Project Evaluation Report

Report title:	Excelling Against the Odds Baseline Evaluation
Evaluator:	Simetrica Ltd
GEC Project:	Excelling Against the Odds
Country	Ethiopia
GEC window	GEC-Transition
Evaluation point:	Baseline
Report date:	April 2019

Notes:

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

Girls' Education Challenge

UKaid from the British people

Contents

Cover sheet

Project: ChildHope

Authors: Christopher Warner, Ali Ladak, Richard Houston, Dora Radosevic, Maria Palma, Daniel Fujiwara

With thanks to ADVA CONSULT and Yared Antonius for quantitative and qualitative data collection and qualitative analysis.

External Evaluator: Simetrica Ltd.

Version number: 2.1

Date: 15/04/19

Executive Summary

Background

This report presents the baseline evaluation of ChildHope's GEC-T project. Among its aims are to: describe the context of the project and the characteristics of its beneficiaries, establish a baseline of comparison for further changes in outcomes, and suggest targets for midline and endline outcomes.

The project is based in Ethiopia in the regional states of Amhara and Oromia. Ethiopia is characterised by its large rural areas, which make providing equitable access to education challenging and contribute to large disparities in enrolment and literacy rates between urban and rural areas. In addition, girls face marginalisation due to the burden of domestic work, poverty, the risk of early marriage and migration, and disability, which may limit their capacity to progress in education. To encourage universal enrolment of pupils, public primary and secondary schools are fee-free, though monetary barriers persist in other forms, for example travel costs, which may also hinder enrolment and transition.

The project seeks to overcome these barriers to learning and transition by providing support through a variety of school-, community- and district-level interventions. The three final outcomes that the project seeks to deliver higher levels of achievement in numeracy and literacy (*Learning*), improved transition of girls to higher grades, tertiary education and/or gainful employment (*Transition*) and increased support for girls' achievement from family, community and government *(Sustainability)*. The project's Theory of Change describes four key intermediate outcomes necessary to deliver these 'final' outcomes: increased school attendance rates, improved quality of teaching, increased self-esteem and the empowerment of girls, and increased skills in entrepreneurship and employability.

The programme's beneficiary population consists of 16,481 girls[1] aged between 7 and 18, who attend school between grades 1 and 12. The evaluation approach draws on five project-level questions intended to quantify the project's impact, investigate which elements of the project intervention are responsible for any effects observed, and understand which subgroups of girls are most effectively reached by the project.

The evaluation utilises a combination of quantitative and qualitative research methods. Quantitative methods are used to measure the project's impact on learning and transition outcomes, in particular a difference in difference analysis which tracks changes in outcomes experienced by girls over time in the intervention group and a comparable control group to serve as a benchmark. This methodology helps ensure that unobserved trends in outcomes over time which are not due to the project (e.g. political instability) or specific group differences (e.g. vulnerability) do not bias the estimated impact of the project[2]. Qualitative key informant interviews and focus group discussions are used to provide information on barriers, enablers, and intermediate outcomes to facilitate deeper insights into what works and why.

Learning Outcomes Findings

To assess progress in literacy and numeracy in grades 4 to 8, an Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA) are used, with a Secondary Grade Reading Assessment (SeGRA) and Secondary Grade Mathematics Assessment (SeGMA) introduced at higher grades. These assessments have been designed to test basic literacy and numeracy skills, identify any gaps in students' knowledge, and calculate a grade level at which students are achieving. Descriptive analysis suggests that there is scope for learning outcomes to improve between evaluation points, as current levels of achievement for project beneficiaries were generally lower than expected, in particular for those in higher grades. Key barriers to learning include: Disability, and living in the Oromia region which has recently experienced political instability

[1] As of March 2019

[2] Assuming the key assumption of parallel trends holds.

Transition Outcomes Findings

For the purposes of the evaluation the key transition point is the move from primary school to post-primary, for which a successful transition includes moving to secondary school, a Technical and Vocational Education and Training (TVET) centre, or fairly paid employment. The evaluation finds a baseline transition rate of 66% for project areas but with considerable variation by group, for example on the basis of region, age. Key barriers to transition include: Being an orphan and finding the journey to school unsafe.

Sustainability Outcomes Findings

For sustainability, the project is awarded an overall score of 1.3 (out of 4) at baseline, indicative of latent changes in the areas of community, school and system. Scores of 1 for community and system level sustainability reflect the relatively low preponderance of households paying the costs associated with schooling (community indicator) and a lack of clarity on whether the engagement of local officials has translated into behavioural changes (system level indicator). School level sustainability is awarded a score of 2, as there is evidence that girls are increasingly enjoying their schooling, that they are treated the same as boys by teachers, and that teachers help students to study effectively after school.

Marginalisation analysis

Our analysis shows that the prevalence of marginalisation in the girls' evaluation sample is broadly aligned with that in beneficiary mapping. The most common characteristics in the intervention group, which may be associated with marginalisation, are living in a household which is: female headed, defined as being poor, or where the head has no education. We may expect girls in poor households (and the subset of these who are 'unable to meet their basic needs', and who 'find it difficult to afford for girl to go to school') to experience worse learning and attendance outcomes if they are required to work to assist their families, or if the costs associated with schooling are unaffordable.

Project girls report having lower potential barriers to learning and transition than girls in the control group: they feel safer, receive more support from their caregivers, have better attendance rates, report better availability and use of the school facilities and have a better opinion of their teachers than do girls in the control group.

Intermediate Outcomes Findings

Our analysis of attendance at baseline shows that 87% of girls attend school most days that their school was open. High levels of attendance are supported by project-level data suggesting rates of as much as 97%, which is also corroborated by evaluation spot checks in classrooms. Qualitative data is less optimistic, as teachers and community representatives suggest that attendance remains a key problem in the educational system.

On teacher quality, 89% of girls consider the quality of teaching they receive to be "good" or "very good". Qualitative data suggests that the ability of teachers to deliver effective teaching may be constrained by inadequate resources and facilities in schools. The evaluation of teacher quality should be supplemented at midline by in-class observations to provide an objective assessment of teaching quality.

Findings from the baseline on self-esteem are mixed and suggest that some girls may be presenting with issues on these measures. Girls presenting with disabilities are particularly likely to show evidence of low self-esteem.

Of the three intermediate outcomes, initial analysis suggests that student-perceived teacher quality is the most closely associated with higher learning outcomes, while the link from attendance and self-esteem to learning outcomes is less clear.

1. Background to project

1.1 Project context[3]

The project is based in the regional states of Amhara and Oromia[4], two of the nine regions which, in addition to two city administrations, make up the Federal Democratic Republic of Ethiopia. Ethiopia is a culturally and linguistically diverse country of around 100 million individuals located in East Africa.

Ethiopia is one of the least urbanised countries in the world, with approximately 80% of the population living in rural areas. The rurality of the Ethiopian population presents a challenge to providing equitable access to education. There is evidence of large disparities between enrolment and literacy rates in urban and rural areas[5]. This is driven by a combination of supply and demand-side factors. On the supply-side, schools may be inaccessible to students living several kilometres or more away. On the demand-side, rural families may be less able to afford the costs associated with schooling[6], and the opportunity cost of a child's time may be higher if she is required to work or carry out household chores. For example, women are often a key source of agricultural labour. The project seeks to overcome these barriers by providing support for transportation and other costs associated with schooling.

In addition to the burden of household chores and poverty, girls may also be marginalised due to the risk of early marriage, violence in the home, migration and disability. Girls from Muslim communities are especially susceptible to societal norms surrounding early marriage even in urban settings. Regional context also plays an important role, given that certain characteristics – which may be associated with additional barriers – are more prevalent. For example, Oromia is approximately 60% Muslim, whereas Amhara is around 30%. Girls in Oromia are likely to have faced additional buffers, particularly in the spring and summer of 2018 due to political unrest. These factors may serve to prevent girls from achieving good learning outcomes and staying on in education. They are likely to be compounded by societal attitudes biased against girls' education. Access to information and knowledge often flows through male household heads, preventing women and girls from making decisions and choices independent of male oversight. The project has a number of interventions targeted at reducing the barriers associated with these factors.

Educational policy in Ethiopia is implemented at the regional level by Educational Bureaus, which are overseen on the federal level by the Ministry of Education. To encourage universal enrolment of pupils, public primary and secondary schools are fee-free, though monetary barriers persist in other forms.

The education and training system, as defined in the Education Sector Development Programme V, has five levels, where key transition points for GEC-T have been identified:

- Key transition point 1: Lower primary (Grades 1-4) to Upper Primary (Grades 5-8)
    - Lower primary typically starts at age 7, and upper primary starts at age 11
- Key transition point 2: Upper primary (Grades 5-8) to Lower Secondary (Grades 9-10)
    - Lower secondary typically starts at age 15
- Key transition point 3: Lower Secondary (Grades 9-10) to Upper Secondary (Grades 11-12)
    - Upper secondary typically starts at age 17
- Key transition point 4: Upper secondary (Grades 11-12) to Higher education
    - University education typically starts at age 19 at bachelor level.
- Key transition point 5: Primary to Non-Formal Vocational or Technical or Employment
- Key transition point 6: Lower Secondary to Formal Vocational or Technical or Employment

As with many industrialising developing countries, the structure of the Ethiopian economy is undergoing rapid change, with the share of services and industry in total output growing in importance. To meet the concurrent increase in demand for

[3] This subsection draws on the Ministry of Education's report on the 'Education Sector Development Programme V (ESDP V)' & a World Bank country study on 'Education in Ethiopia' <Last accessed 30/08/18>

[4] More precisely, the project is based in the zones of South Wollo and South Gondar (in the region of Amhara) and Arsi (in the region of Oromia)

[5] Education in Ethiopia (Table 4.2)

[6] Ethiopian schools do not charge tuition for attendance. The costs associated with schooling may nevertheless present a barrier to attending. These include: uniforms, transportation, registration/examination fees and scholastic materials

skilled labour, the Ethiopian government established Technical and Vocational Education and Training (TVET) programmes[7] in all Woredas[8] for grade 10 children who have completed the first cycle of secondary school. ChildHope's project design reflects the changing nature of the Ethiopian economy by also supporting TVETs and encouraging girls to attend vocational schools.

The language of instruction (LoI) in Ethiopian schools is an additional key contextual factor. It switched in 1994 from a mixture of Amharic and English to one of mother tongue instruction in lower grades and instruction in English at higher grades. Thus, the project schools use Amharic in the regional state of Amhara and Afaan Oromo in Oromia at lower grades, and English at higher grades. The grade at which the language of instruction switches differs in Amhara and Oromia. In Amhara, the switch occurs at grade 7 (final grade of primary), whereas girls in Oromia do not learn in English until grade 9 (second grade of secondary). The reasoning for this is unclear but may be because children in Oromia are required to learn Amharic in addition to English as it is the official working language of the federal government. The change in language of instruction to English is potentially problematic as teachers may lack adequate English skills. For this reason, the project's interventions include training for teachers on effective bilingual pedagogy.

1.2 Project Theory of Change and assumptions

Through the provision of a variety of school-, community- and district-level interventions, the project seeks to yield three 'final' outcomes:

1. Higher levels of achievement in numeracy and literacy (**learning**)
2. Improved transition of girls (**transition)**
3. Increased support for girls' achievement and continuity from family and government (**sustainability**)

These 'final' outcomes are supported by the delivery of four key intermediate outcomes necessary to deliver the 'final' outcomes as part of its Theory of Change:

1. Improved school attendance rates
2. Improved quality of teaching
3. Increased self-esteem and empowerment of girls
4. Increased girls' entrepreneurship and employability skills[9]

The Theory of Change posits that there is an interaction within both intermediate and final outcomes. For example, it is likely that increased self-esteem will result in girls being more employable and entrepreneurial. Similarly, better achievements in numeracy and literacy will mean girls are more likely to transition successfully, as their returns to remaining in education are improved.

The intermediate outcomes in turn rely on a number of direct outputs stemming from the school-based activities:

- To increase school attendance - the project aims to reduce the economic and psychological costs of attending school for girls and their families by contributing to tuition fees and providing guidance on transition. Learning environments are also made more stimulating through the creation of reading corners, which are also intended to increase school attendance. Sanitary corners have been set up and sanitary pads provided so girls can continue to attend school during menstruation, which is believed to be a major barrier to girls attending school. This has a knock-on effect on anxiety which is also linked with poor performance, discouragement, and drop out.
- To improve the quality of teaching through supporting a planned figure of 900 teachers providing them training, mentoring and coaching in subject-specific pedagogy for literacy and numeracy. Changes in their practice will be tracked through classroom observation, reflective practice meetings and one-on-one supportive supervision. The

[7] See Krishnan & Shaorshadze (2013) for a discussion of TVETs in Ethiopia
[8] Woredas, or districts, are third-level administrative divisions below zones, which are, in turn, below regions
[9] Note that this intermediate outcome is not included as part of the MEL.

project has also established teacher resource centres at Amhara and Oromia Regional and Woreda educational offices and offers mentoring to teachers to assist them in providing effective bilingual pedagogy in order to increase the ownership of change brought about by the project. As a result, it is intended that teachers will develop and apply teaching methodologies more effectively, with good quality teaching materials at their disposal.

To increase the self-esteem and empowerment of girls - the project aims to challenge inequitable gender norms and develop new shared beliefs among both boys and girls through the girls and good brother clubs, letter link boxes and sanitary corners to create the spaces to foster these developments to empower girls to recognise and report violence and abuse.

To assist in realising the project's Theory of Change, the project seeks to overcome a number of barriers to education. On the supply-side, girls may have to travel long distances to secondary schools, where teaching, and the resources used to teach, are of poor quality. Poor teacher quality may stem from a lack of institutional capacity at Woreda level to support teachers' professional development and to provide effective teaching resources. On the demand-side, cultural biases negatively shape attitudes towards girls' education. As a result, girls may have low levels of aspiration and also face the risk of early marriage or migration. Moreover, girls' families often cannot reliably meet the costs of education.

The Theory of Change rests on several key assumptions or 'enablers':

1. The provision of high-quality math and literacy teacher training and mentoring should result in an improvement in learning outcomes and will also address the problems associated with English as the mode of delivery in secondary schools. Good teaching and learning materials and resources will also support differentiated learning at upper primary and secondary
2. Woreda education departments must increase their capacity to improve teacher development to ensure sustainability of outcomes, and the project must maintain strong working relationships with these departments to harness and mainstream GEC1 innovations
3. Girls targeted by the project must feel a sense of confidence and positivity about themselves and their experience of learning.

Table 1: Project design and intervention

Intervention types	What is the intervention?	What Intermediate Outcome will the intervention contribute to and how?	How will the intervention contribute to achieving the learning, transition and sustainability outcomes?
Output 1			
Letter link boxes	Letter link boxes are set up primarily to deal with safeguarding incidents. Girls can post incidents, written on a piece of paper, into secured boxes that are emptied on a weekly basis by trained focal teachers	IO1 & indirectly IO3. Providing a safe environment enables girls to attend school. Installation of letter link boxes enables girls to report safeguarding issues and Girls' Clubs provide safe space zones where they can discuss the challenges of inequitable gender norms and develop new shared beliefs among both boys and girls through the girls and good brother clubs	**Outcome 1** By installing letter link boxes attendance will improve as the girls will feel safe(r) within the school grounds. This will have a positive impact on learning, as they will be attending school more regularly, and they will be able to concentrate on learning without the worry or distraction of harm, either from boys within the school or outside of the school. **Outcome 2**

			Letter link boxes reassure girls safety to attend school, particularly in the most vulnerable of transition points (year 7-year 8) **Outcome 3 – community** The letter link box intervention will more regularly involve the Community Care Coalition (CCC) (in cases where referrals are made to the CCC via Kebele administration for immediate action) which supports them to take ownership of safeguarding cases. The CCCs have been selected as stakeholders within the reporting procedure as they are permanent in nature and are set up across government administrative structures down to Kebele level. This will lead to sustaining services for the intended groups in the long run. The letter link box intervention will also strengthen the PTAs (Parent and Teacher Associations) which could see their involvement if a report of harm is received by any of them. The PTAs will be supported under the auspices of the girls' clubs and will promote some of the new social

			norms created by the girls' clubs.
Sanitary corners & provision of sanitary pads	Girls can access the sanitary corner at times of menstruation so they can take rest and feel safe. Sanitary towels can also be taken from the sanitary corner.	IO1 - providing safe spaces enables girls to attend school knowing that if they need private space within the school grounds they can have this within the sanitary corner. This has a direct link to attendance. There is also an indirect link to IO3 as in some schools there have been reports of some girls raising money to buy sanitary towels for school stock through awareness raising activities in the community.	**Outcome 1** By setting up sanitary corners attendance will improve as the girls will be able to attend more regularly throughout the month. This will have a positive impact on learning as they are attending school more regularly.
Set up & furnish reading corners and the purchase of books for reading corners and libraries	Girls can access the reading corners to read and borrow books. Teachers recommend the books they can read in order to better their learning outcomes	IO 3 – indirectly Through improved learning outcomes (supported by access to reading corners and libraries) girls will feel more confident in their abilities to succeed. Taking the initiative to borrow books from the reading corners and libraries is also a sign that girls are carrying out the 'power to act' / agency / learning power, which is a key determiner of self-efficacy, showing the ability to act and exercise a degree of control over their environment and social structures	**Outcome 1** By having access to books girls can improve their learning outcomes in literacy and numeracy **Outcome 2** By having improved learning outcomes girls are more able to transit through each transition stage more easily. Transition to secondary school and TVET (key transition points 2 and 6) will be strengthened by the reading corners and libraries installed in secondary schools. Also transition to the upper grades will be further reinforced through access to libraries
Train CCCs in safeguarding	The training of CCCs is designed to support the Letter link box intervention. The referral of cases to CCCs will be handled within the CH-CHADET's safeguarding	IO1 - Training the CCCs enables a safe environment that enables girls to attend school.	**Outcome 1** By training the CCCs the reporting of cases identified through the letter link box intervention will have a positive impact on attendance as

	guidelines, so that incidents of harm can be dealt with quickly and efficiently.		the girls will feel safe(r) within the school grounds. This will have a positive impact on learning, as they will be attending school more regularly, and they will be able to concentrate on learning without the worry or distraction of harm, either from boys within the school or outside of the school. **Outcome 2** By following through referrals within the guidelines, girls are reassured to attend school safely, particularly in the most vulnerable of transition points (year 7-year 8) **Outcome 3 – community** The letter link box intervention will involve more regularly CCC (in cases where referrals are made to the CCC via Kebele administration for immediate action) which supports them to take ownership of safeguarding cases. Training of the CCCs is therefore required. The CCCs have been selected as stakeholders within the reporting procedure as they are permanent in nature and are set up across government administrative structures down to Kebele level. This will lead to sustaining services for the intended groups in the long run.
Training school counsellors, principals, focal teachers and	School counsellors, principles, focal teachers and education bureau officials are trained in	IO1 – this directly links to attendance by offering guidance to address any worries in relation to	Outcome 1 Where girls are guided through transition stages

education bureau officials in SRH	supporting sexual and reproductive health in girls	SRH, giving SRH information and services to prevent STIs and unwanted pregnancies, while working to break down the barriers faced by pregnant school girls and young mothers. The girls are guided through key transition stages which improves student retention rates.	they are more able to stay in education and achieve their potential.
Tracking and helping girls to resume school	Through working with community workers, key community members and families, the project supports teams to track girls that are at risk of drop-out in order to resume learning.	IO1 – this directly links to attendance. The project has provided a definition of drop out so that it can identify girls at risk of drop out and take relevant measures to prevent drop out.	Outcome 1 Where there is more regular attendance, learning outcomes will improve.
KPO education team trained and mentored in reading strategies	The KPO education team are trained and mentored in improving girls' reading strategies. They are trained to assess the girls' reading progress and evaluate on effort and improvement rather than achievement status. They are also encouraged to employ a universal marking code where class teachers across all subjects use the same marking style which reinforces good writing.	This directly links to IO 2 by strengthening reading strategies through training, new skills and knowledge that contributes to the improvement of mastery in reading and teaching methodologies.	Outcome 1 Improved teacher pedagogy contributes to improved learning outcomes
Set up ICT labs in primary & secondary schools	ICT labs are set up to support girls in developing their literacy and numeracy skills	IO3	Outcome 1 Improved subject knowledge and mastery contributes to improved learning outcomes
Provide assistive devices to girls with disabilities (GWD)	GWD are provided with assistive devices, e.g., glasses and braille kits	IO1 – directly links to attendance as GWD will feel enabled to better participate in class.	Outcome 1 – increased attendance and participation will result in improved learning for GWD. Outcome 2 – increased attendance and improved learning will result in higher transition rates for GWD.
Output 2			

Secondary School Registration fees paid	Girls are supported by the cost of tuition being covered by the project	IO1 - by being able to access learning, attendance rates will be improved. By supporting the girls with fees, the barriers to transition are removed. Secondary school is an enabler of transition into upper secondary, higher education or TVET. This point of transition will support the progression to appropriate courses, settle into college, work or university life, and succeed as higher education learners or business start-ups.	Outcome 1, outcome 2 Where attendance is improved the girls are able to improve their learning outcomes. Transition rates will improve where girls have been able to attend school (due to school fees being paid)
Accommodation fees paid	Girls are supported by the cost of accommodation being covered by the project	IO1 - by being able to access learning (by living nearby the secondary school) attendance rates will be improved. By supporting the girls with fees, the barriers to transition are removed. Secondary school is an enabler of transition into upper secondary, higher education or TVET. This point of transition will support the progression to appropriate courses, as well as settling into college, work or university life, and allow them to succeed as higher education learners or business start-ups.	Outcome 1, outcome 2 Where attendance is improved the girls are able to improve their learning outcomes. Transition rates will improve where girls have been able to access school (due to living nearby).
(Access) transport to Sexual Health and Reproduction Services (SHRS)	Girls are supported by the cost of transport to SHRS being covered by the project	IO1 – this directly links to attendance. By being able to access SHRS, giving SHR girls are able to learn more about how to prevent STIs and	Outcome 1 Where girls are guided through transition stages they are more able to stay

		unwanted pregnancies. They are also guided through key transition stages which improves student retention rates.	in education and achieve their potential.
School uniform provided - primary and secondary	Girls are supported by the cost of school uniform being covered by the project	IO 1 By lifting the additional education costs girls are able to attend school	Outcome 1 By lifting the additional education costs girls are able to attend school and work towards improving their learning outcomes
Books/scholastic materials provided	Girls are supported by the cost of scholastic material being covered by the project	IO 1 By lifting the additional education costs girls are able to attend school	Outcome 1 By lifting the additional education costs girls are able to attend school and work towards improving their learning outcomes
Support secondary schools to organize summer transition camps	Girls are supported in the transition phase from primary to secondary through the attendance of summer secondary school preparation camps	IO1, IO3	Outcome 2, supports the girls during transition from primary to secondary, contributing to an increase in transition rates.
Reward high performing girls	High performing girls are rewarded as an incentive	IO3 Girls are empowered to succeed through recognition for hard work and success. Girls with disabilities are included in the award scheme to embrace a culture of diversity.	**Outcome 1** By supporting success, girls are more motivated to achieve.
		IO 3 – indirectly Through improved learning outcomes (supported by attendance to homework clubs and subsequent progress in literacy and maths) girls will feel more confident in their ability to succeed.	**Outcome 1** By progressing in literacy and maths, girls can improve their learning outcomes **Outcome 2** By having improved learning outcomes, girls are more able to transit through each transition stage more easily. Transition to secondary school and TVET (key transition point 2) will be strengthened by progression in literacy and math.

Provide incentives for homework club teachers	Teachers are given financial incentives for delivering math and literacy homework clubs	**IO 3 – indirectly** Teachers will have more incentive to develop girls' knowledge, mastery and skills in literacy and maths which will have an impact on the girls' confidence in their ability to succeed	**Outcome 1** By progressing in literacy and maths, girls can improve their learning outcomes **Outcome 2** By having improved learning outcomes, girls are more able to transit through each transition stage more easily. Transition to secondary school and TVET (key transition point 2) will be strengthened by progression in literacy and maths.
Output 3			
Cluster supervisors are trained to carry out lesson observations & mentor teachers	Cluster supervisors at Woreda level provide teacher supervision and school improvement plans which support the zonal education office (Quality Improvement Units).	IO2 Through supporting the Zonal level education departments to build capacity with teachers that is directly linked to practice, change is ongoing and school owned.	Outcome 1 Learning outcomes are improved where teacher quality is improved. The work that cluster supervisors carry out with teachers is intended to develop not only teaching strategies and subject knowledge of teachers, but also to shift the mindset of the teachers to embrace a student-centered approach to learning. Outcome 3 – school (teacher quality) - education departments develop stronger capacity to implement the most vital elements of teacher development and tracking of learning outcomes. Through training government staff and sharing technical and human resources with

			education departments capacity can be developed from within school and local system level.
Master trainers are trained and also carry out cascade training	Training for Primary teacher trainers (ToT) concentrates on specific pedagogies around teaching literacy and maths, taking place over a 3 to 5 day period using education officials and specialists. This is then cascaded to all GEC-T Primary teachers with follow up refresher training taking place within one year (to mitigate loss of knowledge due to teacher attrition). Secondary ToT concentrates on specific literacy and maths pedagogy associated with the secondary school curriculum and is subsequently cascaded.	IO2 - by increasing the percentage of teachers who improve their teaching through improved pedagogies and methodologies. Indirectly will support IO1 and IO3. IO1 - Evidence suggests that increasing in-year achievement and the associated rise in self-esteem and efficacy through better pedagogies will increase attendance; by developing an instructional model that focuses on high-quality teaching and learning, and classes that are stimulating and organised, Student engagement will be supported and attendance raised. In addition, developing class and in-class group structures that enable increased connectedness to individual teachers and peers, will further increase attendance IO3 - Better pedagogic techniques and understanding will raise the ability of the girls within the project to be successful within the maths and literacy setting, which leads to	Outcome 1 & Outcome 3

		greater self-esteem and self-efficacy relating to academic achievement (Chan, 2000)	
Trained teachers are mentored by supervisors	GEC-T teachers are coached and mentored as part of ongoing professional development in order to improve the standard of lesson delivery. Trained secondary school teachers will act as coaches through carrying out lesson observations and coaching activities after each lesson observation. Internal coaching/mentoring communities of practice are vehicle through which coaching and mentoring will happen, where peer observations will take place on a continuous basis.	Will directly support IO2 by increasing the percentage of teachers who improve their teaching through improved pedagogies and methodologies Will support the teacher training intervention (see below) and will support and maintain better pedagogic activities from the teachers. Indirectly will support IO1 and IO3 The Coaching and Mentoring strategy is intrinsically inked to the training	Outcome 1, Outcome 3
Offline digital literacy and numeracy resources are installed in ICT labs	Literacy and numeracy digital resources are installed in ICT labs in secondary schools so that girls can have self access to math and literacy and reinforce their skills.	**IO 2**	Outcome 1, outcome 3 Improved subject knowledge and mastery contributes to improved learning outcomes
Run maths and literacy homework clubs to strengthen the government tutorial programme	Homework clubs are run in literacy and maths to strengthen	This indirectly links to IO 3, Girls are able to concentrate on mastery in each subject which boosts confidence and drives the girls'	**Outcome 1** Improved teacher pedagogy contributes to improved learning outcomes

		perception of their ability to succeed academically.	
Output 4			
Strengthen the Girls' and Good Brothers' clubs	The Girls' and Good Brothers' clubs take place, led by trained focal teachers includes creating space for both girls and boys to discuss some of the persistent issues that girls face, which have emerged from learning to date – including sexual harassment and sexualisation of girls, boys' resentment, the persistent burden of heavy domestic labour, hygiene, poverty and sexual and reproductive health, and the educational challenges faced by children with disabilities. Girls are also upskilled to deal with early marriage proposals, sexual violence and gender-related incidents. Child rights and safeguarding are also covered.	IO 3 Raised self-esteem and self-efficacy empower girls to negotiate their journey and transition through different levels at school.	Links to outcome 2 Links to outcome 1 – There is a link between self-efficacy and achievement as self-efficacy contributes toward academic performance. How capable you believe you are is a determining factor to success. By challenging inequalities girls are given the opportunity to succeed in learning. The inequalities that exist between girls and boys affect their educational opportunities and outcomes at every level. The traditional expectations and norms around girls' choices and behaviour determine whether or not girls are even able to access the classroom in the first place and limit the time they are able to dedicate to learning. Their level of participation and confidence they have in learning can also be undermined. The norms and expected behaviours also influence the perceived value of girls' education among others. As girls get older, the gendered norms that they are under pressure to conform to become more pronounced and the opportunities they have to learn often diminish.
Perform theatre and drama to change social norms	Theatre and drama performances showcase in different schools to raise awareness in	IO 3	Outcome 3

	inequalities related to gender and disabilities		
Produce advocacy and communications materials	Advocacy and communications materials are produced by Girls' Clubs addressing negative attitudes and behaviours towards girls' education and disability	IO1 - advocacy and communications materials encourage families to send girls to school so boosting attendance. IO3 – Girls feel empowered by seeing their messages around the community and the impact the messages have on attendance.	Indirectly outcome 1
Train peer leaders in life skills peer education	Peer leaders in Girls' Clubs trained as peer educators in life skills for other girls	IO3 – Girls empowered through training in new skills and subsequently empower other girls through peer education	Outcome 1 – increase girls learning
Output 5			
Support enrolment, registration fee, tuition fees of girls into vocational school		IO1, IO3 By supporting the girls with fees the barriers to transition are removed. TVET is an enabler of transition into paid work and transition into adulthood.	
Cover the cost of accommodation		IO 3 By supporting the girls with fees, the barriers to transition are removed. TVET is an enabler of transition into working life.	
Cover the cost of education supplies		IO 3 By covering the cost of education supplies, the barriers to transition are removed. TVET is an enabler of transition into working life.	
Provide training on employability soft skills including ICT, self-introduction, CV preparation & 'softer' skills - interviewing, work place conduct & ethical behaviour		IO 3 Through providing training on employability, the girls are supported in the transition from learning into working life.	Outcome 2 Key transition point 5: Primary to Non-Formal Vocational or Technical or Employment Key transition point 6: Lower Secondary to Formal Vocational or

			Technical or Employment: Girls will be provided high quality information, advice and guidance on choice of courses, and options for part-time and short-term experience. The project will make stronger linkages with markets and private sector. Girls will be provided information on opportunities for start-up support.
Link girls with employment referral service for private sector			
Cover the cost of COC for girls who complete TVET			
(Access) transport to TVET	The cost of transport to school is covered by the project	IO1 - by being able to access learning attendance rates will be improved. By supporting the girls with transport fees the barriers to transition are removed. TVET is an enabler of transition into paid work and transition into adulthood.	Outcome 1, outcome 2 Where attendance is improved the girls are able to improve their learning outcomes. Transition rates will improve where girls have been able to access school.

1.3 Target beneficiary groups and beneficiary numbers

Box 1: Project's contribution

1. *Describe the project's primary target groups in terms of age range, grades, country/region, characteristics, and expected exposure to interventions over the course of the project.*

Main target groups: Girls
Age range: 7 – 18 years
Grades: Grade 1 – Grade 12
Regions: Amhara & Oromia
Characteristics: Girls at risk of early marriage, risky migration, domestic work. Disabled girls, girls living in poverty

2. *Provide the target number of girls' beneficiaries (direct learning and transition beneficiaries) and the monitoring data that support this number (for example, in-school population numbers, number of schools, number of communities etc.). Describe the method for calculating the number, any assumptions made.*

Target number: 16,481
Monitoring data: In school population numbers per grade, per school collected via in project database
Method of calculation: Database. Regular attendance collected monthly by home room teacher, passed to the principle and reported to the cluster supervisor who report to both the CCC and CHADET local office (where drop out is persistent) and the CHADET local office for regular attendance. Collection of attendance data is spot checked bi-annually by the KMO, cluster advisor or community worker.

