-2.4	-4.5
Caregiver says teaching is of poor quality	1.4	3.8	0.7	-0.7	-2.1	8.5	1.7	2.7	-5.8	-3.4
Girls says teacher does not asks girls and boys an equal amount of questions	7.9	12.5	10.1	2.3	-4.8	12.1	10.2	5.9	-6.2	-6.1
Girl says teacher does not ask girls and boys questions of equal difficulty	7.9	16.2	11.6	3.7	-6.8	14	13.6	5.9	-8.1	0

Household Barriers	**Replaced/Replacements – Intervention**				**Re-Contacted – Intervention**	**Replaced/Replacements – Comparison**				**Re-Contacted – Comparison**
	BL	**ML1**	**ML2**	**BL-ML2 Diff**	**BL-ML2 Diff**	**BL**	**ML1**	**ML2**	**BL-ML2 Diff**	**BL-ML2 Diff**
High chore burden	7.7	5.3	12.7	5	-1.7	15.1	10.2	5.7	-9.4	0.5
Doesn't get support from family to stay in school	1.6	2.2	2.1	0.5	-4.4	2.8	2.5	3.2	0.4	0.7

Girl feels that she has no choice whether to attend or stay in school	85.4	84.6	83.9	-1.4	-0.1	89.2	96.6	88.8	-0.4	-1.1
Girl feels family makes schooling decisions for her	26.6	30.1	28	1.4	10.8	23.2	26.2	43.7	20.5	0.6
Distance to school is greater than 30 minutes	1.4	0.8	1.5	0.1	0.8	0	0.9	6	6	1.9
PCD and family members are not involved in CEC	12.6	17.6	23.1	10.5	5	23.6	15.6	16.9	-6.7	16

TABLE 126: SAMPLE CHARACTERISTICS FOR PANEL AND CROSS-SECTIONAL SAMPLES, ACROSS ROUNDS

Girl Characteristics	Full Panel Sample (No replaced/replacements)				Full Cross-Sectional Sample (Includes replaced/replacements)			
	BL	ML1	ML2	BL-ML2 Diff	BL	ML1	ML2	BL-ML2 Diff
Living without both parents	6.4	8.7	10.1	3.7	7.1	9.4	10.3	3.1
Orphan	0.2	0	0.8	0.6	0.4	0.1	0.6	0.3
HoH is female	42	43.3	53.1	11.1	44.6	46.1	53.8	9.2
HoH has no education	63.7	49.8	44	-19.7	66.6	48.9	45.8	-20.8
HoH has no formal education	63.7	70.5	70.9	7.2	66.6	69.6	71.8	5.2
Primary caregiver has no education	69.4	54.3	51.7	-17.7	71.9	55.8	52.4	-19.5
Primary caregiver has no formal education	78.8	82.2	80	1.1	80.3	82.5	80.7	0.4
Currently married	1.1	0.9	3.1	1.9	1.5	1.4	2.4	0.9
Has ever been married	1.1	0.9	3.4	2.3	1.5	1.5	2.7	1.2
Mother, under 16	0.4	0.9	1	0.6	0.5	1	0.9	0.4
Mother, over 16	0.2	0.6	0	-0.2	0.4	0.4	0	-0.4
Girls with any disability	6.6	14.9	8.9	2.3	6.1	14.3	7.9	1.8
Household Characteristics	**Full Panel Sample (No replaced/replacements)**				**Full Cross-Sectional Sample (Includes replaced/replacements)**			
	BL	ML1	ML2	BL-ML2 Diff	BL	ML1	ML2	BL-ML2 Diff
HH is pastoralist	8.1	8.1	5.3	-2.8	8.6	7.3	7	-1.7
HoH has no wage-earning occupation	40.1	45.4	47.9	7.8	38	46.9	45.9	8

Moved in the past 12 months	0.4	0.2	1.3	1	0.5	0.5	1.5	1
Seasonal migration	4	1.1	3.4	-0.5	3.9	1.3	3.7	-0.2
Owns camels	6.4	10.4	13.4	6.9	7.6	10.7	15.1	7.4
Owns medium-sized livestock	61.4	68.8	69.7	8.2	59.6	67.9	68.7	9.1
House is informal / temporary structure	5	6.6	5.5	0.5	4.5	6.7	5.5	1
Poor quality roof	32	29.4	27.6	-4.4	31.8	29.2	29	-2.8
Mobile	95.8	87	90.3	-5.6	96	87	90.6	-5.4
Household Deprivation	**Full Panel Sample (No replaced/replacements)**				**Full Cross-Sectional Sample (Includes replaced/replacements)**			
	BL	**ML1**	**ML2**	**BL-ML2 Diff**	**BL**	**ML1**	**ML2**	**BL-ML2 Diff**
Goes to sleep hungry many/most days	7.6	11.2	12.8	5.2	8.1	10.4	11.4	3.3
Goes to sleep hungry some days or more often	31.3	28.9	33	1.6	31.3	29.1	31.7	0.4
Gone without clean water many/most days	28.5	30.7	30.3	1.7	27.6	30.1	29.8	2.2
Gone without clean water some days or more often	66.9	67.8	63.2	-3.7	67.3	68.9	60.2	-7.1
School Barriers	**Full Panel Sample (No replaced/replacements)**				**Full Cross-Sectional Sample (Includes replaced/replacements)**			
	BL	**ML1**	**ML2**	**BL-ML2 Diff**	**BL**	**ML1**	**ML2**	**BL-ML2 Diff**
Caregiver believes travel to school is unsafe for girls	5.3	0.0	2	-3.3	5.9	8.1	2	-3.9

Girl feels unsafe travelling to school	5.7	3	0.6	-5.1	5.9	3.1	0.8	-5.1
Doesn't feel safe at school	5.1	1.6	0.4	-4.7	5.5	1.7	0.8	-4.7
Difficult to move around school	19.7	21	13.6	-6.1	18.7	20.2	14.2	-4.6
Doesn't use drinking water facilities	24.5	16.2	10.7	-13.8	27.1	16.9	11.6	-15.5
Doesn't use toilet at school	25.5	22.9	24.2	-1.4	28.2	24.6	26	-2.2
Doesn't use areas where children play/socialize	42.8	41.1	31.1	-11.8	44.5	43.3	29.6	-14.9
Disagrees teachers make them feel welcome	6.9	4.8	2.7	-4.3	6.4	4.5	2.7	-3.6
Agrees teachers treat boys and girls differently in the classroom	40.4	44	32.3	-8	40.5	44.7	32	-8.5
Agrees teachers often absent from class	36.8	22.5	22.2	-14.6	35.9	22.6	22.9	-13.1
Afraid of teacher	56.4	78.5	67.3	10.9	56.9	77.8	67.7	10.8
Does not feel comfortable asking teacher questions	4.6	2	1.6	-3	5.2	2.3	1.8	-3.5
Teacher punishes students who get things wrong in a lesson	77.7	81.2	66.3	-11.4	76.2	83.4	66.2	-9.9
No computers at school	92.6	94.8	92.7	0.1	92.8	94.9	92.3	-0.5
Cannot use books or other learning materials at school	23.3	23.6	15.5	-7.8	23.4	24.4	15.7	-7.7

Not enough seats for all students	24.8	13.4	1.9	-23	25.1	14.9	2	-23.1
Caregiver says principal is of poor quality	4.9	2.3	1.3	-3.6	5.2	3	2	-3.3
Caregiver says teaching is of poor quality	3.7	2	1.1	-2.7	4	2.3	1.3	-2.8
Girls says teacher does not asks girls and boys an equal amount of questions	10.8	8.4	5.4	-5.4	10.5	9.4	6.4	-4.1
Girl says teacher does not ask girls and boys questions of equal difficulty	12.5	12.4	8.7	-3.7	11.9	13.3	8.8	-3.1
Household Barriers	**Full Panel Sample (No replaced/replacements)**				**Full Cross-Sectional Sample (Includes replaced/replacements)**			
	BL	**ML1**	**ML2**	**BL-ML2 Diff**	**BL**	**ML1**	**ML2**	**BL-ML2 Diff**
Doesn't get support from family to stay in school	4	2.6	1.9	-2.1	3.4	2.5	2.1	-1.2
Girl feels that she has no choice whether to attend or stay in school	85.9	89.2	85.4	-0.5	86.3	89.5	85.7	-0.7
Girl feels family makes schooling decisions for her	26.4	35.3	32.5	6.1	25.9	33	33.5	7.5

Distance to school is greater than 30 minutes	0.4	0.4	1.7	1.3	0.5	0.5	2.3	1.8
PCD and family members are not involved in CEC	15.8	17.2	25.3	9.5	16.3	17.1	23.7	7.4

LIMITATIONS AND CHALLENGES

- **Non-random assignment**

The evaluation of SOMGEP-T employs a quasi-experimental design, as discussed earlier in this section. The difference-in-differences design relies on a key assumption around "parallel trends" – that, in the absence of the intervention, intervention and comparison schools would have evolved similarly from baseline to endline. The validity of this assumption is reduced by the fact that the intervention was not randomly assigned to schools/communities; rather, communities were selected for inclusion in SOMGEP-T programming, after which comparison schools were recruited into the sample by CARE's monitoring and evaluation team. Moreover, the baseline evaluation – and the description of the sample in this report – showed that the intervention and comparison groups were not entirely comparable. In the absence of random assignment, we must make stronger, and less tenable, assumptions to draw inferences regarding program impact.

Despite this concern, there are no *specific* reasons to suspect that the parallel trends assumption is violated. Even if intervention and comparison groups started from different baseline values, the evaluation's design controls for this fact explicitly. We spend considerable time, especially in our discussion of re-contact and attrition rates, assessing whether there are valid concerns regarding differential attrition, which would violate the parallel trends assumption. In other sections of the report, we typically limit our analysis to the "true panel" of girls or caregivers, to ensure that the sample has stayed consistent across rounds, and sampling variation cannot account for our results.

- **Panel Attrition**

Perhaps the most important threat to inference in the SOMGEP-T evaluation is the extent of panel attrition from one round to another. At ML1, the attrition rate was 19.7 percent. Many of the girls who fell out of the sample at that time were brought back into the sample at ML2, but attrition nonetheless accumulates in each round: in this round, total attrition since the baseline was 41.8 percent. Attrition was concentrated among OOS girls, but not exclusively so – among girls who were enrolled at baseline, net attrition over the two-year period was 27.4 percent.

Panel attrition tends to be high in Somalia, due to the number of nomadic households and the fact that drought and conflict can displace large numbers of families quickly. Even within the Somali context, SOMGEP-T faces particularly acute challenges, because it targets rural, pastoralist, marginalized areas in eastern Somaliland, western Puntland, and western Galmudug. These areas have higher rates of seasonal and annual migration than agricultural regions in South-Central Somalia or urban areas. The design of the evaluation took into account the potential for significant attrition by increasing the sample size at baseline. Two further steps have been taken: CARE has sought to minimize school-level attrition (i.e. the dropping and/or replacement of entire schools from round to round), and the evaluation team has sought to re-contact girls who fell out of the sample in prior years, even if they were replaced in a previous round. Both approaches have helped to reduce overall attrition. Further efforts are needed at endline, in order to protect the rigor of the evaluation.

- **Bias from Replacement Girls**

More than almost any other factor, panel attrition and replacement of cohort girls in later rounds has the potential to introduce bias into our analysis. As mentioned above, panel attrition in the sample is high, and replacement can introduce differences between the intervention and comparison groups that contribute to violations of the parallel trends assumption. As part of the ML1 evaluation, we analysed the extent to which

replacement girls were similar to the girls they replaced. We repeat this analysis this year, and find that age is the most important predictor of attrition and replacement, implying that the sample is becoming slightly younger than it would, in the absence of replacement, over time.

We take two approaches to mitigating the risk of bias from panel attrition and replacement. First, we minimize attrition as much as possible, and select replacements from the same grade as the replaced girl, wherever possible. However, this approach ensures similarity on only two dimensions (same school, same grade), without accounting for unobserved differences between replaced girls and their replacements. Second, and more importantly, we typically limit our analysis to the "true panel", girls who were successfully contacted at both baseline and ML2. This decision reduces the available sample size and statistical power, but reduces potential attrition bias dramatically.[343]

- **Estimating attendance – inaccuracy of school record-keeping:**

At the outset of the baseline evaluation, the evaluation team and CARE staff recognized that collecting attendance data from school records would be challenging. Past experience conducting GEC evaluations in the region suggested that attendance records would be incomplete and of poor quality, where they existed at all. The baseline and ML1 evaluation included analysis of the reliability of attendance records, which revealed problems with record-based attendance data of this kind. At baseline and ML1, the evaluation team relied primarily on physical headcounts conducted by field teams, and supplemented this analysis with data collected from caregivers and school records.

We follow this approach again in this round, giving pride-of-place to physical headcounts conducted during fieldwork. We also analysed the reliability of the headcounts, and found little reason to question their veracity; headcounts were also subject to extensive quality control during data collection. Given the rich sources of data available to us, we triangulate across multiple indicators, which helps to mitigate the potential bias that might influence any of the measures in isolation.

- **Removal of Outlier and Other Schools**

At the time of the baseline evaluation, five schools in the sample were identified as outliers in terms of their learning scores. As part of the review of the baseline report, CARE, the evaluation team, and the FM agreed that the five outlier schools – all of which were comparison schools – would be removed from the sample going forward. In addition, two schools that were included in the sample at baseline have fallen out of the sample at both ML1 and ML2, for security and accessibility reasons, bringing the total sample to 69 schools (32 comparison, 37 intervention). School-level attrition undermines the research design's statistical power. However, there is no clear relationship between school-level attrition and bias in the difference-in-differences approach, as the removed schools are also retroactively removed from the baseline sample we analyse. In some ways, removal of outliers will actually strengthen the design, insofar as schools that are fundamentally different will likely experience different trajectories over time, which would violate the parallel trends assumption. The cross-sectional sample is less balanced at baseline, but removal of outliers is not an unambiguous source of bias when employing difference-in-differences.

- **Measuring Transition with High Attrition among OOS Girls**

[343] Bias from attrition is still possible, if the girls who fall out of the sample are different between intervention and comparison groups, and this impacts the underlying trends that would have occurred in the two groups under the counterfactual (no intervention) condition.

During the ML1 evaluation, we noted issues related to assessing program impact on transition rates without re-contacting OOS girls. In that evaluation round, OOS girls were not re-contacted, which limited our analysis to the transition pathways (grade progression, drop-out, etc.) that were relevant to girls who were starting from a position of enrolment at baseline.

In this round, both in-school and OOS girls were re-contacted from baseline and ML1. Where OOS girls cannot be re-contacted, they are not replaced. Unfortunately, attrition rates are particularly high among OOS girls, as we document in the next section, resulting in a weaker analysis of transition rates than would otherwise obtain. Our strategy for mitigating this limitation is to gather information about OOS girls who cannot be re-contacted, seeking out their family members, neighbours, friends, and teachers. We collect rudimentary transition information from these individuals, where they have such information, to study crude transition rates among this group of girls. Looking forward to the endline, we recommend dedicating additional time and energy to re-contacting all cohort girls, by using a "mop-up" team that will re-visit schools after the initial visit, to look for the girls a second time. We also recommend using a dedicated team in the office for calling girls and households that have fallen out of the sample, with the goal of conducting a transition-focused interview over the phone with them. This approach will reduce attrition out of the transition cohort, strengthening conclusions that can be drawn regarding transition outcomes.

- **Incomplete Compliance and Contamination of Comparison Sites**

A final limitation of the design is that girls in intervention communities are exposed to SOMGEP-T activities, but their participation is not guaranteed. In methodological terms, this is known as non-compliance, as girls will not necessarily "comply" with the intervention to which they are assigned – they may choose not to participate in GEF activities, they may avoid exposure to the intervention, and so forth. In practice, this concern is over-stated; the validity of our analysis is not contingent on perfect compliance among the intervention group. Rather, our analysis can be interpreted as detecting the average effect of *potential* exposure to the intervention.[344] Such an approach dampens the potential impact estimates, because effect sizes are reduced by girls who are not actually exposed.

Concerns around contamination of the comparison group are also minimal. While migration rates in the region studied are high, most girls who migrate are part of pastoralist, marginalized households, and they are extremely unlikely to enrol in a new school after migrating.[345] Moreover, outside of seasonal or pastoralist migration, the remaining migration tends to consist of girls and their families moving to urban areas in response to environmental shocks, economic shocks (loss of cattle), or to pursue educational opportunities in more populated areas, which might have more secondary schools. Because the SOMGEP-T sample is almost exclusively rural, few migrants are likely to leave one school in the sample and take up residence near a new school in the sample. To the extent that contamination occurs, it is almost certainly far too small to have any tangible impact on our results.

ANNEX 3: CHARACTERISTICS AND BARRIERS

The tables in this annex provide a description of, primarily, the main cohort girls in this evaluation round. Consistent with the approach used throughout this report, we emphasise comparability from baseline to ML2; for this reason, we report provide two sets of tables, employing different samples. In the first table below,

[344] In the econometric literacy on causal inference and program evaluation, this is known as a "local average treatment effect."
[345] This statement is consistent with the vast majority of views expressed in qualitative interviews, both in this and previous rounds.

we report characteristics for girls using the full sample, which includes every girl contacted at baseline *and* ML2, as long as they were drawn from a community that was included in both rounds of the evaluation. Where school-level attrition occurred from baseline to ML2, girls in those schools were excluded from both rounds; all other girls were included, even if they were replaced or fell out of the sample due to individual-level attrition after baseline. The bottom panel of this table reports many of the same sample characteristics for boys.[346]

The second sample that we employ is reported in the second table, below. In this alternative sample, we employ only the "true panel" sample of cohort girls from baseline to ML2. This approach is the most consistent with the remainder of the report, which tended to focus on the true panel – girls who were successfully contacted at both baseline and ML2. This approach is also most consistent with the first midline round, which reported these same tables with the true panel sample utilized in that evaluation round (baseline to ML1).[347]

Beyond the characteristics reported in the two tables below, we also report on common barriers to girls' education in the final two tables in this section. Again, we employ two sets of samples in two separate tables – the first is the full sample, the second is the panel sample. In these tables, we do not report on boys at all, because data on barriers was not collected from boys at either baseline or ML2.

TABLE 127: COHORT GIRLS' AND BOYS' CHARACTERISTICS, FULL SAMPLE

Characteristic	Intervention ML2 (Baseline)	Comparison ML2 (Baseline)	Variable Name in Household Survey
Sample Breakdown (Cohort Girls)			
Demographic Characteristics			
Single Orphan	11.7% (11.1%)	12.8% (12.8%)	partorphan
Double Orphan	0.3% (0.2%)	1% (0.5%)	fullorphan
Living without both parents	10.1% (9.5%)	13.3% (13.7%)	without_parents
Female-headed household	51.9% (46.5%)	53.5% (45.5%)	hoh_female
Married	4.4% (3.4%)	6.6% (4.1%)	married
Motherhood			
Mother, under 18 years old	0.7% (0.9%)	1% (1.2%)	PCG_23g
Mother, under 16 years old	0.1% (0.4%)	0% (0.4%)	PCG_23g
Household Poverty			
Difficult to afford for girl to go to school	7.4% (20%)	7.6% (19.5%)	difficult_afford
Poor quality roof material	25.5% (31.5%)	33.7% (38.3%)	poor_roof
Gone to sleep hungry many days in past year	12.8% (12%)	10.9% (12.1%)	hungry_many

[346] Boys were not tracked longitudinally. Instead, a cross-section of boys was sampled at both baseline and ML2.
[347] For the true panel sample, we do not report characteristics for boys, because no true panel sample exists for boys – they were not tracked over time.

Language Difficulties			
LOI different from mother tongue	9.8% (6.0%)	8.6% (7.6%)	loi_notsomali
Girl does not speak LOI	3.0% (1.8%)	2.7% (3.3%)	PCG_3enr
Parental Education			
HoH has no education	47.8% (72.9%)	44.8% (70.5%)	hoh_noedu
Caregiver has no education	53.7% (74.3%)	53.4% (78.3%)	cgiver_noedu
Additional Characteristics			
Girl has ever been married	5.4% (4.1%)	7.7% (4.5%)	ever_married
HH ate no rich protein source in last 24 hours	24.1% (0%)	17% (0%)	no_protein
HH reduced food expenditures in last 3 months	37.5% (0%)	34.2% (0%)	PCG_12econ1
HH went entire day without food at least one in last 30 days	23.7% (0%)	15.8% (0%)	PCG_13econ
HH reduced meals eaten at least once in last 30 days	32.1% (0%)	28.6% (0%)	PCG_14econ
HoH has no formal education	72.1% (72.9%)	70.5% (70.5%)	hoh_noformal
Caregiver has no formal education	79.8% (82.7%)	80.8% (83.1%)	cgiver_noformal
Sample Breakdown (Boys)			
Demographic Characteristics			
Female-headed household	46.9% (0%)	53.7% (0%)	hoh_female
Household Poverty			
Poor quality roof material	20.3% (0%)	37.1% (0%)	poor_roof
Gone to sleep hungry many days in past year	18.2% (0%)	15% (0%)	hungry_many
Parental Education			
HoH has no education	47% (0%)	41.4% (0%)	hoh_noedu
Caregiver has no education	51.9% (0%)	50.9% (0%)	cgiver_noedu
Additional Characteristics			
HH ate no rich protein source in last 24 hours	20% (0%)	21.3% (0%)	no_protein
HH reduced food expenditures in last 3 months	41.3% (0%)	37% (0%)	PCG_12econ1
HH went entire day without food at least one in last 30 days	28.1% (0%)	23.1% (0%)	PCG_13econ

HH reduced meals eaten at least once in last 30 days	37.5% (0%)	30.6% (0%)	PCG_14econ
HoH has no formal education	68.9% (0%)	65.7% (0%)	hoh_noformal
Caregiver has no formal education	77.5% (0%)	80.6% (0%)	cgiver_noformal

TABLE 128: COHORT GIRLS' CHARACTERISTICS, PANEL SAMPLE

Characteristic	Intervention ML2 (Baseline)	Comparison ML2 (Baseline)	Variable Name in Household Survey
Sample Breakdown (Cohort Girls)			
Demographic Characteristics			
Single Orphan	11.2% (10.3%)	13.1% (13.6%)	partorphan
Double Orphan	0.4% (0.3%)	1% (0.8%)	fullorphan
Living without both parents	9.5% (7.3%)	14.1% (11.8%)	without_parents
Female-headed household	50.4% (45.2%)	52.8% (42.4%)	hoh_female
Married	5.3% (2.4%)	7.5% (4.3%)	married
Motherhood			
Mother, under 18 years old	0.9% (0.5%)	1.2% (0.8%)	PCG_23g
Mother, under 16 years old	0.2% (0.3%)	0% (0.6%)	PCG_23g
Household Poverty			
Difficult to afford for girl to go to school	7.7% (19.8%)	8% (21.2%)	difficult_afford
Poor quality roof material	24.6% (29.9%)	34.3% (37.4%)	poor_roof
Gone to sleep hungry many days in past year	14% (10.1%)	11.4% (11.6%)	hungry_many
Language Difficulties			
LOI different from mother tongue	11.5% (11.5%)	3.7% (3.7%)	loi_notsomali
Girl does not speak LOI	3.7% (1.7%)	3.4% (3.4%)	PCG_3enr
Parental Education			
HoH has no education	45.9% (69.4%)	46.1% (67.7%)	hoh_noedu
Caregiver has no education	52.8% (71.5%)	54.4% (75.8%)	cgiver_noedu
Additional Characteristics			
Girl has ever been married	6.5% (2.8%)	8.9% (4.3%)	ever_married
HH ate no rich protein source in last 24 hours	23% (0%)	16.9% (0%)	no_protein

HH reduced food expenditures in last 3 months	38% (0%)	34.9% (0%)	PCG_12econ1
HH went entire day without food at least one in last 30 days	25.7% (0%)	15.9% (0%)	PCG_13econ
HH reduced meals eaten at least once in last 30 days	32.5% (0%)	29.8% (0%)	PCG_14econ
HoH has no formal education	71.1% (69.4%)	71.3% (67.7%)	hoh_noformal
Caregiver has no formal education	79.4% (81.2%)	81.3% (81.1%)	cgiver_noformal