---

3. *Present and justify any difference with respect to GEC1, with GEC-T proposals and/or the MEL framework:*

GEC	*GEC*	*GEC-T*	*GEC-T*
Primary Schools Targeted marginalised girls in primary schools	✔	✔	**Primary, Secondary schools & TVET** Targets Primary schools, secondary schools and TVET
Secondary Schools Targeted marginalised girls in secondary schools	✘	✔	**Primary, Secondary schools & TVET** Targets Primary schools, secondary schools and TVET
Justification for differences stated below (in relation to community conversations) We see a difference between community focus and school focus from GEC and GEC-T as, based on learning from GEC there has been a transition from community focus to school focus. For example, in GEC there was more of a structured focus on i.e. religious leaders to be included directly into interventions. In GEC-T they are actively present in the CCCs and do not need to be targeted as an intervention. There is more focus on school agents of change in GEC-T than in GEC, i.e. focal teachers. The original design of GEC was community based and GEC-T more school based.			
Output 1 **Community Conversations** took place in central places within the community to mobilise communities to debate some of the most pertinent issues that affect them.	✔	✘	**Community conversations,** which are a structured community engagement process across Ethiopia are not part of GEC-T as an intervention but other types of community engagement take place. Not included in GEC-T, however there is an assumption that the girls' movement will continue driving the change of social norms into the community. It is possible for the girls from the girls' movement to carry forward community conversations in place of the original volunteers in GEC, on a smaller scale. For example, at the end of a drama performance the community are encouraged to discuss their views in relation to the performance they have just seen. Family hubs, which were part of GEC toward the end, and were not built into the design of GEC-T have been considered to be re-introduced.
CC (community conversation) agents were selected to become champions for girls' education and duty bearers including teachers, education officials and female mentors.	✔	✘	This was not targeted as an intervention in GEC-T but happens in many of the schools. Through the boys'/good brothers' clubs many of the boys escort the girls home or have their own social network that feeds back information that prevents early marriage.
Men and boys were also selected as PEER researchers to carry out conversational interviews with others in their social network on issues of girls' education	✔	✘	Not included in GEC-T
Religious leaders were also targeted in relation to the role they play in sanctioning early marriage.	✔	✘	Not included in GEC-T (see justification above)

Male bus station workers were also identified to learn about the issues of child protection and play a critical role in identifying migrant children and street and working girls.	✔	✕	Not included in GEC-T (see justification above)
Good Brothers' clubs – Output 1 Boys were then mentored as 'good-brothers' and tasked to influence at least two peers and two male adults each month as part of the extension of the community conversations. They were also trained to communicate through puppetry and drama.	✔	✕	**Boys Clubs** The Good Brothers' clubs still continue to run but it has not been targeted as a main intervention in GEC-T. (links to sustainability) It's assumed that the impact of the boys' clubs in GEC will continue through to the end of GEC-T. Currently boys collaborate with the girls to perform drama within the school and community to raise awareness on the value of girls' education, street involvement and early marriage.
Girls Clubs and life skills club	✔	✔	**Output 4** - Girls Clubs and the Girls' Movement will continue as a main intervention. This covers life skills and links directly to conversations that take place in sanitary corners. The Girls' movement focuses on developing media messages and campaigning to reach out to key influencers, women's groups and civil society.
This was intended in GEC but did not happen. It was planned that recorded conversations between girls and bus drivers would take place.	✕	✔	**Branded messaging, Media & Communications work – output 4** The girls' movement has grown from the girls' clubs and community conversations. Part of GEC-T plan is for the girls to develop their advocacy and campaigning skills and refine their messaging to a wider community.
Girls receive support through transitional shelters who are reintegrated to their homes	✔	✕	Not included in GEC-T This was low uptake and expensive, so not continued.
Output 2 – financial support (livelihood) Families of marginalised girls (with very insecure livelihoods) supported to develop increased and more stable incomes. Families, with one or more girls not attending school, received business training and supported to develop business plans and subsequently offered livelihoods grants/seed money.	✔	✕	Not included because (i) the same families that were transiting through to GEC-T would build on their businesses and (ii) there was an aspiration in the planning phase, for business development initiatives to be put in the hands of the girls rather than their families (links to sustainability). This also links to output 5, TVET, or where girls are drivers of small businesses.
They also received training in skillful parenting.	✔	✕	Not included
Out-of-school girls receive direct and indirect education financial support, once registered as attending school	✔	✔	Same
In-school girls, that are at high risk of dropping out, receive direct and indirect education support to remain in school	✔	✔	**Output 2** - Scholastic materials, accommodation, transport to school, certificate of completion for TVET (output 5)

Justification for differences stated below. **In GEC there was more focus on the creation of physical environment of schools. In GEC-T the physical environment is maintained but not a focus.**			
Output 3 – focus to keep girls in education that delivers a stimulating, participative and safer learning environment	✔	✔	**Reflected in Output 1, output 3 in GEC-T**
Separate toilets built	✔	✔	Maintained
Additional classrooms built	✔	✔	Maintained
Space made available (not monitored)	✔	✔	Reading corners available and monitored and restocked
Sanitary pads provided (inconsistent and not monitored)	✔	✔	Sanitary pads provided, consistent and monitored, restocked consistently
Not included	✘	✔	Sanitary corners built in both primary and secondary schools
Letter link boxes built	✔	✔	Letter link boxes maintained in all primary schools and built for all secondary schools. All schools now have LL boxes
Output 3 – Tutorials offered to targeted girls	✔	✔	**Output 3** – targeted homework support is offered to girls. This is distinct from tutorials which are a government intervention.
Output 4 – Increase the level of motivation and skills among teachers, school principals and inspectors and education officers to deliver and assure good quality transformative education for marginalised girls. The focus was on generic teaching practice.	✔	✔	**Output 3** - Included in GEC-T with a focus on cascade training in pedagogy. Also coaching and developing teacher competencies through communities of practice (COPs). There is also a focus on literacy and numeracy.
Output 4 - Teachers are trained to deliver more participative and friendly lessons	✔	✔	**Output 3 –** Leadership training offered to school principals, secondary school teachers and project officers. In addition to leadership training there is separate training for primary and secondary teachers on more participative and friendly lessons.
Output 4 Gender equity GWD inclusion Safeguarding training	 ✔ ✔ ✔	 ✔ ✔ ✔	**Output 3** Gender equity GWD inclusion Safeguarding training
Output 5 - Forums set up to support retention of girls into good quality education Forums and action planning meetings set up by school and community members to improve on issues identified	✔	✘	Not included as is part of the planning and improvement process. The woreda education offices and CCCs are more supportive without need for intervention (link to sustainability).
	 ✘ ✘	 ✔ ✔	**Output 5 - TVET** Girls enrolled into TVET provision Girls receive training to start their own business

Extra focus in GEC-T
Working with GWD Including focus on teacher training and awareness raising which is monitored through regular lesson observations and fed back into training modules and communities of practice sessions. GWD are also referred for assessment of impairment and provision of assistive devices, such as glasses, materials in braille and wheelchairs
TVET Business start up training and preparation for work sessions are held for girls within the TVET sector
Transition Girls take part in pre- secondary school summer camps plus post summer camp mentoring. Financial support in relation to transition is offered to girls, covering accommodation, transport to school
Referral to (SRHS) Girls are referred to sexual and reproductive health services.
Safeguarding There has been a shift from project structures (GEC committees) to CCCs when dealing with safeguarding cases (links to sustainability).
Staffing There are fewer staff in GEC-T, fewer volunteers. The change in the staffing is that the safeguarding and inclusion people are now not at HQ but at PCOs. There is now a localisation of specialism.

EE Response

The project uses its own monitoring database to calculate the number of beneficiaries reached by its interventions. This represents a physical head count of the number of beneficiaries in project schools. The database is updated on a monthly basis using attendance data from schools which is collected in-class by teachers and checked for accuracy biannually by CHADET staff. Evaluation collected spot checks (see section 5.1) also support the accuracy of this attendance data, as they very closely align with one another. The figures presented by the project in Box 1 and Annex 4 refer to the period of October-December 2018, which coincides with the second round of data collection. From the end of GEC-1 to the beginning of GEC-T, the project estimates an attrition rate of 12% citing staff turnover due to uncertainty surrounding project continuation to GEC-T, which reduced CHADET's ongoing capacity to intervene where attrition was likely to occur or had occurred. As a result of this relatively high rate of attrition, an additional 2,153 replacement girls were selected in October 2018 to bring the total number of girls back to the targeted amount. These girls were not in the sampling framework used to select treatment (and replacement treatment) girls for baseline, so would not have been included as part of the baseline.

The database collects data on individual-level characteristics, affording the project insights into the barriers faced by girls. Given that the data is collected in this way, it would be expected to be a strong representation of the beneficiary population. Indeed, section 3 shows that the distribution of sub-groups in the evaluation sample is strongly aligned with the project's estimates for the beneficiary population. There are some exceptions to this. For example, disability rates in the evaluation sample seem to be slightly higher than in the beneficiary database, though this may be because of differences in definitions.

Taking into account the methodology used to construct the database, and its alignment with the evaluation sample in most dimensions, we would conclude that the project's database is a reliable account of its beneficiaries.

2. Baseline Evaluation Approach and Methodology

2.1 Key evaluation questions & role of the baseline

The evaluation sets out to respond to five project-level questions[10]. These are as follows:

1. To what extent did the project improve literacy and numeracy outcomes, including for subgroups of girls and sub-elements of the treatment (where possible)?
2. To what extent did the project improve transition outcomes, including for subgroups of girls and sub-elements of the treatment (where possible)?
3. To what extent does the analysis undertaken for the evaluation indicate that changes may be sustainable in the longer term?
4. To what extent has the project been associated with improvements in the intended intermediate outcomes? If possible, how and why? What association is observed between intended intermediate outcomes and final outcomes?
5. To what extent did implementation of the project deliver the intended outputs? How and Why?

Quantifying the project's impact on learning and transition will reveal the extent to which the project has altered outcomes for the population of girls in question. Investigating which elements of the treatment were responsible for any effects observed will provide evidence to the project about potential revisions or refocussing of efforts needed. Understanding which subgroups of girls were affected will shed light on whether girls subject to certain forms of marginalisation are (or are not) being reached by the project.

Questions 4 & 5 will help to test whether the project's Theory of Change is robust, help explain any changes observed in the final outcomes, and help identify whether there are other pathways of interest that the project has not yet considered through which changes in final outcomes can be realised.

The evaluation will consist of three waves: baseline, midline and endline. Midline and endline will take place 1 and 2 years after baseline respectively.[11] A cohort of girls from grades 4 to 8 at baseline will be tracked at each wave, in both 'intervention' (project) and 'control' (comparator, non-project) areas, to understand the impact of the project over time on learning and transition.

At baseline, midline and endline, the tracked cohort of girls will take the in-school survey and the learning tests. At the same time the household survey will be administered to their head of household and primary care-giver[12]. At midline and endline the cohort of girls tracked at baseline will diverge into a learning and transition cohort, given that some girls will likely drop out of school and attrite for other reasons.

If a girl is not able to be found at her school at midline, enumerators will seek updated information as to the girl's whereabouts from their teacher and attempt a home visit, during which three scenarios may occur. Firstly, if the girl is still attending school (i.e. her absence was temporary), the learning test will be applied at home. If the girl is able to be found but is no longer attending school, she will be dropped from the learning cohort (but remain in the transition cohort). She will be replaced in the learning cohort if this is necessary subject to matching based on grade, age and (if possible) marginalisation status. If a girl cannot be found at her listed address, she will be replaced in the learning cohort, subject to the same matching criteria. The approach outlined above for midline will be used as well for endline if attrition from midline to endline is unexpectedly large, which might necessitate a cross-sectional approach.

Qualitative interviews and focus group discussions (FGDs) will be carried out in treatment areas in all waves. The baseline evaluation will also establish benchmark transition and learning outcomes for target setting purposes. For transition

[10] The questions outlined are set out in more detail in the MEL.

[11] Endline was initially due to take place 3 years after baseline. Due to delays incurred when carrying out the baseline survey, there will now only be a 2 year gap between baseline and midline.

[12] Where there is no distinction between the head of household and the primary caregiver, the same individual answers both sets of questions.

benchmarking this will be on a separate sample of girls. For benchmarking of learning, it will rely on data collected for the core sample of treated girls, supplemented by an additional sample of girls in grades 9 to 12 who will take the learning tests but will not be tracked over time. The baseline evaluation will also establish benchmark transition and learning outcomes for target setting purposes. For transition benchmarking this will be on a separate sample of girls and for learning benchmarking it will rely on data collected for the core sample of treated girls, supplemented by an additional sample of girls in grades 9 to 12 who will take the learning tests but not be tracked over time.

As part of the evaluation research plan a difference-in-differences design[13] will be used to assess the causal impacts of the project, which requires baseline data be collected to track outcomes over time. If the assumptions of this research design hold, the baseline evaluation will also shed light on the causal pathways between outputs, intermediate outcomes, and outcomes.

Through a mixture of a qualitative and quantitative research, the baseline will also foster a deeper contextual understanding of project beneficiaries, and the barriers to education they face. The mixed-methods research will facilitate an assessment of the project's theory of change, as well as a commentary on the design and effectiveness of each of the project's activities.

The baseline will also allow targets to be set for the outcomes and intermediate outcomes at midline and endline and provide justification for any necessary changes to the Logframe indicators. Lastly, the evaluation will feed into analysis of the GEC-T portfolio as a whole.

2.2 Outcomes and Intermediate Outcomes

The project has three outcomes and three intermediate outcomes necessary to achieve these.

Outcome 1 - 'Learning' is concerned with improved literacy and numeracy skills and targets the number of girls supported by GEC-T who then secure improved learning outcomes. This will be measured by the number of supported girls who secure improved EGRA/SEGRA results in literacy and improved EGMA/SEGMA in numeracy.

Outcome 2 'Transition' targets the improvement of transition rates, where transition is defined as a girl who successfully passes through a key stage of education, training or employment, for example from primary to secondary education or from secondary education to appropriate employment. Data on transition is expected to be captured as part of the forthcoming household survey.

Outcome 3 - 'Sustainability' concerns the sustainability of improvement in learning and transition outcomes. The measurement of sustainability is described in more detail below.

Intermediate Outcome 1 - 'Attendance' will be measured using two indicators. The first is the percentage of girls whose average school attendance improves in the lifetime of the project, which is captured in the household survey, and by a combination of roster and spot checks in a subsample of schools. This will be augmented by attendance data collected in-project, and by the evaluation as part of spot checks and the household survey. The second indicator is qualitative and seeks to gain an understanding of beneficiaries' views on the barriers that prevent them from attending school regularly.

Intermediate Outcome 2 - looks at improving teacher quality, which is captured through the household survey where each girl and their primary caregiver are asked about the quality of teaching the girl receives.

Intermediate Outcome 3 - targets greater self-esteem and empowerment of marginalised girls. The girls' school survey contains two of the indicators used – self-esteem and self-efficacy. The qualitative beneficiary interviews will be used to gauge how girls' perceptions of their ability to succeed academically alter as a result of the project.

Table 2: Outcomes for measurement

Outcome	**Level at which measurement will**	**Tool and mode of data collection, *e.g.***	**Rationale, *i.e. why is this the most***	**Frequency of data**

[13] See here for a short explanation of the difference-in-differences design <Last Accessed 06/03/2019>

	take place, *e.g. household, school, study club etc.*	***HH survey, school based survey, focus group discussions etc.***	***appropriate approach for this outcome***	**collection, *i.e. per evaluation point, annually, per term***
Literacy (*Improved EGRA/SeGRA score)*	*School*	*EGRA/SeGRA*	Directly captures literacy proficiency	*Per evaluation point*
Numeracy (*Improved EGMA/SeGMA score)*	*School*	*EGMA/SeGMA*	Directly captures numeracy proficiency	*Per evaluation point*
Transition (*Improved transition rate)*	*Household*	*HH survey (Benchmark transition at baseline)*	Allows those dropping out of school to be tracked to their home	*Per evaluation point*
Intermediate outcome 1: Attendance	*Household, school*	*HH survey, on-site roster and spot checks, qualitative beneficiary interview*	Gives qualitative views on causes of low attendance as well as quantitative data on the magnitude of non-attendance	*Per evaluation point*
Intermediate outcome 2: *Teacher quality improvement*	*School, household*	*HH Survey*	Gauges opinion of caregivers on teacher quality	*Per evaluation point*
Intermediate outcome 3: *Greater self-esteem and empowerment of marginalized girls*	*School*	*Girls' school survey, qualitative beneficiary interview*	Asks girls directly and in detail how their self-esteem & self-efficacy has changed	*Per evaluation point*

Sustainability

The likely sustainability of changes in learning and transition outcomes is considered to depend on generating sustainable improvements in girls' attendance and the quality of teaching. The sustainability of these factors is in turn captured by a variety of school-, community- and system-level drivers.

At the school level, the sustainability of teacher quality is measured as part of the girls' school survey and from woreda or school administrative records. The girls' school survey asks girls whether teachers ask more questions to boys, girls, or if this is equal, and whether girls receive suggestions from their teachers on how they can continue to study in the evening after school. Administrative records will also be used to understand the percentage of teachers trained by the project who remain in their post.

At the community level, the sustainability of school attendance is proxied by considering: economic sustainability as part of the household survey, which asks girls' households if they pay for school fees, transportation, meals etc. in the current year; and social sustainability as part of the qualitative caregiver interviewers, which gauge whether parents/carers express a stigma of single or divorced women.

At the system level, the sustainability of teacher quality is captured using woreda administrative records on the number of education officials allocated to conduct supporting supervision activities (e.g. teachers' development).

Table 3: Sustainability outcome for measurement

Sustainability Level	Where will measurement take place?	What source of measurement/verification will you use?	Rationale – clarify how you will use your qualitative analysis to support your chosen indicators.	Frequency of data collection
School	*School*	***Girls school survey – 1)*** *% of girls answering "YES" to "Does your teacher(s) ask more questions to boys / girls / equally",* ***2)*** *Change in % of girls answering "Yes" to "Does your teacher(s) suggest ways you can continue to study after school/at home?"*	*Evidence of changing teacher perceptions towards girls will be complemented by data from administrative records on the number of trained teachers retained*	*Per evaluation point*
Community	*Household*	***Household survey*** *- % of girls' households who pay for school fees, transportation to and from school, school meals, school materials and supplies in the current school year*	*HH survey question for economic sustainability will be complemented by qualitative evidence on changing social norms towards single/divorced women*	*Per evaluation point*
System	*Administrative records*	*Number of trained Woreda education officials allocated to conduct supporting supervision activities (i.e. teacher's development and girls' learning assessments).*	*No system-level qualitative interviews*	*Per evaluation point*

2.3 Evaluation methodology

- **What is the overall evaluation design? Randomised Controlled Trial, Quasi-experimental, or Pre-post? Why has this been chosen?**

The evaluation follows a quasi-experimental approach. For quantitative analysis it uses a difference-in-differences (DiD) design. This tracks the change in outcomes experienced by treatment girls over time (i.e. those girls in schools where the projects were active) and subtracts the change over time experienced by control girls (i.e. those girls in schools where the project was not active). Measurement over time was chosen in preference to a comparison of outcomes between treatment and control girls at one point in time, as the project targets the most marginalised girls and failing to take account of this would result in a downward bias in the estimate of the effect of the project. The change over time for treatment girls is compared with the change for control girls to help ensure that unobserved trends in outcomes over time which are not due to the project (e.g. political instability) are removed. For the approach to be fully valid the treatment and control girls will need to have had the same trends in outcomes in the absence of the programme (the 'parallel trends' assumption). For qualitative analysis, the evaluation intends to track a separate group of treatment girls over time, providing information on barriers, enablers, and intermediate outcomes over time and across a range of girls.

- **What are the identified target beneficiary groups? What other indirect beneficiary groups are included in the evaluation (e.g. teachers, parents, government officials, boys, men etc.).**

The project's main target beneficiaries are marginalised girls whose educational potential may be constrained by various barriers. Boys are a secondary target who will benefit from boys' clubs and indirectly through access to reading corners, improved teaching methodology and greater disability awareness; teachers will benefit as a result of training and greater support from Woreda officials concerning their professional development; and parents and siblings will benefit indirectly as a result of the financial resources made available to families to cover the ancillary costs of girls' schooling, for example for transportation, which lessen the draw on family resources as a whole.

- **How is the cohort of girls being evaluated? Are there different cohorts being evaluated separately? Are the cohorts for learning and transition the same or are they different and why? How did you ensure that the target beneficiary groups and subgroups are represented by the sample?**

Quantitative

A single core cohort of girls in grades 4-8 was evaluated at baseline for both learning and transition outcomes. The objective is to track this linked learning / transition cohort over time, where possible, although the learning and transition samples will inevitably diverge somewhat at midline and endline because of differences in the approach to handling attrition in the two cases.

The evaluation selected a stratified random sample of treated girls using a beneficiary database provided by the project's delivery partner, CHADET. The database allowed us to profile the characteristics of all treated girls together and then to randomly sample girls from the database subject to achieving overall characteristics quotas which replicated the population of treated girls as a whole. Stratification was carried out according to girls' grade, location and marginalisation criteria. If selected girls were no longer in school (the database was from a previous academic year), a replacement strategy was put in place to ensure that the replacement girl was equivalent in terms of grade and marginalisation criteria. This group formed the treatment group.

As the project beneficiary database did not cover control schools, girls were first selected according to their grade as recorded in the relevant school rosters and screening questions were then asked to identify girls similar in terms of marginalisation to treated girls. As part of screening, girls were asked questions on early marriage and migration risk, domestic labour burden, and home residence status (living with parents, relatives or otherwise). In particular, potential control girls were admitted to the control sample subject to stratification criteria which were set to match the criteria used in choosing the treatment sample to ensure sample balance and representativeness. This group was the control group.

Qualitative

The qualitative sample was chosen based on the same sampling framework as the quantitative sample. Sample selection excluded girls already sampled as part of the quantitative data collection to avoid survey fatigue and was restricted to a subset of five of the fifteen treatment schools sampled.

The intention is also to follow the chosen qualitative sample over time, although the ability to do this in practice may prove to be limited given the wider range of participants in the qualitative sample, such as community representatives, who are likely to be more difficult to track. In addition, obtaining the same participants for future FGDs will present further coordination and logistical challenges as it will require bringing groups of people together. If substantial replacement is needed at future waves because of these additional challenges, then stratification will help ensure that the samples are acceptably similar and ensure a consistent range of views are represented. Sample selection for the qualitative sample was stratified (as with the quantitative sample) based on grade and marginalisation criteria. Any replacement of girls selected for the qualitative sample should ensure the sample remains representative according to these characteristics.

Qualitative interviews were conducted in local languages (Amharic and Afaan Oromo), with transcripts translated into English after the interviews had taken place. A thematic coding approach was used with a set of themes and subthemes chosen to capture the prevalence of key areas of interest. The themes chosen were as follows: Attitudes towards girls'

education, attendance, school facilities and quality of teaching, challenges encountered by girls around the home, school and community, and economic strength of community to cover basic needs of girls.

- **What are the respective roles of quantitative and qualitative data in the evaluation and how will you arrive at an integrated set of findings?**

Broadly, the quantitative data will be used to assess the overall impact of the project and will be complemented by qualitative data to better understand what works, how, and why. In particular, quantitative data will allow us to assess how learning and transition outcomes have been affected by the project, and if possible, for which types of girls and which elements of the treatment in particular. Ideally it will also allow us to examine the association between intermediate and final outcomes.

Qualitative data will then provide an elaborated picture of which drivers and barriers are most associated with the patterns observed in intermediate (and, hence, final) outcomes, helping to elucidate the key challenges girls are facing and which parts of the treatment seem to be most effective at present in helping to alleviate these. As a result, the qualitative analysis may also help to assess the validity of the Theory of Change.

- **How are the assumptions concerning the relationship between IO and outcomes going to be evaluated?**

The anticipated relationship between the IOs and outcomes is set out in the Theory of Change. The assumptions made as part of this causal chain can be tested through the lens of the logframe. The theory of change states that girls' achievement of higher order literacy and numeracy skills ('the outcomes') stems from 3 key IOs (improved attendance, improved teacher quality and improved self-esteem/empowerment of girls). Each of these IOs has a testable proxy measure in the logframe. If we observe that the IOs are realised whilst the outcome is not, this provides prima facie evidence that the IOs alone are not sufficient to a realise a change in the final outcome. Conversely, if we observe a significant change in the outcome without a change in the IOs, this may suggest that there is an additional channel through which learning outcomes are realised which has not yet been considered. Alternatively, it may indicate that the proxies used to capture the IOs are in need of revision.

Box 3: Benchmarking for learning and transition (External Evaluator)

Targets for the impact of the project on learning and transition will be calculated from the standard deviations of these outcomes measured on benchmark data collected in treatment areas.

Transition benchmarking involves the collection of a sample of 175 girls randomly chosen from a community in randomly selected *kebeles* (wards). Quotas will be used to ensure that the sample is representative of age and marginalisation in the beneficiary sample. The survey itself will elicit information on current and previous year's enrolment/employment status, and key demographic information.

Learning benchmarking will involve the collection of learning test data for a sample of 210 girls across 3 grades (70 per grade). The evaluator has stratified on the basis of grade and marginalisation using the beneficiary database to ensure that the learning benchmarking sample is representative of the beneficiary sample.

Baseline	Midline (1 year later)	Endline (2 years later)
Project grades		
4	5	6
5	6	7
6	7	8
7	8	9
8	9	10
Benchmark grades		
9	n/a	n/a
10	n/a	n/a
11	n/a	n/a

- **How are GESI minimum standards incorporated into the evaluation that allow measurement of gender sensitivity of the project and efforts to ensure social inclusion of girls across the range of characteristics?** ***MEL Guidance Part 2: Appendix F***

All programmes in the GEC-T are required to be gender and socially inclusive, and to meet the minimum standard of being 'GESI sensitive', defined as either:

- GESI Accommodating - Gender issues are acknowledged, and specific responses designed to address the needs and concerns of girls and boys are included in the project activities and outcomes. The response is more likely to focus on girls' practical needs.
- GESI Transformative - Gender stereotypes and norms are challenged, and the project seeks to transform unequal power relations between boys and girls through changes in roles, status and through the redistribution of resources. The response is more likely to focus on girls' strategic needs.

This project specifically targets the most marginalised girls[14], such as girls at risk of migration, early marriage or with a disability. It also seeks to improve behaviours and attitudes and gain the support of the most pivotal people in the girls' lives, whose views towards girls' education have a direct impact, such as parents, boys, teachers, and other community members, towards girls' education. This is incorporated into the evaluation through a range of measures such as changes in teachers' treatment of girls compared to boys in school, changes in attitudes towards single and divorced women, and changes in disabled girls' perception of community attitudes towards them. Qualitative data are also being collected to measure changes in attitudes towards marginalised girls' education.

[14] The criteria for marginalisation are set out in section 3.1

By targeting the most marginalised girls, supporting their families, schools, and communities, and challenging social and gender norms, this project is actively seeking to transform social and gender inequalities over the long term, in line with MEL Guidance Part 2: Appendix F.

2.4 Baseline data collection process

Pre-data collection

- **How were the sampling frameworks for quantitative and qualitative instruments developed? (Link appropriately with the sampling framework in the annex 10)**

The sampling framework for treatment areas for the core learning-transition sample and the qualitative sample was based on a list provided by CHADET of the 77 schools, across 40 kebeles, where their project is active under GEC-T. A sample of 15 of these schools was selected at random subject to stratification to ensure representation of the three target areas and rural / urban kebeles. Within these schools, attention was then restricted to girls detailed in CHADET's project enrolment database as participating in GEC-T. From the resulting sampling framework of girls, respondents for the linked learning-transition sampling framework were chosen at random but on a stratified basis by grade and marginalisation criteria in order to match the breakdown of the group of treated girls as a whole, the latter as reflected in the full enrolment database. Girls were selected for the qualitative interviews and focus group discussions from the same sampling framework but to limit survey fatigue only girls who had not been chosen for the quantitative work were selected.

For the comparison areas CHADET compiled a sampling framework of 72 non-treated schools (across 54 kebeles), which was intended as an exhaustive list of all schools within the three target zones (South Wollo, South Gondar, and Arsi) which were a distance away from treated schools; in order to minimise the risk of spillover effects that could influence the data, the list did not include non-treated schools in kebeles where other schools are being treated. We then selected fifteen of these schools by applying in the order listed as many of the following criteria as the required sample sizes allowed:

1. Matching – We chose control schools which matched or minimised the discrepancy with treatment schools for all schools in each area. Matching took place on the basis of the rurality of the school, and the zone it is located in.
2. Contamination – We liaised with CHADET to minimise the number of chosen control schools that are in kebeles where it is known there are other educational initiatives working with the same target groups;
3. Spillover effects – Minimised spillover effects by excluding control schools in kebeles which include treatment schools. Ideally, this would have also been done according to geographic proximity however this was not possible due to incomplete data on school locations.

ADVA CONSULT then selected girls on a random basis subject to grade stratification from the school rosters of the schools concerned. Qualitative data collection was not carried out in control areas.

- **How were the research instruments designed?**

The EGRA & EGMA learning instruments were taken from GEC-1, and subject to an additional round of piloting which yielded some minor changes. The SeGRA and SeGMA learning instruments were developed by educational experts[15] in line with fund manager guidance. Simetrica designed the Girls' School Survey and Household Survey instruments for the core girls and benchmark transition samples by drawing on a range of compulsory questions provided by the Fund Manager for these instruments and selecting the sets of additional questions appropriate to the intermediate outcomes targeted by the project, e.g. as the project is targeting teaching quality and girls' life skills as two of its intermediate outcomes, Simetrica included the Fund Manager's modules of questions on teaching quality and girls' self-esteem & self-efficacy in the Girls' School Survey instrument.

The qualitative survey and focus group instruments were designed by Simetrica, based around the intermediate outcomes and the key barriers and enablers drawn from the Theory of Change, developed by ChildHope. The instruments were designed to explore attitudes and behaviour with regard to the intermediate outcomes and girls' education – amongst girls themselves and boys, parents, teachers, and community representatives – and to invite open but structured discussion

15 Dr Charity Limboro, Kenyatta University, Kenya and Dr Desalegn Chalchisa, Addis Ababa University, Ethiopia

about the drivers and inhibiters of these, both contextually and relating to elements of the intervention. We also incorporated questions from the instruments that had been used for the ChildHope project in GEC-1 to ensure no insights were lost. The surveys focused mostly on individuals' attitudes and the focus group scripts more on community views and norms.

- **How did the evaluator prepare for tracking cohorts in future evaluation points?**

Girls' names were recorded to be able to locate girls from the school register at midline and endline. If a girl is not located in the school register, this would be suggestive of non-transition. Information was also collected on girls' addresses (and/or instructions to get to their household) to allow girls not in attendance and dropouts to be tracked to their homes.

- **Which instruments were piloted, when, and with what effects on the final instruments?**

All language and wave versions (9 in total) of EGRA/EGMA were piloted, as were all 3 versions of SeGRA/SeGMA. Piloting took place in March 2018, two months prior to the beginning of expected start time of the baseline. As a result of the pilot, several changes were made to the final instruments.

Firstly, the EGRA/EGMA test version originally allocated to be the baseline Oromiffa test was found to have significantly lower scores across all tasks than the midline and endline versions. Owing to time constraints, it was decided to switch the baseline and midline test versions, such that the midline test was taken at baseline. The former baseline test version will need to be revised to ensure it is comparable prior to be used as part of later waves.

Secondly, floor effects were observed in around a dozen SeGRA and SeGMA items across all test versions. It was identified that these were due to either errors or ambiguities in the mark scheme or phrasing of questions which were corrected after consultation with the SeGRA/SeGMA test developers. The SeGRA/SeGMA tests were also shortened as they were found to be too demanding for girls.

Lastly, the allocation of tasks across grades was altered slightly in light of persistent ceiling effects observed for the lower tasks.

- **How were enumerators recruited?**

The data was collected by ADVA CONSULT, an Ethiopian consulting firm based in Addis Ababa, specialising in research and capacity building for business and development companies where M&E is pivotal to their work. ADVA CONSULT used its core staff and additional enumerators were hired specifically for the project to help meet the relatively tight deadlines, once data collection would be under way. Simetrica required ADVA CONSULT contractually to hire field-researchers with the appropriate skills, by considering the following criteria

- Academic qualifications
- Prior experience in data collection in a related subject matter in Ethiopia
- Ability to fully commit themselves for the whole period of data collection
- Knowledge of local language and custom
- Suitability to collect data from children and others in light of the ethical and child protection requirements.

In addition, ADVA CONSLT were required contractually to:

- pose Child Protection related questions and to take the answers into account when making recruitment decisions to ensure that only safe researchers are recruited;
- complete employment, education and character checks through 2-3 referees who are not related to each potential field researcher;
- require all field-researchers to read through the ethical, child protection and other requirements and protocols of the project and sign a statement of commitment; and
- ensure that all enumerators and ADVA staff complied with these arrangements.

- **What kind of training did the enumerators undertake?**

Enumerators undertook three days of training, delivered by ADVA CONSULT, prior to the pilot, and one day of refresher training prior to the baseline. Training materials for enumerators were developed collaboratively between ADVA CONSULT and Simetrica

The training covered the following:

- An introduction to the project, and its aims, including how each external evaluation related indicator links to the project outcomes and intermediate outcomes
- The principles of data collection and recording of data, using tablet-based software
- An overview and introduction of the survey instruments where enumerators were familiarised with their purpose and administration.
- The administration of the qualitative 1-1 interviews and FGDs
- Data protection, including the management and storage of confidential data, and methods to ensure data quality.
- Child protection protocols – delivered by CHADET

- ***During data collection & provisional dates for midline***

Data collection was carried out by ADVA CONSULT, Simetrica's sub-contractor for the project. Collection began in Amhara on 14th May 2018 and Oromia in the week commencing 21st May[16] for the learning (primary and secondary) and qualitative data and the head of household element of the household survey. Collection ran until approximately the end of June 2018. The cleaned primary learning and head of household data was then provided by ADVA CONSULT to Simetrica in early July 2018 and the secondary learning data provided to Simetrica in early August 2018. Data collection occurred in two segments as several instruments were not collected initially, which were the primary caregiver component of the household survey and the girls school survey. The benchmark transition instruments were also recollected as the primary caregiver component from the first round of data collection was missing. The second round of data collection took place in October and November 2018 once schools had reconvened and the data was received in December 2018 once data entry was complete.

Given the delays experienced, a tentative suggestion for midline data collection would be December 2019, when all data would be collected. This would allow sufficient time to mobilise the necessary resources for data collection, and to make any required changes to the evaluation (e.g. to the logframe).

This date avoids seasonal effects which may stem from time taken for students to acclimatise to being back in school following the lengthy summer break which runs until October. There may be some seasonality as the previous learning tests were administered after summer examinations which may have had a positive (or negative) effect on learning outcomes. However, a delay to Summer 2020 would also push the endline date back beyond its planned delivery.

- **What protocols were followed when collecting the data, particularly to ensure ethical and child protection standards?**

Simetrica has followed, and has contractually obliged ADVA CONSULT to follow, the Child Protection Policy set out in the MEL framework, to ensure ethical and child protection standards are met. For ADVA CONSULT, this includes:

- Enumerators recruited and selected with the appropriate skills through past experience, for example working with high risk, vulnerable or marginalised girls and carrying out background employment, character and education checks.

[16] Data collection in Oromia could not begin on 14th May 2018 as planned due to political instability

- The External Evaluator, their data collector, and all field-researchers have been through an induction and workshop provided by ChildHope and CHADET on Child Protection in Research Ethics and have been inducted into CHADET's policy, procedures and local laws. These policies include the best interests of the children and 'do no harm' principles. ChildHope has also worked with the EE to ensure the interview tools follow these policies. Additional safeguards are in place where sensitive information is given by the children.
- Only children that can provide informed consent (in a manner appropriate for their age and understanding) are included in the programme. Adaptations are made for girls with disabilities, such as accessible buildings, which depend on the type of disability.
- Cultural considerations are also taken into account. As part of the three day training to enumerators, cultural training was provided to all enumerators on interviewing methods, female enumerators were recruited where possible, and girls were able to be accompanied by a person of their choice during interviews. Consent was obtained from girls, parents and teachers.
- Procedures were set in place before interviews to ensure children had someone to support them if they wanted, understood their right to not take part, that the interviewer was sensitive to the child's responses, that differences in gender needs were accounted for, and that permission was taken to take recordings. During interviews it was ensured that agreements were respected, questions were non-leading, responses were non-judgemental, and that body language put the children at ease.