TABLE 129: COHORT GIRLS' BARRIERS, FULL SAMPLE

Characteristic	Intervention ML2 (Baseline)	Comparison ML2 (Baseline)	Variable Name in Household Survey
Sample Breakdown (Cohort Girls)			
Safety			
Fairly or very unsafe to travel to schools in the area	2.5% (5.9%)	1.2% (10.2%)	unsafetravel
Girl doesn't feel safe traveling to/from school	1.4% (4.8%)	0.7% (9.1%)	unsafetravelgirl
Parental/Caregiver Support			
High chore burden (whole day)	14.9% (25.1%)	18.2% (28.7%)	chore_burden
Doesn't get support to stay in school and do well	3.7% (4.8%)	5.2% (2.1%)	nosupport
Attendance			
Attends school half the time	2.8% (1%)	1.6% (1.8%)	PCG_6enr
Attends school less than half the time	0.7% (0.6%)	1.4% (1.5%)	PCG_6enr
Doesn't feel safe at school	0.7% (6.8%)	1.1% (6%)	unsafeschool
School Facilities			
Not enough seats for all students	2% (23%)	2.9% (29.6%)	noseats
Difficult to move around school	14% (17.7%)	14.4% (20%)	diffmoveschool
Doesn't use drinking water facilities	11.2% (21.5%)	15.5% (33.5%)	wontuse_water
Doesn't use toilet at school	26.3% (24.1%)	24.7% (32.6%)	wontuse_toilet
Doesn't use areas where children play/socialize	27.6% (40.4%)	33.6% (49.1%)	wontuse_playarea
Teachers			

Disagrees teachers make them feel welcome	3% (4.8%)	4.3% (7.8%)	notwelcome
Agrees teachers treat boys and girls differently in classroom	32.6% (42.4%)	33.2% (41.6%)	treatdiff
Agrees teachers often absent from class	23.9% (35.2%)	23.2% (38.6%)	absentteacher

TABLE 130: COHORT GIRLS' BARRIERS, PANEL SAMPLE

Characteristic	Intervention ML2 (Baseline)	Comparison ML2 (Baseline)	Variable Name in Household Survey
Sample Breakdown (Cohort Girls)			
Safety			
Fairly or very unsafe to travel to schools in the area	2.2% (5.3%)	1.2% (9.5%)	unsafetravel
Girl doesn't feel safe traveling to/from school	1.3% (4.8%)	0.6% (8.4%)	unsafetravelgirl
Parental/Caregiver Support			
High chore burden (whole day)	15.3% (21.2%)	20.6% (25.7%)	chore_burden
Doesn't get support to stay in school and do well	3.9% (5.1%)	6% (2.2%)	nosupport
Attendance			
Attends school half the time	3.3% (1%)	1.4% (1.8%)	PCG_6enr
Attends school less than half the time	0.9% (0.7%)	1.7% (1.8%)	PCG_6enr
Doesn't feel safe at school	0.2% (6.3%)	1.1% (5.6%)	unsafeschool
School Facilities			
Not enough seats for all students	2% (23.3%)	3.4% (28%)	noseats
Difficult to move around school	14.1% (18.2%)	14.1% (20.2%)	diffmoveschool
Doesn't use drinking water facilities	11.9% (21.5%)	13.8% (31.7%)	wontuse_water
Doesn't use toilet at school	26.3% (25.3%)	21.4% (30.6%)	wontuse_toilet
Doesn't use areas where children play/socialize	27.9% (41%)	35.7% (45.8%)	wontuse_playarea
Teachers			
Disagrees teachers make them feel welcome	2.9% (5.7%)	5% (8.9%)	notwelcome
Agrees teachers treat boys and girls differently in classroom	32.7% (40%)	34.7% (39.7%)	treatdiff

Agrees teachers often absent from class	24.5% (34.3%)	21.1% (36.5%)	absentteacher

DISABILITY PREVALENCE AND IMPAIRMENT LEVELS

We examined the prevalence and severity of disabilities reported in the first and second midline to identify potential discrepancies in reporting of disability. The table below summarizes the frequency and severity of different disability categories among girls who were surveyed in both the first and second midline surveys. Responses for the baseline were not included as disability in the baseline was evaluated using a different measure than in the midlines. For the Washington Group disability questions (seeing, hearing, walking, remembering, self-care, and communicating) a girl was included as having an impairment if they indicated that they had at least some difficulty performing the action. For the questions involving having a serious illness or difficulty moving hands or arms, the girls were counted as having the corresponding impairment if they answered "yes". For questions involving mental health (anxiety or depression), respondents indicated the frequency with which they have felt anxious or depressed (never, daily, weekly, monthly, a few times a year). One important difference to note when comparing prevalence and severity of mental health impairments from ML1 to ML2 is that the information comes from different data sources. In ML1, these questions were asked of the girl's caregiver, whereas in ML2 the girls were asked to report the frequency of feeling anxious or depressed themselves.

While overall reporting of disability remains similar (other than the mental health indicators, prevalence of disability is within 4.1% between ML1 and ML2), it was surprising to find that reported prevalence of disability tended to be lower at ML2. This is surprising considering that given that the girls are one year older, we would have expected that if differences were observed, they would tend to be larger. Drawing definitive conclusions about the variation in disability prevalence between evaluations is difficult as they could be caused by a number of factors. For one, the girls are a year older and may have a different perspective on what it means to have difficulties with these tasks or may be less willing to admit to certain kinds of difficulties due to social stigma or self-consciousness. Alternatively, given that many of the survey participants are quite young, it is conceivable that some may be responding to questions in terms of their normal, early development and not in terms of disability. For example, a 6-7 year old first grader may feel that they have difficulty remembering or concentrating, communication, self-care, or even with physical activities like climbing stairs nd after another year of normal development may be more confident in their abilities and may be less likely to see themselves as having difficulties.

TABLE 131: PREVALENCE AND SEVERITY OF DISABILITIES IN ML1 AND ML2

Disability Category	Round	Prevalence	Severity
Trouble seeing (even if wearing glasses)	ML1	4.1% (n=38)	Some difficulty 92.1% (n =35)
			A lot of difficulty 7.9% (n =3)
			Cannot do at all 0% (n=0)

	ML2	4.4% (n=40)	Some difficulty 85.0% (n = 34)
			A lot of difficulty 15.0% (n = 6)
			Cannot do at all 0% (n=0)
Difficulty hearing	ML1	0.5% (n=5)	Some difficulty 100% (n=5)
			A lot of difficulty 0% (n=0)
			Cannot do at all 0% (n=0)
	ML2	0.9% (n=8)	Some difficulty 75.0% (n=6)
			A lot of difficulty 12.5% (n=1)
			Cannot do at all 12.5% (n=1)
Difficulty walking or climbing steps	ML1	2.1% (n=19)	Some difficulty 89.5% (n=17)
			A lot of difficulty 10.5% (n=2)
			Cannot do at all 0% (n=0)
	ML2	1.9% (n=17)	Some difficulty 76.5% (n=13)
			A lot of difficulty 23.5% (n=4)
			Cannot do at all 0% (n=0)
Difficulty remembering or concentrating	ML1	2.8% (n=26)	Some difficulty 100% (n=26)
			A lot of difficulty 0% (n=0)
			Cannot do at all 0% (n=0)
	ML2	1.2% (n=11)	Some difficulty 72.7% (n=8)

			A lot of difficulty 27.3% (n=3)
			Cannot do at all 0% (n=0)
Difficulty with self-care such as washing all over or dressing	ML1	0.7% (n=7)	Some difficulty 100% (n=7)
			A lot of difficulty 0% (n=0)
			Cannot do at all 0% (n=0)
	ML2	0.3% (n=3)	Some difficulty 100% (n=3)
			A lot of difficulty 0% (n=0)
			Cannot do at all 0% (n=0)
Difficulty communicating when using your usual language	ML1	1.1% (n=10)	Some difficulty 90.0% (n=9)
			A lot of difficulty 10.0% (n=1)
			Cannot do at all 0% (n=0)
	ML2	0.54% (n=5)	Some difficulty 80% (n=4)
			A lot of difficulty 0% (n=0)
			Cannot do at all 20% (n=1)
Had a serious illness in the last year	ML1	12.2% (n=112)	N/A
	ML2	10.2% (n=92)	N/A
Difficulty using your hands and arms	ML1	2.5% (n=23)	Difficulty using hands is severe enough to prevent writing or drawing 4.3% (n=1)

	ML2	4.1% (n=38)	Difficulty using hands is severe enough to prevent writing or drawing 5.3% (n=2)
How often do you feel anxious/nervous/worried?	ML1	Feel anxious/nervous/worried at least a few times a year 19.1% (n=175)	Daily 27.4% (n=48)
			Weekly 8.0% (n=14)
			Monthly 16.0% (n=28)
			A few times a year 48.6% (n=85)
	ML2	Feel anxious/nervous/worried at least a few times a year 11.4% (n=105)	Daily 14.3% (n=15)
			Weekly 12.4% (n=13)
			Monthly 15.2% (n=16)
			A few times a year 58.1% (n=61)
How often do you feel sad or depressed	ML1	Feel sad or depressed at least a few times a year 16.7% (n=153)	Daily 24.2% (n=37)
			Weekly 8.5% (n=13)
			Monthly 13.7% (n=21)
			A few times a year 53.6% (n=82)
	ML2	Feel sad or depressed at least a few times a year 11.1% (n=102)	Daily 12.7% (n=13)
			Weekly 14.7% (n=15)
			Monthly 9.8% (n=10)
			A few times a year 62.7% (n=64)
	ML1	12.9% (n=118)	N/A

Any disability (not including mental health)	ML2	8.8%(n=81)	N/A

ANNEX 4: LOGFRAME

The project's logframe, with values filled in by the evaluation team, is attached as a separate file.

ANNEX 5: OUTCOMES SPREADSHEET

A copy of the Outcomes Spreadsheet, updated for the ML2 evaluation, is attached as a separate file.

ANNEX 6: PROJECT DESIGN AND INTERVENTION

The description of SOMGEP-T's design and intervention is provided below, as reported by the project.

TABLE 132: PROJECT DESIGN AND INTERVENTION

Intervention types	What is the intervention?	What output will the intervention contribute to?	What Intermediate Outcome will the intervention will contribute to and how?	How will the intervention contribute to achieving the learning, transition and sustainability outcomes?
Access	Developing and implementing Alternative Learning Program for Out of School Girls	Output 1	IO- 1	By offering an alternative pathway for girls who may have otherwise dropped out, transition rates will improve. Girls will have increased exposure to higher learning, which will boost learning outcomes. ALP's particular focus on developing life skills will ensure this intervention produces sustainable outcomes, or outcomes that are relevant to the individual and community.
	Provision of partial grants to girls from poor families	Output 1	IO- 1	Increased attendance and retention is expected to improve transition rates and learning outcomes, as girls who are in school and are properly equipped are more likely to succeed. Girls from poor families who may not have otherwise had access to education will be better equipped to participate in decision-making and economic activities.
	Equip and enrol girls in 2 boarding schools	Output 1	IO- 1	
School Governance/management	Capacity building of CEC's to	Output 1	IO- 2	A focus on retention and transition is expected to have a direct impact on transition rates and learning outcomes, as girls will have better

	improve retention and transition			access to higher education levels. The focus on the community level will ensure buy-in and contribute to the project's sustainability at the community level.
Teachers capacity building	Train teachers on improved delivery of literacy and English language supported by digital content	Output 2	IO- 3	Improved teaching quality contributes to enhance learning and transition outcomes, as children are equipped with the literacy skills in Somali, English, numeracy and life skills necessary to progress to higher levels of education. Interventions focused on improving teaching quality are expected to boost transition rates and learning outcomes in a sustainable way, by equipping children with the skills they need to succeed not only in school, but outside school as well
	Train teachers on improved delivery of numeracy	Output 2	IO- 3	
	Train teachers to provide structured remedial support to students at primary and secondary levels	Output 2	IO- 3	
	Train and coach teachers to deliver ALP curriculum	Output 2	IO- 3	
	Train and coach teachers on career guidance	Output 2	IO- 3 & IO- 4	Encouraging girls to think about their futures and how to achieve their aspirations will impress on them the importance of knowledge and education. It will also give them a clear pathway to achieving their goals.
Community-based attitudes and behaviour change	Engage community – level stakeholders including religious leaders, women's groups, men and boys	Output 3	IO- 1	Boosts to attendance and retention are expected to contribute to improvements in transition and learning outcomes. Shifts in gender and social norms are expected to have a long-term, sustainable impact on the communities in which SOMGEP-T will operate.
	Provide adult literacy and financial classes for mothers	Output 3	IO- 1	

Economic empowerment	Support the financial empowerment of mothers through savings groups(VSLA) , business selection and business coaching and mentoring	Output 3	IO- 1	
	Work closely with MoE on NFE for mothers and entrepreneurship skills for girls	Output 4	IO- 1	Enhancing the capacity of MoEs to take action on girls' education will have long-term effects on the communities in which SOMGEP-T operates. It will encourage positive shifts in gender and social norms, and will give MoEs actionable ways to contribute to improving learning and transition outcomes.
Life Skills	Develop girls life skills in upper primary through ALP	Output 1	IO- 4	The project's learning outcomes are focused on literacy, numeracy, and financial literacy. This intervention is designed to boost these specific learning outcomes, as well as increasing the likelihood of transition into ALP or secondary education. Additionally, the focus on leadership skills and other skills relevant to the job market contributes to the sustainability of SOMGEP-T.
	Incorporate life skills and financial literacy training into GEFs and BEFs	Output 1	IO- 4	Financial literacy training is one of the specific learning outcomes SOMGEP-T is expecting to influence. Financial literacy and life skills training will increase the likelihood of girls succeeding in higher levels of education, and will also equip them to contribute to the local economy through income-generating activities. These skills are expected to increase the relevance of education for students and families. Life skills – specifically leadership skills – are expected to boost students' voice and self-confidence, enhancing classroom participation among girls.
Girls self Esteem	Expand and strengthen GEF's and create BEFs to develop leadership and mentorship skills	Output 3	IO- 4	Girls who receive leadership and mentorship skills through life skills development will be better equipped to participate in class, breaking traditional norms that restrict girls' voice; to engage in the local economy; and to contribute to their communities in the future. Additionally, the capacity of GEFs and BEFs to track attendance and retention rates will contribute to improvements in learning and transition outcomes, and will encourage community-based

				organizations to think about how their actions have a direct effect on important student outcomes.
MoE Capacity building	Strengthen Gender Departments' capacity to improve girls' education outcomes through trainings, development of action planning and provision of incentives to retain the gender focal points especially in rural areas	Output 4	All IO	Enhancing the capacity of MoEs to take action on girls' education will have long-term effects on the communities in which SOMGEP-T operates. It will encourage positive shifts in gender and social norms, and will give MoEs actionable ways to contribute to improving learning and transition outcomes.
	Provide support to Regional Education Officers (REOs) and District Education Officers (DEOs) to mainstream improved teaching practices and address retention/ transition	Output 4	All IO	Enhancing the capacity of MoEs to take action on girls' education will have long-term effects on the communities in which SOMGEP-T operates. It will encourage positive shifts in gender and social norms, and will give MoEs actionable ways to contribute to improving learning and transition outcomes.
Construction	Construct additional classrooms in remote primary schools; building water facilities in new secondary schools and [provide solar chargers for mobile devices devices/tablets	Output 2	IO- 2	Boosts to attendance and retention are expected to contribute to improvements in transition and learning outcomes. Infrastructure development will benefit not just the current cohort of students with which SOMGEP-T is engaged, but will also benefit future students

	and sanitary pads to schools			

ANNEX 7: KEY FINDINGS ON OUTPUT INDICATORS

A description of key findings on output indicators has been provided by the project, below.

TABLE 133: OUTPUT INDICATORS

Logframe Output Indicator	Means of verification/sources	Collection frequency
Number and Indicator wording	List all sources used.	E.g. monthly, quarterly, annually. NB: For indicators without data collection to date, please indicate when data collection will take place.
Output 1: Improved access to post-primary options		
Output 1.1: Percentage of project locations with an alternative learning program for upper primary/ secondary	ALP monitoring tool	Monthly
Output 1.2: Percentage of ALP groups providing life skills training to marginalised girls	ALP monitoring tool	Monthly
Output 1.3: Percentage of girls receiving partial grants who remain in school	Partial Grants Fidelity of Implementation	Termly
Output 2: Supportive school practices and conditions for marginalised girls		
Output 2.1: Percentage of teachers not using corporal punishment	Midline Evaluation, GEF/BEF Fidelity of Implementation	Yearly
Output 2.2: Percentage of teachers using the digital learning platform	Classroom Observations	Monthly
Output 2.3 Percentage of Girls' Empowerment Forums providing life skills sessions according to the guidance	GEF Fidelity of Implementation	Bi annual
Output 3: Positive shifts on gender and social norms at community and individual girl level		
Output 3:1 Number of women mentors providing support to marginalised girls	GEF Fidelity of Implementation , Monitoring Visits	Monthly
Output 3:2 Number and percentage of mothers completing literacy courses	NFE Completion records	Annual
Output 3:3 Percentage of active village savings groups in project areas	VSLA Fidelity of Implementation [FOI]	Monthly
Output 4: Enhanced MOEs' capacity to deliver quality and relevant formal and informal education		

Output 4:1 Number of Gender Units conducting activities to promote girls' transition and learning	Gender Units Reports	Monthly
Output 4:2 Percentage of REOs/ DEOs engaged in joint monitoring visits to formal schools/ ALP classes to support teachers	Joint Monitoring Reports	Quarterly

TABLE 134: MIDLINE STATUS OF OUTPUT INDICATORS

Logframe Output Indicator	Midline status/midline values Relevance of the indicator for the project ToC	Midline status/midline values
Number and Indicator wording	What is the contribution of this indicator for the project ToC, IOs, and Outcomes? What does the midline value/status mean for your activities? Is the indicator measuring the right things? Should a revision be considered? Provide short narrative.	What is the midline value/status of this indicator? Provide short narrative.
Output 1: Improved access to post-primary options		
Output 1.1: Percentage of project locations with an alternative learning program for upper primary/ secondary	The ALP sites established by the project offer out of school girls post primary opportunity. Originally the project planned to establish 76 ALP programs by end of Year 2 and an additional 34 by end of Y3. Girls enrolled in ALP are expected to have increased learning outcomes as well as developing essential life skills which will enable them to be productive members of the society. Constant measurement of the coverage of ALP is vital. The indicator is still relevant; no modifications are required.	**Midline Wave 1 Status = 96.05%** functional ALP sites. At the time of the assessment 73 ALP sites were functional out of the targeted 78 ALP sites. **Midline Wave 2 Status = 97.3%** functional ALP sites. At the time of the assessment 107 ALP sites were functional out of the targeted 110 ALP sites. Three ALP centres are not functioning due to prolonged insecurity at the borderline. The project will identify new ALP villages to replace those in conflict areas. ALP classes are ongoing across all 107 project areas, project support provided to the ALP centres includes follow up visits and coaching of teachers and payment of incentives to the teachers/facilitators of the ALP classes.
Output 1.2: Percentage of ALP groups providing life skills training to marginalised girls	Girls are learning relevant life skills that will not only boost their learning outcomes and attendance, but will also enable them to contribute to the local economy once they leave school. This intervention boost learning outcomes, as well as increasing the likelihood of girls transiting into formal schools. Life skills remain a key component of the program it is vital to constantly monitor the delivery of life skills training. The indicator is still valid; modification is not required.	**Midline Wave Status 1 =96.05%** **Midline Wave 2 Status = 97.3%** A total of 3410 girls [180 Galmudug, 1122 Puntland and 2108 Somaliland] were provided with training on life skills as we as training on basic financial literacy meant to inculcate the culture of savings among the girls. `

Output 1.3: Percentage of girls receiving partial grants who remain in school	The provision of partial grants to girls enabled girls who are at risk out of school to continue with their education.	Intervention was completed in Year 2
Output 2: Supportive school practices and conditions for marginalised girls		
Output 2.1: Percentage of teachers not using corporal punishment	Addressing corporal punishment will improve conditions for learning this enable girls to attend schools regularly and improve their learning outcomes. The prevalence of corporal punishment should continue to be monitored. The indicator is still valid.	**Midline Status Wave 1 =76.8%** **Midline Status Wave 2 = 45.3%** Student who reported use of corporal punishment by teachers in their class rose from 23.2 percent of intervention girls at baseline to 30.9 percent at the second midline and student at interventions schools who reported their teacher punishing them for getting things wrong in a lesson increased from 40.8 percent to 54.7 percent. These results are discouraging given the general increase in teaching quality observed during the first midline. These results suggest that without persistent reinforcement of teaching best-practices, the long-term sustainability if program interventions may be diminished.
Output 2.2: Percentage of teachers using the digital learning platform	The digital learning platform is expected to improve the quality of teaching, this will increase student performance and motivation is likely to have a positive effect on attendance and learning.	**Midline Wave 1 Status = 0** **Midline Wave 2 Status = 0** Prior to the midline I the project developed digital content and conducted a pilot for the digital platform in selected schools in SL and Puntland. However, the challenges with the volume have made the platform difficult to be used by beneficiaries. The call in and sms function is also not functioning at the moment. The project is working with Telesom and Cell-Ed to resolve the issues.

Output 2.3 Percentage of Girls' Empowerment Forums providing life skills sessions according to the guidance	Life skills – The girls or boys led activities boost their voice and self-confidence, enhancing classroom participation and improved learning outcomes. The indicator is still relevant	**Midline Wave 1 GEF Status = 43.75%** **Midline Wave 2 GEF Status = 68.57%** This indicator assesses the activities led by girls or boys in their school or communities; these activities are designed to build girl or boys' confidence and participation in the classroom. The data for this indicator was collected through the GEF/BEF fidelity of implementation checklist. The fidelity checklist asked a series of questions to understand the various activities implemented by GEF/BEF; the seven activities assessed include facilitation, debating sessions, competitions, fundraising, sanitation campaigns, community sensitization on girls' education, following up on girls who dropped out of school and participation in other community-related activities. GEF/BEF's who implemented at least 4 out of 7 of the activities, were considered to have met the fidelity of implementation minimum standards. At the end of Year 3, at least 68.57% of GEF met the fidelity of implementation minimum standards as compared to 43.75% at midline 1/FOI wave 1. An analysis of the activities led by GEF show marked improvements in the proportion of GEF's leading various activities at school and community level.

ted by GEF	Wave	
	Round 1	Round 2
	21%	49%
	40%	57%
	48%	57%
	73%	83%
activities	33%	29%

		Sensitization on girls education	46%	63%		17%
Output 3: Positive shifts on gender and social norms at community and individual girl level						
Output 3:1 Number of women mentors providing support to marginalised girls	Girls who receive mentorship skills from women mentors, will be better equipped to participate in class, breaking traditional norms that restrict girls' voice; to engage in the local economy; and to contribute to their communities in the future. This is expected to contribute to improvements in learning and transition outcomes. Indicator still valid.	**Midline Wave 1 Status = 66% of GEFs have active mentors** **Midline Wave 2 Status = 65.7% of GEFs have active mentors[121 mentors from a total of 184 GEFs provided support to marginalised girls]** The results of the fidelity of implementation show that 65.7% of the women mentors provided support to the GEFs to undertake various activities which include; facilitating; debating sessions, competitions, fund raising sanitation campaigns, community sensitization on girls' education, following up on girls who dropped out of school and participation in other community-related activities.				
Output 3:2 Number and percentage of mothers completing literacy courses	Mothers in NFE classes acquire essential literacy skills that enable them to support their girls with homework. This will ultimately improve the girl's learning outcome. Skills learnt from the NFE classes will enable them to venture into business, improving their financial capacity to meet the basic education necessities. Girls with adequate basic education necessities are likely to attend school regularly, learn and improve their learning outcomes. Indicator is still valid.	The project in collaboration with MOEs' NFE department printed and signed off more than 6595 certificates for NFE mothers who successfully completed the numeracy and literacy training. The mothers were given certificates in recognition of their efforts to encourage them to continue providing support to girls and promoting their education				
Output 3:3 Percentage of active village savings groups in project areas	Increased financial capacity of vulnerable households, are expected to contribute to improvements in attendance transition and learning outcomes. Indicator is still valid.	**Midline Wave 1 Status = 100%** **Midline Wave 2 Status = 98 % [134 out 137 VSLA are functional] active village savings established]** A total of 134 VSLA groups are functional, 3 are no longer functional. The non-functional groups collapsed because some of the members, mainly pastoralists moved out from the village.				

Output 4: Enhanced MOEs' capacity to deliver quality and relevant formal and informal education		
Output 4:1 Number of Gender Units conducting activities to promote girls' transition and learning	Enhancing the capacity of MoEs to develop plans, administer trainings, and provide incentives will contribute to all four intermediate outcomes by sending a strong, positive message about the importance of girls' education from the government, and by giving the government clear and actionable ways to contribute to positive changes in girls' education outcomes. Indicator is still valid.	**Midline Wave 1 Status = 3** **Midline Wave 2 Status = 3** The project continued to provide Incentives to 13 all-female Gender Focal Person's (GFPs) in the project zones (6 each in SL and PL, 1 in GM). The GFPs worked closely with the project officers to conduct activities earmarked to promote girls' transition. They have been engaged in coaching of Gender Empowerment Forum (GEF), promoting of inclusive education approaches, coordinate the provision of sanitary wear.
Output 4:2 Percentage of REOs/ DEOs engaged in joint monitoring visits to formal schools/ ALP classes to support teachers	Regular joint field monitoring visits will improve the quality of project delivery more importantly the quality of teaching. This is expected to lead to improvements in attendance, transition and learning outcomes. Project will take timely adaptations to ensure identified gaps in programming there by improving the quality interventions ultimately contributing to project outputs, intermediate and outcomes. Indicator is still valid.	**Midline Wave 1 Status = 60.1 % [89/148 primary schools]** **Midline Wave 2 Status = 47.7% [69/148 primary schools] –** The quality assurance visits were adversely by the Covid Pandemic MoE Supervisors with the support of REOs and DEOs participated in joint monitoring visits/fidelity of implementation assessments with project staff. The Supervisor across the three project zones visited 69 of the 148 project schools/communities. The joint monitoring visits included an in-depth assessment of the program strengths and identification of areas for improvement. Regional and District consultative meetings have continued to play an important role in project's progress, these meetings are the main reason for project's continued good working relationship MoE at regional level. Through the meetings the project gets an opportunity to share with the MOE the project's progress/challenges and jointly plan for steps in supporting planned interventions such as teacher

		training, coaching, mentorship, community liaison and engagement, information sharing and overall leadership in matters pertaining improved curriculum implementation/learning outcome. At each school/community the team had in-depth discussions with various stakeholders including VSLA, CECs members, boys and girls, numeracy, ABE, ALP and head teachers. The information gathered through joint monitoring/FOI assessment inform the realignment of interventions and the bulk of the information was also used to establish the midline status of the output indicators.

TABLE 135: OUTPUT INDICATOR ISSUES

Logframe Output Indicator	Issues with the means of verification/sources and the collection frequency, or the indicator in general?	Changes/additions
Number and Indicator wording	E.g. inappropriate wording, irrelevant sources, or wrong assumptions etc. Was data collection too frequent or too far between? Or no issues?	E.g. change wording, add or remove sources, increase/decrease frequency of data collection; or leave as is.
Output 1: Improved access to post-primary options		
Output 1.1: Percentage of project locations with an alternative learning program for upper primary/ secondary	None	None
Output 1.2: Percentage of ALP groups providing life skills training to marginalised girls	None	None
Output 1.3: Percentage of girls receiving partial grants who remain in school	None	None
Output 2: Supportive school practices and conditions for marginalised girls		
Output 2.1: Percentage of teachers not using corporal punishment	None	None
Output 2.2: Percentage of teachers using the digital learning platform	None	None
Output 2.3 Percentage of Girls' Empowerment Forums providing life skills sessions according to the guidance	None	None
Output 3: Positive shifts on gender and social norms at community and individual girl level		
Output 3:1 Number of women mentors providing support to marginalised girls	None	None
Output 3:2 Number and percentage of mothers completing literacy courses	None	None
Output 3:3 Percentage of active village savings groups in project areas	None	None
Output 4: Enhanced MOEs' capacity to deliver quality and relevant formal and informal education		

Output 4:1 Number of Gender Units conducting activities to promote girls' transition and learning	None	None
Output 4:2 Percentage of REOs/ DEOs engaged in joint monitoring visits to formal schools/ ALP classes to support teachers	None	None

ANNEX 8: BENEFICIARIES TABLES

PROJECT ACCOUNTING OF BENEFICIARIES

TABLE 136: DIRECT BENEFICIARIES

Beneficiary type	Total project number	Total number of girls targeted for learning outcomes that the project has reached by second midline	Comments
Direct learning beneficiaries (girls) – Estimated as 19,635, or 78.9% out of a total of 24,605 intervention girls	24,605	18,215	Total reach is calculated as: -12,241 girls enrolled in primary school (spot-check mid-2019[348]) -1,555 girls enrolled in ABE -2,157 girls enrolled in ALP -1,912 girls enrolled in secondary school -3,454 out-of-school girls -1,768 graduates

The overall total is lower than anticipated at the baseline (26,290 learning beneficiaries). The main reason for the reduction in the expected reach is the lower enrolment in Grade 1, potentially as a result of population displacement for urban areas.

The reach calculations above are conservative and do not include a correction for girls exposed in 2017-18 who have migrated out of the area. For instance, in 2018, 19% of the treatment households had adolescent girls leaving the home. Therefore, it is extremely likely that the actual reach was considerably larger; nonetheless, in the absence of a complete roster of individual in-school girls, it is not possible to obtain the actual total of migrants.

TABLE 137: OTHER BENEFICIARIES

Beneficiary type	Number	Comments
Learning beneficiaries (boys) – as above, but specifically counting boys who will get the same exposure and therefore be expected to also achieve learning gains, if applicable.	11,034	Estimated number of boys enrolled in targeted schools plus total number enrolled in ABE
Broader student beneficiaries (boys) – boys who will benefit from the interventions in a less direct way, and therefore may benefit from	13,793	Boys are primarily targeted in formal school and ABE classes; therefore, the total listed here corresponds to the

[348] As a result of seasonal movements, the enrolment at the beginning of the school year (Nov-2019, when this study was conducted) is lower than mid-year (May 2019, when the previous spot-check was conducted).

aspects such as attitudinal change, etc. but not necessarily achieve improvements in learning outcomes.		total of boys who are not expected to benefit from improved learning outcomes
Broader student beneficiaries (girls) – girls who will benefit from the interventions in a less direct way, and therefore may benefit from aspects such as attitudinal change, etc. but not necessarily achieve improvements in learning outcomes.	9,237	Defined as in-school girls not benefitting from improved learning outcomes (estimated as 30% of total); out-of-school girls not benefitting from improved learning outcomes (estimated as 68% of total[349]); and new intake in Grade 1 (conservatively estimated as equivalent to 2019 intake)
Teacher beneficiaries – number of teachers 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.	618 teachers trained on literacy, numeracy, English and structured remedial classes (36 female) 177 teachers trained to deliver ALP (14 female) 81 teachers trained to deliver ABE (13 female) 103 supervisors and REOs trained on literacy, numeracy, English and structured remedial classes (12 female) 10 supervisors and REOs trained on ABE (2 female)	
Broader community beneficiaries (adults) – adults who benefit from broader interventions, such as community messaging /dialogues, community advocacy, economic empowerment interventions, etc.	6595 mothers receiving NFE training 4,110 community members participating in VSLA	

TABLE 138: TARGET GROUPS – BY SCHOOL

School Age	Project definition of target group (Tick where appropriate)	Number targeted through project interventions[350]	Sample size of target group at Baseline
Lower primary	Yes - Grade 1-4	8854	272
Upper primary	Yes - Grade 5-8	7099	233 (+93 in benchmark)
Lower secondary	Yes - Form 1-2	1,297	12 (benchmark only)
Upper secondary	Yes – Form 3-4	615	

[349] The current pandemic may affect the expected increase in the impact of activities with OOSG, resulting in large increase in migration/ displacement. The estimate is conservative.
[350] Estimated based on the proportions observed in midline data. The overall total for lower primary is inclusive of ABE and the overall total for upper primary includes ALP.

Total:		17,865	[This number should be the same across Tables 32-35]

TABLE 139: TARGET GROUPS – BY AGE

Age Groups	Project definition of target group (Tick where appropriate)	Number targeted through project interventions [351]	Sample size of target group at Baseline
Aged 6-8 (% aged 6-8)			This group will benefit from teacher training, improved school management and conditions at the household. However, the baseline sample included only girls age 10-19.
Aged 9-11 (% aged 9-11)	√		241
Aged 12-13 (% aged 12-13)	√	7988	268
Aged 14-15 (% aged 14-15)	√	8844	192
Aged 16-17 (%aged 16-17)	√	5028	109
Aged 18-19 (%aged 18-19)	√	1997	62
Aged 20+ (% aged 20 and over)	√	713	Benchmark only
Total:		**24,605**	[This number should be the same across Tables 32-35]

TABLE 140: TARGET GROUPS – BY SUB GROUP

Social Groups	Project definition of target group (Tick where appropriate)	Number targeted through project interventions	Sample size of target group at Baseline
Disabled girls (please disaggregate by domain of difficulty)	√	2362	60
Vision impairment	√	172	5
Hearing impairment	√	98	6
Mobility impairment	√	98	6
Cognitive impairment	√	25	7
Self-care impairment	√	0	6

[351] Reflects current proportions based on midline 2 data.

Communication impairment	√	74	9
Mental health impairment	√	1722	45
Disability of arms/ hands	√	467	N/A
Anxiety	√	1599	39
Depression	√	1132	32
Orphaned girls	√	2905	96
Pastoralist girls	√	2724	105
Child labourers	√	1057	13
Poor girls	√	24,605	872
Other (please describe)			
Total:		**24,605**	[This number should be the same across Tables 32-35]

TABLE 141: TARGET GROUPS – BY SCHOOL STATUS

Educational sub-groups	Project definition of target group (Tick where appropriate)	Number targeted through project interventions	Sample size of target group at Baseline
Out-of-school girls: have never attended school	√	1043	92
Out-of-school girls: have attended school, but dropped out	√	2410	275
Girls in-school	√	19,633	505
Total:		**24,605**	

TABLE 142: BENEFICIARIES MATRIX

Outcomes	Direct beneficiaries			Indirect beneficiaries				
	In-school girls (6-10 grade)	OSG (6-9 years)	OSG (18-25)	In-school boys	HT/Teachers	Parents	SMC/PTA	Local government
Learning	✔			✔	✔	✔		
Transition	✔	✔	✔	✔	✔	✔		
Sustainability	✔	✔	✔		✔	✔	✔	
IO 1: Attendance					✔	✔		
IO2: School management and governance	✔				✔	✔	✔	✔

IO3: Teaching quality	✔				✔	✔	✔	✔
IO 4: Life skills	✔	✔	✔					

Definition of Marginalisation

As described in the introductory section of this report, SOMGEP-T is being implemented in some of the most remote areas of Somalia, which have been severely affected by conflict and repeated occurrences of prolonged drought since the project started. Educational marginalisation is experienced by virtually all beneficiaries, with those in school facing dire conditions to continue their education and learn, and limited opportunities for out-of-school children. Only 13% of the schools have reliable access to electricity and 52% have access to water within 1km of the school. During the past year alone, 20% of the schools have been affected by conflict. At the household level, the intersection of multiple factors contributes to absenteeism, poor learning outcomes and dropout. The majority of the caregivers (80%) has never attended formal education, and therefore most of the girls supported by SOMGEP-T are first generation learners. A large proportion of households experiences food insecurity (38%), lacks access to medicine (42%) and does not have access to clean water (62%). These barriers affect student attendance, learning and retention. Overall, 10% of the students faces some form of disability; it is important to note, however, that the proportion of students facing mental health issues increases dramatically during periods of conflict or drought, as reflected in the fluctuation of the prevalence of anxiety and depression in the sample between 2017-2019. The current crisis resulting from the global COVID-19 outbreak further exposes the vulnerability of this population; the National COVID-19 Preparedness Response Plan anticipates that 6,600 people may be affected in the country, but only 15 intensive care unit beds[352] are available in the whole of the country, all of which are located in urban areas. It is anticipated that this crisis will result in a prolonged school break and have a severe negative effect on the economic status of households, as well as adding to girls' vulnerability due to the combination of disease burden at the household, potential loss of caregivers, risk of gender-based violence (including early marriage) and drop-out.

External Evaluator Discussion

The accuracy of CARE beneficiary count was analysed using available data and project documentation provided by CARE. As seen in the table above, the project is expected to reach an estimated 24,605 girls. This number is calculated as the sum of girls enrolled in primary school (taken from a spot check in mid-2019), the number of girls enrolled in ABE, the number of girls enrolled in ALP, the number of girls enrolled in secondary school, and the number of graduates. To verify this estimate, we used the headcount data and surveys with ABE girls to see if the claimed number of beneficiaries is still supported.

During school visits, enumerators looked at the enrolment records for each classroom and evaluated them for completeness and recorded the number of students enrolled for each class. Classrooms that did not keep enrolment records or had records that were incomplete were dropped for this analysis. Total school enrolment was then calculated for each school that had a complete set of records for all classes. The total school enrolment was then calculated for the 31 treatment schools for which we had a complete set of class-level enrolment records. We then calculated an estimate of the total number of boys and girls enrolled in

[352] As per WHO's Situation Report 4 – April 4, 2020.

primary school based by extrapolating the total number of students enrolled in schools with good records to the to the total number of treatment schools.

The estimated number of girls enrolled in primary school, based on our analysis, is 11,638. While slightly lower, this estimate is in line with the 12,241 primary school girls claimed as beneficiaries. Similarly, we estimate that 13,596 boys are currently enrolled in primary schools targeted by the intervention. This estimate is significantly higher than the estimate used above of 11,034, even though it does not include boys enrolled in ABE programs as above. While our analysis is imperfect (it relies on extrapolation from a limited number of schools, assumes that headcounts were completed for all classes within a school, and that the records produced were highly accurate), it does suggest that CARE's estimate of program beneficiaries enrolled in school is likely a good one.

We did not capture direct ABE enrolment data during data collection, so we used an indirect approach to verify CARE's estimate. During data collection, enumerators were tasked with obtaining 15 surveys from ABE girls at each centre, irrespective of their actual enrolment numbers. However, there were instances where there were not enough girls enrolled to meet that requirement. We compared instances where enumerators were unable to obtain 15 surveys from an ABE centre to the estimated enrolment of girls in each centre from the sample frame. Of the 35 ABE centres included in the sample, we were unable to obtain 15 surveys for eight of them. In all but one case, the number of surveys we were able to obtain, was significantly less than the enrolment numbers indicated in the sample frame. While this analysis is very imprecise, it does suggest that the number of ABE girls benefitting from the program is likely lower than what is claimed. This is likely due to girls dropping out of the ABE program since the enrolment numbers were obtained for the sample frame.

Generally, we find that CARE' s approach to calculating overall beneficiaries is sound. At the same time, we view the approach as somewhat conservative, underestimating the number of girls that are benefiting from the program. To some degree, this depends on how one views beneficiaries: conceptually, the program is counting beneficiaries based on whether they have shown an observable increase in learning scores; but it could also be argued that girls exposed to the program – even if they do not improve their learning scores – should be counted as beneficiaries, to the extent they benefit from improved transition rates, higher-quality teachers (e.g., who are less likely to use corporal punishment), and so forth. Taken to its logical conclusion, this view would suggest counting all 24,605 girls as beneficiaries. Our point in this discussion is not that the program needs to fundamentally re-evaluate its approach to counting beneficiaries. Rather, our point is that – regardless of the other assumptions the project makes – the ultimate beneficiary count is almost certainly conservative by (lower than) most standards.

ANNEX 9: MEL FRAMEWORK

The project's MEL Framework, updated is attached as a separate file.

ANNEX 10: EVALUATION INCEPTION REPORT

The Inception Report, submitted prior to the start of data collection for the ML2 round, is attached as a separate file.

ANNEX 11: DATA COLLECTION TOOLS

The quantitative and qualitative data collection tools are attached separately, given their length. Four quantitative tools are provided:

- Household survey, which encompasses the survey module completed by girls, the learning assessments, and the module and assessments completed by boys
- Classroom observation tool
- Headcount tool
- Head teacher (school) survey

The qualitative tools include:

- FGD Mothers
- FGD Teachers
- FGD CEC Members
- KII MOE Officials
- Girls with Disabilites (GWDs)
- Risk Mapping
- Vignette Exercises

In addition, two English-language transcripts of qualitative interviews are provided. Specifically, we provide an FGD conducted among teachers and a risk mapping exercise with girls.

ANNEX 12: DATASETS, CODEBOOKS AND PROGRAMS

The final, cleaned datasets for the ML2 evaluation are attached in a separate annex. To make further analysis or re-analysis possible, we provide the full datasets, which combine baseline, ML1, and ML2 data into a single, "long" format file. To ease replication specifically of the learning and transition results, we provide replication datasets that are specific to those analyses, as the steps to obtain analysis-ready data – especially in the context of the transition outcome – is extensive. These different datasets are clearly indicated in the annex. The datasets are fully labelled.

Stata .do files that replicate our learning and transition analysis are also annexed with the datasets.

Finally, the codebook below is provided to further facilitate replication. It highlights the key variables in each dataset necessary to replicate our core analysis – learning, transition, but also variables of particular importance to some of the intermediate outcome analyses.

TABLE 143: CODEBOOK OF ML2 DATA

Variable Name	Variable Description
Household Survey and Learning Assessments	
zone	Geographic zone (Somaliland, Puntland, Galmudug)
region	Region of interview
district	District of interview
school	School code (labelled with names)
cluster_id	Identifies unique clusters of respondents, accounting for multiple institution types (primary school, ALP, and ABE) in the same "school." School identifies the school, but groups together ALP/ABE centres with the schools associated with them. Cluster_id distinguishes between ALP, ABE and primary schools, even if they were sampled together or considered connected for fieldwork purposes. This is the appropriate variable to use when clustering standard errors.
cohort	Status of girl when first recruited into sample (in-school, OOS, ALP, ABE)
masterid	Master Unique ID. This variable uniquely identifies a girl in each round and across rounds. It identifies a unique girl, *not* her replacement. Therefore, it is suitable for tracking the panel of successfully re-contacted girls across time.
orig_bl_uniqueid	Original unique (master) ID of child recruited at baseline. Matches masterid if girl is not a replacement; distinct from masterid if girl is a replacement.

orig_ml1_uniqueid	Original unique (master) ID of child recruited at ML1.
round	Round of data collection (0 = baseline, 1 = ML1, 2 = ML2)
treatment	Intervention or comparison status
panel_bl	Identifies girls comprising the true panel from baseline to ML2
panel_ml	Identifies girls comprising the true panel from ML1 to ML2
panel_full	Identifies girls comprising the true panel across all 3 rounds (i.e. same girl re-contacted successfully in all 3 rounds)
cs_bl	Identifies girls in the "cross-sectional" sample from baseline to ML2, i.e. girls who appear in both baseline and ML2, but may have been replaced
panel_trans	Identifies girls in the transition panel sample, tracked from baseline to ML2. This is the appropriate sample to use for transition analysis.
num	Numeracy score, 0-100 scale
eng	English literacy score, 0-100 scale. Not available for girls at ML1 (no English assessments conducted)
som8	Somali literacy score, 0-100 scale, using the comparable 8 subtasks from baseline through ML1 and ML2
somfull	Somali literacy score, 0-100 scale, using the full 9 subtasks employed at ML2. Only available for girls at ML2.
memory	Score on working memory test. Only available for girls at ML2
finlit	Score on financial literacy assessment. Not available for girls at ML1.
enrol	Enrolment status at time of data collection
age	Age at time of data collection
grade	Grade at time of data collection (note that grade is given a code for OOS girls., ALP girls, ABE girls, etc.)
Headcount Data	
school	School code/name

panel_bl	Identifies set of schools that are comparable from baseline to ML2
panel_ml1	Identifies set of schools that are comparable from ML1 to ML2
panel_full	Identifies set of schools that are comparable across all 3 rounds
b7	Girls' enrolment count, from school records
b11	Girls' headcount, conducted by field team. Divide by b7 to obtain attendance rate for headcount.
b12	Boys' enrolment count, from school records
b16	Boys' headcount, conducted by field team. Divide by b12 to obtain attendance rate for headcount.
Head Teacher Survey Data	
school	School code/name
panel_bl	Identifies set of schools that are comparable from baseline to ML2
panel_ml1	Identifies set of schools that are comparable from ML1 to ML2
panel_full	Identifies set of schools that are comparable across all 3 rounds
Classroom Observations Data	
school	School code/name
panel_bl	Identifies set of schools that are comparable from baseline to ML2
panel_ml1	Identifies set of schools that are comparable from ML1 to ML2
panel_full	Identifies set of schools that are comparable across all 3 rounds

ANNEX 13: LEARNING TEST DESIGN AND CALIBRATION

PILOT TEST AND CALIBRATION

Revisions to the learning test were conducted in October 2019 in Somaliland, including remote discussions with the Fund Manager. The revisions included two levels of edits: (1) minor changes to the words, short texts, questions and numbers used in the assessments to prevent a potential bias due to pre-exposure; and (2) the development of an additional Somali literacy task, which required students to complete a story. The

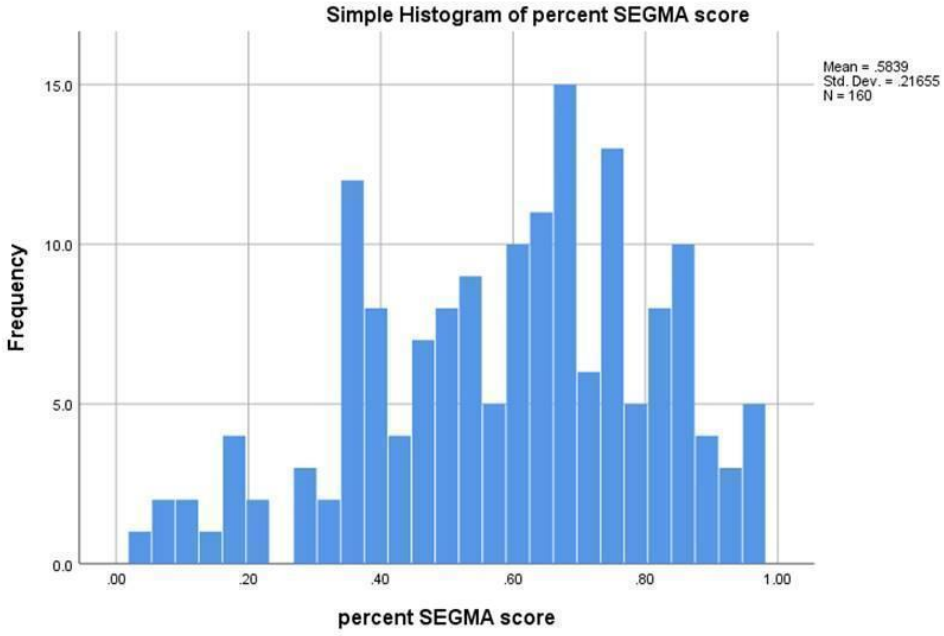
additional literacy task was included to prevent a potential ceiling effect, given the results observed in the Midline 1 study. The revisions were conducted by the MEL team at CARE and included an extensive in-house consultation on the vocabulary and structure of sentences used to ensure (i) equivalent levels of difficulty between the tasks used in different evaluation rounds and (ii) similar levels of understanding for children in different regions of the country, noting variations in accent, spelling and meaning. Therefore, the project seeks to always use a 'neutral' vocabulary in literacy assessments.

Following the development of the revised versions, a team of CARE MEL staff was trained on the administration of the learning assessments. The learning assessments were piloted in five non-sampled schools, with a total of 160 students across grades 6, 7 and 8 (grades the majority of the students are expected to reach by the end of the project, considering the baseline sample distribution). All tests were conducted with girls. Due to limited resources[353], the pilot took place in Somaliland only, between October 28-November 4, 2019.

The results of the pilot were shared with the Fund Manager on November 6, 2019. The following results were observed:

1. Numeracy assessment (SEGMA): No ceiling effect; in total, 29% of the students have scored 75% or above. Among grade 8 students, the proportion of students scoring 75% or above reached 47%.

FIGURE 53: DISTRIBUTION OF SEGMA SCORES, PILOT TEST



2. Somali literacy: No ceiling effect observed. Only 28% of the girls have achieved scores of 75% or above on reading comprehension, while 18% have achieved scores of 75% or above for writing (with the inclusion of the new task). There is no major difference between grades in terms of score distribution.

[353] The project budget could not accommodate contracting an external group of enumerators to conduct the pilot, and therefore the tests were carried out by CARE MEL staff with previous experience on the use of tools. Unfortunately, due to other tasks, a limited number of MEL staff could be engaged in this activity, thus reducing our ability to cover a large number of schools within a limited period of time.