A collaborative ethics panel involving ChildHope's Monitoring and Evaluation Safeguarding Advisor, CHADET's Education Manager, the External Evaluator, and the Data Collector, was developed to ensure all perspectives are taken into account and responsibility is shared for ethics at evaluation points. A reporting procedure was developed, whereby enumerators would report cases of abuse to a relevant party, such as a focal teacher or community worker, who would take the necessary steps to protect the welfare of the child.

- **What was done to ensure the safety of the enumerators during data collection?**

ADVA CONSULT was responsible for the safety of its enumerators. As part of the contractual obligation set between Simetrica and ADVA CONSULT, ADVA CONSULT were required to observe all health and safety rules and regulations and any other reasonable security requirements applicable to the work being carried out. No safety incidents were reported to Simetrica and we understand that ADVA CONSULT took the sensible decision to delay the start of data collection in Oromia by one week due to the impact of political instability there on enumerators' ability to access the area and travel safely within the area.

- **How did sampling of schools/parents/children etc. take place? Differentiate by research instrument as appropriate.**

CHADET works with 77 target schools, of which 47 are primary schools (grades 4 to 8) and 30 are secondary schools (grades 9 to 12). Of the 47 primary schools, 15 were randomly selected for the intervention sample, stratified by zone and rurality. 3,675 girls made up the available sample from these schools for treatment group in the quantitative analysis. From these, 787 were randomly selected (based on the required sample of 780 girls to detect the targeted improvements), stratified by grade and marginalisation status. The sample size is determined based on a statistical power calculation, and an attrition buffer which allows for a maximum attrition rate of 25%[17].

To obtain the control sample, CHADET identified 72 control schools which were matched to the sample of intervention schools based on stratification according to zone and rurality composition of the beneficiary population. These schools were then narrowed down to 66 schools which were not in the same kebeles as the treatment schools to avoid capturing spillover effects – where an intervention has an effect on the population outside of the treatment group. Of these 66 schools, 48 were primary schools, and 46 were, to our knowledge, not contaminated by educational interventions working with the same

[17] This was based on attrition rates observed during GEC-1, where 15.7% and 7.9% of students dropped out from baseline to midline, and midline to endline respectively.

groups. Of these schools 15 were randomly selected, and within these schools, 780 girls were randomly selected, stratified by grade and marginalisation status, to be in the control group.

The sample selection of schools was followed accurately with a few exceptions. In the case of treatment schools, one rural Arsi school and one rural South Wollo school originally selected to be sampled, were not sampled in practice. The Arsi school was replaced by a comparable school in the same area and the South Wollo school was replaced by two rural schools in South Gondar. As such, we may expect South Gondar to be oversampled and South Wollo to be undersampled for treatment schools. For control schools, one rural school in South Gondar and South Wollo were not sampled. The implications of this for sample representativeness are unclear, as the remaining schools were oversampled to obtain the necessary overall sample size. This will be expanded on in section 3.2.

Of the 15 schools in the intervention sample, five schools were randomly selected for the qualitative analysis. There were a total of 1,810 girls in these schools, of which approximately 260 were excluded to avoid survey fatigue as they were already in the quantitative sample. This left 1,550 potential participants, of which 84 girls were randomly selected to take part in the qualitative data collection, stratified by zone and rurality. As per correspondence with ADVA CONSULT, the qualitative sample broadly aligns with that which was agreed upon:

Table 4: Qualitative Sample

Type of Interview	Respondent Type	Actual Sessions	Actual Participants
FGD	Community representatives	5	41
	Girls	5	45
Key Informant Interview (KII)	Girls	30	30
	Boys	15	15
	Parent/care giver	26	26
	Community representatives	10	10

Given the highly dispersed nature of Ethiopian settlements, it was decided to allow household surveys to be conducted in schools where parents consented and were willing to come to the school. If parents were not willing to come to schools, they were to be tracked to households according to the girl's school survey instrument and strategy set out in the quantitative protocol. Enumerators also undertook data collection at the communities' farm land given that the second round of data collection took place during the harvest season. From discussions with ADVA CONSULT, this approach greatly helped data collection as obtaining responses from parents or caregivers would otherwise have been challenging.

- **How was the quality of data assured?**

As part of the quantitative protocol developed for data collection, a range of quality assurance measures were used by enumerators, data management coordinators and field coordinators.

Enumerators were asked to check that data entered on to tablets had been saved correctly after each interview. They were also asked to sample 5% of the data collection after the first few days of data collection for quality assurance purposes, for example to examine whether aspects such as variable coding, follow-on questions and screening-out had been followed correctly. To ensure that any systematic issues with data collection were identified, enumerators were asked to remain in close contact with field supervisors throughout data collection.

The data management coordinator conducted an additional round of quality assurance checks once the data was received from enumerators and ensured that data protection protocols were followed by separating identifying information from the data needed for analysis.

Lastly, the field coordinators were responsible for managing the data collection as a whole and coordinating across enumerators to ensure that any problems encountered were communicated as appropriate.

- **What are the final sample sizes for each of the instruments?**

The sample sizes for the instruments were as follows[18]:

1. Primary learning tests: 1,415
2. Secondary learning tests (grades 7 & 8 - excluding Benchmark learning sample): 498
3. Benchmark learning sample (grades 9-11): 168
4. HH Survey (HoH component): 1,457
5. HH Survey (PCG component): 1,734
6. HH Survey (Girls component): 1,486
7. Girls school survey: 1,560
8. Benchmark transition: 174
9. Qualitative interviews and FGDs: 140

Post data collection

A workflow for cleaning the data was developed by adapting best-practice guidelines from the World Bank's Development Impact Evaluation (DIME) group[19]. After data cleaning took place, identifying variables were stored separately to the main data in an encrypted ZIP file with the unique girl ID to be matched back to the main data as and when necessary. Quantitative data was analysed using the statistical software package Stata (version 13.1). Qualitative data was analysed using NVivo 12.

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

In summary, the challenges encountered in baseline data collection, and limitations of the evaluation design identified are as follows:

- The sample selection of control schools ensuring comparability with treatment schools was constrained by a lack of available data meaning that in practice controls schools were not matched to treatment schools along as many relevant school-level characteristics as would have been preferred. We will investigate the implications for this in terms of sample balance in section three.

[18] Note that we are in the process of resolving an issue related to duplicate IDs in the separate instruments, which affects up to 8% of observations. The reported sample sizes are net of duplicates so will increase if this issue can be partially or fully resolved.
[19] See analysis section of DIME Wiki <Last accessed 06/03/19

- The project data on beneficiaries was out of date when treatment sample girls were selected. Thus, data on girls' marginalisation status in the project data was not fully up to date. This may have impacted on sample stratification to ensure representativeness versus the population as a whole, if the composition of girls in treatment schools had changed since the project data was collected. Non-transition also meant that some girls selected to be sampled would no longer have been in the school, increasing the importance of carefully followed replacement procedures. The project database is in the process of being updated and will be updated in time for the midline evaluation.
- Political unrest resulted in an increase in costs for ADVA CONSULT and introduced delay to commencement of the baseline data collection. Difficulties in securing access to suitable learning tests also caused further delay.
- The recording of girl identifiers was, for a minority of girls, inconsistent across instruments. As matching across instruments for these girls is not possible, the sample size in the final merged dataset is reduced slightly compared to the full sample observed in the EGRA/EGMA learning test data. Of the 1413 girls for whom learning test data was collected, 1218 of these (86.2%) were able to be matched based on student IDs to the girls' school survey. This is lower than would be expected given that the girls school survey was collected as part of a second round of data collection.

3. Transcription, coding and translation of qualitative interviews was carried out by ADVA Consult. Using the thematic coding approach, a saturation point was reached at around 20-25% of transcripts, after which additional analysis did not add to the quality of findings. Midline data collection will likely benefit from further consideration as to the causes behind this, which might include: Survey fatigue, the wording of questions in the qualitative instruments and/or the relative homogeneity of views and opinions in the context. Key Characteristics of Baseline samples

3.1 Project beneficiaries

ChildHope is supporting 16,481 marginalised girls (and 3,362 boys as indirect beneficiaries) in Amhara and Oromia to achieve improved educational outcomes. Marginalised girls were selected for the preceding GEC-1 project and so additional selection for the intervention group for this project was not necessary.

Specifically, girls were recruited for the programme if they experienced at least one of the following marginalisation criteria:

- Poverty (livelihood status)
- Street children
- Early marriage
- Risk of migration
- Risk of work (domestic or other)
- Disability

At GEC-1, treatment kebeles were identified using secondary data about the prevalence of early marriage and risky child migration. Within selected kebeles, one or more schools were then chosen for treatment. Some of the treatment, for example teacher training, was then delivered at school level, benefiting all girls at treatment schools regardless of their individual level of marginalisation. Some was delivered at the individual level, for example psycho-social support or temporary shelter, which was targeted at the most marginalised girls within treatment areas.

In particular, the identification of in-school girls to receive individual treatment (e.g. after class tutorials) was done at school by CHADET Community Workers, school teachers/principals and students' club members. Girls with poor educational performance, especially in reading and numeracy, with low levels of attendance, susceptible to school drop-out (due to early

marriage, risky migration, domestic work, poverty, and lack of basic scholastic materials) were also identified based on the previous year's school records and continuous assessment during the school year. These data were used to decide on the type of services appropriate for the girls and their families.

The identification of out-of-school girls for individual-level treatment was conducted at different levels of the community (household, en-route during migration, and on the streets). The household recruitment was done by house-to-house assessment in the target kebeles by CHADET's community workers and volunteers with close support by kebele GEC Committees (community representatives). The volunteers were provided with questionnaires to collect data on school age girls not going to school and their families. Based on the collected data, decisions were taken on the type of services to provide the girls and their families. School age girls en-route on their own as a result of trafficking were identified by all stakeholders, particularly by bus stations champions (bus drivers collaborating with the project) and by the local police who were trained and supported in child protection. Street girls were identified by CHADET's community workers in the local town where they live and/or work. These data were used to decide on the type of services appropriate for the girls and their families.

Additionally, in the course of the GEC-1 project, the team identified over 500 girls affected by a disability, who were also then included in the treatment. The girls have a range of disabilities from physical to learning. ChildHope, with disability consultants, developed tools for measuring learning difficulties such as dyslexia, dyscalculia and dysgraphia and using Washington Group questions to identify other disabilities.

In the context of gender inequality, the project aims to support boys to develop healthier behaviours and attitudes towards girls. While a specific intervention group is not being tracked, boys in the intervention schools are receiving the project interventions and are being engaged more deeply in this project compared to the GEC-1. For example, they are taking part in tutorials and reading corners, and are also designing their own programmes that empower girls, rather than merely being expected to take part in activities designed by, and for, the girls.

Project beneficiaries have been chosen with the intention to transform inequalities in the long term for all children, despite gender, disability, or any other characteristic, in line with the Gender Equality and Social Inclusion (GESI) guidance on implementing transformative projects[20].

3.2 Representativeness of the learning and transition samples across regions, age groups, grades, disability status and sex of the beneficiaries

The tables below give an idea of the representativeness of the evaluation sample versus the whole beneficiary population.

The sampling strategy specified that data were to be collected on a minimum of 780 girls in both treatment and control schools. provides a breakdown of the evaluation sample across the three administrative zones in which the project is active. The sample sizes reported are after the matching of the girls' school survey and learning test samples. As the data collection took place in two rounds, not all of the girls were able to be matched, meaning the learning test sample drops by approximately 10%, which takes away from some of the extra sample designated to mitigate attrition (the attrition buffer). For this reason, extra efforts will need to be made at midline to avoid significant attrition.

In addition to the minimum sample size, an equal number of girls in the treatment and control groups were to be collected. This shows that this requirement has also been met.

Sampling was to be carried out based on the proportion of treated girls in each zone, which was calculated from a database of enrolled girls provided by ChildHope's delivery partner CHADET. According to this database, 48% of sampled girls should be from South Gondar, 29% from South Wollo and 23% from Arsi.

Table 5 shows that the treatment and controls arms of the sample collected are broadly representative of the population within each region, though there are some disparities. In the treatment group, the proportion of girls from South Gondar is

[20] GESI addendum for Baseline Template- Dec 2017

as expected, but the proportion from South Wollo is slightly lower than the proportion in the population. This is likely to result from miss election of schools discussed previously. Conversely, the proportion of girls from Arsi is slightly higher than would be expected given the population breakdown.

In the control group, the proportion of sampled girls from South Gondar is the largest, as expected, although it is slightly larger than expected based on the population. The proportion of girls from Arsi is a few percentage points lower than expected. This is offset by the proportion of girls from South Wollo being a few percentage points larger than would be expected based on the population.

These differences in geographic dispersion imply some differences in the number of girls sampled between the intervention and control groups within specific geographies. In South Gondar and Arsi, the evaluation sample is slightly biased towards intervention girls (61 more treatment girls in Arsi, and 43 more in South Gondar). In South Wollo, an opposite and slightly stronger bias is observed, with an imbalance of 126 girls occurring in favour of the control group.

Table 6 presents a breakdown of the evaluation sample by grade. The targeted proportional breakdown based on the project's beneficiary database was as follows: Grade 4 - 22.4%, grade 5 – 24.5%, grade 6 – 19.6%, grade 7 – 17.4%, grade 8 – 16.1%[21]. The evaluation sample only approximately tracks the chosen breakdown, though the achieved grade breakdown by treatment status rarely differs by more than 3 to 4 percentage points vis-à-vis the targeted breakdown.

Between the intervention and control groups, an approximate balance between treatment and control girls is achieved across the grades, though there are some imbalances between the intervention and control groups for grades 4 and 6.

Table 7 shows girls in the intervention and control group based on their age group. It can be seen that the majority of girls in the sample are aged between 12 and 15 years old, and that there are no girls under 8 or over 20 years old in any of the groups. Ages are distributed similarly across the intervention and control group, especially in the ranges between 9-11 and 16-19 years old. While for girls aged 12-13 years old, the sample is biased towards the control group, the opposite is true for girls aged 14-15. Representativeness to the total population varies by age range: while girls aged 12-13 and 16-17 are well-represented in our sample, girls aged 14-15 are over represented at the expense of under representing girls aged 9-10. The differences might be driven by the fact that stratification was made based on grade, and age and grade are imperfectly related in our sample.

Table 5: Evaluation sample breakdown (by region)

	Intervention (Baseline)	Control (Baseline)	Population
Sample breakdown (Girls)			
South Gondar (Amhara)	48.0%	51.3%	48%
South Wollo (Amhara)	20.5%	31.7%	29%
Arsi (Oromia)	30.4%	16.9%	23%
Missing region data	1.1%	0.1%	0%
Girls (sample size)	721	692	787 (each)

Table 6: Evaluation sample breakdown (by grade)

	Intervention (Baseline)	Control (Baseline)	Population
Sample breakdown (Girls)			
Grade 4	18.9%	23.1%	22.4%
Grade 5	23.9%	26.3%	24.5%
Grade 6	24.1%	18.5%	19.6%
Grade 7	18.6%	17.5%	17.4%
Grade 8	14.3%	14.6%	16.1%

[21] The intervention population is from grades 3 to 12. For cost and efficiency reasons the sample grades were grades 4 to 8.

Missing grade data	0%	0.3%	0%
Girls (sample size)	721	692	787 (each)

Table 7: Evaluation sample breakdown (by age)

	Intervention (Baseline)	Control (Baseline)	Population
Sample breakdown (Girls)			
Aged 6-8 (% aged 6-8)	0% (0.0%)	0% (0.0%)	2.4%
Aged 9-11 (% aged 9-11)	11.9% (14.0%)	11.1% (12.7%)	32.8%
Aged 12-13 (% aged 12-13)	37.1% (43.8%)	33.2% (38.1%)	33.5%
Aged 14-15 (% aged 14-15)	29.8% (35.1%)	35.0% (40.2%)	22.2%
Aged 16-17 (% aged 16-17)	5.5% (6.5%)	7.2% (8.3%)	7.2%
Aged 18-19 (% aged 18-19)	0.6% (0.7%)	0.7% (0.8%)	1.9%
Missing age data	15.2%	12.9%	N/A
Girls (sample size)	721	629	

Notes: As missing age data is a problem for a non-trivial portion of girls, we report in *brackets* the proportion of girls in age bracket ignoring missing age data. It is these figures which should be compared with the population figures under the assumption that missing age data is uniformly distributed across age groups.

Table 8 breaks down the sample by type of disability. It shows that 10.4% of girls in the control group and 6.1% of girls in the intervention group have at least one of the following impairments: vision, hearing, mobility, cognitive, self-care or communication. Moreover, the proportion of girls with disability in the control group is larger than the proportion of girls with disabilities in the intervention group, for every type of impairment. In the control group cognitive and communication impairments are the most common, with 4.6% of the girls having cognitive impairments, and 3.7% having communication impairments. In the intervention group hearing, visual and cognitive impairments are among the most common ones, affecting 1.6%, 1.4% and 1,4% respectively. The treatment and control samples are slightly less well-aligned in terms of disability, but still broadly comparable. In terms of the overall representativeness of the sample with the population sample as defined in the project database, the proportion of girls with disabilities in our sample is higher than the population figures. While in Table 7 we report 6.1% with disabilities in the intervention group, this figure is estimated at 3% for the whole population.

Table 8: Evaluation sample breakdown (by disability)

Sample breakdown (Girls)	Intervention (Baseline)	Control (Baseline)	Household Survey and Girls School survey – Washington Group and child functioning questions
Girls with disability	6.1%	10.4%	CS_D1s-CS_D6s
Provide data per impairment			
Vision impairment	1.4%	2.1%	CS_D1s
Hearing impairment	1.6%	0.7%	CS_D2s
Mobility impairment	1.3%	0.5%	CS_D3s
Cognitive impairment	1.4%	4.6%	CS_D4s
Self-care impairment	0%	1.8%	CS_D5s
Communication impairment	0.6%	3.7%	CS_D6s

The overall representativeness of the sample is well-aligned with the population sample as defined in the project database. When comparing the above tables with the population estimates, it can be seen that they have broadly consistent distributions among groups of region and grade, whereas the differences in age are explained by considering grade (and

not age) as a variable to stratify the sample. Furthermore, treatment and control samples are also broadly similar between them.

3.3 Educational Marginalisation

Table 9: Characteristics of girls and their households

	Intervention (Baseline) (%)	Control (Baseline) (%)	Difference	Source (Household and Girls School survey)
Sample breakdown (Girls)				
Orphans	10.5	4.1	6.4*	PCG_11g PCG_13g
Living without both parents	8.9	9.2	0.3	PCG_10g PCG_12g
Living in female headed household	42.9	44.0	1.1	HH_8
Married	1.4	1.2	0.2	PCG_22g
Mothers - Under 18 - Under 16	N/A (*see notes)*	N/A *(see notes)*		PCG_23g
Poor households *(see notes for definition)*	87	78.6	8.4*	PCG_7enr PCG_11econ PCG_2econ PCG_5econ_a PCG_7econ
Difficult to afford for girl to go to school	19.2	7	12.2*	PCG_7enr
Household doesn't own land for themselves	9.9	3.6	6.3*	PCG_11econ
Material of the roof: mud or thatch	12.8	13.5	0.7	PCG_2econ
Household unable to meet basic needs	28.6	39.1	10.5*	PCG_5econ_a
Gone to sleep hungry for many days in past year	2.9	6	3.1*	PCG_7econ
Language Difficulties: LoI different from mother tongue*	4	7.1	3.1*	PCG_20g PCG_1enr*
Language Difficulties: Girl doesn't speak LoI	0.1	0.4	0.3	PCG_3enr
Parental education -HoH has no education (%)	66.2	60.4	5.8*	HH_13
-HoH Religion: HoH religion: Christian HoH religion: Muslim HoH religion: Other	 61.8 36.2 2.0	 58.9 41 0.1	 2.9 4.8 1.9	 HH_10 HH_10 HH_10

Note: Data only contains single orphans. Data for girls who are mothers could not be included since it is contained in the HH PCG dataset that could not be matched with the treatment and control variables due to lack of unique identifiers. We could not include the information of girls that are mothers Poor households are considered to have at least one of the following conditions: i) Difficult to afford for girl to go to school, ii) Household doesn't own land for themselves, iii) Material of the roof is thatch or mud, iv) Household unable to meet basic needs or v) Gone to sleep hungry for many days in past year. *PCG_2enr ('Is the main language of instruction at school different from the main language girl speaks at home?'), which is recorded by enumerators, is done so with significant error – as we such we derive this ourselves based on other observed variables on mother tongue and language of instruction. The asterisk represents statistically significant differences with $p<0.05$.

Table 9 presents the characteristics of the sample girls and their households for the intervention and control group. It can be observed that, at the time of interview, 10.5% of girls in the intervention group were single orphans, 8.9% lived with both parents and 42.1% lived in a female headed household. In most respects, the intervention and control group samples are similar in terms of family situation. 4.1% are orphan girls, 9.2% lived with both parents and 45% lived in a female headed household. Furthermore, 1.4% and 1.2% of the girls in the intervention and control group respectively were married.

Regarding poverty-related data, it can be seen that 87% of the households of girls in the intervention group and 79% of households of girls in the control group are defined as a poor household. This means that they meet one or more of the following conditions: i) Difficulty to afford for the girl to go to school, ii) Household doesn't own land for themselves, iii) Material of the roof is thatch or mud, iv) Household is unable to meet basic needs or only able to meet basic needs[22] or v) Primary caregiver reports to have gone to sleep hungry for many days in past year. The main differences in the prevalence of these sub factors between the groups are observed in the numbers facing difficulty to afford girls' school and in number of households reporting as being unable to meet basic needs.

Table 9 also shows that 4% of the girls in the intervention group are taught in a language of instruction which is different to their mother tongue and 0.1% do not speak it. The case for girls in the control group is slightly higher. Girls with a language of instruction different from their mother tongue make up 7.1% of the sample, of which approximately one in twenty (0.4% of the total sample) do not understand the language at all.

Finally, Table 9 shows that 67.3% of the households' heads in the intervention group have no education, while this figure is 58.7% for households' heads in the control group. Also, that the household's heads majority religion is Christian, with 61.8% and 58.9% of the intervention and control group reporting to be part of it, respectively.

To conclude, the most common characteristics in the intervention group are living in a female headed household, living in a household defined as being poor and being in a household where the head has no education. We may expect girls in poor households (and the subset of these who are 'unable to meet their basic needs', and who 'find it difficult to afford for girl to go to school') to experience worse learning and attendance outcomes if they required to work to assist their families, or if the costs associated with schooling are too expensive. Being an orphan, and living without both parents, may also be associated with higher work burdens.

Barriers

Table 10: Potential barriers to learning and transition

	Intervention (Baseline) (%)	Control (Baseline) (%)	Difference	Source
Sample breakdown (Girls)				
Home – community				
Safety:				
Fairly or very unsafe travel to schools in the area	7.7	3.9	3.8*	PCG_9
Parental/caregiver support:				

[22] Thus, we exclude households who are 'able to meet basic needs with some non-essential purchases', and those who are 'able to purchase most non-essential goods'

Sufficient time to study: High chore burden	26.3	50.3	24*	PCG_26g
Doesn't get support to stay in school and do well	4.3	5.8	1.5	HHG_7
School level				
Attendance:				
Attends school half the time	3.5	25.4	21.9*	PCG_6enr
Attends school less than half the time	0.9	11.9	11*	PCG_6enr
Doesn't feel safe at school	3.2	6.6	3.4*	CS_W14s
More than 30 minutes' walk to nearest primary school *(more than 1 hour)*	29.5 *(2.4)*	17.8 *(0.5)*	11.7*	PCG_7
More than 30 minutes' walk to nearest secondary school *(more than 1 hour)*	58.7 *(40.7)*	53.3 *(41.4)*	5.4*	PCG_8
School facilities:				
No seats for all students	8	31.5	23.5*	CS_W5s
Difficult to move around school	3.3	6.5	3.2*	CS_W6s
Doesn't use drinking water facilities	44.3	45	0.7	CS_W7s
Doesn't use toilet at school	10.2	25.1	14.9*	CS_W9s
Doesn't use areas where children play/ socialise	2.6	16	13.4*	CS_W11s
Teachers:				
Disagrees teachers make them feel welcome	1.9	3	1.1	CS_WA
Agrees teachers treat boys and girls differently in the classroom	58	68.8	10.8*	CS_1s
Agrees teachers often absent from class	46	63.7	17.7*	CS_2s

Note: High chore burden is defined as spending half or whole day on helping her family and/or doing work around the house, on a normal school day. The asterisk represents statistically significant differences with p<0.05.

Table 10 presents data on potential barriers to education based on the currently available quantitative and qualitative data. It shows that, on average, girls in the treatment group report having lower potential barriers to learning and transition than girls in the control group: they feel safer, receive more support from their caregivers, have better attendance rates, report better availability and use of the school facilities and have better opinion of their teachers than girls in the control group. Of course, the extent to which this is determined by the sample of schools in each respective group versus the impact the project may be having is not clear at this stage.

Moreover, it can be seen in the parental/caregiver support section that girls in the control group have more demands placed on them from their caregivers than the girls in the intervention group: While 50.3% of the girls in the control group report spending half or a whole day on helping her family on a normal school day, this figure is 26.3% for girls in the intervention group. This may be driven in part by attitudinal differences between control and treatment schools. For example, 82% of primary caregivers from the intervention area agree or strongly agree that a girl is just as likely to use her education as a

boy, whereas only 77% agree or strongly agree in the control area *(PCG_33g)*. Similarly, parents in intervention areas are 5 percentage points more likely to listen to the views of the girl when making decisions about her education than making the decision by themselves alone. Both of these factors are strongly negatively correlated with girls having a high chore burden Furthermore, 5.8% and 4.3% of the control and intervention school girls respectively report receiving no support to stay and do well in school.

The most notable datapoint that runs counter to this trend is the one showing that girls in the control group report feeling safer than girls in the intervention group with regard to travelling to school: 7.7% of girls in the intervention and 3.9% in the control group report feeling "fairly" or "very" unsafe travelling to school from their home area. This may in part reflect the fact that the data shows that girls in the intervention group tend to live further away from school with close to 30% of them 30 minutes or more walk from the nearest primary school, compared to only 17.8% in the control group.

When analysing school-related outcomes, girls in the control group report much lower attendance than girls in the intervention group: 25.4% and 11.9% of the girls in the control group attend school half or less than half the time respectively, while this figure is 3.5% and 0.9% for girls in the intervention group. This finding may be driven by differences in parental expectations and the burden of chores in the treatment and control groups. It may also be explained by the differences in the feeling of safety at school: while only 3.2% of girls in the intervention group report feeling unsafe at school, 6.6% of girls in the control group do so. Further analysis of the data show that 17% of girls in the intervention group (22% of girls in the control group) were absent from school for more than five days in a row last year. This may be due to domestic and farm labour requirements placed on the girls.

With regard to school level barriers to education, girls in the intervention group report better availability and use of school facilities than girls in the control group: a higher fraction of them use school drinking water facilities, toilets and areas to play. Also, a lower fraction report having difficulties to move around school or not having enough seats for all the students.

Regarding girls' opinions on their teachers, the table shows that 1.9% of the girls in the intervention group and 3% of girls in the control group disagree "a little" or "a lot" to the statement "My teachers make me feel welcome in the classroom". Furthermore, around 60% of the girls in both groups agree that teachers treat boys and girls differently in the classroom, though it is not immediately apparent how these differences manifest. Lastly, 46% of the girls in the intervention group and 63.7% of the girls in the control group agree that their teacher is often absent from class.

In summary, Table 10 suggests that the most prevalent barriers affecting girls in project schools are frequent teacher absences from class, non-usage of in-school amenities such as toilets and drinking water facilities, having a high chore burden, feeling that boys and girls are treated differently in class and living a long distance from the nearest primary school. Teacher absences and a high chore burden are likely to lower potential learning outcomes, and the latter may affect attendance as well. Non-use of in-school amenities may lower attendance if this is due to an apprehension about, for example, the use of toilet facilities in an insecure environment. The immediate impact of boys and girls being treated differently in class on outcomes is not obvious unless one assumes that girls are treated worse, or that any difference in the way children are treated has a negative impact on outcomes. There is also qualitative evidence on the barriers to education that girls face. The study mainly looked at the attitudes of community members toward girls' education and the challenges to girls' education from the perspectives of different stakeholders. The study used Focus Group Discussions (FGD) with community representatives and girls and Key Informant Interviews (KII) with boys, girls, teachers, community representatives, and parents/care-givers.

The barriers identified by the qualitative analysis align with the barriers reported in the table above. Both absence from school and teaching quality and competence were reported as key barrier. A high chore burden was reported as one of the key reasons for absence, which also aligns with the findings in Table 10. This also links to the attitudes of community members, which was also reported as a key barrier – while the qualitative data show that most families do prioritise girls' education, there was evidence that some do not. In addition to a high chore burden, reasons included working in the family business and forced migration. Early marriage was also reported as an existing practice which affects educational outcomes.

In addition to these factors, the qualitative analysis found the economic strength of families to be a key barrier to girls' education. Community representatives reported that most families in the study area cannot cover their basic needs, which means that they cannot cover the costs of education for their children. This aligns with Table 9, which shows that 87% of girls in the treatment sample and 79% in the control sample in the project were defined as living in 'poor households'. The qualitative study also found that most families in the community are currently only able to cover the costs of education because of provisions from CHADET, which this suggests a possible sustainability risk for the project.

3.4 Intersection between key characteristics and barriers

Table 11 shows the intersection between barriers to education and the key characteristics of girls in the intervention group. The column headings refer to the characteristics and the row headings refer to the barriers.

Table 11: Barriers to education by characteristic

	Region (Oromia/ Amhara)	Rural / Urban	Physical impairment (Yes / No)	Cognitive impairment (Yes / No)	Mother tongue different to LoI (Yes / No)	Religion (Muslim/Christian)	Living without both parents (Yes / No)	Head of household has no education (Yes / No)
More than one hour to get to school	-	4.3% / 0.4%	-	-	-	-	12.5% / 2.4%	3.8% / 0.5%
Books and learning material not available at school	-	19.3% / 4.2%	-	-	-	6.1% / 16.6%	-	-
Computers not available at school to use	-	-	80.8% / 95.6%	-	-	-	-	-
Seats not available for all students in class	27.4% / 1.9%	11.0% / 2.7%	-	-	65.0% / 19.2%	9.1% / 3.3%	0.0% / 8.6%	-
Not able to move around school easily	7.4% / 2.1%	-	-	-	-	-	-	-
Doesn't use drinking water facilities	65.9% / 40.6%	35.2% / 60.9%	-	-	85.0% / 44.1%	58.9% / 37.0%	21.4% / 46.1%	37.5% / 60.7%
Doesn't use toilets at school	-	5.5% / 19.2%	-	-	-	-	-	7.1% / 14.8%
Doesn't use areas where children play / socialise	7.9% / 1.3%	-	-	-	-	4.1% / 1.4%	-	-
Doesn't feel safe travelling to and from school	-	-	-	-	0.0% / 18.5%	-	-	-
Hasn't used school's library/reading corner in last month	-	2.3% / 14.9%	-	-	35.0% / 15.8%	12.0% / 2.8%	-	4.2% / 12.1%
Doesn't feel safe at school	-	4.5% / 0.8%	-	-	-	-	-	-
Family/guardian does not provide school supplies	-	15.6% / 5.4%	33.3% / 11.6%	-	-	-	-	-
Were you absent from school for more than 5 days in a row last year?	35.4% / 9.4%	15.0% / 21.8%	-	-	-	24.4% / 13.8%	5.4% / 17.8%	-
Absent to support parents in domestic chores	-	70.2% / 31.9%	-	-	-	50.8% / 64.0%	37.9% / 57.5%	67.6% / 41.9%
Absent to be in paid labour	13.0% / 23.2%	10.8% / 30.6%	-	-	-	29.8% / 13.1%	-	13.9% / 29.9%
Risk of early marriage	16.6% / 6.2%	-	-	-	-	-	-	10.7% / 4.9%
Has thought of migrating to another area?	23.2% / 1.1%	-	-	-	-	-	-	-
Disagree teacher makes them feel welcome	4.3% / 1.3%	-	-	-	-	3.6% / 0.8%	-	-
Teacher treat boys and girls differently in the classroom	87.1% / 46.3%	-	-	-	90.0% / 63.0%	-	40.0% / 59.4%	61.0% / 46.7%
My teachers are often absent for class	84.0% / 29.3%	50.7% / 37.9%	-	-	95.0% / 54.3%	-	-	-
Teacher asks more questions to boys	18.9% / 3.9%	8.9% / 2.7%	-	-	-	10.2% / 4.4%	-	8.2% / 2.2%
Teacher asks harder questions to boys	18.9% / 4.9%	10.1% / 5.0%	-	-	-	12.7% / 5.8%	-	10.4% / 2.7%
Teacher asks more questions to girls	9.8% / 5.4%	10.7% / 1.5%	-	-	-	-	-	-

Teacher asks harder questions to girls	12.8% / 3.4%	7.7% / 2.3%	-	-	-	-	-	6.3% / 2.2%
Sample Size	406 / 1,196	1,046 / 700	56 / 1,495	63 / 1,495	20 / 1,792	560 / 877	162 / 1,398	856 / 495 /

Notes: Column headings refer to characteristics, row headings refer to barriers. Only interactions which are statistically significant at the 5% level are shown.
Mother tongue different to LoI characteristic includes treatment and control girls due to small sample size.