FIGURE 54: DISTRIBUTION OF SEGRA READING COMPREHENSION SCORES, PILOT TEST

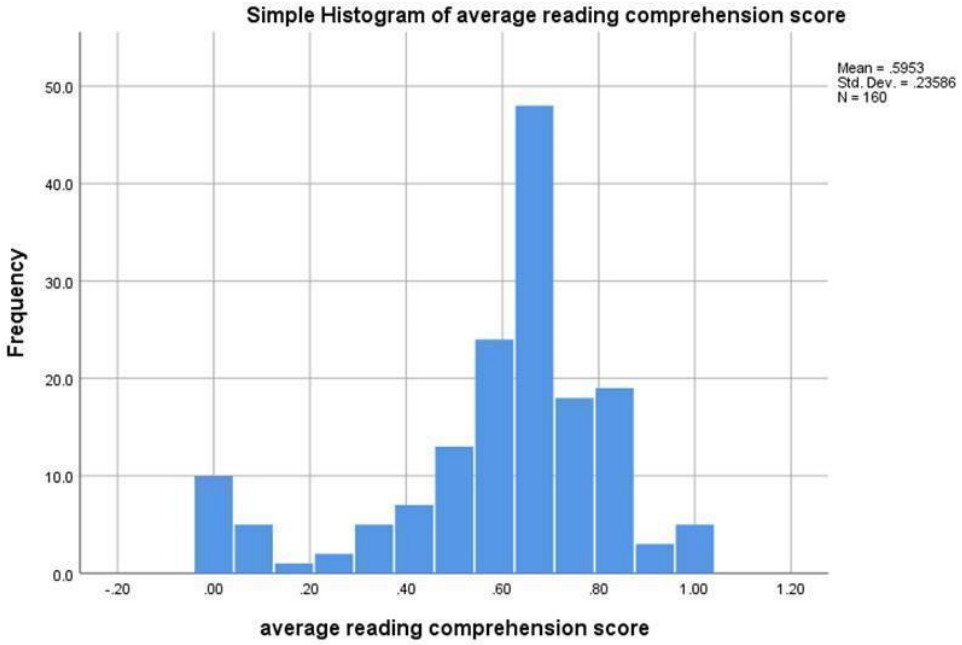
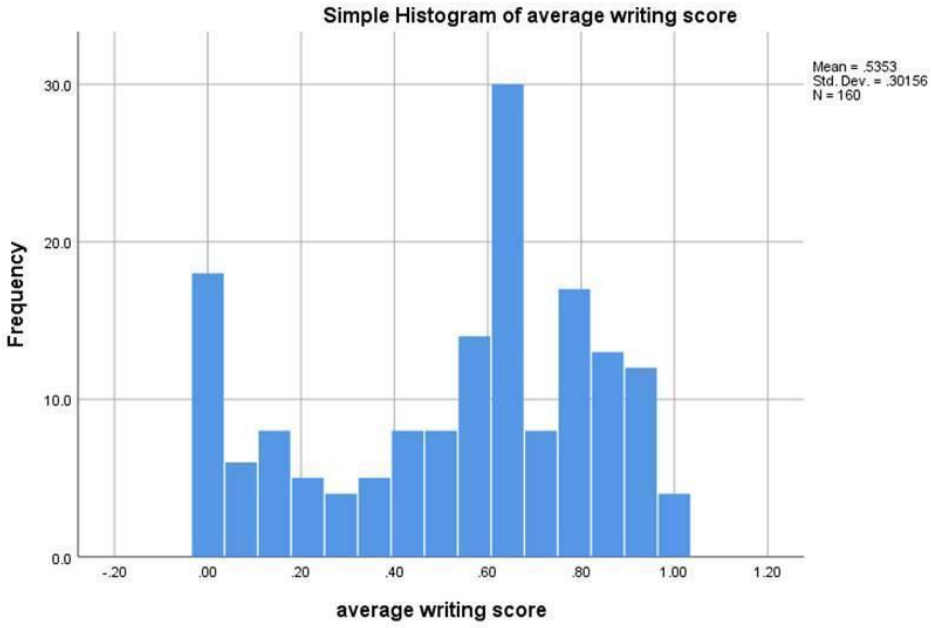


FIGURE 55: DISTRIBUTION OF SCORES, SOMALI WRITING SECTION, PILOT TEST



As described in the methodology section (p.64, *Equivalence of Learning Assessments*) an equating process was used to determine the comparability of the tests across evaluation rounds. The different learning assessments were applied with the same students in a subset of the sample, allowing for an objective comparison of results across rounds and if necessary, for determining an equating coefficient (which was not necessary as none of the tests has shown significant differences in difficulty level across rounds). As discussed with the Fund Manager (communication from Lotte Renault to Alix Clark, November 7, 2019), the project opted for doing

this during the actual data collection to allow for tests with a sample proportionally distributed across project sites, instead of conducting the equating process during the pilot, with a more limited number of schools, which might result in a non-representative / imprecise equating coefficient.

EQUIVALENCE OF LEARNING ASSESSMENTS ACROSS ROUNDS

Learning constitutes one of the core outcomes for SOMGEP-T, and learning scores are measured through adapted versions of the EGMA and EGRA tools, as noted in our previous discussion of the evaluation design. In that discussion, we noted that CARE developed new versions of the learning assessments as part of this midline evaluation, in line with revisions made during the previous midline, in late 2018. Revisions to the data collection tools beg the question of whether the learning assessments are of comparable difficulty from round to round, as changes that made the assessments more or less difficult would influence the learning scores achieved by students.

At the outset, it is important to note that changes in the difficulty of the learning assessment do not, in any way, threaten the inferences we draw regarding program impact on learning. Even if the assessments became markedly more or less difficult from previous years, these changes would not influence or render invalid our conclusions regarding program impact. Thanks to the difference-in-differences design of the evaluation, any change in difficulty over time is applied to both intervention and comparison girls equally, so that the difference does not influence our estimates of program impact.

The result referenced above is subtle and technical, and bears repeating. Recall that the design of the evaluation includes a comparison and intervention group, both of whom are assessed for learning outcomes at two or more points in time. The key advantage of a difference-in-differences design, over that of a simple pre-post design, is that difference-in-differences can account for exogenous shocks or systematic confounding variables that impact both intervention and comparison groups equally.

Consider a hypothetical scenario in which a drought strikes between the baseline and endline of a program. In a pre-post design, a drop in learning scores from baseline to endline could be the result of the program itself, or the result of the drought. In the absence of a more rigorous or nuanced design, adjudicating between these two possible explanations is impossible, because all children at the endline were exposed to both the program and the drought, while girls at the baseline were invariably exposed to neither.

In contrast, a difference-in-differences design makes comparisons in the *trends* in intervention and comparison children from baseline to endline. If the same drought impacted all comparison and intervention children approximately equally, we would still be able to draw conclusions regarding the impact of the program – aside from the impact of the drought. Any change in learning outcomes as a result of the drought would be controlled for explicitly by the design – changes from the drought would be felt in both intervention and comparison areas, while changes due to the program are readily identified because they occur exclusively in treatment areas.[354]

It is reasonable to be skeptical of the idea that a drought or other exogenous shock would apply equally to intervention and comparison communities. If an exogenous shock impacts intervention or comparison areas disproportionately, it violates the parallel trends assumption and estimates based on the difference-in-differences design are no longer unbiased.

[354] This argument does rely on additional assumptions that underlie the difference-in-differences design, such as the parallel trends assumption, and the no interference or no spill-over effects assumption.

Our interest in this section is on the impact of changes in assessment difficulty. Changes in assessment difficulty are exactly equivalent to an exogenous shock that occurs during the endline (or midline, in this case). This is a shock to learning scores that arises not due to the program, but due to some outside occurrence. Importantly, changes in assessment difficulty *precisely meet* the definition of an exogenous shock that applies equally to all girls at the midline – the easier or more difficult assessments are taken by all girls at midline, so any change in the gap between intervention and comparison group learning scores cannot be a function of the change in assessment difficulty, but because of the program itself.[355] Because the change in assessment difficulty applies to all girls equally, the change does not threaten causal inferences we draw in this report.

Although overall conclusions regarding program impact are unaffected by the equivalence of learning assessments from round to round, there are aspects of the analysis where such equivalence is necessary. For instance, our analysis of changes in ALP girls' learning scores does not have the benefit of a comparison group or difference-in-differences design, so small changes in assessment difficulty could produce bias in our analysis of ALP girls.

In order to investigate the equivalence of learning assessments over multiple rounds, the second midline round of data collection included data collection using the previous midline and baseline assessments. Specifically, a sample of 61 cohort girls, who completed the standard learning assessments administered in this round, were recruited into a comparison sample. Following their completion of the standard learning assessments used in this round, they completed the Somali literacy and numeracy assessments from the first midline round (ML1, from late 2018) and the English literacy assessment from the baseline round (in late 2017).[356] By completing both assessments with the same group of girls, we can compare their scores on the two assessments to judge their equivalence; collecting data from the same girls has the advantage of eliminating sampling variation from the analysis, strengthening conclusions regarding equivalence.

The table below reports the mean scores on each assessment among this sample of 61 respondents; the right column reports the difference in means. Across all three tests, scores are very similar from baseline/ML1 to ML2, with the largest gap occurring in Somali literacy. A formal test for differences, using a paired t-test, confirms the similarity in difficulty – none of the differences observed approach statistical significance at any conventional level. In the case of Somali literacy, the associated p-value is 0.5, well above standard and generous thresholds of 0.05 and 0.1, respectively.[357]

TABLE 144: ML2 AND PRIOR ROUND LEARNING ASSESSMENTS, COMPARISON SAMPLE SCORES

	Baseline/ML1 Score	Midline 2 (ML2) Score	Difference in Scores
English Literacy	31.7%	32.6%	0.9
Numeracy	65.1%	64.6%	-0.5
Somali Literacy	68.6%	67.5%	-1.1

[355] Or some other exogenous shock, aside from changes in learning assessment difficulty that differentially impacts one of the two groups.

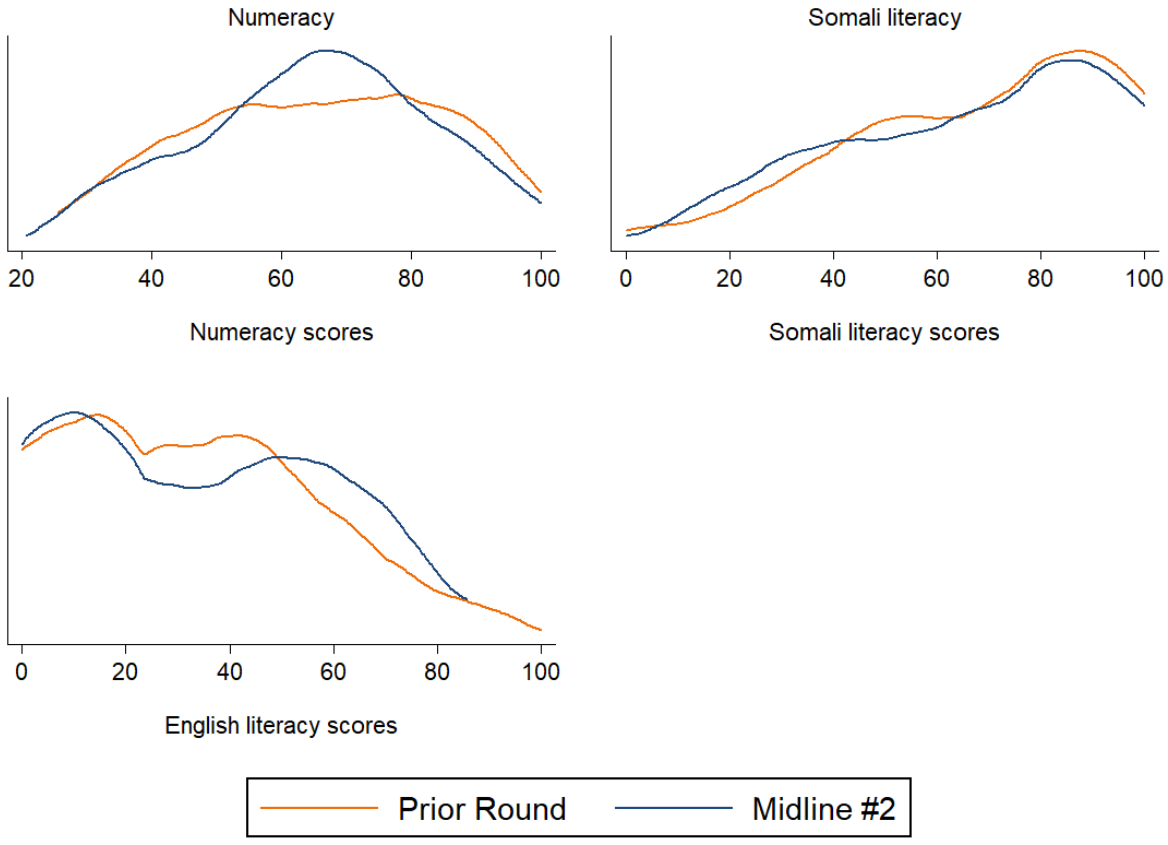
[356] The baseline English literacy assessment is used as the comparison, rather than a midline version, because no English literacy assessment was administered in the first midline round. Any analysis of English literacy will, by necessity, be a comparison of baseline and ML2 scores.

[357] Note that we also employed non-parametric tests, such as the Wilcoxon signed rank test, to ensure that our focus on a simple difference-in-means did not obscure a more nuanced difference in scores between assessments. These tests rarely produce results that differ from a t-test, but they are useful as a robustness check. In each case, the p-value for the signed rank test was similar to those from the t-tests, and well above standard cut-offs for statistical significance.

The finding that the assessments are similar is not surprising, given our understanding of how the learning assessments were adjusted from the previous round. In all cases, the subtasks maintained their fundamental objective and the skills that they tested. In most cases, the subtasks were altered in small ways, such as small changes in the numbers employed in addition or subtraction problems, adjustments sufficiently small that there is no theoretical reason to expect them to be more or less difficult as a result.

To highlight the similarity of difficulties across rounds, we plotted numeracy (top-left panel) and Somali literacy (top-right panel) scores for the sample of 61 girls from both the ML1 and ML2 assessments. As the figure below shows, girls completing the current (ML2) numeracy assessment were clustered more heavily in the middle of the score distribution, but there is no evidence of a systematic shift in scores. Likewise, Somali literacy scores show no evidence of a shift. The lower left panel reports English literacy scores using the baseline and ML2 versions of the assessment. While the two distributions do not match precisely, the shift is substantively small, and there is no reason to believe that the tests of are unequal difficulty overall.

FIGURE 56: PRIOR ROUND AND ML2 LEARNING ASSESSMENT SCORES, COMPARISON SAMPLE



As noted above, differences in assessment difficulty will not affect our conclusions regarding overall program impact on cohort girls, but may influence other findings we report. We suggest caution when it comes to analysing changes in subtask-specific outcomes, as individual subtasks are not necessarily equivalent between evaluation rounds. It is not clear whether this is due to bias from a relatively small sample, but performance

in the comparison sample of 61 girls yielded statistically significant differences on a number of subtasks. For instance, t-tests comparing performance on Somali literacy subtasks 1, 3, and 8 suggested that subtasks 1 and 8 became more difficult in ML2, while subtask 3 became less difficult (all with p-values under 0.10). In practice, these subtask-specific differences balanced each other across the entire assessment; however, it is worth noting that individual subtasks may show differences in difficulty across rounds, even if the assessments overall do not.

ANNEX 14: SAMPLING FRAMEWORK

The sampling framework for selection of ABE centres during the ML2 evaluation round is provided in a separate annex. The sampling framework for primary schools has not been updated or utilized since the baseline evaluation, because no new primary schools have been selected into the sample since that time. We provide the sampling framework – identical to that provided at baseline – in a separate annex, for the sake of completeness.

ANNEX 15: MIGRATION PATTERNS AMONG ADOLESCENTS

The population that SOMGEP-T works with tends to be highly mobile. As noted in the context section of this report (Section 1.1), the project works in rural communities defined by pastoralism and large numbers of nomadic or semi-nomadic households. Even among households that do not engage in nomadic pastoralism, migration tends to be higher than in other areas of Somalia. A higher rate of migration likely stems from the economic and environmental marginalization faced by these communities, and the existence of internecine conflict and political instability.[358] Adolescent migration may be especially high because secondary schools are often not available in the communities where respondents live, and because adolescents moving to live with family in urban centres is a common practice among Somali households.

This section briefly reviews the patterns of adolescent out-migration from SOMGEP-T communities. Part of the goal of this analysis is to provide a deeper understanding of the communities themselves, which are defined, in part, by high levels of mobility. Another goal is to understand the barriers faced by adolescents in terms of educational attainment, as the data documents the reasons for out-migration. The final goal is to understand how migration rates and patterns have shifted over time.

In the table below, we report the number of adolescent girls who have left their households and communities in the past 12 months, broken down by the round of data collection.[359] We provide a count of adolescent girls, as some households report multiple adolescent girls leaving during the reporting period; we also provide the number of out-migrant girls *per household*. In the top panel, we report these findings for the true panel sample of girls recruited at baseline and successfully re-contacted at ML2. Utilizing the panel is helpful in this

[358] The conflict between Somaliland and Puntland for control over Sool, Sanaag, and Cayn, especially, can lead to small-scale conflict, displacement and general disruption of community life. Other areas – such as the western Mudug region of Puntland – are affected by long-running clan disputes that can also drive movement, both within a given area and between areas (with individuals moving to urban areas for better security and economic opportunities alike).

[359] Note that the wording of the question around migration is somewhat unclear, and may imply that girls who left their households but live in the same community should be counted. This is problematic, because girls who have married but remain in the same village are counted as "out-migrants" in our analysis. However, we note that the same question wording has been used consistently across time; thus, we are confident in the broad trends over time reported in this section. Further, the question context clearly implies full migration (not simply leaving one's household and staying in the same community), and nearly all respondents indicated that the girl had either left to live in another village or left the country entirely.

case, because it ensures that the results are not driven by changes in sample composition, which is especially problematic if girls who fell out of the sample are members of less stable households in general. Importantly, these results almost certainly understate total out-migration in the sample, because girls who fell out of the sample entirely are not captured, and these girls are more likely to live in households with high levels of out-migration.[360]

TABLE 145: ADOLESCENT

	Baseline	Midline #2
Count of Adolescent Girls Migrating out of Community in Last Year (Migrants per household) Baseline Panel		
Intervention	76 (0.13)	218 (0.39
Comparison	66 (0.13)	196 (0.39)
Total	142 (0.13)	414 (0.39)
	Midline #1	**Midline #2**
Count of Adolescent Girls Migrating out of Community in Last Year (Migrants per household) Midline Panel		
Intervention	100 (0.28)	129 (0.36)
Comparison	80 (0.27)	125 (0.42)
Total	180 (0.28)	254 (0.39)

The table shows a sharp uptick in movement out of households and communities among adolescent girls from baseline to ML2. At baseline, 0.13 girls per sampled household (in the "true panel") had left their households over the past year. At ML2, this rate had increased to 0.39 girls per household. The bottom panel reports the same analysis for the sample of ML1 girls re-contacted at ML2 – a similar trend is observed, but much less stark. Our interpretation is that out-migration rose steeply from baseline to ML1 and has continued to increase at the time of ML2. A nearly identical trend has occurred among adolescent boys, although we do not report the results here.[361]

Notably, differences between types of girls in the ML2 sample run counter to our expectations in some ways. Out-migration is not higher among households with an OOS girl than among households with an in-school girl, as defined by their status at the time of their recruitment. More importantly, out-migration is actually lower in households with a girl attending an ABE program. It is possible that households that support education enough to enrol their daughter in ABE are also defined by other characteristics that reduce out-migration; regardless, this seems to be an encouraging sign regarding the stability of ABE households.

Beyond the overall number of out-migrants, their reasons for leaving are indicative of the pressures faced in these communities. The table below reports the reasons cited for a girl leaving their household in the previous year, as reported during ML2 data collection. Most notable is the number of girls who have left to get married; however, in the absence of age data for each girl in question, we cannot say whether these marriages are

[360] This analysis also does not capture entire households that have migrated.

[361] An explanation that could undermine the simplistic view of this trend is that the number of girls leaving their households increases naturally as girls (or boys) get older. Older adolescents are more likely to leave the household than their younger siblings; as all household members have age two years since the baseline, could this explain the upward trend in out-migration? We do not believe this is the case, because the age range of interest is specified in the question. Therefore, the mean age of girls being assessed as either present or migrated should not have changed systematically, as older girls leave the eligible age range (11-21 years).

among younger or older adolescent girls. Many girls also have moved to live with another family member, which typically occurs when a girl's household cannot afford to support all of their children, they want to find a suitable spouse for her in a larger town, the household they are joining is in need of domestic help (e.g., watching after young children), or the girl is attending school near where the family member lives.

Reason	Share of Out-Migrated Girls (ML2)
Left because she got married	28.6%
Left to live with another family member	25.0%
Goes to boarding school	23.8%
Other	16.0%
Left to take up work	5.6%

Among boys, the reasons for migrating are slightly different, and tend to emphasize attendance at boarding school and taking up employment. In total. 43.9 percent of boys who had migrated had done so to attend boarding school. Relative to girls, more boys had left home to take up employment, but this was not the most common outcome for either boys or girls – among boys, around 10 percent of out-migrants had left for the purposes of employment. Compared to girls, boys were much less likely to have left to get married (just 6.1 percent of boys left for this reason), which is consistent with a context in which boys tend to marry at later ages than girls.

The data do not shed much light on the upward trend in migration, or in patterns of out-migration across households or communities. Households that report conflict in their village in the last 12 months are more likely to have a girl who has migrated away at ML2, but the share of communities with such outright conflict is not sufficiently high – nor is the migration gap between conflict-affected and unaffected communities sufficiently large – that it would explain a broad upward trend. Out-migration at ML2 tends to be highest in Somaliland, where 0.46 girls per household have migrated in the last 12 months. This trend was not present in the baseline, at which point Galmudug had the highest out-migration rates of girls.

Overall, the discussion in this section underscores the mobile nature of the populations with which SOMGEP-T works, and of the comparison schools included in this evaluation. Beyond the pastoralism that defines many of the communities in Sool, Sanaag, and Mudug, and other regions where SOMGEP-T is operating, there is also significant movement between rural communities and from rural to urban communities, either for the purposes of attending boarding school, helping another family member, easing the burden on one's own household, or getting married. During this evaluation round, for every 5 households in our sample, 2 adolescent girls had left home in the last 12 months, and a similar magnitude obtained among adolescent boys. The extent of individual-level mobility of this kind is surprising, even in this region.

ANNEX 16: EXTERNAL EVALUATOR DECLARATION

Name of Project: Somali Girls Education Promotion Project - Transition

Name of External Evaluator: Consilient Research

Contact Information for External Evaluator:

Consilient House, Masalaha District
Hargeisa, Somaliland
info@consilientresearch.org

Names of all members of the evaluation team:

- Brenton D. Peterson
- Juuso Miettunen, Ph.D.
- Sean Robert

Brenton Peterson certifies that the independent evaluation has been conducted in line with the Terms of Reference and other requirements received. Specifically:

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

Brenton D. Peterson

(Name)

Consilient Research

(Company)

April 10, 2020

(Date)

ANNEX 17: PROJECT MANAGEMENT RESPONSE

Response to Findings

1. Learning Outcomes

The results on financial literacy are a powerful demonstration of SOMGEP-T's impact on girls' ability to apply numeracy skills, particularly considering that this is a core focus of the 'Numeracy Boost' methodology. Major gains in financial literacy were also observed among highly marginalised subgroups – out-of-school girls and girls with disabilities – confirming the effectiveness of SOMGEP-T's approach to inclusion. It is worth noting that the gains in financial literacy for GwDs actually *exceed* those in the general intervention group. A similar pattern was observed among extremely poor girls – for whom difference-in-differences are statistically significant for both numeracy and financial literacy, thus confirming SOMGEP-T's impact on the most marginalised. The fact that girls are acquiring financial literacy skills at a faster pace than boys in intervention schools, while the opposite is observed in comparison schools, highlights how SOMGEP-T's intervention is not only improving learning outcomes but also shifting gendered learning patterns. It is important to note, however, that the impact on financial literacy is dependent on exposure to Girls' Empowerment Forum programming and on the period of participation in GEF activities. Expanding the coverage of GEFs in a sustainable manner – including through girl-led replication – is crucial to maximise future impact on financial literacy.

While no statistically significant difference-in-differences between the intervention and comparison groups have been observed for Somali literacy or numeracy, it is encouraging to see that non-significant differences have emerged in a short period of 12 months. In fact, the intervention group systematically shows a higher performance in relation to the comparison on numeracy; the average numeracy score for the intervention group is 6.4 percentage points higher than the average comparison score.

The positive effect of active CECs on learning outcomes confirms the project's ToC. For this subgroup – 48% of the sample – difference-in-differences are much higher: 10 percentage points higher for numeracy, 13.6 points higher for Somali literacy, 6.4 for English literacy and 4.9 for financial literacy. While the finding is particularly encouraging, it also emphasises the importance of additional training for CECs who are lagging behind and highlights the importance of system investment in participatory governance. Other findings – such as the positive effect on learning of school toilets for girls, in particular when privacy walls are added, and of the presence of female teachers – further confirm the project's ToC.

Gains in English literacy remain limited and non-significant for the true panel (cohort tracked since baseline). The results are not surprising considering the delays in rolling out the e-platform. While the presence of an active CEC maximises such gains by 6 percentage points, it is clear that additional efforts are needed.

As discussed in our response below regarding recommendations for further work on Somali literacy, there is a clear need for adaptation of the approach, but also for emphasising the need to work with a subgroup of girls who have been retained/ demoted to Grade 3. This subgroup has the largest performance gap in relation to the comparison girls and mirrors the difficulties faced by ABE girls, indicating an added layer of marginalisation. The findings on working memory and low protein consumption are of particular importance for revising the approach to these subgroups, as well as for pastoralists (who remain disadvantaged in relation to other groups).

2. ABE

The poor learning outcomes observed at the ABE baseline are not surprising, given the fact that this subgroup lacked previous exposure to education. It is interesting to note that their Somali decoding skills are particularly poor though – which may indicate that even previous exposure to religious classes (*Duqsi*) may have been limited for this subgroup who are mainly nomadic pastoralist drop-outs[362]. As noted above and also in the recommendations section, the findings indicate an added level of vulnerability that may reflect a history of malnutrition, with a potential impact on cognitive development (a hypothesis supported by the low working memory scores observed among this subgroup). It is also important to note the findings on life skills and limited voice, which suggest that GEF participation would be particularly beneficial for this subgroup.

3. Transition

The positive results on transition are very encouraging, in particular noting the impact on dropout and grade progression, and the major impact on the enrolment of former out-of-school girls. In particular, it is important to note how the gains are maximised by the participation in the GEFs, confirming the project's ToC.

4. Attendance

While attendance gains have been modest, the results are not surprising; the timing for SOMGEP-T's evaluation rounds coincides with the end of the dry season, and more often than not, data collection takes place during a drought, when girls are forced to walk longer distances to have access to water and to find pasture for livestock, and at a time of food insecurity, thus increasing the likelihood of illness-related absences. Therefore, gains in attendance are extremely hard to obtain – in particular because SOMGEP-T's schools are located in more remote areas with less resources than comparison schools[363]. For instance, the proportion of households not eating protein-rich foods (staples in a pastoralist diet) is 23% among the intervention group, compared to 17% among the comparison group. It is encouraging, however, to observe the much larger impact on attendance among GEF participants, which further confirms the project's ToC and highlights the importance of the girl's own motivation and vision in boosting attendance.

[362] Children learn to decode in Arabic during religious classes, and there seems to be a positive transfer of skills to the Latin alphabet, as even students in early grades typically do not have a zero baseline in literacy.

[363] It is important to note that SOMGEP-T's quasi-experimental design was built on the top of an existing project (SOMGEP, 2013-2017) which used a pre-post design. SOMGEP-T's intervention schools are the same as those targeted by SOMGEP. These schools were selected in 2013 based on marginalisation criteria, which included remoteness, vulnerability to conflict and limited resources. The comparison schools were selected in 2017, seeking to mirror the characteristics of SOMGEP's original schools; nonetheless, their level of remoteness and exposure to conflict is not as high as SOMGEP's, as extensively discussed in the ML1 study.

5. Teaching Quality

Findings on teaching quality are far more mixed, with gains in terms of the use of respectful language, creating a welcoming environment in school and reducing the use of passive methodologies (copying from the board, repeating after the teacher). On the other hand, progress on the use of formative assessments has stalled, and the use of student-centred games and group work has decreased. These are contradictory results if considering that financial literacy results have increased dramatically – an area where the use of active teaching methodologies is essential for improvement. There is a possibility that more active methodologies are being used primarily in areas seem as relevant (financial literacy is certainly prized in a society where petty trading is a key livelihood) but less so in Somali literacy, for instance, thus leading to poorer performance. As discussed below, the results may also reflect teacher turnover (as our monitoring data indicates that there is a considerable movement of teachers between grades and out of the schools); the findings also highlight the importance of additional training and on-site coaching.

It is interesting to note the dramatic reduction in the use of corporal punishment during classroom observations, indicating a major shift in teacher awareness of the unacceptability of the practice. Although student data indicates that corporal punishment continues to be used, this shift suggests that the increased engagement of CECs in child protection may be having an initial effect in curbing the practice – at least in the presence of observers. The result highlights the importance of additional supervision – by CECs as well as DEOs and QAOs – to increase impact on reducing the use of corporal punishment.

6. School Management and Governance

The proportion of schools with active CECs and school improvement plans has increased over and above the comparison group. The learning results show that the presence of active CECs had a major impact on maximising gains in literacy, numeracy and financial literacy, confirming the project's ToC. On the other hand, progress on multiple areas of CEC engagement seems to have reached a plateau, particularly on attendance tracking and support to students from marginalised subgroups (although awareness of their needs has increased). As discussed in the section below, additional training is needed, but also an added emphasis on coaching (particularly with the engagement of DEOs and QAOs) on following up in cases of absenteeism, dropout and case management.

7. Life Skills

The results confirm the strong correlation between leadership skills development and learning outcomes; increased girls' agency and improved learning results occur in tandem, highlighting the importance of synchronising interventions focusing on life skills and academic skills. GEF participants are outperforming their peers by 16 percentage points, a staggering difference, further confirming previous findings on the positive impact of this intervention.

The positive impact on the acquisition of leadership skills (voice, self-confidence, vision, organisation and decision-making) and the impact of GEF participation on transition and attendance confirm the project's ToC. As further discussed in the section below, it is interesting to see that the impact is more pronounced on girls' voice and 'learning to learn' (organisation, vision) rather than on decision-making. While this is not a surprising finding – girls' voice is typically the first emerging domain of leadership skills development, in our experience in implementing similar programming in other contexts – it is interesting to see the contradiction between the statements of limited capacity for decision-making and the positive effect of GEFs on transition and attendance, which seem to indicate that girls with enhanced leadership skills are indeed making some decisions for themselves on staying in school and attending class, even though social norms on decision-making remain unchanged. Qualitative data from vignettes indicate a mixed scenario,

where some girls feel that girls can make decisions about continuing their education while others consider that the decision falls on parents or that some barriers cannot be overcome, resulting in drop-out. Qualitative data from mothers suggests that while girls may take important decisions, this may be regarded as socially acceptable within the family/home environment but not overtly expressed in public:

"No, it's different within the community and the house. So they can't express anything within the community but except for when the teacher asks them questions or like you come and ask them questions their voice is very low, but they talk in their house too about what they will do after they graduate and how their will education be, and where they would like to be. So they have high ambitions and they talk about it in the house." [Mother, Somaliland]

"The community is full of setbacks, some of them will say, 'How can a girl make a decision for us?' but they are okay to do so in their education." [Mother, Galmudug]

Interestingly, when completing a vignette about a GEF mentor, one of the members and her friend who is struggling in school, girls seem to strongly indicate that the influence of the GEF would result in the struggling girl staying in school and succeeding in her education – showing that peer support may be a critical turning point for transition.

Qualitative data seems to indicate strong parental support for girls' increased ability to express their opinions and participate in school. Mothers participating in focus groups seem to have overwhelmingly positive attitudes about girls' voice, and how girls' activism is inspiring to others. The positive response from parents may explain why this domain emerges in a particularly strong manner, compared to decision-making.

"I'm happy in the way that she expresses things and what she does." [Mother, Somaliland]

"For example, when the girl is presenting something to the school, talking about education, and her teachers, mothers who don't have girls in school will be encouraged to send them." [Mother, Galmudug]

"I have a little girl that goes to the school. If she plays around, girls come to me and tell me to change the way she's acting and that it's not good for her. They advise us." [Mother, Galmudug]

"The school welcomes the girls to express their views. The girls spread awareness in the community and the school welcomes whatever is good for the school." [Mother, Puntland]

It is also interesting to note that Youth Leadership Index (YLI) scores have increased considerably among the GEF participants after an initial drop at the previous second evaluation round. The pattern mirrors the effect previously observed in other contexts, where girls' YLI scores tend to drop once they become more aware of their rights and the barriers for the realisation of such rights; scores start rising again once participants engage in activities that allow them to experience new roles and apply leadership skills at school and community.

Maximising impact

GEFs have emerged as one of the most impactful components of the project. Still, the number of students participating in the Forums or engaging in the activities promoted by members is relatively limited (38% of the sample). Not all locations have mentors and not all GEFs have the same functionality level (for instance,

while 86% had members attending regularly, only 66% have action plans). The project provides coaching support to struggling GEFs and respective mentors based on the results from fidelity of implementation assessments conducted on a biannual basis. Increasing GEF coverage and maintaining efficiency are key priorities as the project finalises its intervention. The main initial focus will be on enhancing fidelity of implementation through on-site coaching and where necessary, training of new mentors. Therefore, the project will prioritize training/ coaching to the 34% GEFs without action plans, including training of new mentors. Whenever possible, the project will train GEF members graduating out of primary school as mentors, while also including in the training a self-replication module – encouraging members to build their own 'off-shoots' of GEFs. This strategy has been successful in other settings, resulting in 'leadership clubs' being formed not only at the school but also at community level. Where this strategy is not possible – for example, in locations where the majority of the graduates tend to move out to attend high school -the expansion of GEF coverage should be, where possible, through the training of female mentors already involved in other project activities (VSL, CECs) to build upon previously acquired GESI awareness and skills.

8. Community Attitudes

Findings on increased support for girls, including positive perceptions of investing in girls' education, are encouraging and reflected in the impact on transition and attendance, confirming the project's ToC. It is not surprising to see that head teachers and teachers continue to consider that community attitudes towards girls' education are negative – and these perceptions are certainly consubstantiated by the presence of a substantial proportion of parents of out-of-school girls within those communities. On the other hand, these perceptions are contradicted by the support to ABE children to attend class vis-à-vis marrying, demonstrating a process of social norm change.

The positive impact of parental participation in VSLA on learning outcomes (resulting in a significant increase of 5.5 percentage points for the children of participants) may be interpreted as resulting from a dual shift – in the economic condition of the family as well as on parental attitudes and practices. Mothers participating in VSLA are exposed to messaging on gender equality, in particular on education, attendance, allocation of chores at home and support for non-traditional roles for girls.

9. School-Related Gender-Based Violence

The increased proportion of parents with positive perceptions of safety on the way to school reflects the project's successful engagement of CECs in this area. It is also positive to observe that girls have identified the school as a safe place during risk mapping exercises. The proportion of children feeling unsafe at school or on the way to school has also decreased. Findings on harassment and unsafe toilets at school are a serious concern, however, particularly considering the added vulnerability of the cohort as they grow older. Findings on learning clearly show how safe toilets contribute to performance (as discussed above, and in line with the ToC). The results emphasise the importance of working with men and boys through BEFs to build different perspectives of masculinity that are supportive of women and girls instead of aggressive towards them, as further discussed below; the need to further support CECs and local leaders to reduce tolerance for harassment at school and on the way to school, potentially resulting in a broader agreement on a code of conduct towards students, enforced at community and school; and also the need for additional qualitative research on social norm change drivers among boys. As the project approaches the end of its intervention, there is also a need to strengthen system-level responses to SRGBV, working with gender focal points and district education officials to build more effective mechanisms for disclosure and redress of cases, as well as to incorporate measurements related to SRGBV in official school monitoring. Last but not least, there is a need to match community-level efforts to improve safety with system-level activities to ensure a better

school environment. The upcoming District Safety Dialogues and School Safety Dialogues, implemented by the MOECHE under the GPE-ESPIG, are designed to bring together education officials, school personnel and community representatives to agree on concrete actions to safeguard students and improve attendance and retention, in particular for girls. SOMGEP-T's findings are informing the design of these activities. Findings on violence in general (not related to gender, but rather to conflict/ armed clashes) and the potential for latent violence are a source of concern, but sadly not unexpected given the level of instability in SOMGEP-T's targeted areas. The economic downturn, ongoing crisis and potential for political instability during elections have the potential to result in an increase in violent attacks/ clashes in the coming months.

10. Sustainability

The project largely agrees with the findings on sustainability, while pointing out that there are certain gains – in particular on transition, reflecting social norm change at community level, and on the positive impact of CECs across all measures of learning (while previously impacting literacy only), potentially due to the new skills acquired. The engagement of GEFs is clearly reflected on the impact on transition, showing a positive trend of girls acquiring skills that will enable them to continue to act with others in the future to shift issues affecting women and female youth. There are gains at system level, with increased monitoring and reports of REOs and DEOs replicating SOMGEP-T approaches in other areas. Also at system level, there is substantial focus on ABE and policy development at the moment, which provides an opportunity for SOMGEP-T to contribute to USAID-funded efforts to develop a policy framework for accelerated education in country. Still, it is a mixed scenario, typical of a transitional period – CECs have reached a plateau in progressing in some areas; teaching quality has progressed in certain areas but not in others, in particular the use of formative assessments; and system gains are limited by structural issues such as funding for logistical costs. At this point, sustainability is at a point between 'sustainable skills' and 'sustained practices', with some practices having changed through time but others not. From a social norm change standpoint, there are promising results at girl (GEF), community (parental support, CECs working on child protection, ABE girls not marrying) and teacher (respectful behaviour) levels.

Response to Recommendations

Recommendations on Evaluation

11. Measurement of Transition

A key difficulty in studying the impact of the program on transition rates is the extent of attrition among the sample of OOS girls. In the absence of a logical mechanism for selecting new OOS girls into the sample, the share of OOS girls in the primary cohort has declined precipitously from baseline (n = 754) to ML2 (n = 368). Given the program's emphasis on enrolment of OOS girls into alternative learning outlets, such as ABE centres, a dropping number of OOS girls in the primary cohort sample is arguably unfair to the program, as it reduces the weight placed on a group within which program impact may be especially significant. Of course, large-scale attrition is also methodologically unsound, presenting inferential problems across many aspects of our analysis. At the endline, an explicit component of the evaluation should be seeking to re-contact any and all girls who fell out of the sample since baseline, even if by telephone. A similar tactic undertaken in this round applied only to in-school girls, as no OOS girls had fallen out of the sample prior to this round; however, this approach should be expanded to all girls who fell out of the sample at any previous round. Moreover, the procedures should be expanded – for girls who cannot be located, their information should be passed to office-based researchers, who can call the given contact numbers for households periodically; in addition, head teachers should be pressed

for additional information about the girls and their families, so that they can be contacted. This effort will take additional time. However, girls will not be replaced at endline, freeing up fieldwork time that would otherwise be spent interviewing replacement girls. By maximising the number of girls who can be tracked—even if only for the purposes of assessing their transition status—the rigor of the endline evaluation will be improved.

A limitation in the transition analysis also arises when girls leave and re-enter the sample, because they cannot be included in the full analysis across multiple rounds of data collection. In the current evaluation, this applied most prominently – but not exclusively – to OOS girls, who were not contacted at ML1. As a result, transition analysis for OOS girls only considered their status at baseline and ML2. However, this problem will become more noticeable at endline, as girls who are successfully brought back into the sample from baseline or ML1 lack one or more rounds of data on transition status. At the endline, a set of transition questions should be added to the household survey to retrospectively capture enrolment status and grade level for each year since the baseline. Capturing this status for each prior round – and cross-checking it against the data collected in the past – will facilitate analysis of transition across all four rounds for a relatively larger sample of cohort girls.

The project agrees with the recommendation to conduct an initial tracking of students – including remotely via phone and through the engagement of head teachers as well as local mosque and GEFs (not proposed by the evaluator, but there is a high likelihood that peers may have additional information about the girl's whereabouts and contact details). This is an activity that can be conducted by the project in advance of the evaluation; if conducted in a participatory manner, it may be a valuable exercise to build local capacity to track dropouts (similar to CARE's previous experience in girl-led action research in Honduras[364] and community-led dropout tracking in Timor-Leste). Regarding the addition of questions to assess enrolment and grades attended during prior years, this can be done relatively easily (while noting for potential errors as the concept of 'year' may be a bit more fluid for individuals to whom this holds limited meaning, such as caregivers).

12. Measurement of Formative Assessment Use

Given that an increase in the use of formative assessments is one of the primary goals under the teaching quality component of this project, future evaluations should develop more effective means of assessing their use. While it may be onerous, asking teachers to produce documentation of the formative assessments (for example asking to see exit tickets) or looking at teacher lesson plans to see if teachers are building in opportunities to elicit and incorporate student feedback may be necessary in order to properly evaluate the use of formative assessments during instruction. Additionally, training enumerators to better understand formative assessments and how to look for teachers using feedback from students during instruction may allow for a more reliable way of assessing the use of formative assessments in classrooms. Lastly, incorporating questions during teacher FGDs about the strategies they use to formatively assess students and how they approach incorporating student feedback into their lessons and lesson plans may help shine a light on if teachers are using formative assessments in their classroom.

For instance, teachers should be asked what proactive steps they take to help students who are falling behind when planning their lessons, during class, and outside of class. They should also be asked how they are incorporating formative assessments

[364] Under the *Renacer* program, CARE trained girls and boys in slums near Tegucigalpa to conduct action research on out-of-school children. The experience had a long-lasting effect in mobilizing the community towards the establishment of a successful accelerated education program (*Amanda Moll & Lotte Renault (2014) Rebirth, empowerment, and youth leading social change: non-formal education in Honduras, Gender & Development, 22:1, 31-47*)

into their instruction and how, specifically, the results are being used to guide instruction and lesson plans. Alternatively, some of this information could be derived from alternative sources, including surveys with head teachers.

While the project agrees in full with the recommendation to ask teachers to produce documentation of the formative assessments and requesting lesson plans, we are surprised that this has been recommended as it has been included as part of the classroom observation since the baseline (variable e11 – "Do you have records of the formative assessment used and the results?" requires observation of the documentation, as the response options are observed/not observed). We do agree, however, that there is a point in providing additional training to enumerators on the formats used and how they are applied during instructional time. There is also value in obtaining qualitative data from teachers on the use of formative assessments and the challenges faced during use – particularly if this is done using a participatory exercise such as storytelling.

The project fully agrees with the recommendation to incorporate the question on the use of formative assessments with students who are struggling. There is also the possibility of conducting follow-up qualitative interviews with teachers in schools with high rates of holding students back/ demoting students at ML2 to explore what has/ has not been done to address the issue, complementing the quantitative results at the endline.

13. Measurement of Community Attitudes

As it stands, the evaluation of SOMGEP-T has sought to use the best possible measures of community attitudes, and triangulate these findings across multiple types of respondents and multiple question types. Measurement of community attitudes has become more nuanced and extensive since the baseline. However, there are still shortcomings in how community attitudes are assessed. Currently, questionnaire design does not make explicit distinctions between high- and low-performing students when asking caregivers about the value of girls' education. For instance, when a caregiver is asked "Is girls' education worth investing in, even when funds are limited?" they can simultaneously agree strongly in principle but advocate removing their own daughter from school because she is performing poorly and they perceive their girls' education to be fruitless. The available evidence suggests that community members and caregivers may be making this distinction implicitly when asked about girls' education – focusing on the highest-achieving girls and emphasising the value of girls' education in those terms.

At the endline, we recommend attempting to distinguish between attitudes toward high-performing girls and toward those who have not performed well or who are not expected to continue into the workforce. Specifically, we suggest adding nuance to existing questions regarding whether investing in girls' education is worthwhile and whether girls are just as likely as boys to use their education, by framing the questions with reference to whether the girl is performing well or poorly in school. We also suggest more targeted probing questions be added to the qualitative tools – alongside training to emphasise how to use these probes – to assess whether broad stated support for girls' education extends to girls who have fallen behind for their age, or are performing poorly. Finally, further analysis of the relationship between girls' performance and community attitudes may be worth considering, including how community attitudes change in response to high performance for just one or two girls, and how caregiver attitudes change in response to positive improvement by their own girls. Efforts of this kind will provide insight into the dynamics of attitudinal change. They may also provide insights that can be useful for a future iteration of GEC or LNGB programming in Somalia.

The project agrees with the proposed recommendation. It is interesting to note, for instance, that ALP students were found in the previous midline study to be low performing students – confirming that poor learning outcomes may contribute to a decision to drop out of school. Parental / teacher perceptions of a student's ability to learn are also likely to have a disproportionately negative impact on GwDs. It is also

worth noting that this is not only a topic for measurement, but also for further sensitisation of community members through platforms such as CECs and VSLAs, as well as through the influence of religious leaders.

14. Improved Use and Collection of Qualitative Data

At the outset of the endline evaluation, specific effort should be dedicated to rewriting the qualitative tools with new purpose. At baseline, the qualitative tools were designed largely to understand the existing situation in communities, identify barriers to girls' education, and understand subgroups that faced the most significant barriers to enrolment and learning. This goal is reflected throughout the qualitative tools. However, as the program has progressed, the goal of the evaluation has shifted from understanding the existing context within communities and guiding targeting of the program toward assessing what changes have taken place since baseline. Before the endline, the qualitative tools should be revised to emphasise this goal, asking interviewees what has changed in their community since baseline, how attitudes toward girls' education have changed and among which groups in the community, whether girls are more likely to enrol and stay in school now than they were previously, and whether this progress is uniform across all girls. Beyond the different goals of baseline, midline, and endline evaluations, the qualitative tools should also be revised simply because conducting qualitative interviews over three rounds has given us more information about what types of questions work and which do not – learning that should be more actively incorporated into tool design in the next round.

Qualitative analysis in this evaluation round focused predominantly on differences across subgroups, i.e. cross-sectional differences in learning or transition, many of which stem from structural or systematic gaps that existed before the program began and are likely to continue – even if in muted form – at endline. However, by reorienting the qualitative tools toward changes in communities, the qualitative analysis at endline can shed light on program impact, differential impacts across groups, and the barriers that have either continued to inhibit progress or have been overcome. In short, the qualitative analysis will be better aligned with the retrospective, evaluative goals of the endline.

The project has consistently revised qualitative tools across all evaluation rounds, so this recommendation reflects an already existing practice. Questions can be revised to ask explicitly if there have been any changes or what changes have occurred at community level; in some cases, this has already been done at ML2 (considering questions such as 'Do you think that there have been any changes during the last year in supporting girls with disabilities in this school?'). It should be noted, however, that by asking about 'change', we may introduce desirability bias – while the current approach of asking about actual practices towards excluded groups presents a more neutral stance. It is also possible to use a different approach – using a 'change map', starting with the changes identified by the participants and then exploring other areas to identify if any practices have been modified, and to what extent these may be perceived as beneficial or not by community members. This is a more time-consuming exercise, however.

.

15. Approach to Assessing Impact

As we have noted repeatedly in this and previous report rounds, the difference-in-differences approach to evaluating SOMGEP-T allows for a rigorous assessment of program impact, requiring comparatively few assumptions to draw valid inferences. However, the nature of interventions like SOMGEP-T is that they are multi-faceted, in the sense that they bring together many different activities, and are geographically clustered. At the same time, many of our most interesting analyses were those that studied the impact of a specific intervention (e.g., GEF participation) on outcomes, because the expected impact of that activity is very closely tied to specific outcomes.

In future GEC-style programs, a more targeted approach to assessing impact may be worthwhile. For instance, if we expect community outreach campaigns to alter local attitudes, a randomised-rollout, dose-response, or other approach to assessing the impact specifically of the campaign on attitudes may be preferable to annual or biannual evaluations, where impacts

are obscured by noise and, in some cases, insufficient statistical power. While such an approach would not capture cross-fertilisation between different activities and their interactive effects, the design may make them more tractable and, in most cases, less expensive to implement.