	Region (Oromia/ Amhara)	Rural / Urban	Physical impairment (Yes / No)	Cognitive impairment (Yes / No)	Mother tongue different to LoI (Yes / No)	Religion (Muslim/Christian)	Living without both parents (Yes / No)	Head of household has no education (Yes / No)
Teacher doesn't use different language if you don't understand	59.6% / 50.3%	-	-	-	-	-	-	-
Teacher doesn't encourage participation in class	-	47.2% / 36.8%	-	-	0.0% / 44.2%	-	-	-
Teacher disciplines/punishes students who gets things wrong	24.0% / 8.4%	12.0% / 17.5%	-	-	-	-	-	-
Teacher used physical punishment in last week	98.4% / 64.8%	61.3% / 76.1%	42.9% / 67.0%	-	-	83.0% / 62.0%	33.3% / 69.4%	-
Does not feel confident answering questions in class	-	-	21.2% / 5.3%	-	-	-	-	-
Does not want to continue studying after this year	-	-	-	-	-	-	5.4% / 1.0%	-
Does not feel able to describe thoughts to other people when speaking	2.4% / 0.2%	-	-	6.7% / 0.7%	-	-	-	-
Does not feel able to work well in groups	-	-	-	-	-	-	3.6% / 0.7%	-
Does not feel able to organise friends to do an activity	3.7% / 0.9%	-	6.1% / 1.4%	13.3% / 1.3%	-	3.0% / 0.6%	-	0.5% / 2.7%
Does not ask teacher if doesn't understand something	-	0.6% / 2.3%	-	6.7% / 1.1%	-	-	-	0.3% / 2.2%
Does not agree their success is due to hard work	-	-	-	-	-	-	-	-
I get nervous when I have to read in front of others	66.5% / 19.5%	38.3% / 26.1%	-	-	80.0% / 40.6%	-	21.4% / 35.0%	-
I get nervous when I have to do maths in front of others	56.7% / 16.7%	32.2% / 21.8%	-	-	85.0% / 37.6%	-	10.7% / 30.2%	-
If I do well in a test it is because I am lucky	89.0% / 42.0%	60.7% / 48.1%	-	-	89.5% / 53.6%	-	-	56.8% / 47.5%
Family decides whether will go to school	-	-	21.2% / 9.9%	-	45.0% / 17.7%	-	-	-
Family decides what age will get married	-	-	36.4% / 13.7%	-	60.0% / 18.3%	-	-	-
Family decides what type of work after finishing studying	14.6% / 4.3%	9.1% / 1.9%	18.2% / 6.2%	20.0% / 6.4%	20.0% / 6.8%	-	-	-
Family decides how much time spent with friends	18.9% / 12.4%	-	-	-	45.0% / 15.8%	-	-	-
Does not feel able to stay focused on goals despite things getting in the way	2.4% / 0.0%	-	6.3% / 0.6%	-	-	-	-	-
Does not recognise choices today about studies affect future	4.1% / 0.0%	-	6.3% / 0.8%	12.5% / 0.8%	-	-	-	-
Does not try to find another way to express self if not understood	5.6% / 1.0%	-	12.5% / 1.6%	12.5% / 1.8%	-	-	-	-
Does not pay attention to body language of others	-	2.8% / 8.4%	-	-	-	9.8% / 2.1%	14.6% / 4.2%	2.5% / 8.9%
Cannot choose whether to stay in school, just accept	-	69.7% / 56.9%	-	-	-	48.8% / 74.3%	-	66.7% / 53.1%

Spend less time reading than male siblings	44.0% / 35.0%	40.7% / 31.7%	-	-	-	-	-	38.8% / 29.6%
Sample Size	406 / 1,196	1,046 / 700	56 / 1,495	63 / 1,495	20 / 1,792	560 / 877	162 / 1,398	856 / 495 /

Notes: Column headings refer to characteristics, row headings refer to barriers. Only interactions which are statistically significant at the 5% level are shown.
Mother tongue different to LoI characteristic includes intervention and control girls due to small sample size.

Table 11 shows that there are several significant interactions between girls' characteristics and the barriers that they face. These are suggestive of the factors which may be responsible for barriers to learning, though these are only simple correlations and further analysis will be needed to understand whether there are causal relationships between them. Statistically insignificant findings should also not be taken as strong evidence that no relationship exists, because in many cases sample sizes are too small to detect a relationship.

The first column looks at differences in barriers between girls living in Oromia compared to girls living in Amhara. There are a number of significant interactions, with girls from Oromia consistently reporting facing greater barriers than girls from Amhara. This may be considered surprising, as poverty levels and other characteristics are similar across the regions[23]. However, this dataset considers barriers at woreda level so the regional comparison is not completely relevant. The data show that girls from Oromia report facing greater barriers regarding their school environment (e.g. availability of seats (27.4% vs. 1.9%), treatment from teachers (87.1% vs. 46.3%), absence (35.4% vs. 9.4%)), external circumstances (e.g. risk of migration (23.2% vs. 1.1%), risk of early marriage (16.6% vs. 6.2%)), and self-esteem and self-efficacy (e.g. confidence reading (66.5% vs. 19.5%) and doing maths (56.7% vs. 16.7%) in front of others, time spent reading compared to male siblings (44.0% vs. 35.0%), family's role in decisions such as what kind of work after studying (14.6% vs. 4.3%)).

The second column compares the responses of girls living in rural areas to girls living in urban (or peri urban) areas. Girls from rural areas are more likely to report facing barriers in a number of areas than girls from urban areas. As one would expect, they report longer journey times to school (4.3% vs. 0.4% reporting greater than one hour). They also report lower availability of books and learning materials (19.3% vs. 4.2%), unavailability of seats for all students (11.0% vs. 2.7%), and are more likely to report feeling unsafe at school (4.5% vs. 0.8%). While rural girls were less likely to have been absent from school for more than five days in a row (15.0% vs. 21.8%), their absence was more likely to be explained by helping with domestic chores (70.2% vs. 31.9%) and less likely to be explained by paid work (10.8% vs. 30.6%).

Girls from rural areas also report different teaching experiences to girls from urban areas – that their teachers are often absent from class (50.7% vs. 37.9%), that their teachers do not encourage participation in class (47.2% vs. 36.9%), and that their teachers ask more (8.9% vs. 2.7%) and harder (10.1% vs 5.0%) questions to boys, though they are also more likely to say that their teachers ask more (10.7% vs. 1.5%) and harder (7.7% vs. 2.3%) questions to girls. A smaller proportion also reported that their teachers discipline or punish students who get things wrong (12.0% vs. 17.5%).

There are also differences between rural and urban girls in self-esteem and self-efficacy barriers. Girls from rural areas are more likely to report feeling nervous reading (38.3% vs. 26.1%) and doing maths (32.2% vs. 21.8%), and are more likely to say that their success in a test is due to luck (60.7% vs. 48.1%). They also report less control in certain situations – their families are more likely to decide what kind of work they do after finishing studying (9.1% vs. 1.9%), they are more likely to say they cannot choose whether to stay in school and just have to accept the outcome (69.7% vs. 56.9%), and they are more likely to report that they spend less time reading than their male siblings (40.7% vs. 31.7%).

The next two columns compare the responses of girls with physical and cognitive impairments compared to those without. The sample sizes for these characteristics, particularly cognitive impairments, are small, so results should be interpreted cautiously. The key differences for these characteristics are in self-esteem

[23]See Ethiopia Demographic and Health Survey

and self-efficacy. More girls with physical disabilities reported feeling unconfident answering questions in class (21.2% vs. 5.3%). Girls with cognitive impairments were more likely to report that they feel unable to describe their thoughts to others (6.7% vs. 0.7%), and a higher proportion also reported not asking teachers questions if they do not understand something (6.7% vs. 1.1%).

For girls with physical or cognitive impairments there was also a difference in decision making. Girls with physical impairments reported at higher rates that their families make decisions about whether they will go to school (21.2% vs. 9.9%), when they will get married (36.4% vs. 13.7%), and what work they will do after finishing studying (18.2% vs. 6.2%). A higher proportion of girls with cognitive impairments also reported that their families will decide what type of work they will do after finishing studying (20.0% vs. 6.4%). Both girls with physical and cognitive disabilities were more likely to report that they do not recognise that their choices today about their studies will affect their futures (physical impairment: 6.3% vs. 0.8%, cognitive impairment: 12.5% vs. 0.8%) and that they do not try to find other ways to express themselves if they are not understood (physical impairment: 12.5% vs. 1.6%, cognitive impairment: 12.5% vs. 1.8%).

The fifth column looks at the experience of girls whose mother tongue is different to their school's LoI. The intervention sample for this characteristic is particularly small (less than 10 girls), so the reported correlations for this characteristic include girls from the control group. The table shows that girls whose mother tongue is different to their school's LoI are more likely to report facing a number of barriers. Most relevant are those related to confidence in school – they report feeling more nervous reading (80.0% vs. 40.6%) and doing maths (85.0% vs. 37.6%), and are more likely to believe that luck explains good performance in tests (89.5% vs. 53.6%). Their families are also more likely to make educational related decisions for them compared to other students. However, due to the small sample size these results should be interpreted cautiously.

The next column compares the barriers faced by girls from Christian and Muslim families (the two main religions in Amhara and Oromia)[24]. Girls from Muslim families are more likely to report facing a number of barriers than girls from Christian families. For example, girls from Muslim families report in higher proportions that seats are not available in class (9.1% vs. 3.3%), that they do not use drinking water facilities at school (58.9% vs. 37.0%), that they do not use the areas where children play and socialise (4.1% vs. 1.4%), and that they haven't used the school's reading corners in the last month (12.0% vs. 2.8%). More girls from Muslim families also report that their teacher do not make them feel welcome (3.6% vs. 0.8%), that teachers ask more (10.2% vs. 4.4%) and harder (12.7% vs. 5.8%) to boys, and that teachers have used physical punishment in the last week (83.0% vs. 62.0%). They were also more likely to have been absent from school for more than five days in a row in the last year (24.4% vs. 13.8%).

The next column looks at the barriers faced by girls living without both of their parents compared to those living with at least one of their parents. Responses across the barriers are mixed for this characteristic. On the one hand, girls living without both parents report longer journey times to school (12.5% vs. 2.4% greater than one hour to school), lower desire to continue studying after the current year (5.4% compared to 1.0%), feel less able to work in groups (3.6% vs. 0.7%) and pay less attention to others' body language (14.6% vs. 4.2%). On the other hand, girls living without both of their parents report feeling less nervous reading (21.4% vs. 35.0%) and doing maths (10.7% vs. 30.2%) in front of others, and also report a better school environment in terms of availability of seats (0.0% vs. 8.6%) , usage of drinking water facilities (21.4% vs. 46.1%), treatment in class (40.0% vs. 59.4%) and physical punishment from teachers (33.3% vs. 69.4%). This mix in responses may be because this characteristic has a differential impact on factors affecting the

[24] These reported differences may be driven by a third factor which is correlated with religion, such as region, which is not captured in these bivariate correlations.

barriers, for example reducing motivation but increasing resilience, though such hypotheses would need to be explored in further analysis.

The final column compares the responses of girls whose head of household has no education compared to some education. The correlations are also mixed on this characteristic. These girls tend to report longer journeys to school (3.8% vs. 0.5% longer than one hour), perceive girls and boys being treated differently in class (61.0% vs. 46.7%), that teachers ask more (8.2% vs. 2.2%) and harder (10.4% vs. 2.7%) to boys (though also that teachers ask more and harder questions to girls (6.3% vs. 2.2%), that their success in a test would be due to luck (56.8% vs. 47.5%), that they cannot choose whether to stay in school (66.7% vs. 53.1%) and that they spend less time reading than male siblings (38.8% vs. 29.6%). They were also more likely to explain absence from school as a result of carrying out domestic chores (67.6% vs. 41.9%) and less likely to explain it in terms of engaging in paid work (13.9% vs. 29.9%)

On the other hand, girls whose head of household have no education are less likely to report not using the school's drinking water facilities (37.5% vs. 60.7%), using the toilets at school (7.1% vs. 14.8%), not being able to organise activities with friends (0.5% vs. 2.7%), not asking teachers if they do not understand something in class (0.3% vs. 2.2%), and not paying attention to others' body language (2.5% vs. 8.9%).

In summary, the analysis reveals the presence of significant correlations between barriers and characteristics of girls. There are consistent regional differences – girls from Oromia are more likely to report facing barriers than girls from Amhara. This is also true for girls living in rural areas compared to girls living in urban or peri urban areas. There are also consistent differences between girls from Muslim and Christian families, though these results are plausibly driven by a third variable which is not captured by these correlations. As would be expected, girls with either cognitive or physical disabilities also consistently report facing greater barriers, particularly regarding self-esteem and self-efficacy. Girls whose mother tongue is different to their school's LoI also report facing more barriers, though the sample for this characteristic is small and so the correlations should be interpreted cautiously. For girls living without their parents and girls whose head of household has no education, the correlations are inconsistent across the barriers. This may reflect different responses as a result of the girls' experiences – for example, girls living without both parents may develop more resilience, and thus report feeling less nervous compared to other students. This kind of explanation would need to be tested in further analysis.

3.5 Appropriateness of project activities to the characteristics and barriers identified

3.5.1 Project identified characteristics in the data

We start by considering the marginalisation criteria which the project has identified, and which have played a key role in shaping the project's activities. The project considers five key 'marginalised groups' which it aims to target through its interventions. From Annex 4, they are the following: Girls at risk of early marriage, girls at risk of forced migration, girls with a high domestic work burden, girls in poverty, and girls with a disability.

Beginning with disability, Table 8 breaks down the sample by type of disability. It shows that 10.4% of girls in the control group and 6.1% of girls in the intervention group have at least one of the following impairments: vision, hearing, mobility, cognitive, self-care or communication. Moreover, the proportion of girls with disability in the control group is larger than the proportion of girls with disabilities in the intervention

group, for every type of impairment. In the control group cognitive and communication impairments are the most common, with 4.6% of the girls having cognitive impairments, and 3.7% having communication impairments. In the intervention group hearing, visual and cognitive impairments are among the most common ones, affecting 1.6%, 1.4% and 1,4% respectively. The treatment and control samples are slightly less well-aligned in terms of disability, but still broadly comparable. In terms of the overall representativeness of the sample with the population sample as defined in the project database, the proportion of girls with disabilities in our sample is higher than the population figures. While in Table 7 we report 6.1% with disabilities in the intervention group, this figure is estimated at 3% for the whole population.

Table 8 that 6.1% of the intervention sample have at least one of the six disability-types considered. This definition is broadly aligned with that of the project[25]. The project forecasts that 3% (500 out of 16481) of targeted girls are disabled. In relative terms, this represents a reasonable underestimate. Extrapolating from a sample proportion of 6.1%, we would estimate that 1000 girls are disabled in the population database, which is double what the project assumes.

Secondly, the project estimates that 2270 girls (13.8%) are subject to risky migration. In the girl's school survey, 3.1% of girls responded that they had thought of migrating to another area. This figure is considerably lower than the project-level estimate. Being 'at risk of migration' is positively associated with age, and the age profile of the sample is slightly older than the population meaning we may expect to observe higher prevalence of risky migration than the project. The estimate of 3.1% is, however, likely to be an underestimate as it does not include girls who have already migrated. It also does not account for girl's subject to forced migration who have themselves not considered it. It seems that the latter two effects outweigh the expected difference due to the difference in age profile. As such, we consider the project's figure for risky migration to be plausible. Subsequent evaluation points may benefit from a revision to the instruments to provide further clarity on this criterion of marginalisation.

On early marriage, the project estimates that 1265 (7.7%) of girls in the beneficiary population are at risk of early marriage. We define 'at risk of early marriage' as a girl who: Is currently married or has been in the past, has faced an attempt for an arranged/forced marriage, feels that early marriage can happen to them. This is well-aligned with the project's definition but does not capture girls who may have been abducted for marriage, or who have fled their community to escape early marriage. Under this definition, 10% of girls are considered at risk of early marriage. Based on this proportion, approximately 1650 girls are at risk of early marriage, which is slightly higher, but still consistent, in terms of magnitude, to the project's figure.

The fourth marginalisation criterion considered by the project classifies girls as marginalised if they face a high burden of domestic labour, defined as more than 5.5 hours a day. As per the project's estimates, 12690 (76.9%) beneficiaries are in this category. We align our definition with that of the project's by considering the following activities as domestic labour: caring for other family members, domestic tasks (fetching water, firewood, cleaning etc.), tasks on a family farm or business, activities for pay or money outside of the household. When considering the same threshold of 5.5 hours, the survey data concludes that 57.2% of the evaluation sample are subject to a high burden of domestic labour. This is considerably lower than the project's estimate. There are, however, several artefacts in the data that suggest the figure

[25] As per the project's GEC-T proposal, a girl is considered disabled if she experiences *"difficulties with seeing, hearing, walking, remembering, self-caring or communicating. If the girl presents any difficulty in one of those areas she is considered disabled."*

of 57% is measured with some noise[26]. For the purposes of verifying that the project' s activities are targeted correctly, it is sufficient to note that we also conclude a majority of girls are subject to a high work burden as evidenced by both quantitative and qualitative data.[27]

Lastly, the project targets girls in poverty, which is assumed to be the entire beneficiary population. The project has a clear definition for poverty based on girls whose families are defined as 'struggling' or 'impoverished' on the basis of family members' employment, income, assets (land, farm animals, in-house assets), housing and whether a father is present as a caregiver. Girls whose families are in a third category 'doing-well' are not offered assistance by the project. Using the household survey data, we attempt to align our definition with that of the project's, as laid out in Table 9. Using this definition, 87% of households are defined as in poverty. Given that we lack accurate information on household heads' professions and incomes[28], we cannot comment on these two important aspects of poverty considered by the project. As poverty is also both an absolute and relative concept, some households who we would define as 'not in poverty' by virtue of being able to afford some or many non-essential goods may not consider themselves in poverty given comparisons to others around them but may be defined as in absolute poverty per national (and international standards). Given what the quantitative data finds, it seems reasonable to assume that an overwhelming majority of girls' households are in poverty, even though this figure may be slightly less than 100% depending on its exact definition.

The above analysis suggests that the marginalisation criteria outlined by the project are present in the evaluation sample to an extent which is broadly comparable with the project's own estimates. There may however be further marginalisation criteria which the project has not considered in its beneficiary mapping and when designing its supporting activities.

3.5.2 Unsupported groups at risk of marginalisation

One potential group at risk of marginalisation identified by the quantitative data are girls whose mother tongue differs to the language of instruction in schools. Table 9 shows that this applies to around 1.1% of girls in project schools, and around a tenth (0.1% of the total sample) of these do not speak the language of instruction at all. The extent to which this may present a barrier is difficult to assess. Analysis of the interactions between characteristics and barriers show that girls whose mother tongue is different to the language of instruction are more likely to become nervous when reading (80% vs 41%) and doing maths (85% vs 38%) in front others. This is suggestive of a barrier to communication and lower self-esteem and may have a knock-on effect on learning and transition outcomes.

[26] For example, it seems many respondents (and enumerators) interpret the time component to refer to hours, but there are a minority who interpret it is as minutes. Where we observe values greater than 24, we divide by 60 to convert the minute-level measure to hours. For values below 24, we cannot know whether the value refers to minutes or hours, but we assume that a girl at most spends 14 hours on activity. By treating all values below 14 as hours, we may overestimate the true proportion of burdened girls.

[27] The chore burden placed on girls is mentioned in key informant interviews of community representatives, boys and girls. For example, a community representative in Dodota, Arsi states: "Most of the time girls do not attend classes as often as boys. The reason is girls are engaged at home in different businesses/works. Because of this, girls could do not do their homework and come on time to school. This also leads them to be absent from school. Some could not follow their lesson effectively and this leads them to choose not to attend classes."

[28] The profession variable was coded incorrectly rendering it unusable. Income data was not collected.

If we assume that 1% of the beneficiary population have a mother tongue different to the language of instruction, and that for half of these girls, their knowledge of the language of instruction is weak enough to restrict their ability to learn, then around 80 beneficiaries may be affected by this. This is a small, but potentially marginalised cohort of the beneficiary population.

3.5.3 Project & evaluator identified barriers and how these map to the Theory of Change

Alongside the population characteristics outlined above, the project also identifies the following key barriers as part of its theory of change (referred to as ToC Barriers below):

1. Attitudes and identities biased against girls' education
2. Concerns about girls' safety
3. Girls' families insecure livelihoods mean they struggle to meet costs of education
4. Long distances to secondary schools
5. Poor quality teaching
6. Poorly designed teacher and learner resources
7. Low capacity of Woreda level educational departments to support and supervise teacher's development

Next, we analyse the extent to which barriers suggested by the project are present in the data, and discuss other barriers identified by the evaluation which the project may not have considered.

Based on analysis carried out to this point, the extent to which attitudes and identities bias against girls' education *(ToC Barrier 1)* is unclear. A majority of girl's state that teachers treat girls and boys differently suggesting that gender is still a relevant factor in determining how pupils are treated, though the extent to which this translates into a bias against girls is unclear. For example, girls and boys are treated equally in terms of the number of questions they are asked in class[29]. Qualitative data also shows that parents generally view girls' education positively, with exceptions to this being a small minority.

Concerns about girls' safety *(ToC Barrier 2)* appear to be justified as a relevant minority of students report feeling unsafe on their journey to school (8%), and in school (3%). Qualitative data supports this as a minority of girls report feeling unsafe, particularly those living far from school[30]. Insecure livelihoods which mean girls' families struggle to meet the costs of education *(ToC Barrier 3*) are likely to be a significant barrier, as one fifth of households find it difficult to afford for girls to attend school, even with the project's support. Long distances to secondary schools *(ToC Barrier 4)* are a significant problem in project areas, with over 40 percent of households stating that they live more than an hour's walk from the nearest secondary school. This figure is considerably lower for primary schools at 2 percent but may present a barrier to attending even primary school for a small subset of the population.

A very small portion of girls (0.1%) report that they receive poor quality teaching[31] *(ToC Barrier 5).* This suggests that girls' satisfaction with their quality of teaching is good. Whether their subjective assessment of their teachers' would align with a more objective assessment of the teachers' didactic methods remains unclear, so we cannot rule out that this is a significant barrier. Moreover, the analysis to this point does not

[29] Discussed as part of **Table 29**

[30] An 11-year-old girl from Dengors primary school is quoted as saying "Yes I feel safe here [in school] but afraid of coming to school alone since it is far from our village. It is not safe for anyone to come alone."

[31] See section 5.2

allow us to assess whether poorly designed resources *(ToC Barrier 6)* and low capacity of Woreda level educational departments *(ToC Barrier 7)* are significant barriers in project areas.

From the analysis thus far, we have identified that the most prevalent barriers and characteristics likely to impact learning and transition outcomes are as follows (mapped to ToC barriers and/or marginalisation criteria where there is a link):

1. Frequent teacher absences from class (potentially correlated with *ToC Barrier 5)*
2. Non-usage of in-school amenities such as toilets and drinking water facilities *(possibly linked to ToC Barrier 2)*
3. Having a high chore burden (*Marginalisation Criteria 4)*
4. Feeling that boys and girls are treated differently in class *(ToC Barrier 1)*
5. Remoteness of secondary schools *(ToC Barrier 4)*
6. Household is poor and/or finds it difficult to afford schooling *(ToC Barrier 3)*

The key barriers and subgroups identified in the analysis have a strong overlap with the barriers and characteristics identified by the project as part of its theory of change and beneficiary mapping. One exception to this is the barrier experienced by girls whose teacher is frequently absent from classes. This is not identified as a barrier to learning and transition outcomes by the project but is highly prevalent in the data.

3.5.4 Project interventions and key barriers

Having identified the key barriers and subgroups, and the interactions between these, as well as their overlap with the barriers and subgroups identified by the project, we now look at the extent to which the projects' interventions are likely to address these.

The project identifies attitudes and identities which are biased against girls' education as a key barrier, for which there is some evidence in the data collected. A key focus was placed on this barrier, and several interventions were designed to address it, including the following: An anonymous reporting system for addressing violence and abuse (known as 'Letter link boxes'), girls and boys (Good Brothers') clubs which include discussions of issues faced by girls and education of girls on issues such as early marriage, and theatre and drama performances to raise awareness of gender- and disability-based inequalities. These interventions directly address the issue of bias against girls' education, and in doing so also attempt to deal with potential stigmas towards disabled girls and the education of disabled girls. We do note however that direct support for disabled girls' learning appears to be lacking from the programme's suite of interventions.

From qualitative data, the number of students using letter link boxes seems to be low, possibly as they do not understand the benefit of the boxes. There is, however, evidence of students using the boxes to communicate cases of early marriage in their community, and the girls that do use them state that it helps them to freely express their feelings, concerns and challenges and that it is friendly and easy to use. As such, emphasis should be placed on raising awareness of letter link boxes as their use does seem to be effective.

The qualitative data also shows that many interviewees were members of girls' clubs, as club members and teachers promoted the benefits of the club. Some respondents state that being a member of the girls

club has changed the way they feel about going to school in a positive manner[32], and helped to advise girls on how to use their time properly, and how to support each other in school. There is, however, suggestive evidence of limits placed on intake into girls' clubs that may limit their impact[33]. The 'Good Brothers' Clubs' are also well-attended according to qualitative data. Boys report joining the clubs to study together with other club members, to gain knowledge, and to help girls in school.

If the domestic work burden placed on girls is also driven by differing expectations placed on them by their family compared to boys, then changing parental expectations will work to reduce this. As part of Girls' clubs, girls produce advocacy and communication materials to address negative behaviours and attitudes in their families and communities. This is the only intervention which targets community and familial attitudes directly.

Concerns about girls' safety, reflected in the data by a minority of girls feeling unsafe during their journey to school, and when in school, is a further key barrier addressed by the project. By helping girls to report instances of violence and abuse, the letter-link boxes also promote girls' safety, as do the boys clubs if boys behaviour contributes to girls feeling unsafe in school. It is unclear how the project's interventions address girls feeling unsafe during their journey to school.

The barrier of households being unable to afford the costs of education is targeted by several project interventions. Some costs associated with education such as registration fees, scholastic materials, uniforms and transport are subsidised, thus reducing the burden placed on girls' families which may serve to lower transition rates. By subsiding transportation costs (to eligible grade 11 and 12 students), the barrier presented by the distance to secondary school is also reduced, as it makes those schools more accessible for girls.

Though we have no evidence to suggest that teaching quality is poor, as perceived by girls, the project targets this as a key barrier, and the subjective nature of the evaluations' data in this area means it cannot be discounted as a potential barrier. The project has a number of interventions targeted at improving teacher quality, which include the following: training of cluster supervisors to carry out lesson observations and mentoring and training of 'master trainers' in specific pedagogies for teaching math and literacy which cascades down to all GEC-T teachers.

A potential gap in the project's interventions may pertain to the barrier presented by frequent teacher absences, reported by girls. The project does incentivise teacher attendance directly by providing financial compensation for teachers who deliver homework clubs, and indirectly through training and career development, but neither of these target day-to-day attendance for all project teachers directly.

Lastly, whilst we cannot evaluate the extent to which teacher and learning resources are poorly designed, nor the capacity of Woreda educational departments to support and supervise teacher's development, we can assess how well these potential barriers are supported by the project. To improve teacher and learning

[32] In Oromia, a grade 6 student responded, "The attitude I have towards my classmates in particular and staying in general in school is totally changed [by the girls club]." Another grade 6 student from Oromia stated that "After I became a member of the club, my attitude of going to school is positively changed… because of the advice I got from teachers and students who are members of the club"

[33] In Amhara, a KII interviewee stated that "during registration time [for the girls' club], the CHADET focal teacher asked us to raise of hands if we need to be members of girls club, then all girls in our class raised their hands. She registered all students until she reached the number she wanted and left the rest of us."

resources, offline digital literacy and numeracy resources are installed in project-sponsored ICT Labs in secondary schools to facilitate self-access to math and literacy resources. To improve Woreda-level capacity in supporting and supervising teacher's development, the project trains cluster supervisors to carry out lesson observations and mentor teachers.

In summary, the range of interventions put in place by the project seem to be broadly aligned with the barriers identified both by the project, and the EE. For certain barriers, the range of interventions seems to be limited, and in a minority of cases does not seem to be present. Direct support for girls identified as having disabilities seems to be missing, and interventions aimed at changing attitudes and behaviours with communities are also limited.

Box 2: Project's contribution

The project should respond to the external evaluators comments on the above questions. In particular the project should respond to:

• Why the evaluators sample characteristics may differ from any mapping the project has done fir its wider beneficiary population.

• Why the projects theory of change may not correspond with some of the key barriers identified.

• Whether the project plans to review their Theory of change in light of these findings.

Why the evaluators sample characteristics may differ from any mapping the project has done fir its wider beneficiary population?

Disability: In addition to the higher number of girls that have been identified with disabilities than specified in the project proposal, we also notice that there are marked differences between the treatment and control groups on cognitive impairment and communication impairment which we are interested to investigate further. This will have an impact on how we tailor our support interventions to GWD.

A possible reason for an increase in numbers of GWD identified at baseline may be due to a more realistic representation of the full beneficiary cohort has been reflected through the Washington Group Questions, in an environment where respondents feel more comfortable to respond to the questions asked.

Why the projects theory of change may not correspond with some of the key barriers identified

Teacher absenteeism – Data collected regarding teacher absenteeism relates to general lessons, not the project intervention tutorials/homework sessions. Teacher absenteeism has also not been included as a main barrier in our TOC and is a national issue, especially with rural schools (in relation to teachers who deliver general lessons), interventions particular to GEC will only be able to address factors within our control as we do not have jurisdiction to reduce teacher absenteeism. Socio-economic factors, such as increases in teacher incentives outside of GEC, the development of private school provision or government driven interventions may still pose as the barrier of teacher absenteeism. Currently we drive incentives for teacher attendance by increasing their motivation through strong relationship building within the school and between CHADET's education officers project co-ordination offices, through school-based communities of practice and post lesson observation feedback. In addition to this, Education Officers work closely with school principals and the project leadership and management training looks to address teacher absenteeism alongside other school leadership and management areas (listed as activity 3.9 in the original project proposal). We are also developing the leadership capacities of school masters, Woreda education officers and supervisors.

Non-usage of in-school amenities such as toilets and drinking water facilities – The reason for minimal use of toilets in school is due to lack of access to water which is an external issue to GEC-T, often out of the control of the school and surrounding area.

Language of instruction different to mother tongue - For the 1.1% of girls whose mother tongue is different from the language of instruction, where some 80 girls may be negatively affected, so that all girls facing this barrier can be supported, we are interested to know the exact number of girls and what geographical areas this pertains to, so further characteristics or a trend can be identified. For example, do these girls all speak the same mother tongue, or different - as there are as many as 90 different ethnic groups in Ethiopia. The reasons for girls who are more likely to become nervous when reading (80% vs 41%) and doing maths (85% vs 38%) in front others when reading in their L2 (second language) needs to be further explored and cannot solely be linked to barriers in communication and lower self-esteem. Lack of practice in reading aloud and/or testing stress could also be a factor. We also note that there is a large difference between treatment and control group responses which is surprising, given our teacher development intervention is developing engagement levels of students through a more student-centred approach in learning. Once we know more about this category we will be able to factor in additional support. All girls who have LOI as a barrier will need intervention.

'Concerns about girls' safety, reflected in the data by a minority of girls feeling unsafe during their journey to school, and when in school, is a further key barrier addressed by the project …. It is unclear how the project's interventions address girls feeling unsafe during their journey to school' - the report says it is unclear how the project interventions address girls feeling unsafe during their journey to school. Below is our recent insert from our Q7 report that directly relates to the comment.

The life skills education and girls' and boys' clubs' activities have increased the awareness of girls and boys about safety on the way to and in school. Both girls and boys are able to think more deeply about their safety when they travel from home to school, and from school to home. For example, girls are going in groups, along with boys' club members knowing 'bad people' may be on the route. Girls are very much aware of possible risks and carefully watch their surroundings. They are also becoming more confident to control situations if something should happen.

The impact of the girls' and boys' clubs intervention has led to many cases of boys escorting girls to and from school, or networking with the local community to keep a watchful eye on girls as they travel to school.

TOC barrier 1 – ***The extent to which attitudes and identities bias against girls' education is unclear.*** We know that this is a factor but how much it translates into bias has been identified as unclear. This may be due to the limitation of items within the tools that have identified these factors, or other factors. Further discussion needs to take place regarding how this can be assessed internally, with in-project tools.

TOC barrier 4 – Long Distances to Secondary Schools. We are aware that long distances to secondary schools pose a significant problem which is addressed by interventions and support under output 2 covering 'assistance with transportation costs' & 'accommodation costs for girls in secondary schools'. We are also aware that this figure is considerably lower in primary schools.

TOC barrier 5 – **Poor Quality Teaching**. We query whether respondents are sufficiently qualified to make this judgement where beneficiaries have not been subjected to other teaching methodologies, apart from the methodology that has been prevalent for years in Ethiopia plus our recent revised approaches. We question how beneficiaries are equipped to make this judgment. 'Satisfaction' does not imply 'good', or that

'change may be needed or wanted'. Further discussion regarding how we can explore this in project is needed through the development and triangulation of in-project tools.

TOC barrier 6&7 – 'Poorly designed teacher and learner resources & Low capacity of Woreda level educational departments to support and supervise teacher's development. More work in-project needs to take place to investigate barriers 6&7 'poorly designed resources' and 'low capacity of Woreda level educational departments.' Action planning sessions take place to develop each area's own plan in liaison with CHADET's project co-ordination office's education officers. Education officers work directly with Woreda education departments over time.

IT labs stationed at Woreda offices not only support teachers through using specific resources but also develop the direct link between teachers and Woreda education departments.

Responding to both questions 1 & 2

Point 2 – Migration - It has been identified that further characteristics of migration are prevalent in the data under the 'not considered' criteria, which has already been factored into our cohort under the 'at risk of migration' category. However, the tools that have been used to capture the response, as has now been identified, do not lend themselves to capture responses for girls who have already migrated or girls who have been subjected to forced migration (within the 'not considered' criteria). As we see these as barriers within the sub category of migration, revision to the instruments to capture migration behaviour will need to be considered for the next two evaluation points so that we can be confident of more solid data that captures characteristics of migration. We will also be discussing the project's response via intervention to these additional (sub category) barriers. Currently tracking 'girls for truancy' and 'girls who are risk of drop out' are successful interventions and both the girls' and boys' clubs are active in awareness raising and protecting girls of external dangers.
Point 3 – Items in the baseline tools covering 'girls who have been abducted for marriage' or 'who have fled their community to escape early marriage' have not been included in the tools. Further revision is required here to assist beneficiary mapping.

Responding to Question 3 – Do we plan to review their Theory of change in light of these findings

Support to enable disabled girls learning: Girls with disabilities, as a marginalised sub group, have been included in the TOC in the areas of raising awareness around disability through the delivery of drama performances targeted at impacting on family and community attitudes and to support girls with disabilities (originally detailed in activity 1.8 in the project proposal). Until now the project has focused on raising awareness on disability and identifying the level of support needed so that we are targeted in offering assistive devices and support with transition costs. Reviewing full baseline findings will enable us to refine these activities but we do not feel the TOC needs to be amended in relation to this area. We are also in the process of considering a SEN training workshop for CHADET Education Officers, who can disseminate learning to teachers.