While this recommendation is not directed to the project, it is a point worth discussing in our response. At the time CARE designed the evaluation methodology, we were fully aware of the challenges involved in assessing the contribution of specific intervention components to the project's impact; in a similar manner, we were aware of the fact that a complex intervention implemented in a conflict-affected context is bound to have different levels of impact for different components. This is a costly and complex exercise, and not as straightforward as a randomised-rollout of independent components. The randomised-rollout of individual components would be a valid approach to test individual innovations; on the other hand, the individual components combined into SOMGEP-T's intervention have been rigorously tested in other settings and were selected for this intervention precisely in an attempt to maximise their impact on the marginalised girls targeted by the project. It is crucial to understand if this is indeed occurring and the synergies between those components.

SOMGEP-T is designed for complexity, acknowledging that the intersectionality of the factors affecting ultra-marginalised populations is a core reason behind the limited impact of more punctual interventions. In this case, some components are clearly emerging as powerful interventions with a broader impact despite the limited sample size (e.g. GEFs) while others clearly require not just adaptations of the approach, but also additional time to gain traction (teacher training). This exercise also allows us to better understand how the intervention components interact through time and contribute to our learning on designing better intervention packages for complex, conflict-affected settings.

16. Refinement of logframe indicators

Several of the logframe indicators present difficulties in operationalization. First, there is a lack of clarity in the definition of the learning indicator, specified as "number of girls sampled who demonstrate increase in their SEGRA/SEGMA scores." This definition is unclear because nearly every girl improves over time with regard to these scores, due to maturation effects; moreover, if a comparison were to be made to a specific girl in the comparison group, each girl should have been assigned into a matched-pair design. We recommend reframing the indicator with explicit reference to the comparison group, by making the indicator the difference (in percentage points) in SEGRA/SEGMA scores between intervention and comparison groups. For each evaluation round, the relative gap between intervention and comparison groups can be set with specific targets.

Since these are standard GEC indicators pre-loaded into the logframe, the project will defer to the Fund Manager on the response to the recommendation. It should be noted, however, that the evaluation reports provide a much broader analysis of the results, allowing for longitudinal tracking of the differences through time between intervention and comparison groups in SEGRA/ SEGMA scores.

More problematic, in many ways, are the qualitative-oriented indicators of attendance, teaching quality, life skills, and school management. In general, these indicators were not explicitly addressed in the baseline report, and it does not seem that the tools were generally designed to target these qualitative indicators. At times, proxy indicators have been used to fill the logframe (e.g., a quantitative proxy indicator for girls' feeling empowered to speak up). However, agreement should be made going forward on whether these proxy indicators will constitute the actual logframe indicator, or whether qualitative evidence will be used to assess if the target has been reached. If the qualitative indicators are to be used at endline, we recommend a longer-than-normal tool development period, with either CARE or the external evaluator developing qualitative questionnaires that more concretely address these indicators, and with additional review by the other parties and the FM. Whereas the qualitative tools currently employed are extremely useful for understanding barriers to

girls' learning and transition, they are less suited to measuring changes in the IOs over time, and this will need to be remedied prior to the endline evaluation.

The project agrees that the qualitative indicators for IOs – in particular for life skills - are not being adequately contemplated in the qualitative tools used during the last evaluation cycle. While qualitative work would ideally be expanded to cover all IOs, this is a difficult choice as it has major financial implications. The project made changes to qualitative data collection allowing for a greater focus on sensitive topics for which quantitative data offers limited clarity – such as social norm change and school-related gender-based violence[365], as well as an in-depth understanding of the experiences of GwDs. The project has also prioritised the use of heavily participatory tools that put girls' voice front and centre, such as risk mapping and vignettes. This approach strengthened the depth of our understanding of those hidden factors, but it is also time-consuming and costly, leading to adjustments in other areas of inquiry. The project is, however, open to the idea of further expanding the scope of the qualitative work at the final evaluation, provided sufficient financial resources are available within the budget to support higher evaluation costs.

Response to Programming Recommendations

17. Learning Outcomes - Literacy

Both analysis of aggregate learning outcomes as well as the sub-group analysis of learning outcomes seem to suggest that the program will be able to show positive impact—at the endline—in terms of financial literacy. The program has also made significant progress in the field of numeracy, as much improvement in learning scores has taken place since ML1. As it pertains to literacy outcomes, however, no impact can be detected at this stage. Prior to the endline, we recommend that the program undertake a final push of activities designed to boost literacy outcomes, alongside additional, targeted pedagogical training for teachers focused on literacy. This recommendation relates to those, below, regarding teaching quality and strategies for addressing foundational gaps in literacy, such as letter recognition.

A second recommendation regarding literacy concerns future investments in girls' education programming in Somalia. We recommend investigating why the program seems to not have achieved the same level of impact in the field of literacy without any evidence of neglecting it in the programming. This might require a more comprehensive profiling of the comparison schools to identify whether and what kind of other education investments are taking place in those areas.

The project agrees with the recommendation to increase efforts on training and coaching teachers specifically on literacy. Two key areas of effort emerge from the data:

(a) Emphasising the development of more advanced skills, as there is a clear polarization of results for writing skills, with a large proportion of non-learners as well as a large proportion of proficient learners, potentially indicating the presence of a large subgroup of students who has not acquired such skills and is now lagging behind;
(b) While data on repeaters/ students held back is not included in this report, the largest differences in performance are observed among Grade 3 students, mirroring the pattern observed in the second evaluation round (Midline 1). Students in this grade include those who repeated grades or were demoted. While the project has already implemented remedial classes, there is a clear need for further emphasis on

[365] SRGBV is not a selected IO at the project's logframe but is closely associated with other IOs (life skills and community attitudes) and an important hidden factor affecting retention and sustainability.

literacy during those, as well as potentially assessing students for dyslexia and working memory issues (further discussed below).

These activities will start immediately as classes resume in September. In case in-person coaching is not possible due to the restrictions posed by the pandemic, the project will provide coaching to teachers via Whatsapp and phone.

Regarding differential impact in relation to comparison schools, it is important to note that other teacher training initiatives are currently in place under the Education Sector Program Implementation Grants (ESPIGs) funded by the Global Partnership for Education in both Somaliland and Puntland. The project conducted an assessment of other interventions being implemented in comparison schools at the end of the 2019 school year (after the second evaluation round). The results indicated that five schools in Somaliland had received support from UNICEF through its Emergency Education project (implemented by local NGOs YOVENCO and HAVOYOCO); another five schools had received support from the GPE-ESPIG in Somaliland; three schools receive support from the EU-funded system strengthening project, *Horumarinta Elmiga;* and three schools are receiving more structured diaspora support (not limited to individual support to students). In addition to those school-specific interventions, the ESPIGs provide general, centralised trainings to teachers, CECs and officials. It is therefore correct to assume that there is a high likelihood of a similar component being implemented in comparison schools, particularly since those schools are more accessible and less affected by conflict than SOMGEP-T's (which were originally selected exactly due to their remoteness and fragility).

18. Learning Outcomes – Hidden Factors: Working Memory and Protein Consumption

New data in this evaluation round revealed that there is a correlation between the cognitive function – in terms of working memory – of children and their dietary diversity, and consistency of food consumption.[366] *While a full-scale school feeding intervention is well beyond the scope of the program, targeted supplementation with protein sources may be worth considering, either over the next year or in any future stages of GEC-T implementation in Somalia. Supplementation of diet with a daily protein source would be cost-effective and potentially yield gains in cognitive development.*

ABE girls report experiencing indicators of poverty and an inability to meet basic needs (low quality roof, going to bed hungry, and lacking clean water), have the highest likelihood of having an unemployed head of household, and are more likely to report having a medical condition that prevents them from attending school. If future program interventions and programs target economic well-being or health/nutrition, ABE girls may be the group that benefits most from these interventions.

The project agrees with the recommendations. CARE opted for including these tools (working memory and dietary diversity) in the Midline 2 round based on learning from other programmes where we have recently piloted working memory assessments in combination with EGRA and a dietary diversity survey. In doing so, we sought to gain a better understanding of the factors affecting the lowest performers in the cohort and how to address those.

It is worth noting that the lowest working memory scores were observed among girls in early grades -in other words, former OOSG or girls who repeated grades/ were demoted (such as the girls in Grade 3, who show the largest gap in performance in relation to the comparison group). It is possible and likely that the limited acquisition of literacy and numeracy skills might be related to mild cognitive disabilities, captured in the

[366] This finding is consistent with at least some of the existing literature. See, e.g.: Whaley, et al. 2003. "The Impact of Dietary Intervention on Cognitive Development of Kenyan School Children." *The Journal of Nutrition* 133 (11): 3965S-3971S.

working memory test. For instance, G3 students in the treatment group have an average working memory score of 45%, compared to 58% among those in the comparison group – a difference of 13 percent points. This may also reflect a history of micronutrient deprivation and malnutrition, with a direct impact on cognitive functions. Therefore, CARE fully agrees that interventions on health/nutrition may be of great benefit for this subgroup as well as ABE girls, and will explore the possibility of doing so through other partnerships (as the cost and scope of such interventions may fall outside GEC-T programming).

19. CEC Trainings

Many interviewees continue to call for training of the CECs. The lack of CEC member capacity was identified as a source of contention between the CECs and the head teachers during the ML1 evaluation round. In this round, some community members, teachers and CEC members themselves continue to lament the lack of committee member awareness and capacity as a barrier for their work being more effective. The program could thus conduct another round of training of committee members before the end of the project.

The project agrees with the recommendation and will conduct another round of trainings, in particular considering the potential for additional community engagement in child protection and support to marginalised subgroups, including struggling students. Given the limited resources of CECs, it will also be important to follow up with VSL members on suggestions for the use of the social fund established by the groups to ensure support to schools and marginalised students in the community – an approach that proved to be successful in Malawi and Zimbabwe. On the other hand, as discussed in the sustainability section, it is also important to consider the potential effects of the ongoing COVID-19 crisis on the ability of the diaspora and local community to contribute to schools; there is a high likelihood of a major economic downturn as a result of lower diaspora contributions and income reduction due to the decrease in livestock sales for export to the Gulf countries and Saudi Arabia (as a result of the ongoing cancellation of religious pilgrimages to the Holy Sites).

20. Positive Deviants at System Level

As it pertains to addressing broader barriers in education, we suggest exploring the current adoption of best practices among MoE respondents, and promoting their broader adoption into institutional mandates. Namely, in some areas in Somaliland, the REOs are ensuring that when a pastoralist child migrates, the school records of the child are sent to the new school in order for the child to continue from where they had left off. While sustainable and systematic change might require a more comprehensive elimination of the educational barriers faced by pastoralists and children with disabilities, such efforts require more resource-intensive programming, while the efforts noted above by MoE employees can be implemented on a smaller scale and with a smaller footprint.

The project agrees with the recommendation. There is a potential positive effect of 'cross-fertilisation' of best practices by identifying positive deviants (as in the case noted above) and building platforms for them to share experiences with peers. The project will facilitate cross-fertilisation meetings (virtual or presential) wherever possible, while also considering the potential political sensitivities of such exchanges. The project will also positively reinforce such practices by publicly acknowledging the efforts of officials who are replicating/ expanding positive practices. It is worth noting that REOs and DEOs have also been observed to replicate SOMGEP interventions in non-targeted areas. The project will also highlight such practices to officials and specialists involved in the preparation of the upcoming new ESSPs, noting that the existing self-replication indicates the potential for broader adoption.

21. Disability

In each round of the SOMGEP-T evaluation, the most common disability indicated by girls and their caregivers relate to frequent anxiety and depression, with anxiety reported by a slightly higher share of caregivers than depression. Although we are not able to provide strong empirical evidence that experiences of conflict or specific instances of violence are correlated with anxiety rates in this sample, that relationship is both theoretically justified and supported in the broader literature.[367] *While providing mental health counselling to students in all program schools is likely cost-prohibitive, some aspects of mental health counselling could be integrated into GEF activities, and efforts could be made to target more specific counselling services to students in communities known to have been impacted by violence (e.g., the ongoing clan conflict in Ceerigaabo). Similarly, girls in schools known to have been impacted by widespread hunger—which can be identified partially based on data collected during this evaluation—could be targeted for mental health support designed to ease the anxiety that arises from such long-run uncertainty and stress.*

The project agrees with the recommendation and in fact has already started to integrate counselling in GEF practices (noting the limitations re: capacity of mentors and lack of professional counselling support in country). We also consider that there is a need for synchronisation of counselling and efforts to address root causes of anxiety (vulnerability to shocks, hunger, child protection issues, systematic discrimination as a result of traditional gender norms), such as safety net components and the already ongoing efforts on child protection. This is particularly true in the context of the current crisis, which is likely to exacerbate the vulnerability of households and may result in a major escalation of anxiety and depression rates. Part of this response may be conducted through SOMGEP-T, but we also anticipate the possibility of overlaps with the planned emergency response.

22. ABE

Large skill gaps in both Somali literacy and numeracy were observed between ABE girls and their ISG and ALP counterparts. However, the skills gap in Somali literacy is especially troubling given that a high proportion of ABE girls struggle with the most basic skills needed to recognise words (a precondition for reading comprehension or writing). Aggressive approaches to teaching or remediating basic literacy skills may be necessary for ABE girls to be able to achieve significant progress and close skill gaps (especially given the low rates of formal education and literacy among ABE HoH/caregivers). There is a large body of research on the link between developing foundational literacy skills at a young age and future academic success.[368] *Future interventions with administrators and teachers may focus on training educators in teaching and reinforcing foundation literacy skills such as phonological awareness and developing curriculum that addresses the specific needs of girls struggling with literacy.*[369]

Educators have become increasingly aware of the importance of teaching literacy in all subjects, not just language arts. For example, when common core standards were adopted in the United States, literacy requirements were included into all subject area teaching standards (for a closer look at K-12 literacy standards under common core in the United States, see: http://www.corestandards.org/ELA-Literacy/*). Given the lack of foundational knowledge in the most basic reading skills, working with school administrators and teachers to help teachers of all subjects to incorporate lessons which teach and reinforce basic literacy skills may help narrow the gap.*

[367] Ayazi, Touraj, et al. 2014. "Association between exposure to traumatic events and anxiety disorders in a post-conflict setting: a cross-sectional community study in South Sudan." *BMC Psychiatry* 14; Bronstein, Israel, and Paul Montgomery. 2011. "Psychological distress in refugee children: a systematic review." *Clinical Child and Family Psychology Review* 14: 44-56.
[368] Catts, H.W., Fey, M.E., Zhang, X., & Toblin, J.B. (1999). Language basis of reading and reading disabilities: Evidence from a longitudinal investigation. Scientific Studies of Reading, 3,331-361.
[369] Most people who struggle with recognising printed words struggle with phonological awareness (decoding and encoding) written words. See: Moats, L.C. 2001, *When Older Students Can't Read,* Educational Leadership: Journal of the Department of Supervision and Curriculum Development, N.E.A 58(6) 36-40

The project agrees with the recommendation and will provide support to ABE teachers to reinforce foundational literacy skills. As noted above, however, the project also considers necessary to take into account the possibility that the learning gaps observed among ABE students may be connected to working memory issues/ mild cognitive disabilities as well as potentially associated with dyslexia, and will invest in equipping teachers with the skills to identify where students may be struggling to decode words due to such conditions. Furthermore, the project will explore the possibility of partnering with other initiatives for supplementary nutritional support (which falls outside the scope of SOMGEP-T's work). We consider that the evidence in this study may generate interest from other development actors to invest in targeted nutritional support for this subgroup as well as for other vulnerable children in country.

23. Teaching Quality

The index of active and participatory classrooms decreased from the first to the second midline rounds in intervention schools. While certain indicators of an active classroom remain high such as not spending most time copying from the board (64.6 percent), teachers asking open-ended questions (60 percent), and not spending most time repeating words aloud (80 percent) other indicators of the student-centred approach like students working in groups (16.9 percent), teachers asking for student opinions (47.7 percent), students instructing one another (47.7 percent), and the use of student-centred games/activities (13.8 percent) remains low. Refresher trainings and providing teachers with example curriculum/lessons that incorporate a more student-centred approach may be necessary to ensure that teachers do not fall back into the teaching strategies they may be most comfortable with.

Reported use of formative assessments increased drastically in both comparison and intervention schools from baseline to midline (more than 90 percent of teachers in both schools reported using formative assessments). However, documentation of formative assessments remains low (only 38.5 percent of teachers at intervention schools report that they would be able to provide documentation of formative assessments). Given the infrequent use of student-centred activities that would lend themselves to teachers eliciting feedback from students and assess their knowledge during learning, there is reason to be cautious about the use of formative assessments to guide instruction. Additional training on the importance and utilisation of formative assessments as well as strategies for incorporating formative assessments into instruction may be necessary.

The project agrees with the recommendation. It is particularly important to understand the reasons for the decrease in the use of student-centred games/activities and group work, and to explore if this is related to teacher turnover. SOMGEP-T's data suggests that there is considerable teacher turnover; additionally, trained teachers tend to be moved to upper grades. The situation clearly calls for additional training as well as more frequent coaching, with an emphasis on teachers who are not applying student-centred practices instead of a blanket approach. The use of formative assessments has clearly reached a plateau (potentially explained by turnover, but also by challenges in adoption of the approach) and requires further coaching effort. The project also needs to further explore why the training is having a positive effect on certain areas (respectful interactions with students; lesson planning; providing encouraging feedback to students) but not in others (use of games, group work), despite all of those requiring changes not only on practices but also on social norms.

24. Girls' Empowerment Forums

Girls' Empowerment Forums stand out as one of the most effective SOMGEP-T activities. While it is difficult to attribute improvements in life skills and attendance directly to girls' participation in GEF activities, as girls can self-select into participation or not, it is clear that girls participating in GEFs have gained greater self-confidence than other girls. Particular effort should be made to ensure the sustainability of GEFs after the end of SOMGEP-T programming should be encouraged. If possible, their expansion through recruitment of additional girls should also be encouraged.

The project agrees with the recommendation and also considers that there is a need for further investment in building the foundations for self-replication of the approach. In other contexts, graduates from leadership groups such as GEFs supported the mobilisation and training of new GEFs, self-replicating the approach through girl-led action. Sharing the evidence generated by SOMGEP-T with the MOEHE staff (particularly Gender Focal Points and Director-Generals) and consultants engaged in the preparation of the new Education Strategic Sector Plans will also open the possibility for GEFs to inform a systemic approach to incorporating life skills development in national strategies.

25. Traveling to school

For schools in rural areas with open spaces or dangerous roads connecting the school to its respective community, or in which girls travel long distances, organizing a way for them to travel to school in safety is needed. There are many mechanisms by which this could be achieved. Teachers could walk with students, though this asks teachers to put in additional time and effort that they may resist. Parents could also be organized, via the CEC or the Girls Empowerment Forums, to walk with students who live in their part of the community, perhaps on a rotating basis. This could serve to reinforce the value of the GEF as an institution and give it longevity after the conclusion of SOMGEP-T's intervention. It is also a genuine exercise in empowerment, fitting with the mission of GEFs. The ideal approach might be to have the BEFs and GEFs at the same school coordinate this effort, to improve boys' understanding of the risks girls face and promote cross-gender camaraderie.

While students walking in groups to school are already a common scene in target areas, the idea of formalizing group walks through the CEC is valid and will be pursued. The engagement of BEFs is a particularly interesting approach that could be used, particularly if it focuses on having BEFs and their mentors addressing harassment at community level. CARE used a similar approach in Bangladesh through the ARSHI-ITSPLEY project, working with groups of positive deviant boys to address harassment ("eve teasing") on the way to school.

26. CEC role in safeguarding

While it is not part of the school management and governance target, CECs are also responsible for safeguarding children and ensuring child protection policies are upheld within their schools. Currently, a majority of CECs reinforce the use of nonviolent discipline in schools and address child protection issues. However, the trend since ML1 has only been marginally positive. Given the CECs' role as the link between parents and the school, their participation in child protection issues should be emphasized going forward. The evidence suggests that active CECs are able to influence outcomes in the areas where they focus (e.g., attendance) and their active participation is appreciated by parents, based on qualitative interviews. A child protection group could be formed in each school, bringing together the head teacher, 2-3 CEC members, and one of the traditional elders in the community, which would allow them to resolve conflicts that occur in school or spill over from the community into the school. They could also raise awareness among the community on child protection issues and the importance of reporting issues that occur in the school.

The project agrees with the recommendation. Child protection committees had a positive impact on the disclosure of cases in the first phase of the IGATE project in Zimbabwe. Moreover, additional training on child protection and reporting mechanisms aimed at traditional elders and VSLAs may be beneficial; and the engagement of religious leaders in campaigning against harassment may be particularly effective. The recommendation is particularly relevant in the current crisis context, when the occurrence of abuses is likely to increase.

27. Household chores

Efforts to improve attendance should focus on the role of domestic work in reducing girls' available time and energy for schoolwork and their ability to arrive on time and stay in the classroom consistently. Community awareness around the importance of girls' education appears to have improved; at this stage, it is important that adults recognize that sacrifices beyond the financial may need to be made to ensure their girls stay in school; namely, household chores can be divided more evenly. If communities shift toward this way of thinking – emphasizing greater equity in the distribution of household work among children and between children and adults – it could generate social pressure to avoid saddling adolescent girls with too many responsibilities.

The recommendation is in line with the project activities; the sensitisation of mothers on the equal allocation of chores at home is a key topic of messaging in VSLAs, literacy courses and religious leaders' sermons. It is likely that girls' chore burden will further increase in the current crisis, given the need to care for sick household members and the strained financial resources of households. The project will also use participatory activities with VSLA members (building upon their GESI knowledge and influence in the community) to generate dialogue on equitable allocation of chores. This activity will build upon CARE's previous experience of conducting participatory mappings of household/ agricultural tasks in Tanzania, Zimbabwe and Cambodia, leading to dialogues on gendered workload and community-led advocacy on equitable practices.

It is important, however, to acknowledge that while there are inequalities in the gendered allocation of chores and resources, there is also a large proportion of female-headed households – nearly half – where girls remained with their mothers/ grandmothers while fathers and boys have migrated or spend long periods away on pastoral activities. Previous qualitative data shows that extremely poor mothers are often making difficult choices – sending one daughter to school on one day and another on the following day – to reach an acceptable compromise between their livelihoods and children's schooling. While the project works to address the realities of gender discrimination in Somalia, we also acknowledge the burden of extreme poverty on female-only households, particularly in the current combination of crises (conflict, locust swarms and COVID). Qualitative data shows a mixed picture; while some mothers explained how girls are usually responsible for chores, others explained how they have taken the responsibility for chores to allow girls to attend school.

"The way that I know it is that I send them to their class on time and stop pressuring them to do housework. I also call the teachers and principal to ask them whether she was there when the class was starting and when it ended. And she comes to me during break times so I can know that she went to school." "Yes, I have given her a chance to study, the housework is minimal or not given to her at all during exam season. We encourage them to be the best. We call the teacher and tell him to repeat the lessons for her if they get hard." [Mother, Galmudug]

"I help them by asking them not to do the housework because she is in Grade 8 and the Grade 8 students take separate classes in the evening." [Mother, Puntland]

"All mothers love sending their daughters to schools. They wash and cook for their daughters in order to make sure they go to school happily. Also, low-income mothers work extremely hard to get some income to cover their daughter's needs." [Mother, Somaliland]

"Also, we as parents reduce our girls the domestic work and we love to see our girls making a huge progress and achieving their dreams and goals" [Mother, Somaliland]

"I don't tell her to do anything when she comes home, I tell her to study and her understanding is very good." [Mother, Somaliland]

"I follow them up, and girls work really hard. When they are focused in their studies, I do the work. Also, I check their handwriting and know when they're not focused in their studies. I do the housework, I herd the livestock and I get the woods." [Mother, Somaliland]

Men's absence / lack of engagement with the household is a recurrent theme in focus group discussions, along with boys' migration. Mothers describe themselves as the 'backbone' of villages.

"As mentioned before, mothers face tons of challenges. They work hard to get income for her family while their husband is chewing khat and doing nothing for the family, children and the home. Only mothers work hard to finance the family and also cook and wash for their family, so they face tremendous challenges." [Mother, Somaliland]

28. Grade progression

The four biggest pathways in which transition rates have improved as a result of the program are enrolment of OOS girls into school, movement of OOS girls into informal education, reducing the number of in-school girls who drop out and move into informal education, and increasing the share of girls who remain enrolled and progress a grade. At this stage, additional attention should be placed on increasing grade-to-grade progression for girls who remain in school. As learning scores improve, this should naturally be tied to higher pass rates; increased progression can also motivate girls to continue their education, and assuage their family that progress is being made. While it is important that standards are not reduced to accommodate this, increasing grade progression can also reduce dropouts and may have positive downstream effects in other areas of programming.