'If the domestic work burden placed on girls is also driven by differing expectations placed on them by their family compared to boys, then changing parental expectations will work to reduce this. As part of Girls' clubs, girls produce advocacy and communication materials to address negative behaviours and attitudes in their families and communities. This is the only intervention which targets community and familial attitudes directly' - Building on from GEC, the focus in GEC-T has concentrated its interventions more around schools rather than community and has put the emphasis on the Girls' Clubs activities to address negative behaviours of family and community members, already factored into our TOC. There is discussion currently regarding how the re-introduction of family hubs and community conversations that existed in GEC

can further shift the attitude of parents towards girls' education but with a slightly different focus on the girls driving these initiatives. Further work also within the Girls Clubs aims to make their work more explicit (while also considering 'Do No Harm').

4. Key Outcome Findings

4.1 Learning Outcome

The first primary outcome of this project is concerned with improved literacy and numeracy skills and targets the number of girls supported by GEC-T who then secure improved learning outcomes. This is being measured by the number of supported girls who secure improved EGRA/SeGRA results in the case of literacy and improved EGMA/SeGMA in the case of numeracy.

The EGRA and EGMA are widely used in assessing educational progress in developing countries. They have been designed to assess basic literacy and numeracy skills and to identify any gaps in students' knowledge which need to be addressed[34].

For students in higher grades the SeGRA and SeGMA are used[35]. These assessments have been designed specifically for the GEC to assess students' learning as they progress to higher grades. The tasks are intended to progress beyond the EGRA and EGMA to assess competencies acquired at higher grades.

The specific tasks within these tests have been adapted in light of guidance provided by the Fund Manager for GEC-T concerning language and version requirements.

Provision of the tests by enumerators and completion of the tests by students were piloted in February 2018. The pilots took place in both of the project's regions and included three sets of tests to ensure comparability between baseline, midline and endline results. The pilot enabled the EE to communicate feedback on the suitability of the tests to the Project, facilitating necessary changes.

The final test administered included two sets of literacy and numeracy tests – EGRA and EGMA, taken by all students (grades 4 to 8) and SeGRA and SeGMA tests taken by students in grades 7 and 8 only at baseline. Table 12 shows a full breakdown of the tasks administered to students in each grade, along with the testing language used.

Table 12: Task-Language breakdown by Grade

Grade	EGRA/EGMA	SEGRA/SEGMA	Language
4	All	N/A	Local language
5	All	N/A	Local language
6	All	N/A	Local language
7	All	Task 1 (both)	Local Language for EGRA/EGMA and English for SeGRA/SeGMA 1
8	All excluding EGRA/EGMA 1	Task 1 (both)	Local Language for EGRA/EGMA and English for SeGRA/SeGMA 1

[34] GEC MEL guidance part 2, p21-28
[35] GEC MEL guidance part 2 p29-32

9, 10, 11	EGRA 3 & EGMA 6	SEGRA 1, SEGMA 1 & 2	English

EGRA and EGMA were scored 'on-the-fly' by enumerators, with built-in capacity on the tablet software used to incorporate a child correcting themselves. The marking scheme was designed according to EGRA/EGMA guidelines. As SeGRA and SeGMA are written tests, they were scored out of the field. The tests were produced with a marking scheme which was revised following the pilot. The marking scheme allocated points according to the difficulty of each question and allowed for partial credit if working out was shown. The enumerators grading the tests were instructed to give full marks to a pupil if the correct answer was given, even if no (or limited) working out was shown.

4.1.1 Early Grade Reading Assessment (EGRA)

EGRA is the literacy learning assessment which was taken by students in all grades in the sample. The version given to the students consisted of four subtasks:

- letter sound identification, which requires students to identify the sound of written letters
- invented sound identification, which requires students to read a list of made up words
- oral passage reading, which requires students to read a short story aloud
- reading comprehension, which requires students to read a short story and then answer five questions about it.

Due to errors in the coding of the oral passage reading subtask, this has been excluded from the aggregated scores in Table 13.[36] A simple average of the percentage scores on each of the other three subtasks was taken to calculate the overall average EGRA scores presented in Table 13.

The scores in Table 13 are acceptably similar between intervention and control groups providing initial indication that matching of the two groups based on background characteristics has served to reduce learning differences between the groups. However, the scores do not reflect the level of progression across grades that might be expected. This lack of grade progression is consistent across regions. Explanations for this, including the possibility of a cohort effect for grade 7, will be explored in subsequent sections.

Table 13: Literacy (EGRA)

Grade	Intervention Group Mean	Control Group Mean	Standard Deviation in the intervention group
Grade 4	75.7	78.1	17.4
Grade 5	73.7	74.3	19.5
Grade 6	71.4	75.1	21.1
Grade 7	74.8	74.4	19.9
Grade 8	74.1	76.9	20.2

Notes: Scores exclude subtask 3 – oral passage reading – due to discrepancies in scores.
The observed lack of grade progression is consistent across regions (analysis available upon request)

[36]Coding of the EGRA 3a task was carried out inconsistently with both raw WPM scores and converted percentages being coded in the same variable. As such, it is not possible to distinguish between the reported score representing a WPM or a percentage for scores equal to and below the total numbers of words in the task

4.1.2 Secondary Grade Reading Assessment (SeGRA)

SeGRA is the literacy learning assessment which was taken by students in grades 7 and above. The version used consisted of one task - a more difficult comprehension task. To obtain the combined EGRA/SeGRA scores in Table 14, the simple average of the percentage scores on the three EGRA subtasks (excluding oral passage reading) and the SeGRA task was calculated.

Table 14 shows that there is an approximate balance between the treatment and control groups though there is a slight difference in grade 8 scores, and there is little difference between combined scores of students in grades 7 and 8. The standardised scores are lower in Table 14 than in Table 13 as students' SeGRA scores are considerably lower than their EGRA results, as is expected given the greater level of difficulty of the secondary tests.

Table 14: Literacy (EGRA/SeGRA combined)

Grade	Intervention Group Mean	Control Group Mean	Standard Deviation in the intervention group
Grade 7	61.5	60.1	15.2
Grade 8	60.3	63.2	15.4

*Notes: Scores exclude subtask 3 – oral passage reading – due to discrepancies in scores [due to language of completion]

4.1.3 Early Grade Mathematics Assessment (EGMA)

EGMA is the numeracy learning assessment which was taken by students in all grades in the sample. The version given to the students consisted of six subtasks:

- number identification, which involves identifying written numbers
- quantity discrimination, which involves determining the relative size of pairs of numbers
- missing numbers, which involves determining a missing number in a group of numbers
- addition, which involves a set of addition questions
- subtraction, which involves a set of subtraction questions
- written exercises, which involves a set of more difficult arithmetic questions

A simple average of the percentage scores on these tasks was taken to determine the overall average EGMA scores presented in

Table ***15***.

Table 15 show a balance between intervention and control groups, but again there appears to be a lack of progress across grades, which will require further investigation. There is a drop off in grade 8 which is

driven by a lower average score for girls in South Wollo (intervention group: 61.8, control group: 70.2). The overall lack of grade progression is again consistent across regions.

Table 15: Numeracy (EGMA)

Grade	Intervention Group Mean	Control Group Mean	Standard Deviation in the intervention group
Grade 4	68.1	69	11.5
Grade 5	67.1	67.6	12.4
Grade 6	66.6	67.2	11.3
Grade 7	66.7	65.9	12.1
Grade 8	64.5	68	13.6

*The observed lack of grade progression is consistent across regions (analysis available upon request)

4.1.4 Secondary Grade Mathematics Assessment (SeGMA)

SeGMA is the numeracy assessment which was taken by students in grades 7 and above. For the purposes of the baseline, it consisted of one additional task:

- advanced multiplication, division, etc., which consists of a set of more difficult arithmetic and geometry questions

The combined EGMA/SeGMA scores presented in Table 16 were calculated by taking the simple average of the percentage scores on the six EGMA subtasks and the two SeGMA subtasks.

Table 16 shows the intervention and control groups are approximately balanced but there is a slight reduction in scores from grade 7 and 8 in the intervention group. This is not the case for the control group, for which there is a slight progression in scores. Note that scores in Table 16 are lower than for the equivalent grades in Table 16, which as above is because students perform less well on the SeGMA than EGMA tests, which lowers their overall average grade in the combined table.

Table 16: Numeracy (EGMA/SeGMA combined)

Grade	Intervention Group Mean	Control Group Mean	Standard Deviation in the intervention group
Grade 7	61.7	60.9	9.8
Grade 8	57.5	62.3	10.4

4.1.5 Benchmarking sample

The benchmarking sample consists of students in grades 9 to 11. These students took the SeGRA and SeGMA subtasks taken by students in grades 7 and 8, and an additional SeGMA subtask on algebra ('SEGMA Subtask 2'). The low average test scores can be partly explained by the switch in testing language from local languages to English, which coincides with the switch in the language of instruction in the Ethiopian education system.

The table below shows the grade breakdown of the SeGRA and SeGMA tests taken by the benchmarking sample. It also includes students in grades 7 and 8 for comparison. While previously we did not find the expected progression across grades, the table shows that there is progression in test scores when the secondary learning tests are looked at separately. In each of the subtasks there is a positive trend between grade and test score with the exception of grade 10 to 11 for SeGRA subtask 1.

Table 17: SeGRA and SeGMA test scores (%)

Categories	SeGRA subtask 1	SeGMA subtask 1	SeGMA subtask 2	SeGMA total
Grade 7	15.9	22.8	n/a	n/a
Grade 8	18.5	24.1	n/a	n/a
Grade 9	19.9	30.9	13.2	22
Grade 10	37	41.1	13.1	27.1
Grade 11	31.2	42.8	18.9	30.9

4.1.6 Foundational literacy and numeracy skills gap

Table 18 and Table 19 below assign girls in the intervention group into four bands ranging from non-learner to proficient learner based on their performance on each of the literacy and numeracy subtasks. Based on the GEC-T MEL guidance (part 2), these bands are defined as:

- 0% - non-learner
- 1 to 40% - emergent learner
- 41 to 80% - established learner
- 81% and above – proficient learner

Table 18: Foundational numeracy skills gaps

Categories	Subtask 1 (EGMA1) Number Identification	Subtask 2 (EGMA2) Quantity Discrimination	Subtask 3 (EGMA3) Missing Numbers	Subtask 4 (EGMA4) Addition	Subtask 5 (EGMA5) Subtraction	Subtask 6 (EGMA6) Word problems	Subtask 7 (SeGMA 1) Advanced multiplication, division etc.
Non-learner 0%	0%	2.9%	3.6%	1.1%	4.0%	28.0%	14.0%
Emergent learner 1%-40%	0.4%	2.8%	47.2%	6.9%	13.3%	54.9%	76.3%
Established learner 41%-80%	5.4%	22.2%	37.3%	36.2%	49.0%	13.2%	9.7%
Proficient learner 81%-100%	94.2%	72.1%	11.9%	55.8%	33.7%	3.9%	0%

100%	**100%**	**100%**	**100%**	**100%**	**100%**	**100%**

Table 18 shows the proportion of students in each band for the numeracy subtasks. As can be seen for Subtask 1 (EGMA1) the majority students are performing at the highest band. There is a general downward trend as the tasks get more difficult. Subtask 3 (EGMA3) is a slight exception, with fewer students in the highest band than expected based on the trend. We will explore the possible reasons for this to determine whether there is a gap in this subject area that needs to be addressed.

There are no floor or ceiling effects for EGMA or SeGMA. For the hardest EGMA subtask (EGMA6) 28.0% of students performed at the lowest band and only 3.9% of students performed at the highest band. This suggests there is scope for improvement in this test at midline and endline. For SeGMA, no student performed in the highest band. However, the majority of students did correctly answer some of the questions on the subtask, suggesting there is scope for improvement in this test as well.

Table 19: Foundational literacy skills gaps

Categories	**Subtask 1 (EGRA1)** Letter Sound Identification	**Subtask 2 (EGRA2)** Invented word	**Subtask 3* (EGRA3a)** Oral passage reading	**Subtask 4 (EGRA3b)** Reading comprehension	**Subtask 5 (SeGRA1)** Advanced comprehension
Non-learner 0%	0.3%	0.6%	n/a	15.3%	33.8%
Emergent learner 1%-40%	3.2%	6.0%	n/a	26.8%	53.6%
Established learner 41%-80%	19.7%	37.4%	n/a	37.7%	12.6%
Proficient learner 81%-100%	76.8%	56%	n/a	20.2%	0%
	100%	**100%**	**n/a**	**100%**	**100%**

Notes: Scores for subtask 3 – Oral passage reading – excluded due discrepancies in scores

A similar pattern can be seen in Table 19 – more pupils are proficient in the earlier subtasks and performance levels fall as the tasks get more difficult. There are no students in the highest band for Subtask 5 (SeGRA1). Some students in grades 9 to 11 did perform at the highest level, suggesting that progress in this subtask improves at grades beyond the core sample, and that there are no floor or ceiling effects for SeGRA. This is also the case for EGRA – there is no grouping for Subtask 4 (EGRA3b) at either the highest or lowest band.

4.1.7 Grade level achieved

To determine the grade level achieved, the following table maps subtasks to the grades by which a girl should be able to complete them by, as per the Ethiopian national curriculum.

Table 20: Grade-subtask mapping[37]

Subtask name	Skill tested	Grade to be achieved by

37 Mapping is based on the grades for which the tests were designed (based on the Ethiopian curriculum)

EGRA 1	Letter sound identification	4
EGRA 2	Invented sound identification	4
EGRA 3	Oral passage reading + reading comprehension	4 (Established Learner), 5 (Proficient Learner)
SEGRA 1	Longer comprehension passage in English	7[38] (Established Learner), 8 (Proficient Learner)
EGMA 1	Number identification	4
EGMA 2	Quantity discrimination (x>y)	4
EGMA 3	Missing numbers (in a sequence)	4
EGMA 4	Addition	4
EGMA 5	Subtraction	4
EGMA 6	Written exercises (more complicated arithmetic)	4 (Established Learner), 5 (Proficient Learner)
SEGMA 1	Advanced multiplication, division, geometry	7 (Established Learner), 8 (Proficient Learner)

Based on the mapping in Table 20 we have estimated the grade levels achieved by each girl. These are summarised, by grade, in Table 21 and Table 22 below.

Table 21: Grade Level Achieved - Literacy

Categories	Grade level - less than 4	Grade level 4	Grade level 5	Grade level 6	Grade level 7	Grade level 8
Grade 4	57.4%	23.3%	19.3%	0.0%	0.0%	0.0%
Grade 5	65.4%	20.0%	14.6%	0.0%	0.0%	0.0%
Grade 6	62%	23.8%	14.2%	0.0%	0.0%	0.0%
Grade 7	60.4%	22.0%	16.1%	0.0%	1.6%	0.0%
Grade 8	56.4%	23.5%	18.6%	0%	1.5%	0%

Notes: Scores for subtask 3 – Oral passage reading – excluded due discrepancies in scores

Table 21 shows the Grade Level Achieved for literacy. The lack of grade progression is observed again – girls in higher grades are no more likely to have achieved higher grade levels than girls in lower grades. The table shows that the majority of girls, around 60%, are performing at lower than grade level 4. Between one-fifth to one-quarter of girls are performing at grade level 4. Most of the remaining girls are performing

38 There were floor effects in SEGRA 1 and SEGMA 1 in grade 6 meaning it was decided to introduce these as of grade 7.

at grade level 5. While a handful in grade 7 and 8 are performing at grade level 7, generally the girls did not achieve higher than grade level 5 in literacy.

Table 22 Grade Level Achieved - Numeracy:

Categories	Grade level - less than 4	Grade level 4	Grade level 5	Grade level 6	Grade level 7	Grade level 8
Grade 4	99.7%	0.3%	0%	0%	0%	0%
Grade 5	99.7%	0.3%	0%	0%	0%	0%
Grade 6	99.3%	0.7%	0%	0%	0%	0%
Grade 7	99.6%	0.4%	0%	0%	0%	0%
Grade 8	100%	0%	0%	0%	0%	0%

Table 22 shows the Grade Level Achieved for numeracy. The lack of grade progression is observed again – girls in higher grades are no more likely to have achieved higher grade levels than girls in lower grades. The table shows that almost all girls performed at lower than grade level 4. Less than 1% of girls achieved grade level 4, and no girls achieved above grade level 4.

This underperformance can be explained by two factors. First, girls struggled significantly with EGMA3 (missing numbers). To achieve grade level 4, girls must achieve at least 80 on this sub-task (equivalent to being categorised as a proficient learner, see Table 18). Only 11% of girls achieved this score. However, excluding this subtask from the analysis only marginally improves the number of students achieving grade level 4. The second factor is the number of tasks where good performance is needed to achieve above grade level 4. Students must achieve proficient learner status on each of EGMA1 to EGMA5. While many students achieved proficient learner status on these tasks individually, very few students were able to achieve this level consistently across the five subtasks.

4.1.8 Reflections on learning test data (Adequacy of control group etc.)

When comparing treatment and control group means for the learning test scores in Table 13 and

Table 15, we see that the control group broadly matches the treatment group in terms of test scores. Scores are, on average, slightly higher in the control group for all grades except grade 7, though not all of these differences in scores are large enough to be statistically significant. It may be the case that learning levels are slightly higher in the control group, though we cannot conclude this categorically. The lack of progression in learning test scores across grades may be due to the data only being available on the subtask level, which would mask any systematic errors in aggregation from item to subtask level. This is contrary to the pilot data which did show progression across grades, where data was available on the item level and aggregation was carried out by the EE post-data collection. If item level data cannot be recovered, pilot data could be used for the purposes of target setting for midline. Subtask EGRA 3a was also graded inconsistently, possibly due to the introduction of a modified subtask after the pilot, meaning it could not be

used in the construction of learning test scores. At midline, clarity will need to be sought with enumerators to ensure consistent marking of EGRA 3a.

4.2 Subgroup analysis of the Learning Outcome

Table 23: Learning scores of key subgroups

	Average literacy score (aggregate)	Average numeracy score (aggregate)	Sample size
All girls	73.8	66.7	721
Living without both parents	74.6	66.5	44
Region: Oromia	72.4	65.9	217
Region: Amhara	74.3	67.0	497
Mother tongue different to LoI	84.7	71.9	5
Physical impairment	69.2	67.8	26
Cognitive impairment	68.9	65.5	12
Rural	73.3	66.7	446
Religion: Muslim	74.1	65.7*	205
Religion: Christian	74.0	67.7*	367
HoH has no education	72.6	66.8	367

*Notes: *, **, *** indicates t-test for difference in means of scores statistically significance at the 10%, 5% and 1%.*

The table shows that across all girls in the treatment group the average standardised literacy score is 73.8 and the average numeracy score is 66.7. The table shows that across the sub-groups identified as important in understanding marginalisation and barriers faced by girls, there is only one significant difference in learning test scores. This is for religion, though the average score is only marginally different (girls from Muslim families: 65.7, girls from Christian families: 67.7). However, this is expected to occur by chance when this number of significance tests are carried out (since this is significant at the 10% level, we would expect to find a difference of this size by chance in 1 out of every 10 tests carried out).

Contrary to expectation, the scores for girls whose mother tongue is different to their school's LoI. However, the difference is not statistically significant and is likely driven by the small sample size (five girls) which is not representative of girls in this group.

The consistency in scores across the sub-groups is less likely to be explained by these characteristics having no effect on girls learning test scores, and more likely to be due to data issues which also explain the lack of progress across grades (see Table 13 to Table 16). The breakdown in grades by sub-group will need to be reassessed at midline and endline once these data issues have been resolved.

Table 24: Learning scores of key barriers

	Average literacy score (aggregate)	Average numeracy score (aggregate)	Sample size
All girls	73.8	66.7	721
More than one hour to get to school	78.4	65.9	19
Books and learning material not available at school	73.7	68.0	86
Computers not available at school to use	73.8	66.6	567
Seats not available for all students in class	69.5	64.3	54
Not able to move around school easily	62.9**	65.4	21
Doesn't use drinking water facilities	73.0	66.2	291
Doesn't use toilets at school	73.6	68.7	59
Doesn't use areas where children play / socialise	66.8	63.8	19
Doesn't feel safe travelling to and from school	85.7*	75.7**	7
Hasn't used school's library or reading corner in last month	72.5	65.4	41
Doesn't feel safe at school	77.4	69.9	17
Family/guardian does not provide girl with school supplies	74.6	66.6	81
Were you absent from school for more than 5 days in a row last year?	73.6	67.5	102
To support my parents in domestic chores	72.9	67.5	223
Risk of early marriage	75.4	68.7	56
Have you ever thought of migrating to another area?	75.2	70.2**	44
My teachers treat boys and girls differently in the classroom	73.4	66.9	356
Teacher doesn't use different language if you don't understand	73.6	66.8	326
Teacher doesn't encourage participation in class	72.5	66.8	271
Teacher doesn't suggest ways to continue study after school	75.1	69.3*	59
Teacher used physical punishment in last week	73.4	66.4	296
Does not feel able to do things as well as friends	75.3	64.0	10
Does not want to do well at school	79.1	70.7	5
Does not feel confident answering questions in class	74.4	67.9	39
Does not want to continue studying after this year	87.3*	66.9	6
Does not feel able to describe thoughts to other people when speaking	83.7	74.0	5
Does not feel able to work well in groups	83.9	65.5	7
Does not feel able to organise friends to do an activity	66.7	55.5***	10
Does not ask teacher if doesn't understand something	84.6	73.5	7
I get nervous when I have to read in front of others	71.3*	66.4	201
I get nervous when I have to do maths in front of others	71.7	66.9	170
If I do well in a test it is because I am lucky	71.5**	66.6	337
Family decides whether will go to school	78.1**	69.2*	64
Family decides whether will go to school beyond this grade	78.2**	69.0*	76
Does not recognise choices today about studies affect future	75.7	62.0	5

*Notes: *, **, *** indicates t-test for difference in means of scores statistically significance at the 10%, 5% and 1%.*

Table 24 shows the learning scores of girls broken down by key learning barriers (many of the learning barriers cited in the earlier sections have been excluded for ease of interpretation). The table shows that

there are some statistically significant differences across the barriers, though given the number of significance tests carried out, and due to the data issues discussed previously, any interpretations of these differences should be strongly caveated.

The data itself presents a mixed picture – for example, the sample of girls who do not feel safe travelling to school performed better than other girls in both literacy and numeracy (though the sample size is only 7), girls whose teachers do not suggest ways to continue studying have better performance in numeracy, and girls whose families decide whether they go to school performed better on both literacy and numeracy.

Despite these unexpected trends, there are also some plausible correlations. Girls who reported not being able to move around their school easily performed worse on literacy, girls who reported feeling nervous reading in front of others performed less well on literacy, and girls who reported not feeling able to organise activities performed less well in numeracy.

Nevertheless, the majority of the interactions do not show any significant trends and the breakdown in grades by barriers may need to be reassessed at midline and endline.

4.3 Transition Outcome

Table 25 outlines transition pathways and defines successful and unsuccessful transitions based on the stages of girls' education. Regardless of their current stage, in-school progression (i.e. advancing a grade) is always defined as successful transition, whereas repeating a grade is deemed unsuccessful. Dropping out of school (or an alternative educational institution) is also deemed unsuccessful, and conversely re-enrolling, having previously dropped out, is considered to have been successful. Girls who were not in school last year and remain unenrolled are still considered to have transitioned unsuccessfully.

The legal minimum working age in Ethiopia is 14 years of age[39], and education is technically compulsory up to this point[40] to coincide with the end of primary school. As such, moving into work whilst in lower or upper primary would be considered unsuccessful transition. From the age of 15 upwards, moving from primary school education into employment is considered successful if a girl's work is paid what is considered by their primary caregiver to be 'a fair wage'[41].

Having completed primary school, girls may also enrol into technical & vocational education & training (TVET) as an alternative to continuing education in school. Attending a TVET after primary is also considered to be a successful transition

Table 25: Transition pathways

	Baseline point	Successful Transition	Unsuccessful Transition
	Enrolled in Grade 1, 2 ,3, 4	In-school progression	Drops out of school Remains in same grade

[39] See https://www.ilo.org/ifpdial/information-resources/national-labour-law-profiles/WCMS_158894/lang--en/index.htm <Last accessed 31/01/19>
[40] See 'Education system' under http://uis.unesco.org/country/ET <Last accessed 31/01/19>
[41] As Ethiopia does not have a minimum wage, this cannot be used as a benchmark for fair compensation.

Lower primary school			Moves into work, but is below legal age
Upper primary	Enrolled in Grade 5, 6, 7, 8	In-school progression Moves into secondary school	Drops out of school Remains in same grade Moves into work, but is below legal age
Lower Secondary	Enrolled in Grade 9, 10	In-school progression Enrols into technical & vocational education & training (TVET) Gainful employment	Drops out of school Remains in same grade Moves into employment, but is paid below what is deemed a 'fair wage'
Upper Secondary	Enrolled in Grade 11, 12	In-school progression Enrols into technical & vocational education & training (TVET) Gainful employment	Drops out of school Remains in same grade Moves into employment, but is paid below what is deemed a 'fair wage'
Out of school	Dropped out	Re-enrol in appropriate grade level in basic education	Remains out of school

Benchmarking

Table 26: Benchmarking for the Transition Outcome

Benchmark Group							
		Benchmark transition pathway					Transition Rates
Age Group	Sample size (#)	In education progression, non-TVET	In education progression, TVET	Drops out of school	Repeats grade	Not enrolled last year and enrolled this year	Successful transition rate per age
% aged 8 to 10	45	53.3	0.0	0.0	22.2	24.4	77.8
% aged 11 to 13	45	55.6	0.0	0.0	17.8	26.7	82.2
% aged 14 to 16	26	11.5	0.0	0.0	57.7	30.8	42.3
% aged 17 to 19	32	9.4	12.5	3.1	50.0	25.0	46.9
% aged 20 to 23	21	9.5	23.8	4.8	28.6	33.3	66.7
Overall	169	33.7	5.3	1.2	32.5	27.2	66.3

*Note: Educational progression is defined as i) not repeating course, ii) not reporting not being enrolled last year, iii) being currently enrolled in the same or higher level than last year, for primary, secondary and tertiary levels excluding TVET.

Table 26 contains transition pathways for the 169 girls[42] in the benchmark group by age range. Pathways are mutually exclusive so that each girl in the sample only belongs to one. Furthermore, this table contains all the possible transitions that girls undergo[43]. Transition is considered to be successful if the girl progresses in education or is enrolled in education at the moment of the survey, and the primary caregiver does not state that the previous grade is being repeated. Due to data inconsistencies, we cannot distinguish among primary, secondary and tertiary levels of education progression.

The table suggests that successful transition rates have a non-linear relation with age. Girls until 13 years old have successful transition rates of around 80%, which are the highest among all. Between the ages of 14 and 19 there is a distinct cut off point where the proportion of girls with successful transition pathways diminishes to just above 40%. Transition rates have a slight improvement for girls aged 20 years or more, where 66% of the girls in that age range achieve successful pathways.

The drop in successful transition pathways coincides with the age that girls usually make the transition between primary and secondary education, which suggests that girls face high barriers to continue in school after completing their primary education. This decrease is mainly driven by more than half of the girls repeating the same course from last year, but this should be cautiously interpreted since the measure used to capture drop outs may underestimate drop out percentage. This is consistent with, and may be driven by, pupils performing poorly in the national examinations which occur at grade 8 to determine entrance to lower secondary, and grade 10 to determine entrance to upper secondary[44]. This may suggest that the project should focus efforts on preparing girls for these risky transition points.

The proportion of drop-outs observed in the data is very low. This may be due to inconsistencies in the benchmark data[45], or reflect the fact that the data cannot differentiate between students who drop out during the course of a year, only to re-enrol the next year in the same grade, and those who repeat the grade having attended for the entire year. This is a pattern which has been observed elsewhere in Ethiopia. In 2002, 22% of all students repeated grade 8, and of these students, 21% had dropped out during the 2001-2002 school year and re-enrolled in the same grade in 2002-2003[46].

The rates of successful transition observed as part of the Benchmark transition sample are considerably lower than has been observed in secondary data. Recent data from the Ethiopia Ministry of Education (MoE)[47] for grades 1-8 estimates grade-level repetition rates at between 6.4% (for grade 6) at its lowest[48]

[42] Note that the full sample contains 174 girls, but 5 of these have been omitted from analysis as they do not contain age data.
[43] We do not observe any instances in the data of girls moving into work, so we exclude this from the table.
[44]For an outline when national exams occur, see:
http://www.nafsa.org/_/file/_/ac12/ac12_teachered_ethiopian_ed.pdf <Last accessed 31/01/2019>
[45] The first question of the BT instrument, PCG_1tc asks primary caregivers if the girl in question is currently enrolled in school, to which 100% of the sample is assigned 'Yes'. Question PCG_2tc which asks for school name contains two responses as follows: "she is not registered" and "she is not enrolled". As these directly contradict PCG_1tc, we suggest interpreting the results of Table 26 with caution
[46] See http://siteresources.worldbank.org/INTAFRICA/Resources/wp86_ethiopia_edu.pdf (particularly Table 2.7) <Last accessed 31/01/2019>)
[47] Data comes from the MoE's Annual Abstract on Education Statistics for 2016/2017 (available on request)

and 8.3% (at grade 8) at its highest. The peak at grade 8 is consistent with what we find (though we have to extrapolate based on age) but differs greatly in magnitude. The MoE estimate drop-out rates as being around 9% per grade, with the exception of grade 1 where dropouts are estimated to be 18%. This is higher than what is observed in the benchmark transition sample but may, as mentioned, be due to inconsistencies in the data. Combining these two main determinants of unsuccessful transition suggests that successful transitions rates are, on a country-level average, in the region of 80-85%. This is consistent with what is observed in ages 8 to 13.

The MoE report does not estimate repetition or drop-out rates at higher grades (and therefore ages). As such, we lack an appropriate benchmark with which to compare results.

4.4 Sub-group analysis of the transition outcome

Table 27: Barrier to successful transitions

Barriers to Successful Transitions		
	Successful Transition (%)	
Categories	**Yes**	**No**
Orphan	48.0	71.1
Living without both parents*	57.1	71.0
Girl is married*	90.0	64.8
Girl is a mother*	100.0	65.0
Fairly unsafe or very unsafe to travel to school	56.0	68.1
Carer did not complete any school level	78.2	48.5
Region (Oromia)	44.0	75.6
Region (Amhara)	75.6	44.0

*Notes: * implies sample size <10 for those for whom category applies – interpret results with caution.*

Table 27 attempts to draw out descriptive information as to the importance of certain barriers in determining successful transition among girls in the benchmark transition sample. In the case of girls who are orphans it can be interpreted as such: On average, a girl who is an orphan has an average successful transition rate of 48% (and hence 52% unsuccessful), compared to non-orphans who have an average successful transition rate of 71% (and hence 29% unsuccessful). This would suggest that being an orphan has a strong effect on the likelihood of transition successfully. Finding the journey to school fairly or very unsafe is also associated with higher levels of unsuccessful transition, though the difference is not as large as in the comparison between orphans and non-orphans.

Girls whose primary caregiver did not complete any schooling exhibit higher levels of transition, suggestive of an encouragement effect from caregivers who did not have the same opportunities afforded to girls in the program. Regional breakdowns of transition suggest that transition is lower in Oromia than Amhara, though this may be driven by school-level rather than region-level effects given that the benchmark

transition sample only sampled one area (and hence one school) in Dera, Oromia. Alternatively, it may be due to recent political instability in Oromia. Secondary data provides evidence that Oromia does perform worse than Amhara in transition. Rates of survival in education to grade 5, which are likely to be highly correlated with unsuccessful transition at any given grade, are considerably higher in Oromia than Amhara (54% vs 26%). This suggests that special attention may need to be focussed on girls in Oromia, where transition is lower.

The analysis is constrained by the limited number of questions relating to barriers and characteristics asked as part of the benchmark transition instrument. It suggests that being an orphan and having an unsafe journey to school are both important determinants of transition. Interpretation of additional channels such as whether a girl is married, or a mother, are limited by the very small sample sizes.

4.5 Cohort tracking and target setting for the transition outcome

Cohort tracking will follow the process outlined in the MEL. At baseline, location data is collected to allow girls to be tracked to their homes at midline should they no longer be found at school. At midline, the transition cohort will diverge from the learning cohort as girls drop out from project schools. Girls who cannot be found at school will be tracked to their homes using location data collected at baseline.

Targets for midline and endline are generated from the Outcomes Spreadsheet and included in

Table 28 below.

Table 28: Target setting

	Evaluation point 2	Evaluation point 3
Target generated by the outcome spreadsheet	**8%**	**10%**
Alternative target proposed by project (if applicable)		

Adapt as required

4.6 Sustainability Outcome

Table 29 presents the Sustainability Indicators outlined in the logframe and discussed in 2.2. The indicators attempt to gauge sustainability at the community, school and system level using data from the household survey, girls school survey, qualitative data and administrative records from schools and Woredas collected on the project-level. The indicators will allow us to measure sustainability on a broad basis and assess the extent to which delivering sustainability has been successful across different areas. For quantitative measures of sustainability, captured as part of the household survey, the control group serves as a useful comparison and tentative indication as to the impact the project (or its predecessor project) may be having.

Table 29: Sustainability indicators

		Community	School	System
Indicator 1	**Wording**	% of girls' households who pay for school fees, transportation to and from school, school meals, materials and supplies in the current school year (control schools in brackets) *Appropriate*	% of girls answering "Boys" to "Does your teacher(s) ask more questions to boys / girls / equally" (control schools in brackets) *Requires consideration*	Number of trained Woreda education officials allocated to conduct supporting supervision activities (i.e. teacher's development and girls' learning assessments). *Requires revision*
	Status	School fees: 47% (35%) School meals: 7% (0.2%) Transportation: 10% (3%) School materials: 61% (37%)	6% Boys vs. 7% Girls (12% Boys vs. 13% Girls)	12 (6% of all officials)[49]50
	Score	2	2	1
Indicator 2	**Wording**	% of primary caregivers who state 'single and divorced women in the community face negative attitudes' sometimes/often or very often (control schools in brackets) *Requires consideration*	% of girls answering "Yes" to "Does your teacher(s) suggest ways you can continue to study after school/at home?" (control schools in brackets) *Requires consideration*	N/A
	Status	64% (67%)	90% (80%)	N/A
	Score	0	2	
Indicator 3	**Wording**	N/A		

% of trained teachers who remain in their post after being trained by the project.