The project agrees entirely with the point. Furthermore, there is a need to work with teachers to address the specific needs of girls whose poor performance resulted in being demoted to early grades. Further investment on remedial support will be of paramount importance in the current crisis, as children who are struggling to learn are likely to be more affected by prolonged school closures, increasing the risk of repetition and dropout.

29. WASH

(a) At the school level, there is a small but meaningful relationship between attendance and availability of gender-segregated toilets. Access to separate toilets is substantially more impactful on girls' attendance than it is on boys' attendance. This trend is supported by qualitative data, in which respondents cite a deficiency in the availability of facilities required for girls during menstruation. While infrastructural inputs such as the provision of latrines across all targeted schools may remain beyond the scope of CARE activities, highlighting the importance of gender-segregated facilities within CEC's and MOE departments may encourage local resource mobilization. Should CEC members (and other local proponents of girls' education) better understand and articulate the significance of access to female-only bathrooms on girls' attendance (and in turn education outcomes), this particular challenge may be addressed through locally driven action.

The provision of gender-segregated toilets was part of the scope of the first phase of SOMGEP and the impact noted is a result of the renovations conducted by the project in a large proportion of the schools. Nonetheless, this was not included in the second phase of the project precisely due to the potential for mobilisation of local as well as government resources. The Puntland 2017-2021 ESSP includes investment in school construction and WASH in 95 primary schools, while the Somaliland 2017-2021 ESSP plans for the refurbishment of toilets in 340 primary schools. Both plans also include grants to CECs to increase equity in access. The large-scale projects implemented through GPE funding (ESPIGs), implemented across all three States, include rehabilitation / construction of gender-segregated toilet facilities. It is important to note that these activities

have been further expanded after the onset of the COVID-19; therefore, this recommendation might not be applicable given the potential for duplicating efforts already underway.

Evaluator's Approach to GESI

The Midline 2 study provides a comprehensive analysis of how gender affects education outcomes, from the individual level (girls' agency) to the household (perceptions of girls' education), community (social norms, gender-based violence) and school (gendered practices in class, perceptions about female teachers). The project had requested the evaluator to provide additional analysis on the impact of GEFs (developing girls' agency) on learning outcomes, given the results observed at Midline 1 and the impact on self-confidence and transition. The newly included findings on attendance gains and social capital are particularly encouraging, while considering the need for further expansion of the GEF approach. It is particularly interesting to note the dimensions of life skills (learning to learn, self-confidence) that seem to be positively affected by the project's approach to girls' empowerment, while not having a clear impact on others (decision-making). The finding may reflect the extent to which social norms have or have not changed: girls and women are exercising voice in public spaces (self-confidence) and organising themselves to take new paths (education) but key decisions (marriage) are primarily made by parents and relatives. Nonetheless, it is interesting to see that when opportunities emerge (ABE), decisions are made to remain in school instead of marrying, signalling norm change for this subgroup.

Qualitative data on gender-based violence continues to provide important information to the project on dynamics related to harassment and abuse, as well as to violence in general and its disproportionate effect on girls. The results provide important evidence to reassess construction/ maintenance needs, as well as to work with the broader community – men and boys, as done through BEFs – to shift social norms underlying harassment. Regarding violence in general, and the latent potential for violence outlined in the findings, it is important to consider how governments may act to prevent situations such as having soldiers stationed in schools.

Disability is extensively discussed in the study, while also acknowledging the limitations posed by the relatively small sample of children with disabilities. The report includes a unique analysis of working memory patterns, conducted for the first time in Somalia, providing information on the prevalence of a mild type of cognitive disability and its implications for learning. It is interesting to note the qualitative findings showing that children do not define themselves by disability; although there are certainly many negative aspects of the experience of disability in Somalia, the qualitative data also allows for a second narrative. It is notable that learning gains in financial literacy have emerged in a similar manner for GwDs and for the broader group of intervention girls in relation to the comparison group; in fact, the difference-in-differences for the GwD subgroup is larger than for the overall intervention group, suggesting that this subgroup has benefitted more than others. It is also fascinating to observe that the prevalence of anxiety and depression has decreased considerably since the baseline (which was conducted under a major drought); while this is not necessarily a program effect, it is an important finding to consubstantiate the theory that the exposure to major shocks exacerbates mental health issues among adolescents. This is particularly important in the current circumstances, given the ongoing COVID-19 outbreak and the potential impact on mental health.

In terms of social inclusion, the report includes a broader analysis of poverty indicators, in particular food insecurity. While the overall prevalence of extreme food insecurity seems relatively low at first glance, a more in-depth analysis shows a higher prevalence, as well as the major impact of lack of protein in the diet on cognitive development, and its implications for marginalised groups, such as ABE students. This analysis provides important information for the Education Sector Analyses as well as for future programming in the

sector. The report also provides information on how subgroups of the population – pastoralists, ABE students – are reacting to the interventions. The analysis of adolescent migration indicates that nearly 50% of the households had adolescents leaving the home last year. While some adolescent migrants – more boys than girls – leave for boarding school, the majority of the girls are leaving home to live with relatives (most likely as domestic workers) or to marry. While migration may have positive outcomes, particularly if girls are able to attain higher levels of education and better employment, it is likely that this will not be the case for the majority of the migrants, highlighting the importance of building skills for self-employment and financial literacy as a means of equipping potential migrants for other opportunities beyond domestic work.

The additional analysis on the impact on subgroups, added to this report, is highly encouraging, indicating that the project is having a particular impact on orphans, pastoralists and the extremely poor. At the upcoming endline study, it will be crucial to determine to what extent these gains have been maintained in face of a major shock such as the ongoing COVID-19 crisis.

Positive findings on progress in acquisition of financial literacy show that girls are acquiring skills needed for economic empowerment; this is particularly important considering an extremely low baseline, indicating that girls (and potentially women as well) lacked basic skills for self-employment and profit-making in income-generation activities. There is a likelihood that the positive results may be related to the engagement of mothers in savings' groups, as well as to increased teacher capacity on teaching applied numeracy skills. It is encouraging to note that mothers' participation in savings' groups is indeed improving learning outcomes – potentially due to a combination of better economic conditions and a shift in social norms related to the exposure to gender equality messaging in VSLAs, but also to the opportunities for girls to practice financial literacy at home, while supporting their mothers.

The study also provides a valuable analysis of the intersection between gender, disability and social inclusion, noting the added vulnerabilities of subgroups such as ABE students in terms of gender and poverty indicators, as well as between disability and experiences of violence and food/water insecurity.

Proposed Changes to the Logframe

At this stage, the inclusion of additional indicators/ revision of existing indicators is conditional to data availability (as the final evaluation will compare results with existing data from previous rounds) and therefore limited in scope. There is, however, the possibility to include additional indicators on teacher use of student-centred games/activities; teacher use of group work; and students reporting feeling welcomed by teachers. These variables are being systematically tracked by the project and could provide a broader perspective of progress on teaching quality. There is also potential for the inclusion of an indicator on social norm change under sustainability, using a qualitative approach – for instance, on the level of social tolerance for corporal punishment and/or acceptance of harassment.

ANNEX 18: SUPPLEMENTAL OUTCOMES ANALYSIS

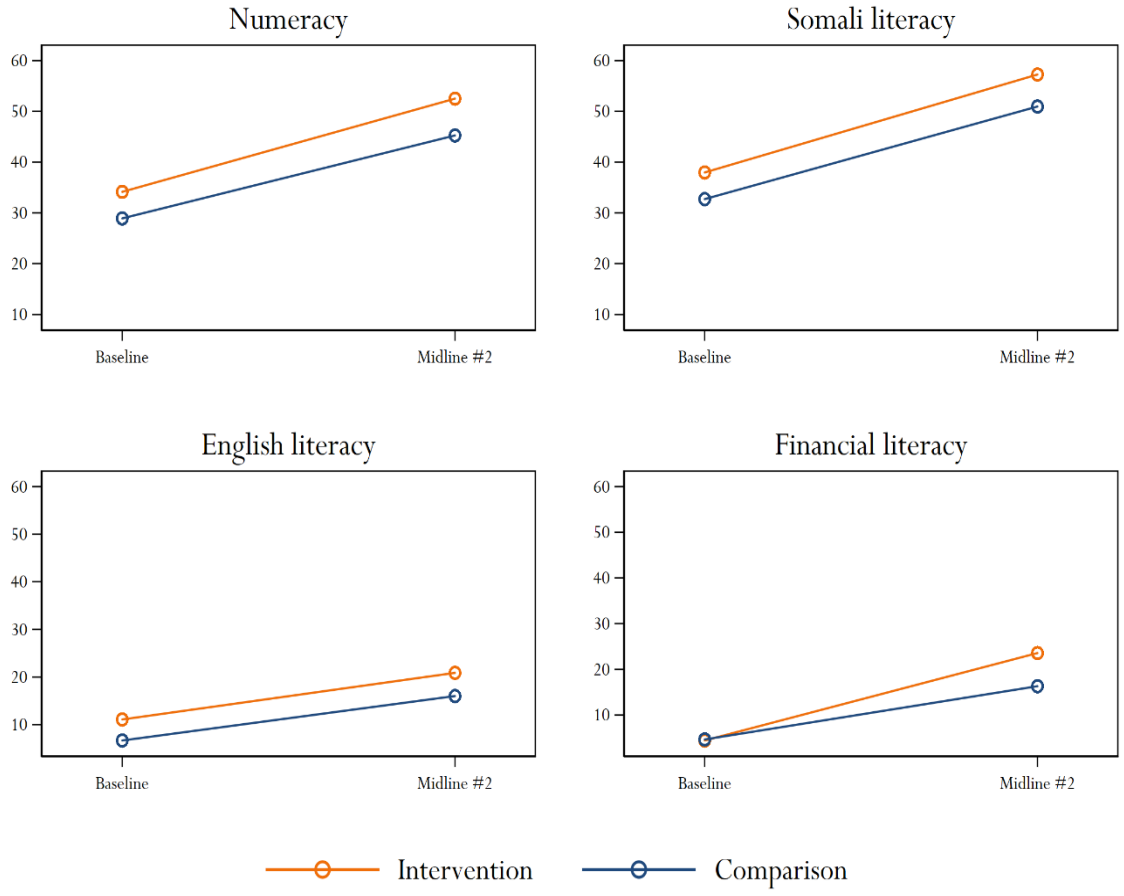
SUPPLEMENTAL LEARNING RESULTS

For the benefit of the reader, in the section on aggregate learning outcomes findings are presented for the primary panel only – that consisting of the individuals who overlap between each wave of data collection. Additional analysis using different panels was conducted, however. This section presents the findings of the analysis with different panels.

Cross-sectional analysis: Baseline to Midline round 1

First, we look at learning results using a cross-sectional panel between baseline and midline round 2. As with the true panel used for analysis in the main body of the report, Midline Round 1 is excluded because at that time no out-of-school girls were assessed. In contrast to the true panel, the cross-sectional panel includes substitutes for 133 girls who could not be re-contacted. As such, while the true panel had 100 percent overlap in terms of the individuals in it, this panel contains more individuals as those who could not be re-contacted have been replaced with similar individuals. The sample size is thus slightly higher than with the true panel. Some more statistical power has been derived from the increase in sample size deriving from the addition of the 133 substitute respondents. Yet, some statistical rigor has been lost as the panel does not contain the exact same respondents that would make it a longitudinal study. The sample size is 1,217 girls with 643 girls in intervention schools and 574 in comparison schools.

FIGURE 57: AVERAGE LEARNING SCORES, BL TO ML2 – CROSS-SECTIONAL ANALYSIS



The results, however, are very similar to those acquired using the true panel, presented in the main body of the report. This is not altogether surprising since almost 90 percent of the respondents in the cross-sectional panel are the same individuals that make up the true panel. Looking at numeracy outcomes with this sub-set of the dataset we see that intervention school students have increased their average score from 34.1 percent to 52.5 percent. Meanwhile in comparison schools the increase in scores between Baseline and Midline Round

2 has been from 28.9 percent to 45.3 percent. This gives a difference in differences of 2.0 in favour of intervention schools. This estimate, again, is not statistically significant.

TABLE 146: DIFFERENCE-IN-DIFFERENCES, LEARNING SCORES AMONG THE COHORT CROSS-SECTIONAL SAMPLE

	Intervention schools			Comparison schools			Difference in Differences	
	BL	**ML2**	**n**	**BL**	**ML2**	**n**	**DID**	**p**
Numeracy	34.1	52.5	643	28.9	45.3	574	2.0	0.466
Somali literacy	38.0	57.3	643	32.7	51.0	574	1.0	0.693
English literacy	11.1	20.9	643	6.7	16.0	574	0.5	0.872
Financial literacy	4.4	23.6	643	4.7	16.3	574	7.5	0.014**

Meanwhile, for Somali literacy, the scores in the intervention group have risen from an average of 38.0 at baseline to 57.3 at Midline round 1. In the comparison group the scores changed from 32.7 to 51.0 between Baseline and Midline Round 2, respectively. The difference of differences estimate for Somali literacy, comparing the intervention cohort to the comparison group is 1.0 which implies that on average the scores in intervention schools improved by one point more than in the comparison group. However, this estimate is not statistically significant, and as such we cannot say with confidence that the program has had an impact on Somali literacy learning scores.

As it pertains to English literacy, the findings are much in line with the other learning assessments. Intervention group scores increased from 11.1 to 20.9 while comparison group saw an improvement from 6.7 to 16.0 on average. The difference in differences estimate, although again slightly favouring the intervention group, is not statistically significant.

Again, while in all assessments the intervention girls have improved their results slightly more, only financial literacy this difference is marked enough to be significant at the 5 percent level. The difference of differences estimate is high, at 7.5 which implies that the program impact on financial literacy seems to have been 7.5 points – girls in the intervention group have improved their scores in financial literacy by on average 7.5 percentage points more than those in comparison schools, between Baseline and Midline round 1.[370]

We also looked at the program's impact on aggregate learning scores by adding up the learning scores for numeracy, Somali literacy and English literacy[371] and conducting and running a regression test with the aggregate score as the dependent variable. When a program has multiple impacts that run in the same direction, combining those outcomes into a single measure can produce greater statistical power. The correlation, while positive with the coefficient for the interaction between round and treatment being 3.6, is not statistically significant (p=0.530). As such, the program is not able to show impact on the aggregate learning outcomes at this stage.

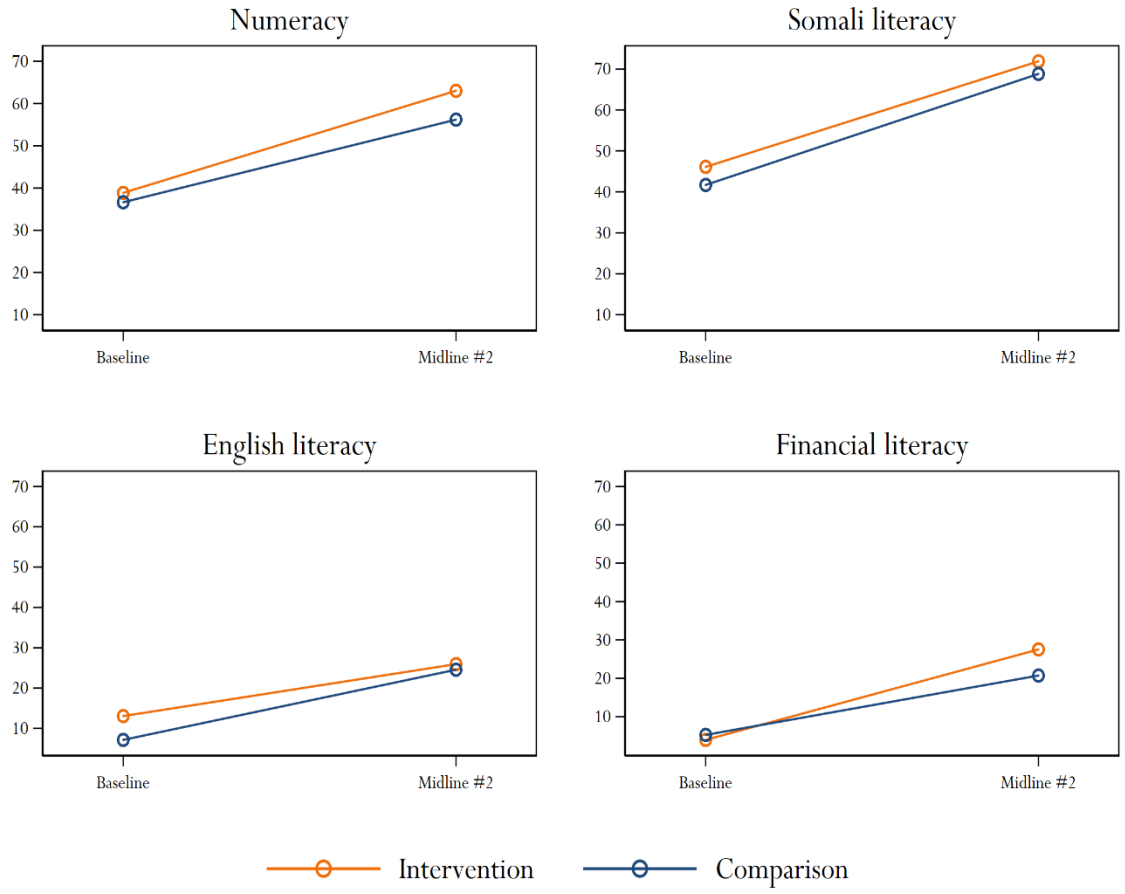
[370] Again, the results are fundamentally unchanged when survey weights are used.

[371] Financial literacy was not included in this aggregate score due to its overlap with numeracy in terms of the skillsets that would risk compounding the same effect.

Girls enrolled Since Baseline

We now turn our attention to a specific group to further explore the program's impact on learning, namely those girls who have remained enrolled throughout the project, including all three evaluation rounds. This will allow us to estimate the impact of continued exposure to the program's activities in intervention schools compared to the education in comparison schools as this is the group most exposed to the teaching done through the program. Given the respondents' continued engagement with the teaching activities of the program, this is where we can best see the difference made by the teaching element of the program. The sample size for this panel is only 477 girls, where 210 girls are in comparison schools and 267 in intervention schools.

FIGURE 58: AVERAGE LEARNING SCORES, BL TO ML2 – GIRLS ENROLLED IN BOTH PERIODS



The above graph and the below table show the results. The trend identified above continues. Namely when it comes to the literacy outcomes, the program does not show impact. In fact, for Somali literacy, the scores in the intervention group have actually risen on average slightly less for those who have remained enrolled through the program. The change is from 46.1 to 71.9 percent. Meanwhile, in comparison schools the scores increase from 41.7 to 68.8 between Baseline and Midline Round 2, respectively. The difference of differences estimate for Somali literacy, comparing the intervention cohort to the comparison group is -1.4 which implies

that on average the scores in intervention schools improved by 1.4 points less than in the comparison group. However, this estimate is not statistically significant, and as such these differences are not conclusive.

TABLE 147: DIFFERENCE-IN-DIFFERENCES, LEARNING SCORES AMONG GIRLS ENROLLED IN BOTH BASELINE AND ML2

	Intervention schools			Comparison schools			Difference in Differences	
	BL	**ML2**	**n**	**BL**	**ML2**	**n**	**DID**	**p**
Numeracy	38.9	63.0	267	36.6	56.2	210	4.6	0.126
Somali literacy	46.1	71.9	267	41.7	68.8	210	-1.4	0.658
English literacy	13.1	25.9	267	7.2	24.6	210	-4.5	0.305
Financial literacy	3.9	27.5	267	5.2	20.7	210	8.1	0.067

For English literacy the findings are largely similar whereby the comparison group has in fact increased their scores slightly more than those in the intervention schools. Cohort group scores increased from 13.1 to 25.9 while comparison group saw the change from 7.2 to 24.6 percent on average in the English literacy assessment. In essence, thus, it seems that for both English and Somali literacy, the initial lower levels of literacy in comparison schools have almost disappeared by Midline Round 2. However, these findings are not conclusive as the difference in differences estimates are not statistically significant.[372] Yet, the fact that the difference is in favour of more visible increases in comparison schools, it will be difficult to make the case that the program is positively impacting language learning in its schools.

Meanwhile, with numeracy scores, we see that intervention school students who have remained enrolled throughout the program have increased their average score from 38.9 percent to 63.0 percent. Meanwhile in comparison schools the increase for those similarly enrolled between Baseline and Midline Round 2 has been from 36.6 percent to 56.2 percent. This results in DID of 4.6 in favour of intervention schools. This estimate is somewhat closer to statistical significance when compared to the other learning differences analysed above.[373]

Meanwhile, as has been the case with other sub-samples of the data that we have looked at thus far, the most positive findings can be observed when looking at the development of financial literacy scores. The increases

[372] It should be noted that with a relative small sample size, as in this case, the statistical significance of a measure is more difficult to achieve.

[373] p=0.126
Many of the sub-groups that will be analysed in the following section are in fact showing positive changes in the intervention group, and likely the development in the numeracy learning between the two Midlines is much more positive than between Baseline and Midline Round 1. As such, it seems that the program is starting to have an effect. We expect this effect to become significant going into Endline.

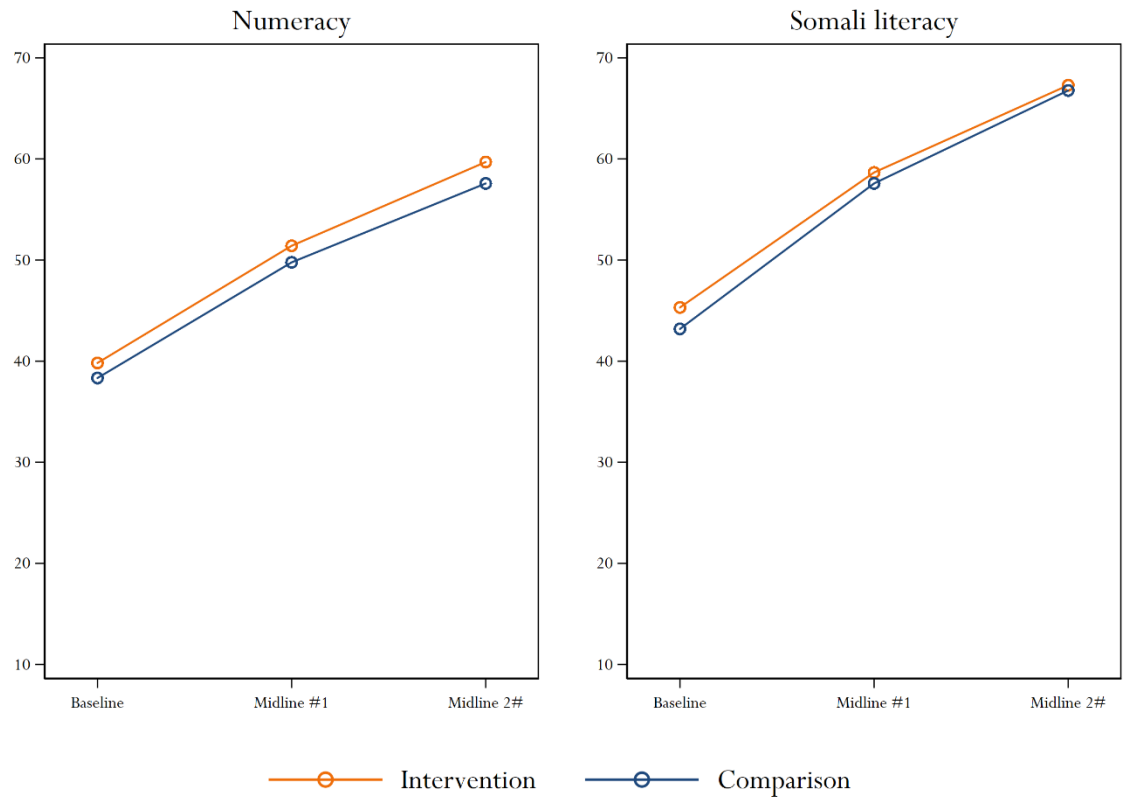
among the intervention cohort are on average 8.1 percentage points higher. The difference in differences estimate is marginally significant (p=0.067).

As a robustness check, we also conducted the analysis using a panel of respondents who were enrolled at Baseline and Midline Round 2. This sample consists of 642 girls, 373 in treatment and 269 in the comparison group. The findings are very much in line with what was presented above.