Appropriate

[49] Project data shows that there are 206 educational experts operating across all project Woredas.

N/A				
	Status	N/A	71.7% (South Wollo: 62%, South Gondar: 80%, Arsi: 87%)	N/A
	Score		1	
Baseline Sustainability Score (0-4)		1	2	1
Overall Sustainability Score (0-4, average of the three level scores)		1.33		

The community aspect of sustainability is concerned with establishing changes in perceptions towards girl's education at the community level and bringing about behaviours conducive to sustainable improvements in girls' attendance and learning. The sustainability score is derived from two quantitative indicators[51].

Firstly, the percentage of girls' households paying for the associated costs of attending school is used as a proxy for household's endorsement for girls' education and their ability to pay for it. Generally, we observe low levels of self-financing of girls' education in project schools, though depending on the component, there may be different reasons for this. The low percent of households paying for school meals and transportation may be driven by households' inability to pay for these, or by demand factors (e.g. a girl who walks to school does not pay for her transport by definition). Perhaps more concerning are the relatively low proportions of girls' households paying for school materials (61%) and school fees (47%). This might be suggestive that a large contingent of households are dependent on project support to cover these costs. There is evidence that the project is having an impact compared to control schools, as parents are more likely to pay in every category of school-related expenditure.

The project's sustainability plan states that *'the project will meet these costs with a decreasing frequency'* (by reducing the number of girls receiving support) under the assumption that changing social norms and attitudes would make families more willing to contribute towards their daughter's educational costs and take on the cost previously borne by the project. Discussion with project staff at CHADET responsible for sustainability suggests that meeting costs with a decreasing frequency poses a problem, as it could results in dropouts as there is no economic element to the project to generate higher incomes in households. As such, it appears to mostly be a problem of the financial capacity of households to meet the costs, rather than a reluctance due to any negative attitudes towards girls' education, which is further supported by qualitative evidence below.

Qualitative data from FGDs held with community representatives in treatment kebeles in both Oromia and Amhara shed further light on the sustainability of community-level shifts in attitudes and behaviour. Focus

[51] The second of these indicators (survey question *DIV_2)* represents a deviation from the logframe, which asks for qualitative data. It appears that the qualitative instrument for FGDs was not properly aligned with the logframe which originally targeted the '% of parents or carers of girls expressing stigma of single or divorced women'. The deviation here stems from the fact that parents are not asked directly whether they stigmatize single and divorced girls but are asked for their view on community-wide stigma.

groups broadly agreed that gender is not a decisive parameter when deciding which of their children to send to school, for example one community representative indicated that "*Most parents send their children to school regardless of sex or disability. Everyone has understood the value of education".* Despite this broad consensus, a minority indicated there are still families who want girls to stay at home and support house work.

On the ability of households to pay for school-related costs, the evidence complements the low numbers of households financing girls' education directly, and the comments of CHADET staff. As part of the community FGDs, one community representative from the Haik kebele mentioned that families will make children stay at home once the support from CHADET stops given their reliance on this support. Whether this view is representative of the community is hard to establish from qualitative data. Currently, 14% of primary caregivers in programme areas state that it is difficult to afford for their child to attend school *(PCG_7enr)*.

Secondly, the percentage of primary caregivers who state that 'single and divorced women in the community face negative attitudes' either sometimes, often or very often, captures attitudes towards delaying marriage that may inhibit girls pursuing education beyond secondary, and potentially even primary-level. 64% of primary caregivers believe that single and divorced women face stigma in the community. Note that the question does not directly elicit opinions on stigma but asks for the individual's view on community-level stigma. If individuals' views do not correspond with their perceived view of community-level stigma, and if this perception is incorrect, then the figure of 64% reported may be in accurate. Nevertheless, a figure of 64% is high enough to suggest that this is likely to be a barrier to girls' pursuing education beyond the compulsory amount.

Given the findings of the quantitative data, the qualitative data showed a surprising lack of discussion on the stigma surrounding single and divorced women. This may be because the prevalence of unmarried woman seems to be low in the communities surveyed[52]. The practice of early marriage was discussed as occurring within communities which is likely to go together with a stigma towards single and divorced women, though the extent of this practice is unclear. The qualitative instrument may need to be revised at midline to directly elicit (in an unbiased manner) whether stigma towards single and divorced women is a significant issue in Ethiopian communities.

Taking the findings from quantitative and qualitative data, along with discussion with project officials, community-level sustainability is awarded a score of 1 to indicate latent changes in sustainability. This score was decided on as changes in attitudes towards girls' education appear to be occurring in project communities, but there is no clear evidence that this is being accompanied by changes in behaviour. The issue of dropout once financial support from the project is wound down also remains unsolved. For the project to be awarded a score of 2, there should be evidence of activities developing to mobilise non-project funding and resources to support girls' families with the financial costs associated with schooling. The project may also consider adding an additional indicator to capture community-level sustainability in terms of the general opinion of community members towards girls' education.

[52] Of those respondents who did not answer 'Do not know' 89% of respondents stated that there were very few or few single or divorced women in their communities (survey question *DIV_1*)

Sustainability of in-school behaviours and attitudes will stem from changes in attitudes and behaviours amongst school-level stakeholders such as teachers, and school stakeholders. This is tracked by three key indicators, two from household survey data, and one from project administrative records.

The first indicator stems from the girls' school survey, and asks girls whether teachers generally ask more questions to boys, girls, or whether they ask questions to both equally. There is strong evidence of teachers asking questions to students equally irrespective of gender. In intervention schools, 87 percent of girls state that teachers ask questions equally to both boys and girls, with the remaining 13 percent being almost equally split between boys and girls.[53]

The second indicator also stems from the girls' school survey and asks girls whether their teacher suggests ways in which they can continue to study at home after school. In intervention schools, 90% of girls state this to be the case which, when compared to control schools, provides preliminary evidence that the project may be having an impact in promoting continued learning at home.

These two indicators provide first evidence that the project's activities may be realising important changes in attitudes and behaviours among school stakeholders, resulting in girls feeling more equally treated in class, and encouraging them to continue studying at home. Qualitative evidence broadly supports the quantitative data here as interviewed students expressed that their enjoyment of schooling has grown compared to the year before as the teaching and learning process has shown improvement.

Lastly, the third indicator for sustainability in schools stems from administrative records on the proportion of teachers who remain in their post after being trained by the project. Teacher retention rates are 72% on average but exhibit considerable heterogeneity across schools. The average is driven down by the district of South Wollo where retention rates are 62%, compared to 80% and 87% in South Gondar and Arsi respectively. From discussions with the project coordinator in South Wollo, this seems to be driven by transfer from remote to urban areas, teachers moving out of the education sector, and teachers being promoted to leadership positions elsewhere. These factors are likely to be present elsewhere, though South Wollo contains two large cities (Dessie and Kombolcha) so the opportunities, and therefore the 'brain drain' effect into other industries may be especially strong there compared to Arsi and South Gondar. These aggregate figures also hide exceptional cases of very remote schools where teacher retention is likely to be much more problematic.

School-level sustainability is awarded a score of 2, as there is evidence of improved support for girls' by teachers and a general 'buy-in' into the project ethos among project schools. To reach a sustainability score of 3, preliminary evidence should be observed that schools are beginning to take on more responsibility to deliver project outcomes themselves such that there is potential for the project work to be phased out.

The project should consider revising indicators 1 and 2 as they exhibit scores which suggest limited scope for further improvement. There is likely to be a floor (ceiling) for indicator 1 (indicator 2), meaning that the project cannot demonstrate increases in sustainability. New indicators may need to be developed to capture school-level sustainability, or could be taken from questions administered at baseline, if appropriate indicators can be found.

[53] A follow-up questions asks girls whether teachers ask **harder** questions equally/more to boys/more to girls. The evidence here was supportive of that of the previous question as 84% state that the balance is equal. There was, however, a 3 percentage point gap in favour of boys suggesting that this may warrant further investigation.

System level sustainability captures the extent to which officials engage with and support the fundamentals of the project in their work, as well as the degree of knowledge and capacity development among educational authorities. The key project indicator to assess system level sustainability is the number of Woreda officials trained by the project who are allocated to conduct supporting supervision activities. Qualitative data suggests that slight changes have been observed due to the support and follow-up by Woreda education officials.

As we can only draw very limited conclusions based on the single indicator for system level sustainability, a score of 1 is awarded. There is evidence of knowledge development through training of Woreda officials, though this training is limited to one or two officials per Woreda. It is unclear from the evidence available whether this training has resulted in changes in attitudes and behaviours, and whether it has been effective in generating knowledge spill overs to untrained officials. The evaluation recommends consideration of how further indicators capturing system-level sustainability can be built into subsequent evaluation points.

The following sub-section and

Table 30 should be completed by the project.

1) Set reasonable expectations: At each of the three levels of sustainability, what changes need to take place to ensure that attitudes, behaviours or approaches are established which provide for ongoing learning and successful transition for future cohorts of girls and boys? Who are the stakeholders involved in these changes? What are the factors that help or hinder changes? Refer to your sustainability plan, theory of change and logframe. Be brief in the table and provide narrative analysis below the table that refers back to the mixed-methods analysis under 1)

Table 30: Changes needed for sustainability

	Community	School	System
Change: what change should happen by the end of the implementation period	The project should be considering a decrease in contributions to households for the provision of school related assistance	Teacher retention to be monitored more regularly for attendance to improve, annual top up training to mitigate brain drain	Woreda officers to be more regularly engaged with teachers
Activities: What activities are aimed at this change?	Livelihood assistance through the CCCs	Leadership and management training, regular updates with the local Woreda, more regular tracking of teacher attendance	More planned and consistent networking and liaison with Woreda officials
Stakeholders: Who are the relevant stakeholders?	All members of the CCCs	Teachers, principals, education officers, woreda officers, M&E personnel	Woreda officials, teachers, Education officers (CHADET)

Factors: what factors are hindering or helping achieve changes? Think of people, systems, social norms etc.	Economic factors external to the project	Attrition	None currently

Provide narrative analysis here of the points raised in the table above. Explain the change the project intends to achieve. Highlight cross-cutting activities, stakeholders and factors, but also those that relate to only one level of sustainability. Link the analysis here with that under section 1) drawing on the scores given for each level. Link the analysis to the other Outcomes and Intermediate Outcomes.

5. Key Intermediate Outcome Findings

5.1 Attendance

Improved attendance is chosen as an intermediate outcome as a necessary condition for improvements in learning outcomes to be realised. To realise improvements in attendance, the project is attempting to reduce the economic and psychological barriers that girls (and their families) face which may lower their attendance. For example, girls' families may not be able to support the costs associated with sending children to school.

'Attendance' is being measured using two indicators. The first is the percentage of girls whose average school attendance improves in the lifetime of the project, which will be captured via the household survey and school attendance registers collected as part of the evaluation. The second is qualitative and seeks to gain an understanding of beneficiaries' views on the barriers that prevent them from attending school regularly.

The primary caregiver component of the household survey gives an insight into the state of attendance in project and control schools at baseline, and how attendance is likely to relate to key characteristics and barriers.

Table 31: Attendance (IO1)

Question *(name)*	Total	Age <=13	Age>14	Has disability+
Since the start of the most recent school year, has [GIRL] attended her school on **most** days that the school was open? *(PCG_5enr)*	87% (92%)	86% (91%)	89% (93%)	71% (100%)
Do you feel safe at school? *(CS_W14s)*	96.8% (93.4%)	98.3%** (94.1%)	95.1%** (91.9%)	100.0% (90.1%)
Have you used the school's library or reading corners in the last month? *(CS_W1A)*	93.5% (69.8%)	92.2% (64.5%**)	94.8% (74.1%**)	92.7% (44.4%***)
Were you absent from school for more than 5 days in a row last year? *(CS_M3)*	16.9% (22.4%)	19.4%** (23.2%)	14.1%** (21.8%)	9.3% (23.9%)

Notes: Figures in brackets refer to the control group. Use of the word 'most' in the survey instruments may have created different interpretations across survey respondents, as the definition of most ('more than half of school days') was only defined if respondents queried this. + Low sample size means results should be interpreted with caution

Table 32: Attendance and key barriers and characteristics

Question *(name)*	Total	Age <=13	Age>14	Has disability+
Does [chore burden] stop [girl] from going to school? *(PCG_27g)*	31.5% (38.6%)	37.2% (46.8%)	24.0% (32.1%)	85.7% (59.25%)
Under which of the following conditions do you think it is				

acceptable for a child to not attend school? (*% who think it is acceptable)*				
- The child is married/is getting married	24.6% (25.0%)	20.2% (24.9%)	31.5% (24.4%)	0.0% (4.6%)
- The child is too old	15.8% (18.1%)	12.1% (14.0%)	20.8% (21.1%)	0.0% (3.1%)
- Education is too costly	11.2% (24.8%)	8.2% (24%)	15.4% (25.4%)	3.4% (1.5%)
- The child is a mother	25.2% (25.3%)	23.9% (25.8%)	24.8% (27.3%)	3.4 (4.6%)
- The child may be physically harmed or teased at school or on the way to/from school	52.9% (25.6%)	46.9% (22.2%)	61.2% (28.1%)	37.3% (63.1%)

Notes: PCG_27g is binarised to be equal to 1 if a PCG answers girl is either: i) Not enrolled because of this ii) Stops her often iii) = Stops her sometimes. + Low sample size means results should be interpreted with caution

This is supplemented by attendance data collected by the project, and spot check data taken in classrooms as part of the external evaluation. Attendance data from the project shows the following trends in attendance across 3 schools (one from each Zone):

Table 33: Attendance in three schools for November and December 2018

School	November	December
Dera #1	98.6%	98%
Koley Primary School	95.8%	97.8%
Tado Mender Primary School	98.3%	96.9%

Notes: Attendance rate measured as 1 minus the absentee rate, where the absentee rate is calculated as the number of days schooling missed in a month divided by the total number of school days in that month (21 for both November and December)

Data collected by the evaluation can be used to corroborate this in-project data. We collected attendance data, both from the school registers, and from spot checks in the classroom. This showed that the school register data was broadly accurate, with a tendency to slightly overstate attendance rates[54].

[54] On average, attendance rosters overstated attendance by 2.6 girls. This effect was observed in all 3 schools, and was largest in Dera #1 and smallest in Tado Mender (3.6 vs 1).

The quantitative data presented above appears to suggest that high levels of attendance are already being achieved across the board. When looking further into instances of high levels of absence, we observe that only 13 girls (2.3%) have attendance rates lower than 80% equivalent to missing a day of school a week, according to the in-project attendance data. Very high levels of absence (defined as greater than or equal to 50%) occur for only 1.6% of girls according to household survey data, which is consistent with project data. When looking at the disaggregated quantitative data in Table 31, age does not seem to be a significant determinant of attendance. Disability, particularly in treatment schools, does seem to result in significantly lower attendance rates, which suggests that this may deserve particular attention from the project.

The quantitative evidence presented above suggests that attendance in project schools is at high levels, with the exception of girls with a disability, who have somewhat lower attendance rates. Taking the results of the household survey, and the evaluation roster checks, we would tentatively conclude that attendance rates are in the order of five to ten percentage points lower than the in-project attendance data suggests. Despite this, they are suggestive that improved attendance as an IO is being achieved.

We also estimated the relationship between attendance and some of the barriers in Table 10 through multivariate regression. This showed that having a high core burden or feeling unsafe when travelling to school is significantly associated with lower attendance rates. These results are broadly confirmed by Table 32 containing household survey evidence which sheds light on how key characteristics and barriers are likely to interact. This shows that around a third of girls are likely to have their attendance reduced due to high chore burden. Table 32 also presents data on parental views towards education and whether certain conditions warrant a child not attending school. Generally, potentially limiting characteristics such as being (or becoming) a mother, getting (or being) married, and being 'too old' are only viewed as acceptable reasons not to attend school by between 15 and 25% of caregivers. Caregivers in project areas are less than half as likely (11.2% on average) to suggest that the cost of education is a valid reason not to attend school, suggesting that attitudinal changes may already becoming realised. The two reasons seen as 'most valid' by caregivers were a high chore burden, and feeling threatened whilst traveling to (or in) school. On the basis of Table 32 and the multivariate regression conducted, any project impacts on reducing the chore burden on girls, and improving their journey to school, are likely to realise improvements in attendance.

In contrast to the quantitative data, the qualitative interviews cite attendance as a key problem – community representatives reported that children are frequently absent and drop out of school. The reason cited is that families want children to engage in activities at home. For girls the main tasks are fetching water, cleaning the home, taking care of children, sick, and elderly. Forced migration was also reported as a key reason for absenteeism or dropout. As such, it may be the case that the high levels of attendance reported in the quantitative data are a function of selection bias given that the evaluation would have been less likely to select absent girls.

The responses in the qualitative study on access and attendance for children with disabilities are mixed. Some respondents in rural areas reported that there were no problems with attendance for children with disabilities, but that extra help was sometimes not available for these students. Others said that children with disabilities do not attend school at all, citing lack of special needs teachers as the reason for this; however, others reported that special needs resources were available. More analysis is necessary to understand the schooling experiences of students with disabilities.

5.2 Quality of teaching

Intermediate Outcome 2 looks at improving teacher quality, captured using the household survey which asks each girl's primary caregiver about the quality of teaching the girl receives, and through the girl's

school survey. An in-project tool[55] has also been used to capture the key components of teaching practice, linked to the core content of the teacher training modules run by Education Development Trust (EDT), who have been contracted to deliver the teacher training component of the project. Data from the lesson observation tool is also used to feed into areas of development covered in communities of practice. The tool has been piloted and therefore not yet used to feed data into baseline.

The table below shows the proportion (of either girls or their caregivers) answering as indicated to each of the survey questions related to teaching quality. The percentages reported outside of the brackets are for the intervention group, inside the brackets refers to the control group. For example, the first cell of the main body of the table shows that 6.9% of all girls in the intervention group reported that their teachers ask more questions to boys in class. For the control group, 11.6% responded similarly. The responses are broken down further by age (13 years and under, 14 years and older) and disability (defined as at least some degree of cognitive or physical impairment). For example, in the intervention group, 5.4% of girls aged 13 or under, 8.4% of girls aged 14 or over and 8.5% of girls with a disability, responded similarly.

Table 34: Indicator for IO2 – Quality of teaching

Intermediate outcome indicator	All girls	Aged 13 or under	Aged 14 or over	Disabled
Teacher asks more questions to boys	6.9% (11.6%)	5.4%* (13.6%***)	8.4%* (5.9%***)	8.5% (8.6%)
Teacher asks harder questions to boys	8.5% (13.7%)	6.9% (14.6%*)	10.0% (10.0%*)	8.5% (19.8%*)
Teacher asks more questions to girls	7.4% (12.6%)	9.3%** (17.4%***)	5.4%** (7.5%***)	14.9%** (16.0%)
Teacher asks harder questions to girls	5.8% (10.4%)	6.4% (14.3%***)	5.1% (5.9%***)	4.3% (12.3%)
Teacher asks questions equally to boys and girls	85.1% (74.5%)	85.0% (67.8%***)	85.4% (85.0%***)	76.7% (73.2%)
Teacher asks hard questions equally to boys and girls	83.8% (75.3%)	85.3% (70.7%***)	82.4% (83.2%***)	90.7% (69.0%)
Teacher doesn't use different language if you don't understand	51.2% (52.4%)	48.8% (57.0%**)	54.3% (48.3%**)	40.4% (60.5%)
Teacher doesn't encourage participation in class	43.9% (44.2%)	40.9%* (47.2%**)	47.3%* (39.6%**)	37.0% (61.7%***)
Teacher suggests ways to continue studying after school / at home	90.5% (80.8%)	89.9% (78.9%**)	91.1% (85.0%**)	83.7% (76.1%)
Teacher disciplines/punishes students who gets things wrong	13.6% (6.7%)	17.8%*** (6.8%)	9.1%*** (5.3%)	19.1% (11.3%*)

[55] See *Annex 7: Data collection tools used for Baseline*

Teaching quality 'good' or 'very good'	89.0% (92.4%)	91.8%* (93.4%)	87.3%* (91.4%)	89.0% (92.4%)

*Notes: *, **, *** indicates t-test statistical significance at the 10%, 5% and 1% levels*

The table shows that, for the intervention group, 51.2% of all girls reported that their teacher does not use another language to help them if they do not understand something and 43.9% reported that their teachers do not encourage participation in class. These outcomes are consistent across sub-groups and for the control group. Over 60% of girls with disabilities in the control group responded negatively to these questions, though this was not the case for girls with disabilities in the intervention group. However, this may still indicate an additional barrier faced by girls with disabilities.

The data also suggest that while a significant minority of girls report that their teachers ask more and harder questions to boys, a comparable proportion reported that their teachers ask more and harder questions to girls.

The next two rows look at the proportion of girls who report that their teachers suggest ways that they can continue study after school and the proportion that report that their teachers discipline or punish students who get things wrong. Between 76.1% (girls with disabilities, control) and 91.1% (girls over 13, treatment) responded that their teachers do suggest ways to continue studying after school or at home. Between 5.3% (girls over 13, control) and 19.1% (girls with disabilities, treatment) report that their teachers discipline or punish students who get things wrong.

The final row shows the proportion of girls who report the quality of teaching they receive to be 'good' or very good' as opposed to either 'poor' or 'neither good nor poor'. Of all girls in the treatment group, 11% responded negatively to this question. However, most of these responded that the teaching quality was 'neither poor nor good' – only 0.1% of the total sample (two students) rated their teaching quality as poor. There is a significant difference between the age groups – older students reported lower quality of teaching compared to younger students.

The qualitative survey supports that teaching quality is a key factor in girls' education. While most students reported that they enjoyed school more in the last year due to better teaching and learning processes, most teachers and community representatives reported that schools did not have adequate facilities such as textbooks, computers, water sources, and separate male/female toilets. These factors can be expected to reduce quality of teaching and the learning experiences of students. Respondents cited improvements in the last year with respect to these issues, with support from the education office and CHADET recognised as contributing factors. However, they also reported that the improvements made are still insufficient.

The logframe also contains an indicator on the 'percentage of trained teachers who improve their teaching techniques (disaggregated by literacy and numeracy skills)'. This is collected in-project using classroom observations. The approach for this has been piloted but will not have been fully completed until midline.

To further investigate intermediate outcomes, Table 35 shows the interaction between girls' characteristics and potential barriers, and indicators for teaching quality. Each cell shows the percentage of girls included and excluded from each category that report certain indicators. Only differences that are statistically significant are shown. It can be seen that 8.9% and 47.2% of girls in rural settings report that teachers ask more questions to boys and that teachers do not encourage participation in class, while these figures are 2.7% and 36.8% for girls living in urban settings. These indicators provide suggestive evidence that teaching is more inclusive in urban schools.

Girls that feel that early marriage can happen to them, and girls who have thought of migrating to another area are more prone to report that the teacher does not use a different language if they do not understand.

Table 35: *Quality of teaching and key barriers and characteristics*

	Teacher asks more questions to boys	Teacher doesn't use different language if you don't understand	Teacher doesn't encourage participation in class	Poor teaching quality
Rural/Urban	8.9% / 2.7%	-	47.2% / 36.8%	13.0% / 7.4%
Amhara/Oromia	3.9% / 18.8%	50.2% / 59.6%	-	-
Books and learning material not available at school (Yes/No)	-	63.0% / 49.4%	65.7% / 40.4%	-
Computers not available at school to use (Yes/No)	-	54.7% / 31.6%	46.1% / 23.7%	9.7% / 22.6%
Doesn't use toilets at school (Yes/No)	0.0% / 7.7%	-	55.7% / 42.6%	-
Doesn't use areas where children play / socialise (Yes/No)	20.0% / 6.7%	-	-	29.4% / 9.9%
Are you or have you ever been married? (Yes/No)	-	-	-	-
Have you ever faced an attempt for an arranged or forced marriage? (Yes/No)	-	-	-	-
Do you feel that early marriage can happen to you? (Yes/No)	-	70.8% / 49.4%	62.1% / 42.1%	-
Have you ever thought of migrating to another area? (Yes/No)	14.3% / 6.4%	-	-	-

5.3 Girls' self-esteem

Intermediate outcome 3 targets greater self-esteem and empowerment of marginalised girls. The girls' school survey contains questions on two of the indicators used – self-esteem and self-efficacy. There are 24 questions in the girls' school survey which relate to this outcome. The qualitative beneficiary interviews will be used to gauge how girls' perceptions of their ability to succeed academically change as a result of the project.

The table below shows the responses of girls to each of the self-esteem and self-efficacy survey questions.

Table 36: Indicator for IO3 – Girls' self-esteem

Intermediate outcome indicator	All girls	Aged 13 or under	Aged 14 or over	Disabled

Does not feel able to do things as well as friends	1.9% (2.2%)	2.2% (3.1%**)	1.6% (0.9%**)	2.1% (2.5%)
Does not want to do well at school	0.6% (2.6%)	0.2% (3.5%)	1.1% (1.6%)	0.0% (11.1%***)
I get nervous when I have to read in front of others	34.0% (48.2%)	33.4% (50.7%***)	34.4% (39.9%***)	36.2% (53.1%)
I get nervous when I have to do maths in front of others	28.8% (47.6%)	28.5% (51.2%***)	28.7% (38.1%***)	29.8% (48.1%)
Does not feel confident answering questions in class	5.9% (4.0%)	4.9% (4.5%)	7.0% (3.7%)	14.9%*** (9.9%***)
Does not want to continue studying after this year	1.3% (1.2%)	1.7% (1.6%)	0.8% (0.6%)	2.1% (4.9%***)
Does not feel able to describe thoughts to other people when speaking	0.8% (3.0%)	0.5% (3.3%)	1.1% (2.8%)	2.1% (7.4%**)
Does not feel able to work well in groups	0.9% (2.6%)	0.5% (3.8%**)	1.4% (1.2%**)	0.0% (9.9%***)
Does not feel able to organise friends to do an activity	1.5% (3.3%)	1.2% (3.8%)	1.9% (3.1%)	6.4%*** (7.4%**)
Does not ask teacher if doesn't understand something	1.2% (7.1%)	0.7% (7.5%)	1.6% (7.2%)	2.1% (10.0%)
Does not agree their success is due to hard work	0.4% (2.2%)	0.5% (3.3%**)	0.3% (0.9%**)	0.0% (7.4%***)
If I do well in a test it is because I am lucky	55.8% (52.4%)	56.0% (54.9%***)	55.2% (44.5%***)	57.4% (48.1%)
Family decides whether will go to school	10.4% (25.6%)	9.1% (27.9%)	11.9% (24.6%)	19.1%** (21.0%)
Family decides whether will go to school beyond this grade	11.1% (23.1%)	9.6% (23.7%)	12.8% (24.1%)	17.0% (39.5%***)
Family decides what age will get married	14.5% (23.2%)	14.3% (26.1%)	14.6% (21.3%)	29.8%*** (38.3%***)
Family decides what type of work after finishing studying	6.7% (7.3%)	4.7%** (7.7%)	8.7%** (7.2%)	17.0%*** (23.5%***)
Family decides how much time spent with friends	13.0% (19.4%)	10.8%* (23.0%**)	15.2%* (15.9%**)	19.1% (30.9%***)
Does not feel able to stay focused on goals despite things getting in the way	0.8% (2.7%)	0.0% (2.4%)	1.1% (2.8%)	4.3%** (17.9%***)

Does not feel able put a plan in place and stick with it	0.6% (8.2%)	0.0% (13.3%***)	0.8% (5.6%***)	0.0% (7.1%)
Does not recognise choices today about studies affect future	1.0% (1.8%)	0.0% (1.8%)	1.4% (1.9%)	8.7%*** (8.9%***)
Does not try to find another way to express self if not understood	1.9% (3.9%)	0.7% (7.2%***)	2.4% (2.2%***)	13.0%*** (8.9%**)
Does not pay attention to body language of others	5.0% (10.5%)	2.7% (15.8%***)	6.0% (7.8%***)	4.3% (10.7%)
Feels lonely at school	31.9% (38.3%)	31.1% (36.7%)	32.3% (38.9%)	21.7% (48.2%)

*Notes: *, **, *** indicates t-test statistical significance at the 10%, 5% and 1% levels. Figures refer to percentage of people answering 'agree' or 'strongly agree' to question, or strongly disagree/disagree where we select a positive response, but present its negative inverse.*

The table shows that a significant number of girls responded negatively to many of the self-esteem and self-efficacy questions. Of all girls, 41% reported feeling nervous reading in front of others and 38% reported feeling nervous doing maths in front of others. This suggests that nervousness may have been a factor in determining girls' performance in the learning tests. A large proportion of girls - 54% - report that they believe if they do well it is due to luck, although the previous row suggests that only 1.3% do not agree that their success is due to hard work. These findings are consistent across sub-groups.

A significant proportion of girls reported that their families make key decisions for them (rather than the girls making decisions either alone or jointly with their family). 18% of girls reported that their families decide whether they will go to school and 17.1% reported that their families will decide whether they will go to school beyond the girls' current grade; 18.9% reported that their families will decide what age they will get married, and 16.2% of girls reported that their families decide how much time they spend with friends.

For girls with disabilities the proportion of these decisions being made by the girls' families rather than themselves increases substantially: 29.4% of girls with disabilities reported that their families decide whether they will go to school, 67.6% reported that their families decide whether they will go to school beyond the current year, 61.8% reported that their families will decide what age they get married, and 50% reported that their families decide how much time they spend with friends. 47.1% of girls with disabilities also reported that their families will decide what kind of work they do after finishing studying.

In general, girls with disabilities responded more negatively to these questions – the percentages tend to be larger than the other groups and statistically significant. For example, while most girls report wanting to do well in school, a significant proportion of girls with disabilities, 14.7%, responded negatively to this question. 17.2% of disabled girls reported not recognising that their choices today about their studies will affect their future, compared to 1.4% of all girls. While the qualitative evidence was mixed on the experience of girls with disabilities, the quantitative results reported here suggest that there are significant self-esteem and self-efficacy challenges for girls with disabilities.

Older girls (aged 14 or over) generally report more positive responses than younger girls (aged 13 or under). For example, the ability to do things as well as friends, nervousness working in front of others, a desire to continue studying, the belief that their own hard work and not luck leads to success, and the ability to stick to a plan, all improve with age.

Finally, a significant proportion of girls across all sub-groups report feeling lonely at school. 35% of all girls report feeling lonely. There is no significant difference between older and younger girls but again, this number is higher for girls with disabilities at 48.3%.

The qualitative analysis provides some further insights into girls' levels of self-efficacy. This is mainly related to families' roles in girls' educational decisions. While the analysis suggests that parents generally send girls to school regardless of their gender, a small minority of respondents indicated that some girls stay at home to carry out domestic work. There were also reports of girls being required to stay at home to help with farming, and of girls being forcefully sent to other countries. These factors reduce girls' capacities to make decisions and achieve their goals. These reports support the quantitative findings on the significant minority of girls who have key decisions made for them by their families.

Table 37 shows analysis between intermediate outcomes and key characteristics and barriers. It shows that 21.4% and 14.6% of girls living without both parents report they feel nervous when reading in front of others and feel lonely at school, compared to 35% and 33.4% of the girls that live with one or both parents.

Girls living in rural areas are more prone to report they get nervous when they read in front of others and that they do well in their tests because they are lucky than girls living in urban areas. Urban life may be correlated with greater confidence and a sense of self-determination. Girls living in Amhara follow the same pattern compared to girls living in Oromia.

93.9% of girls that have thought of migrating think that if they do well in a test it is because of luck, compared to 53.2% of girls who have not thought of migrating. More than 20% of the girls that thought of migrating report their family decides whether they go to school, at what age she gets married and how much time she spends with her friends. These figures are around 10% for girls who have not thought of migrating. Thus, there is seemingly a clear link between being at risk of migration and having less of a say in decision-making.

Finally, there are higher rates of girls that feel that early marriage can happen to them reporting that they do well in tests because of luck, compared to girls that do not feel early marriage can happen to them or have faced no such attempts.

Table 37: Girls' self-esteem and key barriers and characteristics

	I get nervous when I have to read in front of others	If I do well in a test it is because I am lucky	Family decides whether will go to school	Family decides what age will get married	Family decides how much time spent	Feels lonely at school

					with friends	
Living without both parents	21.4% / 35.0%	-	-	-	-	14.6% / 33.4%
Rural	38.3% / 26.1%	60.7% / 48.1%	-	-	-	-
Amhara	19.5% / 66.1%	42.1% / 89.0%	-	-	12.4% / 18.8%	19.7% / 53.2%
Books and learning material not available at school	24.8% / 35.6%	-	-	-	-	-
Computers not available at school to use	32.0% / 57.9%	52.5% / 89.5%	-	-	-	-
Seats not available for all students in class	64.5% / 31.5%	87.1% / 52.9%	22.6% / 9.4%	-	-	46.3% / 30.2%
Doesn't use drinking water facilities	29.6% / 37.6%	-	-	-	-	-
Doesn't use toilets at school	45.6% / 32.6%	68.4% / 54.3%	-	-	-	51.7% / 29.3%
Doesn't use areas where children play / socialise	60.0% / 33.8%	-	-	-	-	60.0% / 31.6%
Doesn't feel safe travelling to and from school	65.2% / 33.1%	-	-	-	-	-
Do you feel that early marriage can happen to you?	-	75.8% / 53.9%	-	-	-	56.0% / 29.4%
Have you ever thought of migrating to another area?	-	93.9% / 53.2%	24.5% / 9.5%	30.6% / 13.3%	26.5% / 12.1%	47.6% / 30.6%

5.4 Learning scores by Intermediate Outcomes

Table 38: Learning scores by intermediate outcome indicator

	Average literacy score (aggregate)	Average numeracy score (aggregate)	Sample size
All girls	74.7	67.2	1413
Good teaching quality	75.1**	67.4**	1133
Do you feel safe at school?	74.4	67.1	1165
Does your teacher(s) suggest ways you can continue to study after school/at home	73.9**	66.7***	1025
Have you used the school's library or reading corners in the last month?	73.6	66.5	832
Teacher asks more questions to boys	75.5	67.4	128
Teacher asks more questions to girls	75.6	68.8*	126
Teacher asks questions equally to boys and girls	74.1	66.8*	954
Were you absent from school for more than 5 days in a row last year?	75.1	67.8	249

Feels confident answering questions in class	74.3	67.1	1096
Feels lonely at school	74.1	66.8	277

*Notes: *, **, *** indicates t-test statistical significance at the 10%, 5% and 1% levels*

Table 38 shows the aggregate EGRA and EGMA scores by the key IO indicators. The table shows that there are some significant differences (marked by the asterisks) in the learning test scores of girls for the teaching quality outcome indicator, but not for either the attendance or self-esteem indicators. Girls who respond that they receive good teaching quality scored higher on both EGRA and EGMA compared to girls who did not. However, girls who reported that their teachers suggest ways they can carry on studying after school or at home earned lower test scores in both literacy and numeracy. This may be because teachers are less likely to give this kind of advice to girls who are high achievers, though this hypothesis would need to be tested further.