Cross-Sectional Analysis - Including Midline Round 1

As with the cross-sectional panel used in the second part of this learning section, we then looked at the changes over the three moments of evaluation, this time using a panel that includes replacement girls for those cohort girls who could not be contacted again in the waves of data collection after the Baseline. This panel, again, excludes the out-of-school cohort by virtue of including the Midline Round 1 results. As is the case with cross-sectional analysis compared with using a pure longitudinal sample, some statistical power is gained by increasing the sample by including the substitutions. However, some is conversely lost as the panels are not identical over time as individuals have been replaced.

FIGURE 59: AVERAGE LEARNING SCORES ACROSS ALL ROUNDS, CROSS-SECTIONAL SAMPLE



The sample is 798 individuals, which is somewhat larger than the true panel. Out of these respondents, 429 are in intervention schools and 369 in comparison ones.

TABLE 148: DIFFERENCE-IN-DIFFERENCES ACROSS ALL ROUNDS, CROSS-SECTIONAL SAMPLE

	Intervention schools				Comparison schools				Difference in Differences		
	BL	**ML1**	**ML2**	**n**	**BL**	**ML1**	**ML2**	**n**	**BL - ML2**	**ML1 - ML2**	**p**
Numeracy	39.8	51.4	59.7	429	38.3	49.8	57.6	369	0.5	0.6	0.857
Somali literacy	45.3	58.7	67.3	429	43.2	57.6	66.8	369	-0.6	-1.6	0.605

The results when looking at the cross-section of in-school girls over the three waves are very similar for both intervention and comparison cohorts. In both cohorts the results have improved in a very similar fashion, resulting in a DID estimate of 0.5 and 0.6 for numeracy between Baseline and Midline Round 2 and between Midline Round 1 and Round 2, respectively. For Somali literacy, again, the comparison group in fact fared better since baseline. Yet both of the estimates lack statistical significance and can thus be thought of as indicative rather than anything else.

Naturally, the sample size being relatively small – comparison group, for example has only 369 individuals – this panel is much more vulnerable to potential skew deriving from substitutions of the cohort girls and other variations.

Out-of-School Girls

As a final panel for aggregate learning outcomes we looked at the out-of-school girl cohort. The focus here is on girls in the cohort who were not enrolled when the baseline was conducted. The analysis here is somewhat different from what has been done above as we focus mainly on the effect on learning once the girl has enrolled at school. This section, essentially, then, captures the program's effect with out-of-school girls who have since enrolled in one of the program schools. It is important in order to assess whether the program is able to help those girls catch up who enrol at an older age.

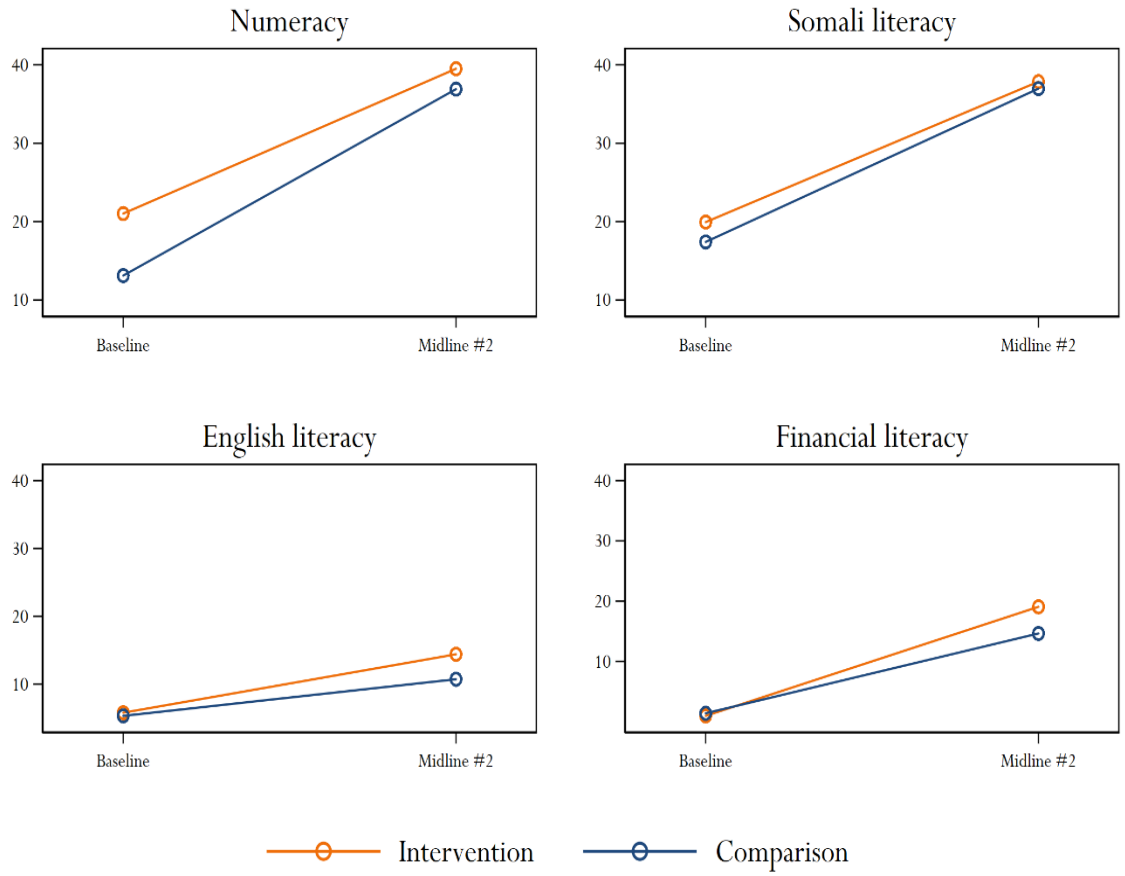
The total sample in this panel is 136 girls who were out of school at baseline and have since enrolled. Out of then sample 71 are in comparison schools, while 65 are in intervention schools.

TABLE 149: PROGRAM IMPACT ON OOS GIRLS WHO ENROLLED POST-BASELINE

	Comparison			**Intervention**			**Difference in differences**	
	BL	**ML2**	**n**	**BL**	**ML2**	**n**	**DID**	**p**
Numeracy	13.1	36.9	71	21.0	39.5	65	-5.3	0.323
Somali literacy	17.4	37.0	71	19.9	37.9	65	-1.6	0.778
English literacy	5.3	10.7	71	5.8	14.4	65	+3.2	0.567
Financial literacy	1.4	14.6	71	1.0	19.1	65	+4.9	0.579

When looking at the table above, no clear program impact can be discerned. For numeracy and Somali literacy, in fact those in comparison schools had improved their learning scores on average more than those in intervention schools. For example, girls in comparison schools raised their numeracy average from 13.1 to 36.9 compared with the intervention schoolgirls whose score went from 21.0 to 39.5. For English and financial literacy, the reverse is true, whereby girls in intervention schools improved their learning assessment scores overall more than those in comparison schools. However, none of the difference in differences estimates are statistically significant, and as such the findings are not conclusive one way or the other.[374] Based on this analysis it cannot be argued, thus, that the program has been especially effective with this group of girls. Yet, as it comes to financial literacy, the intervention girls average 19.1 percent which is almost at the level of the overall sample average of 19.8. As has been shown above, the program tends to have more impact in financial literacy.

FIGURE 60: LEARNING OUTCOMES AMONG OOSG WHO HAVE ENROLLED SINCE BASELINE



[374] The sample sizes, however, are very small and as such not easily conducive to achieving statistical significance.

LEARNING TARGET REGRESSIONS

Section 4.1 of the report provided an assessment of whether the program has met its learning targets, including evaluating progress toward the learning target in the context of linear regression models. These models are difference-in-differences models estimated in a regression framework, which allows for the inclusion of sampling or analytical weights, and additional control variables. In the tables below, we expand on the relatively straightforward models reports in Section 4.1 to include additional control variables (including school and individual fixed effects) and weights that correct for imbalance in the grade levels of intervention and comparison girls at baseline.[375] Each row in the table represents a distinct regression model, though we restrict reporting to the outcome of interest, the program's impact on learning scores in intervention versus comparison schools. We do not report other regression coefficients or standard errors, simply because many of the regression models include dozens or even hundreds of coefficients (in the case of those models using individual-level fixed effects).

The third column describes how the model differs from our standard model, by indicating which control variables, fixed effects, and weights are included in the analysis. Note that the inclusion of time-invariant control variables – those that do not change from round to round – does not have a tangible effect on difference-in-differences estimates of program impact, because the difference-in-differences model already explicitly controls for time-invariant factors that might influence learning scores. Therefore, the main robustness checks in this section are the inclusion of school and individual fixed effects, as well as the use of weights.

TABLE 150: SOMALI LITERACY CHANGE FROM ML1 TO ML2, ROBUSTNESS CHECKS

Result	Details	Comments
Somali Literacy Baseline – Midline # 1 - Midline #2 (ML1 to ML2 coefficients reported)	Beta = 1.06 Standard error = 2.9 Two-tailed p-value = 0.72 Target = 7.3-point increase above comparison group Performance against target = 14.5%	Results based on a model including clustered standard errors at school level, without additional controls, and without weights to adjust for differential sample sizes across grade levels. Sample includes in-school girls tracked from baseline, through ML1, to ML2.
Somali Literacy Baseline – Midline # 1 - Midline #2 (ML1 to ML2 coefficients reported)	Beta = 0.26 Standard error = 2.9 Two-tailed p-value = 0.93	Baseline age and grade level as controls
Somali Literacy Baseline – Midline # 1 - Midline #2 (ML1 to ML2 coefficients reported)	Beta = 0.6 Standard error = 3.0 Two-tailed p-value = 0.84	Baseline age and grade level as controls School fixed effects
Somali Literacy Baseline – Midline # 1 - Midline #2 (ML1 to ML2 coefficients reported)	Beta = 0.55 Standard error = 3.5 Two-tailed p-value = 0.88	Baseline age and grade level as controls Individual fixed effects
Somali Literacy Baseline – Midline # 1 - Midline #2 (ML1 to ML2 coefficients reported)	Beta = 1.52 Standard error = 2.9 Two-tailed p-value = 0.60	Weighted to correct for unbalanced grade level at baseline between intervention and comparison schools

[375] The intervention group had a slightly different balance of grade levels than comparison schools at baseline, though the extent of this imbalance is very small.

Somali Literacy Baseline – Midline # 1 - Midline #2 (ML1 to ML2 coefficients reported)	Beta = 0.79 Standard error = 2.9 Two-tailed p-value = 0.79	Baseline age and grade level as controls Weighted to correct for unbalanced grade level at baseline between intervention and comparison schools
Somali Literacy Baseline – Midline # 1 - Midline #2 (ML1 to ML2 coefficients reported)	Beta = 1.1 Standard error = 3.0 Two-tailed p-value = 0.71	Baseline age and grade level as controls School fixed effects Weighted to correct for unbalanced grade level at baseline between intervention and comparison schools
Somali Literacy Baseline – Midline # 1 - Midline #2 (ML1 to ML2 coefficients reported)	Beta = 1.0 Standard error = 3.5 Two-tailed p-value = 0.77	Baseline age and grade level as controls Individual fixed effects Weighted to correct for unbalanced grade level at baseline between intervention and comparison schools

TABLE 151: NUMERACY CHANGE FROM ML1 TO ML2, ROBUSTNESS CHECKS

Result	Details	Comments
Numeracy Baseline – Midline #1 - Midline #2 (ML1 to ML2 coefficients reported)	Beta = 3.1 Standard error = 2.9 Two-tailed p-value = 0.29 Target = 6.4-point increase above comparison group Performance against target = 48.4%	Results based on a model including clustered standard errors at school level, without additional controls, and without weights to adjust for differential sample sizes across grade levels. Sample includes in-school girls tracked from baseline to ML2, excluding ML1.
Numeracy Baseline – Midline #1 - Midline #2 (ML1 to ML2 coefficients reported)	Beta = 2.9 Standard error = 2.8 Two-tailed p-value = 0.31	Baseline age and grade level as controls
Numeracy Baseline – Midline #1 - Midline #2 (ML1 to ML2 coefficients reported)	Beta = 3.1 Standard error = 2.9 Two-tailed p-value = 0.30	Baseline age and grade level as controls School fixed effects
Numeracy Baseline – Midline #1 - Midline #2 (ML1 to ML2 coefficients reported)	Beta = 3.1 Standard error = 3.5 Two-tailed p-value = 0.38	Baseline age and grade level as controls Individual fixed effects
Numeracy Baseline – Midline #1 - Midline #2 (ML1 to ML2 coefficients reported)	Beta = 4.0 Standard error = 3.0 Two-tailed p-value = 0.18	Weighted to correct for unbalanced grade level at baseline between intervention and comparison schools
Numeracy Baseline – Midline #1 - Midline #2 (ML1 to ML2 coefficients reported)	Beta = 3.8 Standard error = 2.9 Two-tailed p-value = 0.19	Baseline age and grade level as controls Weighted to correct for unbalanced grade level at baseline between intervention and comparison schools
Numeracy Baseline – Midline #1 - Midline #2 (ML1 to ML2 coefficients reported)	Beta = 4.0 Standard error = 3.0 Two-tailed p-value = 0.19	Baseline age and grade level as controls School fixed effects Weighted to correct for unbalanced grade level at baseline between intervention and comparison schools
Numeracy Baseline – Midline #1 - Midline #2 (ML1 to ML2 coefficients reported)	Beta = 4.0 Standard error = 3.5	Baseline age and grade level as controls Individual fixed effects

	Two-tailed p-value = 0.26	Weighted to correct for unbalanced grade level at baseline between intervention and comparison schools

TABLE 152: ENGLISH LITERACY CHANGE FROM BL TO ML2, ROBUSTNESS CHECKS

Result	Details	Comments
English Literacy Baseline - Midline #2	Beta = -0.72 Standard error = 3.8 Two-tailed p-value = 0.85 Target = 8.7-point increase above comparison group Performance against target = -8.3%	Results based on a model including clustered standard errors at school level, without additional controls, and without weights to adjust for differential sample sizes across grade levels. Sample includes in-school girls tracked from baseline to ML2, excluding ML1.
English Literacy Baseline - Midline #2	Beta = -0.71 Standard error = 3.8 Two-tailed p-value = 0.85	Baseline age and grade level as controls
English Literacy Baseline - Midline #2	Beta = -0.64 Standard error = 3.9 Two-tailed p-value = 0.87	Baseline age and grade level as controls School fixed effects
English Literacy Baseline - Midline #2	Beta = -0.71 Standard error = 5.4 Two-tailed p-value = 0.90	Baseline age and grade level as controls Individual fixed effects
English Literacy Baseline - Midline #2	Beta = -0.83 Standard error = 3.82 Two-tailed p-value = 0.83	Weighted to correct for unbalanced grade level at baseline between intervention and comparison schools
English Literacy Baseline - Midline #2	Beta = -0.77 Standard error = 3.85 Two-tailed p-value = 0.84	Baseline age and grade level as controls Weighted to correct for unbalanced grade level at baseline between intervention and comparison schools
English Literacy Baseline - Midline #2	Beta = -0.70 Standard error = 3.92 Two-tailed p-value = 0.86	Baseline age and grade level as controls School fixed effects Weighted to correct for unbalanced grade level at baseline between intervention and comparison schools
English Literacy Baseline - Midline #2	Beta = -0.81 Standard error = 5.4 Two-tailed p-value = 0.88	Baseline age and grade level as controls Individual fixed effects Weighted to correct for unbalanced grade level at baseline between intervention and comparison schools

SUPPLEMENTAL TRANSITION RESULTS

In the main body of the report, the transition analysis uses a split sample of in-school girls and OOS girls to assess transition outcomes. OOS girls in that analysis are those who were tracked from baseline to ML2.

Critically, the decision was made to focus the in-school girls analysis on those in-school girls who were successfully tracked from baseline, through ML1, to ML2 (the so-called "full panel"). This approach produces a smaller set of cohort girls, but three data points for each. In this supplemental analysis, we take a different approach, reporting transition outcomes from baseline to ML2 for the combined sample of OOS girls and in-school girls tracked in those two periods, regardless of their inclusion at ML1.

We identified the panel of girls who were successfully recruited into the cohort at baseline and were successfully re-contacted at ML2. Importantly, this includes some girls who we attempted to contact at ML1 but who fell out of the sample at that time. The panel sample of 1,013 respondents includes 116 such girls, who were either unavailable, refused to participate, or could not be located at ML1, but who were re-introduced into the cohort at ML2. Of course, the sample also includes OOS girls – those girls who were OOS at the time of baseline recruitment – none of whom were sought out at ML1. Using this panel ensures the most rigorous possible comparison between intervention and comparison groups over time, as the set of girls in both rounds is identical.

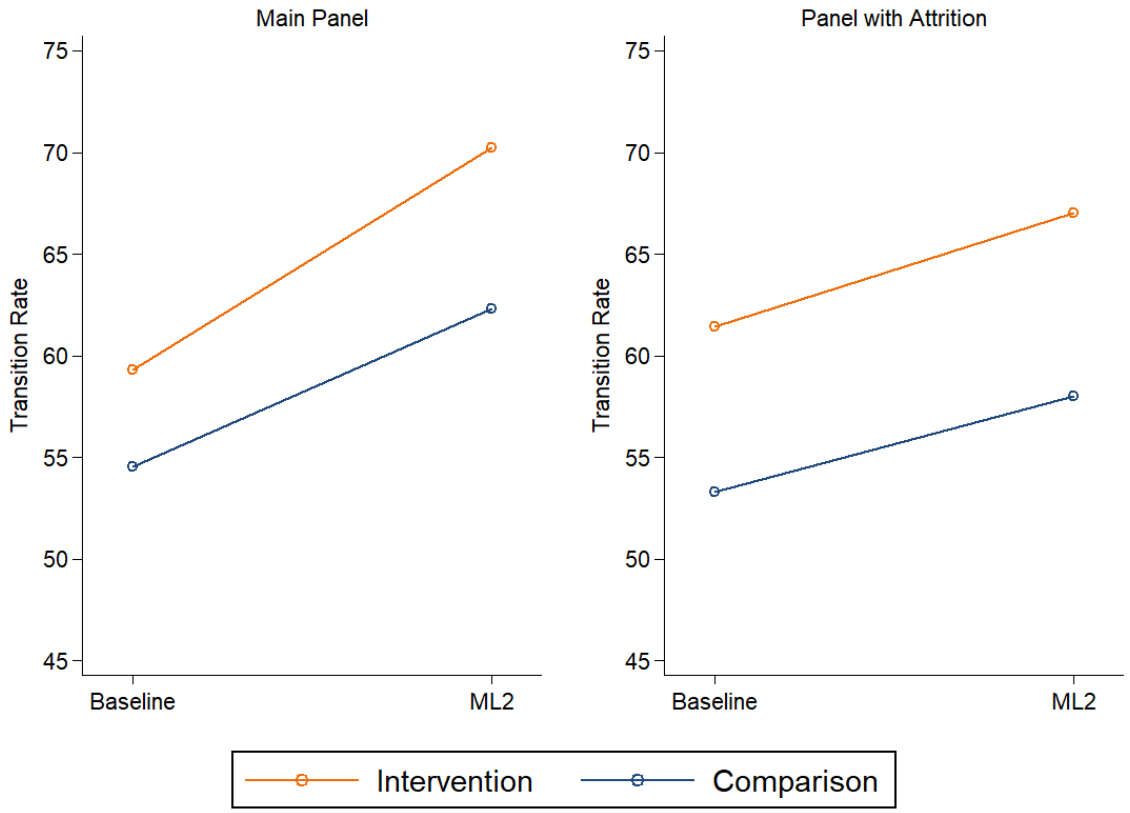
The figure below reports the topline trends in transition rates among the cohort, by intervention status and round. The left panel of the graph utilizes the sample of 1,013 girls mentioned previously. In this analysis, we define transition as a binary outcome, with positive outcomes including all those identified as positive transition pathways in the pathways table above. Specifically, this includes re-enrolment or transition into gainful employment, vocational training, or alternative/informal education for girls who were previously out-of-school, and progression through the grades, or transition into gainful employment for girls who were enrolled previously. Unsuccessful outcomes for girls who were already enrolled include being held back a grade, dropping out of school into non-gainful employment, or dropping out of school into alternative education.[376] For girls who were out-of-school to start, unsuccessful outcomes include remaining out-of-school, or transitioning into either non-gainful or underage employment.

In the left panel, it is clear that transition rates have increased across the board from baseline to ML1, in both types of communities. Importantly, the gains in intervention communities outpace those in comparison communities, represented by the steeper slope of the line for the intervention group. In comparison communities, transition rates rose from 54.5 percent at baseline to 62.3 percent at ML2. As noted above, it is important note to over-interpret this number, as the period over which transition is measured has changed from baseline to ML2. However, we do suspect that this represents a true improvement in comparison community transition rates over time.

The critical finding in the left panel is that transition has improved more in intervention communities. In these areas, transition has improved from a higher starting point (59.3 percent) to 70.3 percent from baseline to ML2, an 11-point gain, compared to a 7.8 point gain in comparison communities.

[376] While enrolling in alternative education is a positive outcome for OOS girls, it is seen as a negative outcome – though by no means is it the most negative possible outcome – for girls who were enrolled in formal primary school previously, as these girls should ideally continue their formal education.

FIGURE 61: TRANSITION RATES OVER TIME, BY INTERVENTION STATS



The right panel of the figure above reports the same analysis using a more expansive sample than the "true panel" reported on thus far. As part of data collection during ML2, we incorporated a module into the household survey that captured information about girls who could not be located or were not available for interviews. The original panel sample analysed on the left includes 1,013 girls re-contacted from baseline, out of a possible 1,581; this more expansive panel sample includes all of the same girls, and 207 additional girls.[377] When a girl could not be located at ML2, enumerators were asked whether they were able to identify anyone in the community who knows the girl and can provide information about her whereabouts and enrolment status. If so, they recorded enrolment information, but no other details about her grade level, employment status, or other transition activities.[378] The right panel of the figure reports transition rates over time using a sample that captures transition information of this kind, incorporating an additional 207 girls in the panel who were recruited at baseline, could not be located at ML2, but whose enrolment status could be determined from a family member or teacher. We limited inclusion to those girls for whom we collected information from a parent, another family member, or a teacher, and excluded those whose information was drawn from a friend or neighbour, for the sake of maintaining reliability.

The coding of transition in this sample is coarser. We do not know, for example, whether girls who are still enrolled progressed to the next grade, or whether girls who are out-of-school have enrolled into alternative

[377] Although the baseline evaluation included more than 1,581 girls, this smaller number represents the set of girls who are in schools that were not dropped from the sample wholesale between the baseline and ML2.

[378] More fine-grained data could be collected, but would be of questionable reliability in many cases, given that our informants were often head teachers or extended family members.

education programs. For the purposes of the analysis, we code *all* girls in the sample using coarser categories, which ensures the girls are comparable to one another but simultaneously obscures nuance in the data. The four outcomes we are able to identify for this group are:

- Enrolled at baseline, remain enrolled at ML2 (positive)
- Enrolled at baseline, dropped out at ML2 (negative)
- Out-of-school at baseline, enrolled at ML2 (positive)
- Out-of-school at baseline, remain out-of-school at ML2 (negative)

Among the girls who fell out of the overall sample but for whom we were able to collect enrolment information (n = 207), the vast majority had either remained out-of-school (68.6 percent) or dropped out of school (23.2). Just 8.2 percent of these girls, combined, remained in school or enrolled in school since baseline.

The findings in this sample are much less positive than in the main sample; while both intervention and comparison communities show improvement over time, the trend lines are very similar, implying that the project did not have any appreciable impact. We emphasize the limitations of this approach, as it can easily obscure important differences between baseline and ML2, especially in the context of girls who remain enrolled in school but who do not advance to the next grade. Additionally, girls who were out-of-school in previous years but have moved into acceptable transition paths – such as non-formal education – are included in the sample, but their transition status is mischaracterized in the data as negative (non-enrolment) rather than positive (enrolment in non-formal education), because the data collected from family members of these girls is insufficiently granular. In general, our preference is to rely on the sample of girls who were successfully re-contacted, primarily because it allows for finer-grained distinctions in transition status to be drawn, which is essential to understanding the program's true impact.[379]

[379] Many of the transition pathways targeted by the program are more nuanced than simply enrolment in formal school. It is also worth noting that the bias from excluding girls who fell out of the sample – the vast majority of whom were out-of-school in previous rounds – is likely less than the bias that stems from coarsening the transition status of all girls in the sample, in an effort to include "attrition girls" also.