Girls who reported that their teachers ask more questions to girls earned higher test scores, though only the difference for EGMA is statistically significant. Girls who reported that their teachers ask more questions to boys also earned higher test scores, though these differences are not statistically significant. The bulk of respondents who said that teachers ask questions equally received lower grades.

The table shows that a number of the intermediate outcomes being measured by the project are correlated with learning scores. It is most clear that the project's theory of change holds for teacher quality, whereas the evidence is not as clear for self-esteem and attendance. However, it should be noted that the primary learning test scores have shown some inconsistencies throughout this document, and so these findings should be interpreted cautiously.

Project Checks on Intermediate Outcomes

- Ensure that the IO analysis is relevant to the logframe

6. Conclusions & Recommendations

6.1 Conclusions

1. *What is the profile of the project's beneficiaries and what are the barriers to learning and transition through stages of education that they face? (use findings from Section 3)*

The evaluation sample, which was broadly consistent with project level data on the mix of beneficiaries, indicates that on the whole **the project has reached the groups intended.** Sample beneficiaries are concentrated for example around the ages of 12-15 and the vast majority meet the project's definition of being a 'poor' household. Also, whilst the sample indicates that the project beneficiaries may slightly under-represent those with disabilities, in particular cognitive disabilities, coverage of other marginalised groups targeted by the project is good.

The **barriers identified by the evaluation largely confirm those posited in the project's Theory of Change**. We find that a significant minority (26%) of girls are subject to a high chore burden, impacting negatively on the time they have to study. Also, 41% of girls have to travel more than 1 hour to secondary school, which may negatively impact on their likelihood to transition to secondary education. Evaluation data shows that risky migration affects fewer girls than the project estimates, though the evaluation estimate is likely to be an underestimate as it only captures girls who themselves had considered migration. As identified by the project, we find some evidence of concerns about girls' safety, as a minority feel unsafe while travelling to and from, and while in school.

In addition to the barriers highlighted by the project, **the evaluation highlighted two further barriers to learning** which the project may need to consider: **language of instruction and teacher absence**. Around 1.1% of girls in project schools have a mother tongue different to the language of instruction in their school, and this is shown to be correlated with the likelihood that a girl becomes nervous when reading and doing maths in front of others. Moreover, 46% of girls state that teachers are often absent from class suggesting that teacher absence is a further barrier.

2. *What are baseline learning levels of the project's beneficiaries? What foundational literacy and numeracy skills do they master or lack? Do the learning levels vary by region, age, gender or any other subgroup? (use findings from Section 4)*

The **baseline level of learning outcomes for the project beneficiaries was generally low**, in particular for those in higher grades, across both literacy and numeracy. For example, overall 60.8% fail to achieve even grade 4 in either literacy or numeracy. In particular, whilst in general there appears to be reasonable proficiency at the lower level tasks (with the majority classed as established or proficient learners for the first two tasks in the numeracy test, for example), this drops off with more complicated tasks. Due to the requirement for performance across a range of tasks, the number of students achieving higher grade levels in both literacy and numeracy is therefore relatively low.

However, **the learning level outcomes also do not appear to vary with grade** as would be expected. Table 13 and

Table **15** show that the progression that would be expected across grades 4 to 8, which was evident in the pilot, was not present in the data. However, when the secondary learning tests are looked at separately and the benchmarking sample is included, the data do show significant improvements in test scores as grade increases. The lack of progression in primary does not appear to be driven by any subtask in particular and may therefore be due to the aggregation from item to subtask level data. It will be possible to explore this issue in more detail as part of the midline evaluation.

The baseline data also suggest that **average test scores are slightly higher in Amhara than in Oromia**. This is consistent with the results in the pilot, where results for tests taken in Oromiffa were found to be systematically lower than those for tests taken in Amharic. It was noted in the pilot that this effect may be due to differences in the language of instruction rather than differences in the ability of students across regions. Other than these small regional differences, we do not observe any subgroup characteristics playing a significant role in determining learning test scores.

Overall, the data suggests that there is **the potential for learning outcomes to improve between baseline and midline and endline**. The extent to which learning outcomes may be observed to improve is limited by the reduced gap of one year between baseline and midline.

3. *What are the baseline transition rates in project's areas? What are the key transition points that project beneficiaries will go through? Do the transition rates vary by region, age, gender or any other subgroup? (use findings from Section 4)*

Project beneficiaries encounter a number of key transition points as they progress through the education system. For the purposes of the evaluation which tracks a sample of girls in grades four to eight at baseline, the key transition point is the move from primary school to post-primary, for which a successful transition includes moving to secondary school, a TVET, or fairly paid employment.

The evaluation finds a baseline transition rate of 66% for project areas. However, the aggregate figure masks considerable variation between subgroups. **Age is shown to be a significant determinant of transition**, with rates of around 80% from ages 8 to 13 but around 45% from ages 14 to 19. This may indicate that transition from primary to secondary school is a particular hurdle for students, perhaps reflecting cultural norms, the long distances which some students need to travel to get to a secondary school, or student difficulties in completing the final grade before secondary school.

We also find that being an orphan and finding it unsafe to travel to school are both strongly associated with lower levels of transition and that transition rates in Oromia are considerably lower than in Amhara (44% vs 75%).

4. What is the baseline Sustainability Score of the project at community, school, and system level? What factors are likely to hinder/support the sustainability of the project's activities and results? (use findings from Section 4)

The project is awarded an overall sustainability score of 1.3 at baseline, indicative of latent changes on average in the key areas of community, school and system. The scores for sustainability in community, school and system are 1, 2 and 1 respectively.

Community sustainability is hindered by the relatively low preponderance of households paying the costs associated with schooling. This is important in light of **qualitative data which suggests that a lack of financial support can lower transition rates substantially**, as many girls would drop out from school. There is, however, evidence of widespread endorsement of girls' education in communities, which is a positive and significant indicator.

School level sustainability is the strongest of the three sustainability components, as there is evidence suggesting that attitudes to girls' education are shifting and of teachers helping students to study effectively after school. School level sustainability could be raised further by increasing teacher retention rates, particularly in in the Arsi region.

At a system level there is evidence of the engagement of local officials in delivery of the project and of knowledge development, although it is unclear whether this has translated into changes in behaviour yet. **The midline and endline evaluation points could benefit from additional indicators to capture system level sustainability.**

5. What are the baseline levels of the intermediate outcomes indicators? Are the assumptions made by the project around the linkages between the intermediate outcomes and the outcomes clear and likely to hold? (use findings from Section 5)

The three intermediate outcomes selected by the project are attendance, quality of teaching, and girls' self-esteem. The project's Theory of Change posits that changes in these will lead to material changes in learning and transition outcomes.

Evidence for attendance is that it is generally high, with 87% of girls attending school most days that the school was open. Whilst secondary data from the project suggests attendance rates of between 95% and 99%, corroborated somewhat by evaluation data comparing the attendance registers to spot checks in class, this may be optimistically high and qualitative data suggests **we should exercise caution in assuming baseline attendance rates are unequivocally high**. Interviewees cite attendance as a key a problem, and community representatives report that children are frequently absent and drop out of school. As such, we may be observing a level of selection bias in the data available on attendance.

On teacher quality, some 89% of girls consider the quality of teaching they receive to be 'good' or 'very good'. Furthermore, qualitative data suggests that teaching quality has improved in the last year and that pupils are enjoying school more compared to the previous year. Teachers and community representatives do however report that inadequacy in resources, for example textbooks and computers, and of facilities, such as water sources and separate toilets for boys and girls, may reduce the capacity of teachers to oversee quality teaching. **The midline evaluation would benefit from a measure of the quality of teaching based on in-classroom observation**, for which the project is currently developing a protocol.

There is **mixed evidence around girls' self-esteem and self-efficacy and we consider that some girls may be presenting with issues on these measures**. For example, around a third of girls report that they get nervous when reading in front of others, with a slightly smaller proportion reporting the same for mathematics. Moreover, roughly a third of girls report feeling lonely at school, perhaps suggesting that they may not be properly integrated. Nevertheless, most girls report that they want to do well at school, that they are just as capable as their peer group, and that they want to continue studying, all of which may be positive indicators of their potential progress at future evaluation points.

Table 38 shows that of the three intermediate outcomes, **student-perceived teacher quality is most associated with higher learning outcomes**. The link from attendance and self-esteem to learning

outcomes is less clear, although this may be because the measures used to proxy those intermediate outcomes are imperfect and it does not in itself refute the position of those outcomes in the project's theory of change.

6. What is the projects' approach to addressing gender inequalities? Is the project design gender sensitive? How is the project promoting gender equality through interventions? Are there aspects where the project is being gender transformative? Are there any risks associated with in their approach to gender? Respond to the extent possible.

The project **specifically targets and designs its interventions around the most marginalised girls**, such as girls with disabilities or at risk of migration or early marriage, which ensures the project is gender and disability sensitive. By targeting the most marginalised girls, supporting their families, schools, and communities, and challenging social and gender norms, the project is actively seeking to transform social and gender inequalities over the long term, in line with MEL Guidance Part 2: Appendix F.

The project seeks to improve the behaviours and attitudes and gain the support of key stakeholders in the girls' lives, such as parents, boys, teachers, and other community members. This is being achieved through several key activities, primarily those linked to Outputs 1 and 4.

On Output 1, project data on project girls is disaggregated by age, grade, location, religion and disability and fed through to inform future project interventions. Alongside this, data is further analysed in relation to girls with disabilities and the different types of impairments within the cohort.

As part of Output 1, the Girls' Clubs identify and discuss the negative behaviours linked to social norms in their region, including early marriage, the unequal distribution of chores and migration for work. They then lead activities designed to challenge the social norms in school, at home and in the wider community. Boys' Clubs have been invited to Girls' Club events to discuss negative behaviours. Other areas of support include setting up buddy systems to support girls on the journey to and from school, acting as collectors of information to prevent abduction, forced marriages, teenage pregnancies and early marriages and route the information back to focal teachers and boys clubs leaders. There is also a two-way help system in place where girls also help boys who are more marginalised. School- and community-led child safeguarding mechanisms have received safeguarding and inclusion training and facilitate school and community discussions on topics such as early marriage and gender-based violence and advocate with the local authorities to take action against such behaviour. Keeping in mind Do No Harm, the girls have been trained in negotiation techniques, for example to be assertive but not aggressive and to avoid human rights terminology, so as not to receive negative responses from the Ethiopian authorities.

Activities linked to output 4 also seek to improve behaviours and attitudes where girls use peer education approaches in Girls' Clubs to share their learning with other girls on a range of topics. The Clubs set their own agendas, take minutes, prepare their own action plans and lead on the implementation of the plans. Through school radio and drama performances to the wider community, the girls challenge negative social norms through broadcasting messages and raising awareness on barriers to girls' education. Girls have been actively involved with the setting up and maintenance of 'sanitary corners', including fund raising for materials. The corners have not only been used to support the girls during the menstrual cycle, but the presence of the corners has been used for discussion purposes in boys and girls clubs, to remove the stigma attached to menstruation and to raise awareness of the need for good menstrual hygiene. Following the principle of Do No Harm, the project has avoided human rights language and terminology as this could potentially put the girls, boys and project staff at odds with the local authorities.

Given that the project's interventions have been shown to **actively target areas of inequality in gender and other characteristics**, and that there is **preliminary evidence that attitudes and behaviours are changing** in project areas, the project is **GESI transformative** in its approach.

Recommendations

The evaluation team recommends that:

Monitoring, evaluation and learning

1. The project should consider the inclusion of additional indicators in the logframe to more fully capture system-level sustainability. For the project to advance beyond a score of 1 for system-level sustainability, it must show evidence of improved capacity of local officials and of engagement by government with the project. The sole indicator currently tracked does not allow for these aspects to be measured.
2. The project should also consider revisions to the indicators used to capture school and community-level sustainability. This would allow the project to have a more holistic oversight of sustainability on these levels and to fully capture changes in sustainability at midline and endline.
3. The project should ensure that classroom observations are in place for midline and that these observations allow for inferences to be drawn not just on the current levels of teacher quality but also on whether teaching is improving as the result of the project.
4. The project should consider tracking teacher attendance to ascertain the extent to which this may be a barrier to learning and whether any project-level intervention is necessary. A majority of girls report this to be an issue, but quantitative data would allow for a more thorough understanding.

Design

1. The project should consider the following adaptations to its design, to improve learning and transition outcomes at midline:
    - Qualitative data suggests that the project's letter-link boxes are a useful and effective tool for reporting cases of early marriage and to encourage girls to express their feelings, concerns and challenges. To maximise their impact, the project should consider raising awareness of their benefits and uses.
    - Benchmark transition data shows that transition rates drop sharply around the age where students should transition from primary to secondary school. The project should consider whether adaptation of project interventions is needed to provide extra support to girls at this risky stage of transition. The project should also seek to develop an understanding of why transition rates seem to be lower in Oromia than other areas and whether mitigation strategies may be needed.
    - There is evidence that a small proportion of girls (4%-7%) may be marginalised as a result of having a mother tongue which is different from the language of instruction of the school which they attend. The project should consider whether additional support is needed for this group.

Scalability and sustainability

1. The project's approach to community-level sustainability relies on reducing the number of girls receiving support to pay for school-related costs over time with households themselves picking up the slack. This operates under the assumption that households will be willing to do so as a result of the project's influence in shifting attitudes and behaviours in favour of girls' education. Evidence shows that this assumption is highly unlikely to hold, as households are unable rather than unwilling to pay. The project should consider possible approaches to address this, such as promoting the economic wellbeing of the worst-off households.
2. Sustainability in schools relies on project-trained teachers remaining in post, to continue to provide improved teaching to pupils and pass on their knowledge to teachers who have not received project-training. Low levels of teacher retention in South Wollo, possibly due a 'brain drain' effect given other opportunities available locally, is likely to inhibit this. The project should consider approaches to increasing teacher retention, particularly in South Wollo.

Project contribution: Response to conclusions and recommendations

- The recommendations above should come from the External Evaluator. The project should add a short response to the recommendations in light of the conclusions of the Baseline Evaluation Report in Annex 13.
- Project response to evaluators' comments on gender approach used and how well gender is integrated through the project.

Annexes

Annex 1: Logframe

091017_CH5170_ Logframe.xlsx

Annex 2: Outcomes Spreadsheet

GEC-T Outcomes Spreadsheet v3.0.xlsx

Annex 3: Key findings on Output Indicators

This annex should be completed by the project.

Table 39: Output indicators

Logframe Output Indicator	Means of verification/sources	Collection frequency
Number and Indicator wording		E.g. monthly, quarterly, annually. NB: For indicators without data collection to date, please indicate when data collection will take place.
Output 1: Safer and stimulating learning spaces for girls		
Output 1.1: % of girls reporting that they feel safe in school (yes/no)	HH survey (girls questionnaire) - with FGDs and Individual Interviews with girls at baseline, midline and endline for further explanation as appropriate. Project's Reporting formats from letter link information annually.	Baseline, midline, endline
Output 1.2: % of girls at primary school who report accessing reading corners in the last month	HH survey (girls questionnaire) with girls at baseline, midline and endline.	Baseline, midline, endline
Output 2: Effect of transition costs (financial, material and psychological)		
Output 2.1: % of girls receiving transitioning cost support (tuition fees and accommodation) transitioning to TVE,	School and TVE Centres enrolment registers collected bi-annually (collected by project) and monitoring data provided by project detailing which parents / girls received transition cost support (reception checklist)	Bi-annually quarterly

and secondary education		
Output 2.2: % of girls who receive scholastic materials who have completed primary/secondary education at the end of the project	% of girls who receive scholastic materials who have completed primary/secondary education at the end of the project	Quarterly
Output 2.3: % of girls who receive scholastic materials who have completed primary/secondary education at the end of the project	HH survey (girls questionnaire) with girls at baseline, midline and endline. Project's tutorials register and attendance formats collected by project annually and monitoring data provided by project detailing which parents / girls received academic support	Baseline, midline, endline
Output 3: Well-trained, mentored and supported teachers confident in adapting more effective mathematics and literacy (spoken language, reading and writing) pedagogies for the remote contexts		
Output 3.1: Number of teachers trained by project	Monitoring data provided by project	1 – 2 monthly. Same teachers tracked over time to observe improvement,
Output 3.2: % of trained teachers who receive mentoring by "supervisors" (cluster supervisors, principals, and education bureau experts) to implementing the learning	HH survey (girls questionnaire) - with FGDs, and Individual Interviews with girls for wider explanation as appropriate - at baseline, midline and endline. FGDs with girls run by project annually	Baseline, midline, endline
Output 4: Empowered girls with key life values, skills and challenged norms (output will be further adapted following the work on social norms component).		
Output 4.1 % of girls who report a more equitable division of household chores (male siblings have increased their collaboration in domestic chores)	HH survey - with FGDs, and Individual Interviews with girls for explanatory purposes as appropriate - at baseline, midline and endline. Project's FGD with girls	Baseline, midline, endline
Output 4.2: % of girls with disabilities who report improved attitudes in the community and school (peers) towards them	HH survey - with FGDs, and Individual Interviews with girls for explanatory purposes as appropriate - at baseline, midline and endline. Project's FGD with girls	Baseline, midline, endline
Output 5: Girls who choose vocational education or independent adulthood develop entrepreneurship & employability skills		

Output 5.1: % of girls attending TVET who completed TVET education	TVE Centres enrolment registers (collected by the project), HH survey (for triangulation)	Bi-annually Baseline, midline, endline
Output 5.2: % of girls attending TVET who completed COC (certificate of competence)	TVE Centres enrolment registers (collected by the project); HH survey (for triangulation)	Bi-annually Baseline, midline, endline
Output 5.3: % of girls (drop-out girls or who attended TVET) who start up their own business	Project's follow up questionnaires collected annually (collected by the project); HH survey (for triangulation)	Baseline, midline, endline

Report on the Baseline values/Baseline status of each Output Indicator in the table below. Reflect on the relevancy of the Output Indicator for your Intermediate Outcomes and Outcomes and the wider Theory of Change based on the data collected so far. Are the indicators measuring the right things? What do the Baseline values/Baseline status mean for the implementation of your activities?

Table 40: Baseline status of output indicators

Logframe Output Indicator	Means of verification/sources	Collection frequency
Number and Indicator wording	What is the contribution of this indicator for the project ToC, IOs, and Outcomes? What does the Baseline value/status mean for your activities? Is the indicator measuring the right things? Should a revision be considered? Provide short narrative.	What is the Baseline value/status of this indicator? Provide short narrative.
Output 1: Safer and stimulating learning spaces for girls		
Output 1.1: % of girls reporting that they feel safe in school (yes/no)	Links to attendance and learning outcomes. Girls need to feel safe in school so that they can attend. Has direct impact on outcome 1	
Output 1.2: % of girls at primary school who report accessing reading corners in the last month	Has direct impact on outcome 1. Targeted reading links to an increase in learning outcomes.	
Output 2: Effect of transition costs (financial, material and psychological)		
Output 2.1: % of girls receiving transitioning cost support (tuition fees and accommodation) transitioning to TVE, and secondary education	Links directly to outcome 1, learning outcomes, and attendance – IO1. Also feeds through to future transition points.	

Output 2.2: % of girls who receive scholastic materials who have completed primary/secondary education at the end of the project	Links directly to outcome 1, learning outcomes, and attendance – IO1. Also feeds through to future transition points.	
Output 2.3: % of girls receiving academic support (accelerated tutorials) who transitioned to secondary education	Links directly to outcome 1, learning outcomes, and attendance – IO1. Also feeds through to future transition points. HH survey (girls questionnaire) with girls at baseline, midline and endline. Project's tutorials register and attendance formats collected by project annually and monitoring data provided by project detailing which parents / girls received academic support	
Output 3: Well-trained, mentored and supported teachers confident in adapting more effective mathematics and literacy (spoken language, reading and writing) pedagogies for the remote contexts		
Output 3.1: Number of teachers trained by project	Links directly to outcome 1 and 3 (school sustainability), has a direct impact on transition. Links to IO2	
Output 3.2: % of trained teachers who receive mentoring by "supervisors" (cluster supervisors, principals, and education bureau experts) to implementing the learning	Links directly to outcome 1 and 3 (system sustainability), has a direct impact on transition. Links to IO2.	
Output 4: Empowered girls with key life values, skills and challenged norms (output will be further adapted following the work on social norms component).		
Output 4.1 % of girls who report a more equitable division of household chores (male siblings have increased their collaboration in domestic chores)	Links directly to IO3	
Output 4.2: % of girls with disabilities who report improved attitudes in the community and school (peers) towards them	Links directly to IO3.	
Output 5: Girls who choose vocational education or independent adulthood develop entrepreneurship & employability skills		

Output 5.1: % of girls attending TVET who completed TVET education	Links directly to outcome 1 and 2 and IO1	
Output 5.2: % of girls attending TVET who completed COC (certificate of competence)	Links directly to outcome 1 and 2 and IO1	
Output 5.3: % of girls (drop-out girls or who attended TVET) who start up their own business	Links directly to outcome 1 and 2 and IO1,	

List all issues with the means of verification/sources or the frequency of data collection which require changes or additions.

Table 41: 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: Safer and stimulating learning spaces for girls		
Output 1.1: % of girls reporting that they feel safe in school (yes/no)	HH survey (girls questionnaire) - with FGDs and Individual Interviews with girls at baseline, midline and endline for further explanation as appropriate. Project's Reporting formats from letter link information annually. No issues	
Output 1.2: % of girls at primary school who report accessing reading corners in the last month	Database information can link to learning outcome data, i.e. are girls that attend the reading corners high achievers?	
Output 2: Effect of transition costs (financial, material and psychological)		
Output 2.1: % of girls receiving transitioning cost support (tuition fees and accommodation) transitioning to TVE, and secondary education	The proposed way to monitor needs to be revised. Collection of registers is not sufficient.	
Output 2.2:	HH survey (girls questionnaire) at baseline, midline and endline. School enrolment registers collected	

% of girls who receive scholastic materials who have completed primary/secondary education at the end of the project	bi-annually and monitoring data provided by project detailing which parents / girls received scholastic materials (reception checklist) School enrolment registers needs to be revised. Mapping needs to take place.	
Output 2.3: % of girls receiving academic support (accelerated tutorials) who transitioned to secondary education	HH survey (girls questionnaire) with girls at baseline, midline and endline. Project's tutorials register and attendance formats collected by project annually and monitoring data provided by project detailing which parents / girls received academic support Data on results in in-class assessments to be linked to national tests/school tests so that trends in learning can be identified (project capacity to be assessed to carry out this)	
Output 3: Well-trained, mentored and supported teachers confident in adapting more effective mathematics and literacy (spoken language, reading and writing) pedagogies for the remote contexts		
Output 3.1: Number of teachers trained by project	Lesson observation tool - adequate	
Output 3.2: % of trained teachers who receive mentoring by "supervisors" (cluster supervisors, principals, and education bureau experts) to implementing the learning **Review indicator**	Links directly to outcome 1 and 3 (system sustainability), has a direct impact on transition. Links to IO2. Needs to be reviewed	
Output 4: Empowered girls with key life values, skills and challenged norms (output will be further adapted following the work on social norms component).		
Output 4.1 % of girls who report a more equitable division of household chores (male siblings have increased their collaboration in domestic chores)	Participatory focus group interviews could be conducted	
Output 4.2: % of girls with disabilities who report improved attitudes in the community and school (peers) towards them	Participatory focus group interviews could be conducted	
Output 5: Girls who choose vocational education or independent adulthood develop entrepreneurship & employability skills		
Output 5.1:	TVE Centres enrolment registers (collected by the project), HH survey (for triangulation)	

% of girls attending TVET who completed TVET education	Database can provide attendance regularity	
Output 5.2: % of girls attending TVET who completed COC (certificate of competence)	TVE Centres enrolment registers (collected by the project); HH survey (for triangulation) Database can provide attendance regularity	
Output 5.3: % of girls (drop-out girls or who attended TVET) who start up their own business	Project's follow up questionnaires collected annually (collected by the project); HH survey (for triangulation) Database can provide attendance	

Annex 4: Beneficiary tables

This annex should be completed by the project.

Please fill in the tables below. Individuals included in the project's target group should be direct beneficiaries of the project.

Table 42: Direct beneficiaries

Beneficiary type	Total project number	Total number of girls targeted for learning outcomes that the project has reached by Endline	Comments
Direct learning beneficiaries (girls) – girls in the intervention group who are specifically expected to achieve learning outcomes in line with targets. If relevant, please disaggregate girls with disabilities in this overall number.	**[This should align with the total beneficiary numbers reported in the outcomes spreadsheet]** **As of October 2018**	**[This may equal the total project number in the outcomes spreadsheet and in the column to the left, or may be less if you have a staggered approach]**	**[Projects should provide additional information on who they are and the methodology used. If the numbers have changed since Baseline, an explanation should be provided]**
Marginalised girls – Consisting of ….	16,481	Approach not staggered	Methodology used is monthly recording of attendance data that is inserted into the database after collection Project staff monitor at risk of drop out categories (within the month) outside of the monthly database input, and connect with

			project, school and community staff to intervene with the girl and family members where they see risk of drop out. In our Q7 report we have logged this under 'track and help truant girls to resume school' intervention. Below is an example of this activity, that impacts on attendance rates.

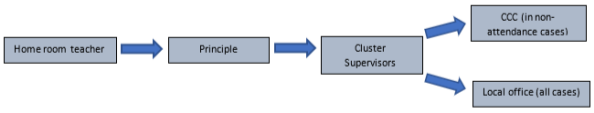
The project team with kebele officials and community volunteers have made house to house visits and had discussion with girls and their parents to persuade and allow the girls to return to school. As a result of tracking activity in Q7, 391 girls (380 in Arsi, 4 in SW & 7 in SG), who were previously reported as dropped out in the 2010 Ethiopian Academic (2017-18 European calendar year) have returned to school. This is in addition to the 586 girls who returned to school in Q6 as a result of project interventions. Further work is also planned to sub categorise drop out categories so that 'return to school' interventions can be adapted as required.

I. Education

1. **Regular attendance:**

Function: collect attendance data

- Collection frequency: monthly
- Triangulation: quant. triangulate with spot checks
- Trend analysis/correlation: analyse by school, teacher



Homeroom teachers submit attendance on a monthly basis using paper report format form generated by the database. Local office staff are the KMO, Community Worker.

2. **Spot checks attendance:**

Function: check that attendance data collection is being collected accurately

- Collection frequency: bi-annually (twice a year)
- Triangulation: quant. triangulate with attendance collection
- Trend analysis/correlation: analyse by school, teacher

Currently the spot checks on attendance are collected twice a year, which check for accuracy on regular attendance. In cases where anomalies are detected, the local office staff meet with the school to inform of the inconsistency and select required staff for closer monitoring.

Primary school girls	11,601	Approach not staggered	Numbers based on database entries
Secondary school girls	2,851	Approach not staggered	Numbers based on database entries
Girls in alternative institutions - CBE/ALP / TVET	307	Approach not staggered	Numbers based on database entries
Out-of-school girls that have enrolled into alternatives	0	Approach not staggered	

Table 43: 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.	0	Boys do not benefit from GEC-T learning interventions.
Broader student beneficiaries (boys) – boys who will benefit from the interventions in a less direct way, and therefore may benefit from aspects such as attitudinal change, etc. but not necessarily achieve improvements in learning outcomes.	3,362, although this may be more as those 3,362 boys may influence their peers who are not actively involved.	Boys attending boys' clubs. Boys do not have the same exposure as girls but are impacted by the attitudinal changes in GEC-T girls, Teaching staff, school staff (including leadership staff), community workers all impact on attitudinal changes in boys. It is not sufficient to say that all school boys will be impacted directly and is not evidenced throughout the programme.
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.	Total of girls in GEC targeted schools, including those that are not in the program (assuming non-program girls benefit indirectly) This totals to 51,139 (34,658 are indirect beneficiaries)	Non-GEC girls, who do not partake in GEC-T interventions are impacted by the attitudinal change in GEC-T girls. It is not sufficient to say that all school girls will be affected and is not evidenced throughout the programme.
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.	Direct training - Women: 167 Direct training – Men: 315	
Broader community beneficiaries (adults) – adults who benefit from broader interventions, such as community messaging /dialogues, community advocacy, economic empowerment interventions, etc.	Unknown It's expected that all GEC communities are affected.	although all communities in connection with schools will be affected by targeted messaging, drama and theatre awareness raising, home based conversations. Arsi – 20 communities South Gondar – 35 communities South Wollo – 28 communities

- Tables 3-6 provide different ways of defining and identifying the project's target groups. They each refer to the same total number of girls, but use different definitions and categories. These are girls who can be counted and have regular involvement with project activities.
- The total number of sampled girls in the last row of Tables 3-6 should be the same – these are just different ways of identifying and describing the girls included in the sample.

Table 44: Target groups - by school

School Age	Project definition of target group (Tick where appropriate)	Number targeted through project interventions	Sample size of target group at Baseline
Lower primary	Grade 1 – grade 4 Age 7 – 10	2407	
Upper primary	Grade 5 – grade 8 Age 11 – 14	9194	
Lower secondary	Grade 9 – grade 10 Age 15 – 16	2252	
Upper secondary	Grade 11 – Grade 12 Age 17 – 18	599	
Total:			[This number should be the same across Tables 3, 4, 5 & 6]

Table 45: Target groups - by age

0	Project definition of target group (Tick where appropriate)	Number targeted through project interventions	Sample size of target group at Baseline
Aged 6-8 (% aged 6-8)	✔	337	
Aged 9-11 (% aged 9-11)	✔	4641	
Aged 12-13 (% aged 12-13)	✔	4732	
Aged 14-15 (% aged 14-15)	✔	3139	
Aged 16-17 (%aged 16-17)	✔	1020	
Aged 18-19 (%aged 18-19) -	✔	264	
Aged 20+ (% aged 20 and over)			
Total:			[This number should be the same across Tables 3, 4, 5 & 6]

Table 46: Target groups - by sub group

Social Groups	Project definition of target group (Tick where appropriate)	Number targeted through project interventions – based on figures in original proposal	Sample size of target group at Baseline
Disabled girls (please disaggregate by disability type)	✔	500	
Orphaned girls		0	
Pastoralist girls	✔	14711	
Child labourers		0	

Social Groups	Project definition of target group (Tick where appropriate)	Number targeted through project interventions – based on figures in original proposal	Sample size of target group at Baseline
Poor girls	✔	16,481	
Early marriage	✔	1,265	
Risky migration	✔	2,270	
High domestic work burden	✔	12,690	
Total:			[This number should be the same across Tables 3, 4, 5 & 6]

Table 47: 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			
Out-of-school girls: have attended school, but dropped out			
Girls in-school	✔	**16481**	

Annex 5: MEL Framework

MEL Framework

Annex 6: External Evaluator's Inception Report (where applicable)

It was agreed with the fund manager earlier in the evaluation process that given the level of detail provided in the MEL an Inception Report is not required for the ChildHope evaluation.

Annex 7: Data collection tools used for Baseline

Learning Tests

EGRA/EGMA (Amharic, English, Oromiffa)


Protocols EGRA EGMA Amharic Baseli


Stimulus Booklet EGMA EGRA Amharic


Protocols EGRA EGMA English Baselin


Stimulus Booklet EGMA EGRA English E


Protocols EGRA EGMA Oromiffa Basel


Student Stimulus EGRA EGMA Oromiffa

SEGRA/SEGMA


SeGMA Test (BL).docx


SeGMA Mark Scheme (BL)


SEGRA Test/Mark scheme (BL)

Quantitative Instruments


Girls School Survey.xlsx


Household Survey (Core cohort).xlsx


Household Survey (BT).xlsx

Qualitative Instruments (KII= Key informant interview, FGD= Focus group discussion)


Boys KII


Community Rep KII.docx


Girls KII


Primary Caregiver KII


Teacher KII


Community Rep FGD


Girls FGD.docx

Example Transcripts


Boy 1 - Sire.docx


Girl 1 - Dera.docx


Teacher M - Dera.docx


CR F - Dera.docx

Observation tools (*Not used as part of Baseline)*


1. Observation Tool (Lessons)


2. Observation Tool (Homework Support)

Annex 8: Datasets, codebooks and programs

i. Input Data:


inputs.zip

ii. Do Files:

Master Do file:


master_baseline.do

Other Do files:

iii. Read-me File:

iv. Codebooks:

Annex 9: Learning test pilot and calibration

A pilot of all learning test instruments was carried out in February 2018. In the case of EGRA/EGMA, all 9 language and data point combinations were piloted. For SeGRA and SeGMA, which is tested only in English, the baseline, midline and endline tests were piloted. Piloting was carried out with the intention of detecting floor and ceiling effects in subtasks and items, to investigate comparability across languages and test versions, and to show progression over grades (or lack thereof).

As the EGRA/EGMA tests had already been used in GEC-1, it was expected that they would require little to no revision. Nevertheless, piloting did lead to a number of proposed changes. Firstly, evidence of ceiling effects in the lower EGRA/EGMA tasks were observed in higher grades. This is to be expected as EGRA/EGMA is designed to be taken by pupils in grades 1-4. As a result, the task allocation was altered slightly so that grade 8 no longer took EGRA/EGMA 1 as was initially planned. Secondly, the test to be administered at baseline for students learning in the Oromiffa language yielded a significantly lower grade average than other test-language versions. Consequently, it was decided that students would take a different version of the Oromiffa test, with the baseline version being reviewed by a professional with suitable expertise at a later date.

The SeGRA/SeGMA test versions were newly created for GEC-T so it was expected that they may require more detailed revisions, which was indeed the case. Firstly, item-level floor effects were observed for 11 items across 3 test versions owing, for the most part, to inaccuracies in the mark scheme and/or question phrasing. Secondly, the mark schemes were found to be unclear in the awarding of partial points for workings out. The tests were also found to be too long, exceeding the test development guidance which stated that subtasks should be around 10 minutes. Furthermore, it was decided that SeGRA subtask 3 should not be administered, as many girls refused to answer it, possibly as it necessitated a long piece of prose to be written in the English language. This decision could be confidently made without incurring possible ceiling effects, as the achieved percentage in SeGRA subtask 2 was very low.

Thus, as a result of the pilot, a round of revisions was undertaken to correct the defect items, to make the mark scheme more consistent and clear and also to condense the subtasks in line with the guidance whilst still testing the necessary range of competencies. The grade-task allocation for baseline was also revised in light of the low completion for SeGRA 3.

Following revisions to the learning tests after the pilot, 8 EGRA/EGMA test versions[56] and 3 SEGRA/SEGMA test versions were calibrated to be of the same difficulty.

Annex 10: Sampling Framework

Sampling Framework

Annex 11: Control group approach validation

The approach to selecting the intervention and control girls was described in Section 2.4. Control schools were matched to the sample of intervention schools based on stratification according to zone and rurality composition of the beneficiary population. The sample selection of schools was followed accurately with a few exceptions. These exceptions mean that in practice controls schools were not matched to treatment schools along as many relevant school-level characteristics as would have been preferred.

This can be seen in Table 5 to Table 10, which compares intervention and control sample by region, grade, age, characteristics, and barriers. Some minor differences between intervention and control group on these categories are described in Section 3, and from it we conclude that the treatment and controls arms of the sample collected is broadly representative of the population within each region, age group, grade and other characteristics. These imbalances between both groups do not translate to differences in baseline learning levels – Section 0 does not reveal any differences between the intervention and control groups.

The differences between the intervention and control group on these characteristics should not affect the difference-in-difference evaluation, provided the same sampling strategy is used at midline and endline. This is because this evaluation approach measures the change in learning scores rather than absolute levels. It is therefore not affected by differences in baseline levels between intervention and control groups, only by the trend in the counterfactual scenario (where no intervention occurred). As long as the trend in the counterfactual scenario after the point of intervention is constant, that is, the gap between the intervention and control group remains constant after this point, the difference-in-difference approach will estimate an unbiased causal effect.

To mitigate the potential issues associated with the difference-in-difference approach, we will ensure the same sampling strategy is used at midline and endline in order to obtain appropriate comparison samples at these points.

[56] The 9th (and 3rd) Oromia test version to be calibrated at a later date.

The treatment sample is identified using the project database of girls being supported by the project (referred to as the *'CHADET database'* in the MEL Framework). Control girls are selected from control school rosters. Stratification is conducted to ensure balance between control and intervention samples in terms of grade, age and region to ensure comparability at midline. Girls in the treatment sample were also stratified according to a set of marginalisation criteria. A procedure was laid out in the MEL to select control girls on a similar process using pre-screening questions. In practice, this was not followed strictly, such that they we may not observe balance in terms of marginalisation criteria across treatment and control schools. Sample weights can be used to correct for this as part of the difference-in-differences approach.

Moreover, the sample selection of control schools ensuring comparability with treatment schools was constrained by a lack of available data meaning that in practice controls schools were not matched to treatment schools along as many relevant school-level characteristics as would have been preferred.

When comparing the treatment and controls group for similarity, it is most appropriate to start with characteristics which are not influenceable by the project itself. We should also consider the sample composition as discussed previously, and whether this may distort findings. Table 5 shows that the treatment and control samples are broadly aligned by region, except in the cases of South Wollo where 21% of the treatment girls are located versus 34% of control girls. To the extent that South Wollo differs in terms of underlying trends in learning and transition, this may violate the parallel trends assumption inherent to the difference-in-differences approach, but could be corrected using weighting as above. Table 6 presents the evaluation sample breakdown by grade, and Table 7 by age. These show the evaluation sample to be balanced across these two dimensions. Table 8 breaks down the sample by type of disability. It shows that 10.4% of girls in the control group and 6.1% of girls in the intervention group have at least one of the following impairments: vision, hearing, mobility, cognitive, self-care or communication. Moreover, the proportion of girls with disability in the control group is larger than the proportion of girls with disabilities in the intervention group, for every type of impairment. In the control group cognitive and communication impairments are the most common, with 4.6% of the girls having cognitive impairments, and 3.7% having communication impairments. In the intervention group hearing, visual and cognitive impairments are among the most common ones, affecting 1.6%, 1.4% and 1,4% respectively. The treatment and control samples are slightly less well-aligned in terms of disability, but still broadly comparable. In terms of the overall representativeness of the sample with the population sample as defined in the project database, the proportion of girls with disabilities in our sample is higher than the population figures. While in Table 7 we report 6.1% with disabilities in the intervention group, this figure is estimated at 3% for the whole population.

Table 8 breaks down the sample by disability prevalence. Girls in the treatment sample are less likely to present with a disability than girls in the control group, which is mostly driven by differences in the prevalence of cognitive and communication impairments.

Table 9 shows that treatment and control girls are broadly similar in terms of baseline characteristics, suggesting that treatment and control samples are generally comparable along observable characteristics. Differences in the proportion of girls orphaned, living without both parents, and living in female headed households are small. The absolute proportion of girls who are married in treatment and control schools is very low (0.1% T vs 0.2% C), as is the proportion of girls who do not speak the language of instruction (0.1% T vs 0.4% C). The proportion of girls whose mother tongue is different from the language of instruction is lower in treatment schools than in control schools (4% T vs 7.1% C). This may be driven by differences in sample breakdown across treatment and control samples. For example, treatment girls were somewhat oversampled in the region of Oromia, and thus undersampled in Amhara. Given that the language of

instruction is Oromiffa in Oromia, this difference may be picking up girls speaking Amharic as a mother tongue.

There is mixed evidence on whether girls attending treatment schools are poorer on average. 19% of primary caregivers in treatment schools say that it is difficult for them to afford sending their girl to school, whereas this only applies to 7% of girls in control schools. Households from treatment schools are also less likely to own land (10% T vs. 4% C). In contrast, households with girls attending control schools are more likely to be ‘unable to meet their basic needs (29% T vs 39% C). Control girls are also more likely to have ‘gone to sleep hungry for many days in the past year’ (3% vs 6%).

Annex 12: External Evaluator declaration

Name of Project: ChildHope

Name of External Evaluator: Simetrica

Contact Information for External Evaluator: See here for contact details

Names of all members of the evaluation team: Richard Houston, Christopher Warner, Ali Ladak, Dora Radosevic, Maria Palma, Yared Antonius (ADVA), Daniel Fujiwara

I, Richard Houston (Name) certify that the independent evaluation has been conducted in line with the Terms of Reference and other requirements received.

Specifically:

- All of the quantitative data was collected independently ((Initials: RH)
- All data analysis was conducted independently and provides a fair and consistent representation of progress (Initials: RH)
- Data quality assurance and verification mechanisms agreed in the terms of reference with the project have been soundly followed (Initials: RH)
- The recipient has not fundamentally altered or misrepresented the nature of the analysis originally provided by Simetrica (Company) (Initials: RH)
- All child protection protocols and guidance have been followed ((initials: RH)
- Data has been anonymised, treated confidentially and stored safely, in line with the GEC data protection and ethics protocols (Initials: RH)

Richard Houston
(Name)

Simetrica
(Company)

7th March 2019
(Date)

Annex 13: Project Management Response

This annex should be completed by the project.

Project Response to Characteristics and Barriers

Attitudes toward girls' education

The project is pleased to see that attitudes to girls' education are shifting positively and that communities are supporting girls' education reflected in data such as 'higher chore burden' placed on girls in the control group (table 10), where caregivers put higher demands on their girls (50.3% v 26.3%). We are also eager to develop in-project tools to have a more accurate figure across project, as the data in table 10 is drawn from a sample.

Language of instruction

The project feel that the difference between school grades, where English becomes the language of instruction in Amhara and Oromia, will definitely have an impact on learning outcomes (Amhara, switch at grade 7 and in Oromia, switch at grade 9). With girls from Amhara having an advantage over girls in Oromia as they are stronger in English up to grade 7. This would apply to tests that use the language of instruction in English and assessments that test English language ability.

investigation is required to know more about this area so we can factor in additional support. All girls who have L2 (second language acquisition) and LOI as a barrier will need intervention.

Transition

In light of the findings relating to transition we feel that further work needs to take place to work with girls who travel long distances to school in order to reduce the risk which should impact on transition. The project interventions we already have in place address safety through awareness raising, which can be more targeted in areas of need and to make sure the girls are travelling accompanied or in groups.

We also notice that the report states the decrease in transition is "mainly driven by more than half of the girls repeating the same course from last year" (P69). Further investigation will take place on this via our data collection on transition.

Risky migration

Girls at risk of abduction is very much outside the control of the project, where approximately 3 out of 6 kebeles in Oromia have reported facing this risk. The work we currently do with the girls' and boys' clubs can be continue to address these issues, while also considering 'Do No Harm.'

Sustainability

School sustainability - The project has observed examples of these such as in Oromia there has been a challenge in reinstating teachers after political unrest which may have affected retention data. In the case of teachers who live away from their families and teach in schools in and around Dera, attendance drops at the beginning of the week after a slow return back to work after visiting family over the weekend. Further comments regarding how the project intends to respond to this are detailed in our response to the recommendations and table 30 that gives the project response to sustainability.

Community Sustainability

Whilst we are pleased to see that the attitudes to girl's education are shifting positively and the project agrees that sustainability at community level needs further attention. Where there has been a shift in attitudes in favour of the value of girls' education, girls are still at risk of being pulled away from school due to households not being able to afford the cost of schooling (addressed in output 2). The fact that there is a change in attitudes to girls schooling does not address the issue of lack of affordability. It has also been observed in-project that orphans and most vulnerable mothers from the most vulnerable families are in need of livelihood support, amongst others. Lack of focus on this indicator could jeopardise the success of this outcome.

The livelihoods intervention was a successful feature in GEC-1 but was not continued through to GEC-T as there was an assumption that this would continue through to GEC-T, which has not held.

School Sustainability

It is also encouraging to read that school sustainability is the strongest and will increase if teacher retention rates improve. Teacher attendance is already collected at school level and the project will continue to assess the accuracy of this data through more external monitoring to improve reliability on data returns. Annual top up training is currently planned to prevent 'brain drain' on current school teams. This is also scheduled to be discussed in our leadership and management training, and management meetings at school level.

System Sustainability

We agree that further discussion needs to take place around system sustainability and how the project plans the gradual release of activities (relating to teacher quality improvement) to system level. Further monitoring of this also needs to be considered . .

With regard to attendance. Attendance data will continue to be collected and fed through to the database, in particular to investigate the external evaluator's response that attendance rates are 'unequivocally high' which could be a result of selection bias. However, it should not go unnoticed that the project goes to great lengths to track girls who are at risk of drop out, with CCC, community worker and focal teacher involvement, to prevent drop out and reinstate girls where necessary. Benchmark setting, and the sharing of standard operating procedures and process flows to establish consistent practice are some of the actions included in the project's approach to improve the monitoring of attendance.

Differences in data between Oromia and Amhara

Generally, we find it is not surprising we see a difference in barriers to education by characteristic (table 11) between Oromia and Amhara as it has been widespread knowledge that beneficiaries in Oromia face more challenges that beneficiaries in Amhara, plus there have also been setbacks due to political instability which have had an impact on a range of factors. We also observe the differences between Muslim and

Christian families, there is a higher concentration of Muslim families in Oromia, alongside other sub-characteristics noted in the report that we would like to explore further. This information will better able us to tailor our interventions better. We also notice that as the learning outcomes between Amhara and Oromia are different, with Amhara scoring higher than Oromia. We would also like to explore resilience and the role it plays in contributing to self-esteem, self-efficacy and learning.

Girls With Disabilities

As mentioned in section 3.5, a higher number of girls have been identified with disabilities than originally anticipated alongside a marked difference between the treatment and control groups on disability in general – (intervention) 6.1% v 10.4% (control) in table 8, and cognitive impairment being more than three times as prevalent in the control group as in the treatment group (4.6 % vs 1.4%). These findings will have an impact on how we tailor our support interventions to GWD.

A possible reason for an increase in numbers is a more realistic representation of the full beneficiary cohort that has been reflected through the Washington Group Questions in an environment where respondents have felt more comfortable to respond to the questions asked. This will be addressed via our scheduled interventions.

Page 45 also mentions that a large number of girls with disabilities report that decisions are made for them on whether or not they will attend school, when they will marry and what work they will do after study. In terms of our targeted interventions relating to girls with disabilities (output 4) further work with the girls should take place to develop their levels of confidence to discuss such issues at home, and also to raise awareness on the challenges that GWD face with their families and communities. This does not change our Theory of Change or planned interventions.

Letter Link Boxes

It has been suggested that further awareness raising should take place around the benefits of letter link boxes and its uses as the qualitative data findings suggest that the number of students using letter link boxes is low and that a reason for this could be that they do not understand their benefit. In contrast to this there is evidence of girls using the boxes to communicate cases of early marriage in their community. Data collected from in-project (under output 1) suggests that letter link boxes are widely used, and in some cases, girls trust the case reporting system enough to verbally report cases to their focal teacher or principal. However, as letter link boxes are not as established in secondary schools we will focus awareness raising on secondary school use more, in addition to monitoring the reporting of cases.

Learning Outcomes

We acknowledge that there is lack of grade progression in both literacy and numeracy particularly as tasks become more difficult. Where it is logical that progression tapers off as tasks get harder we feel that further mapping on the cognitive requirements linked to each sub task or at item level (depending on the difference in cognitive requirements of each item) is required in order for teachers to target the areas of need linked to the curriculum. This happens on a large scale, with the national exams, but tends to happen at class and group level, not individual level.

We cannot comment on the lack of progression in learning test scores across grades being linked to subtask data only being available as we do not have responses at item level so are unable to remark further on the progression. We agree that if item level data cannot be recovered, pilot data should be used for us to a)

target set for midline and b) feed direction through to education officers and teachers on areas of development required for the girls to work on in order to boost their level of performance and overall ability to tackle items of higher difficulty, (P65). These findings do not change our theory of change. The suggestion that there is significant room for learning outcomes to improve between baseline and midline and endline is pleasing and should give sufficient time to the project to target improvement.

What is the project's response to the conclusions and recommendations in the report?

- The management response should respond to the each of the External Evaluator's recommendations that are relevant to the grantee organisation (see Section 6). The response should make clear what changes and adaptations to implementation will be proposed as a result of the recommendations and which ones are not considered appropriate, providing a clear explanation why.

1. The barriers identified by the evaluation largely confirm those posited in the project's Theory of Change.

High Chore burden - There is a possibility that the amount of girls who are subject to a high chore burden % may be higher as this has been reported as evident in project. Further investigation is required to ascertain a more accurate figure.

Risky Migration – We feel that the risk of abduction is very much outside the control of the project, an example of this is where approximately 3 out of 6 kebeles in Oromia have reported facing this risk. The work we currently do with the girls' and boys' clubs can be considered to address these issues, while it is also important for us to consider 'Do No Harm' and not put the girls at risk whilst trying to address the issues.

ADAPTATION: Amendment to tools that capture risky migration - The tools also used also need to be amended for midline and endline as they do not capture responses linking to girls at risk of abduction.

Language of Instruction

The project feels that the difference between school grades, where English becomes the language of instruction in Amhara and Oromia, will have an impact on learning outcomes (Amhara, switch at grade 7 and in Oromia, switch at grade 9). With girls from Amhara having an advantage over girls in Oromia as they are stronger in English up to grade 7. As mentioned on page 116, this would apply to tests that use the language of instruction in English and assessments that test English language ability.

For the 1.1% of girls whose mother tongue is different from the language of instruction, where some 80 girls may be negatively affected, so that we can ensure all girls facing this barrier can be supported we are interested to know the exact number of girls and what geographical areas this pertains to, so further characteristics or a trend can be identified. For example, what is the diversity in mother tongue across groups? As there are as many as 90 different ethnic groups in Ethiopia.

The project observes that dual languages operate in some towns, where families operate in a different language to their surroundings. This occurs particularly in the Oromia region where families converse in a different language to their surroundings. Families may converse in a different mother tongue to the operational and official language of the towns they reside, whereas in Amhara the mother tongue is Amharic. There are also cases in Dera where children speak two languages, Amharic and Afaan Oromo.

Further investigation is required to know more about this so we can factor in additional support. All girls who have L2 (second language acquisition) and LOI as a barrier will need intervention.

ADAPTATION: No adaptations to the Theory of Change are required, however, interventions need to be designed specifically to meet the needs of this newly identified cohort.

Teacher Absence

We note that teacher absenteeism is a barrier that had not originally been factored into our Theory of Change. Despite this, the control group (table 10) does report higher levels of absenteeism than the intervention group, 63.7% v 46%. As teacher absenteeism is seen as a national issue and in some cases lying outside of the control of the project our project interventions will only be able to address factors within the project control. We also do not have jurisdiction to reduce teacher absenteeism.

ADAPTATION: Addressing teacher attrition - An amendment to the Theory of Change will therefore need to take place with more structured interventions that reduce teacher retention in place. Currently we drive incentives for teacher attendance by increasing their motivation through strong relationship building within the school and between CHADET's staff who offer guidance within school-based communities of practice and through post lesson observation feedback. Education Officers also have a close working relationship with Woreda offices, school principals with the project leadership and management training addressing teacher absenteeism, alongside other school leadership and management areas. We are also developing the leadership capacities of school masters, Woreda education officers, and supervisors. Regular meetings are held at regional level to feed into government policy on their management of teacher attrition. Annual top up training is currently planned to prevent 'brain drain' on current school teams.

Learning Outcomes

In response to the findings that there is lack of grade progression in both literacy and numeracy (shown in table 21 & 22) particularly with more complex tasks - *'girls in higher grades are no more likely to have achieved higher grade levels than girls in lower grades …. the table shows that the majority of girls, around 60%, are performing at lower than grade level 4*, ' '*The lack of grade progression is observed again …' (P64)*, '*… girls struggle with EGMA 3 – missing number task, where they must achieve at least 80 on this subtask*' and '*the number of tasks that requires good performance was only consistently achieved by a few students across the five subtasks*' (P.63).

ADAPTATIONS: LEARNING OUTCOMES Where it is logical that progression tapers off as tasks get harder we feel that further mapping on the cognitive requirements linked to each sub task or at item level (depending on the difference in cognitive requirements of each item) is required in order for teachers to target the areas of need linked to the curriculum. This will further inform teachers on which areas girls are lacking in so that they can develop mastery to tackle items of higher difficulty within the homework club intervention. Also through our intervention to improve pedagogy work is done with teachers to tailor the correct support to the correct students through using learner led data, whereby the girls that need extra help in certain areas receive targeted support. This is already identified within the grade 8 national exams results. Further work is also taking place, through the professional development of teachers, to interpret in-class data that better targets support students need.

We cannot comment on the lack of progression in learning test scores across grades being linked to subtask data only being available as we do not have responses at item level so are unable to remark further on the progression. We agree that if item level data cannot be recovered, pilot data should be used for us to a)

target set for midline and b) feed direction through to education officers and teachers on areas of development required for the girls to work on in order to boost their level of performance and overall ability to tackle items of higher difficulty, (P65). These findings do not change our theory of change but will have an impact on our level of detail and focus of intervention work.

The difference in scored between Amhara and Oromia may well link to the differences in language instruction across the regions of which further scrutiny in test pilot data, and/or live data should take place in order to ascertain trends in item response.

Transition

The evaluation finds a baseline transition rate of 66% for project areas, yet the aggregate figure masks considerable variation between groups which we would like more information on in order to better project interventions.

We observe that i) the barriers to transition increase from the age of 14, where the rate drops from 80% – 45%, ii) being an orphan is a barrier to transition, iii) unsafe travel to school is linked to lower levels of transition and iv) the transition rates are lower in Oromia than in Amhara. On further reflection and discussion with project staff we observe that:

i) Lower transition does not begin only at age 14, but earlier. It has been observed that transition dips at the latter grades of upper primary (which partly coincides with baseline data).

ii) Orphans are observed as needing to work in order to make a living which is financially driven, especially in rural areas, which is a factor outside of the control of the project.

iii) Unsafe and long distance travel does impact on transition. There are three categories of secondary school girls in relation to the distances they travel to school. **Category 1** covers girls who are residing near to the secondary school so are not subjected to long distances of travel. In **category 2** girls travel a long distance to get to school, in rural areas they travel more than two hours to get to school. **In category 3,** girls reside near the school during the week and travel a long distance to return home at the weekends. For example, in South Gondar, one of the secondary schools, which is more than 5 km away from the kebele, the girls take two hours to reach the school, which is located in the Woreda town.

In saying this, further work needs to take place to work with girls from category 2 in order to reduce the risk and increase transition, particularly in the Oromia region as findings suggest that the risk is higher in Oromia. Ways of doing this are through more targeted awareness raising, which already takes place, to ensure the girls are not travelling alone, are buddied-up by fellow school boys, and or travel in groups. Baseline data will also feed into government policy so that new secondary schools can be built to address this issue as there is clearly a supply and demand issue with secondary schools in some Woredas. Regular review meetings with regional government take place where this information can be fed forward alongside regular updates.

iv) In response to the baseline findings that some girls feel unsafe whilst travelling to school, tailoring our interventions that maintain girls' safety is already an intervention that takes place through the letter link boxes, girls' and boys' club activities, buddy and escorting systems within the girls' clubs and between the boys' and girls' clubs. However, given that it has been identified that a higher number of girls in Oromia are at risk, further focus will take place on identifying suitable interventions in this area, that can be addressed within the control of the project.

Outside of baseline findings it has been observed that transition rates are lower in Oromia due to migration to The Middle East, the area of intervention (Arsi) is predominately Muslim. It has also been observed at project level and zonal level that girls do not focus on continuing their education due to lack of hope in finding employment after returning from university. Many girls who have attended university return home after unsuccessful employment seeking. Their investment in education, they feel is lost. They cannot see the power of education so prefer to not pursue it in the first place. This is particular of areas of high employment, of which Arsi is one. These two factors are out of the project control and despite this South Gondar is currently running a small scale project to connect TVET with local employment and business which other areas will observe as model to follow.

ADAPTATION: Whilst there is no adaptation needed to the Theory of Change, the project feel that the triangulation of data, when capturing transition rates should continue to be collected through our database so we can pinpoint the exact risks in the project. Also in areas where transition is at risk, targeted support is required. Further data needs to be collected on the communities that face higher risks.

Sustainability

Community Sustainability

Whilst we are pleased to see that the attitudes to girl's education are shifting positively, the project agrees that sustainability at community level needs further attention. Where there has been a shift in attitudes in favour of the value of girls' education, girls are still at risk of being pulled away from school due to households not being able to afford the cost of schooling. The fact that there is a change in attitudes to girls schooling does not address the issue of lack of affordability.

ADAPTATION: Interventions that will address this need to be discussed which requires the project to re-assess the Theory of Change, particularly in the areas of livelihood support where we feel CCC involvement is key.

It has been observed in-project that orphans and most vulnerable mothers from the most vulnerable families are in need of livelihood support, amongst others. Mapping needs to take place to address this issue via liaison with the CCCs and the Kebele administration as their capacity to mobilise resources is high. Support can be generated for families in need, for example, the provision of food and scholastic materials which will assist families in affording the cost of education.

Coaching and training of CCCs can be considered where beneficiary's families can be coached on budgeting and resource saving strategies.

The livelihoods intervention was a successful feature in GEC-1 but was not continued through to GEC-T as there was an assumption that this would continue through to GEC-T, which has not held. It would be beneficial for further investigation to take place to identify any positive case studies that can be used as models in GEC-T.

School Sustainability

It is also encouraging to read that school sustainability is the strongest and will increase if teacher retention rates improve. Teacher attendance is already collected at school level and the project will continue to assess the accuracy of this data through more external monitoring to improve reliability on data returns.

System Sustainability

Further discussion needs to take place around system sustainability and how the project plans the gradual release of activities (relating to teacher quality improvement) to system level. Further monitoring of this also needs to take place. Currently two Woreda officials per Woreda have been trained on safeguarding and case management and teachers are further supported through ICT centres. The project is also undergoing the inclusion of Woreda officials in lesson observations. An integration or gradual release plan needs to be drawn up that is tailored to Woreda capacity. There is currently no difference identified in level of capacity across areas, despite different languages in different areas. Strong networking and relationships exist for sharing and handover of tasks such as lesson observations to take place. Further development of the indicators and tools that monitor progress on Woreda involvement in teacher quality also needs to take place.

Teacher quality

The project observes an improvement in teaching quality, particularly within year 2 of the project. Lesson observations are carried out, covering a combination of 25% of trained and non-trained teachers, so as to ascertain the difference in standard of practice. Findings are then recorded and trend analysis undertaken that can be fed into teacher professional development through the school based communities of practice. The design of the lesson observation tool has taken place to capture the progression in competencies covered in the EDT teacher training. We also raise the question in relation to respondents indicating that, student-perceived teacher quality is most associated with higher learning outcomes. This finding may pertain to higher levels of engagement with higher performing students in the classroom which will need to be investigated. It is also addressed in the lesson observation tool in questions 22 – 25, 40 – 41. Where a negative trend would be identified here, further work in relation to teacher-student engagement would take place through communities of practice.

ADAPTATIONS: There are no adaptations to the Theory of Change or amendments to current interventions however further work on logframe indicators and tools that monitor progress on Woreda involvement in teacher quality need to take place.

Self-esteem and self-efficacy

As these findings are mixed the project is not able to respond fully on how it intends to further improve the girls' levels of self-esteem and self-efficacy and its impacts on learning outcomes (other than the project interventions already in place). Further work will need to take place to investigate the domains of self-esteem and self-efficacy in relation to marginalised communities within the Ethiopian context. Some of the findings in table 11, such as the underpinning drivers of nervousness, anxiety, loneliness and even the role of luck in success need further exploration via a more participatory method to unpick the domains that underpin implicit self-esteem and self-efficacy, alongside the key drivers of resilience and the relationship resilience plays. For example, girls that feel nervous reading in English does not necessarily point to low self-esteem. There may be other variables that contribute to a girl feeling nervous when reading, i.e. lack of practice in reading, lack of practice in reading in another language or fear of punitive correction.

In contrast to some of the confusing findings in table 11, the project does observe a positive trend in the girls' participation and levels of motivation in the girls' clubs. Girls report they feel confident enough to

make decisions, act as role models to other girls, decline offers of marriage and support other girls, outside of the GEC project in declining offers of marriage by taking an active role in negotiating with parents. These are factors which may reflect behaviours underpinned by high self-esteem and self-efficacy.

ADAPTATIONS: Further work within output 4 and IO2 will take place to learn more about this. The questions pertaining to self-esteem and self-efficacy in the qualitative tools will also need reviewing for midline and endline. Family Hubs

The project would like to reinstate the family hubs as an extension to the community conversations that took place in GEC-1. Box 1 also illustrates this. Family hubs will feed into a range of areas – 1) shifting of social norms and attitudes toward early marriage and girls' education, 2) awareness raising on safety in travelling to school, particularly in the areas of risk that have been highlighted, 3) livelihood, budgeting and resource saving, in conjunction with CCC focus.

It has not been decided exactly how the family hubs will run, however further discussion is under.

Gender equity

The project continues to assess its approach to gender equity and disability through the application of the GESI tool, which is used 2 – 3 times a year to assess gender and disability inclusion in relation to its planning and delivery of interventions. Key stakeholders involved in activities are the girls themselves, their role models, focal teachers, school staff and community members.

Recommendations

The evaluation team recommends that:

Monitoring, evaluation and learning

• The project should consider the inclusion of additional indicators in the logframe to more fully capture system-level sustainability. For the project to advance beyond a score of 1 for system-level sustainability, it must show evidence of improved capacity of local officials and of engagement by government with the project. The sole indicator currently tracked does not allow for these aspects to be measured.	At the moment the indicator only focuses on numbers of Woreda officials rather than the quality of their support/inputs. Additional indicators need to be included into the logframe that can accurately measure system level sustainability by a) Inserting a measure to capture how capacity at Woreda level has improved b) Inserting a measure to capture improvement of quality of practice. The logframe currently states that system sustainability is monitored by an increase in the *'number of trained Woreda education officials allocated to conduct supporting supervision activities (i.e. teacher's development and girls' learning assessments).'* Further work needs to take place to ascertain how the increase in capacity of education officials will be monitored, not only quantitatively.

• Classroom observations need to be in place for midline that are linked to suggestions for improved practice so we can know that teaching is improving as a result of the project	Classroom observations are already in place with data collection underway for our first main reporting phase to be completed by April 2019.
• Measurement of teacher attendance tracking needs to take place.	Collection of teacher attendance data currently takes place but needs to be shared with Woreda officials on a regular basis.

Design, including calculation of beneficiary numbers

Qualitative data suggests that the project's letter-link boxes are a useful and effective tool for reporting cases of early marriage and to encourage girls to express their feelings, concerns and challenges. To maximise their impact, the project should consider raising awareness of their benefits and uses.	Data collected from in-project suggests that letter link boxes are widely used. In some cases, girls trust the case reporting system enough to verbally report cases to school staff. As letter link boxes are not as established in secondary schools we will focus awareness raising on secondary school use more, in addition to monitoring the reporting of cases.
Benchmark transition data shows that transition rates drop sharply around the age where students should transition from primary to secondary school. The project should consider whether adaptation of project interventions is needed to provide extra support to girls at this risky stage of transition. The project should also seek to develop an understanding of why transition rates seem to be lower in Oromia than other areas and whether mitigation strategies may be needed.	Targeted focus is required in the following areas: • In Oromia • For girls at the latter stages of primary school to the end of secondary school • For girls that have to travel a distance to school
There is evidence that a small proportion of girls (4%-7%) may be marginalised as a result of having a mother tongue which is different from the language of instruction of the school which they attend. The project should consider whether additional support is needed for this group.	Further investigation is needed to ascertain the location of this cohort so that targeted intervention can be designed.

Scalability and sustainability

The project's approach to community-level sustainability relies on reducing the number of girls receiving support to pay for school related costs over time with households themselves picking up the slack. This operates under the assumption that households will be willing to do so as a result of the project's influence in shifting attitudes and behaviours in favour of girls' education. Evidence shows that this assumption is highly unlikely to hold, as households are unable rather than unwilling to pay. The project should consider possible approaches to address this, such as promoting the economic wellbeing of the worst-off households.	Mapping needs to take place to address this issue via liaison with the CCCs and the Kebele administration as their capacity to mobilise resources is high. Support can be generated for families in need, for example, the provision of food and scholastic materials which will assist families in affording the cost of education. Coaching and training of CCCs can be considered where beneficiary's families can be coached on money and resource saving strategies.
Sustainability in schools relies on project trained teachers remaining in post, to continue to provide improved teaching to pupils and pass on their knowledge to teachers who have not received project-training. Low levels of teacher retention in South Wollo, possibly due a 'brain drain' effect given other opportunities available locally, is likely to inhibit this. The project should consider approaches to increasing teacher retention, particularly in South Wollo.	Further external monitoring on data returns will need to take place to ensure reliability. Annual refresher training takes place to accommodate for teacher attrition Trend analysis will take place to identify drops in teacher retention Regular meetings will be held at Woreda level to factor in support from Woreda to schools Attendance data will be fed through to school leadership and management teams.

- Does the external evaluator's conclusion of the projects' approach to gender correspond to the projects' gender ambitions and objectives?

  Yes

What changes to the logframe will be proposed to DFID and the Fund Manager?

- The management response should outline any changes that the project is proposing to do following any emergent findings from the baseline evaluation. This exercise is not limited to outcomes and intermediate outcomes but extends also to outputs (following completion of Annex 3 on the output indicators).

ADAPTATIONS

Capturing risky migration - The tools also used also need to be amended for midline and endline as they do not capture responses linking to girls at risk of abduction.

Addressing teacher attrition - An amendment to the Theory of Change will therefore need to take place with more structured interventions that reduce teacher retention in place. Currently we drive incentives for teacher attendance by increasing their motivation through strong relationship building within the school and between CHADET's staff who offer guidance in school-based communities of practice and post lesson observation feedback. Education Officers also have a close working relationship with Woreda offices, school principals with the project leadership and management training addressing teacher absenteeism, alongside other school leadership and management areas. We are also developing the leadership capacities of school masters, Woreda education officers, and supervisors. Regular meetings are held at regional level to feed into government policy on their management of teacher attrition.

Learning Outcomes - Where it is logical that progression tapers off as tasks get harder we feel that further mapping on the cognitive requirements linked to each sub task or at item level (depending on the difference in cognitive requirements of each item) is required in order for teachers to target the areas of need linked to the curriculum. This will further inform teachers on which areas girls are lacking in so that they can develop mastery to tackle items of higher difficulty within the homework club intervention.

Also through our intervention to improve pedagogy work is done with teachers to tailor the correct support to the correct students through using learner led data, whereby the girls that need extra help in certain areas receive targeted support. This is already identified within the grade 8 national exams results.

Further work is also taking place, through the professional development of teachers, to interpret in-class data that better targets support students need.

Livelihood support - Mapping needs to take place to address this issue via liaison with the CCCs and the Kebele administration as their capacity to mobilise resources is high. Support can be generated for families in need, for example, the provision of food and scholastic materials which will assist families in affording the cost of education.

Girls' self-esteem and self- efficacy - Further work within output 4 and IO2 will take place to learn more about this. The questions pertaining to self-esteem and self-efficacy in the qualitative tools will also need reviewing for midline and endline.

Transition - whilst there is no adaptation needed to the Theory of Change, the project feel that the triangulation of data, when capturing transition rates should continue to be collected through our database so we can pinpoint the exact risks in the project. Also in areas where transition is at risk, targeted support is required. Further data needs to be collected on the communities that face higher risks.

LOI - No adaptations to the Theory of Change are required, however, targeted interventions need to be designed.