

# Project Evaluation Report

<b>Report title:</b>	Endline Evaluation
<b>Evaluator:</b>	Consilient
<b>GEC Project:</b>	Wasichana Wetu Wafaulu (Let Our Girls Succeed)
<b>Country:</b>	Kenya
<b>GEC window:</b>	GEC-Transition
<b>Evaluation point:</b>	Endline
<b>Report date:</b>	March 2023

## Notes:

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

# WASICHANA WETU WAFULU (LET OUR GIRLS SUCCEED) PROGRAMME

March 2023

## ENDLINE EVALUATION

*Prepared on behalf of:*



Wasichana Wetu Wafaulu



**Authors:**  
Payce Madden  
Ryan Geitner  
Iñaki Alvarez Camps  
Nicolas Wicaksono

# Project Evaluation Report

<b>Report title:</b>	Endline Evaluation of <i>Wasichana Wetu Wafaulu</i> Programme
<b>Evaluator:</b>	Consilient Research  Authors: Payce Madden, Ryan Geitner, Iñaki Alvarez Camps, Nicolas Wicaksano
<b>GEC Project:</b>	<i>Wasichana Wetu Wafaulu</i>
<b>Country:</b>	Kenya
<b>GEC window:</b>	GEC Transition
<b>Evaluation point:</b>	Endline
<b>Report date:</b>	March 30, 2023

## Table of Contents

List of Tables .....	iv
List of Figures.....	vi
Acronyms .....	vii
Executive Summary .....	1
Introduction.....	5
Project Context .....	5
Theory of Change.....	12
Project Outputs, Activities, and Outcomes .....	13
Project Beneficiaries.....	17
Barriers and Assumptions .....	17
Methodology.....	20
Evaluation Questions.....	20
Evaluation Design.....	21
Data Collection Tools .....	23
Target and Achieved Sample .....	23
Limitations and Assumptions.....	31
Learning Outcomes.....	35
Literacy Outcomes.....	35
Numeracy Outcomes.....	45
STEM Outcomes .....	52
Subgroup Learning Outcomes.....	54
Reflections on Learning Outcomes .....	63
Transition Outcomes .....	64
Subgroup Transition Rates.....	67
Alternative Transition Pathways .....	69
Reflections on Transition Outcomes .....	72
Sustainability Outcomes.....	73
Community Level.....	75
School Level .....	78
System Level .....	81
Reflections on Sustainability Outcomes .....	82
Intermediate Outcomes.....	83
Attendance .....	83
Teaching Quality.....	87
Life Skills.....	96
Household Support.....	102

Community Behaviour and Support .....	108
Auxiliary Outcomes of Interest .....	112
Economic Empowerment.....	112
Child Protection .....	113
Sexual and Reproductive Health.....	116
COVID-19 Impact .....	122
Attitudes Towards STEM.....	124
Project-specific Outcomes .....	126
Relevance/Appropriateness .....	128
Effectiveness .....	129
Short-term Impact.....	131
Efficiency and Value for Money .....	132
Sustainability .....	134
Gender Equity and Social Inclusion .....	135
Conclusion.....	139
Primary Outcomes.....	139
Intermediate Outcomes .....	140
Concluding Remarks and Recommendations .....	140

## List of Tables

Table 1: Project outputs .....	15
Table 2: Student beneficiaries, endline .....	17
Table 3: Overview of barriers faced in project .....	18
Table 4: Overview of project assumptions and reasoning .....	19
Table 5: Data collected within each evaluation round .....	23
Table 6: Target sample by tool.....	24
Table 7: Target sample for learning assessments by county .....	24
Table 8: Target sample for schools by county .....	25
Table 9: Achieved sample by county and class/form, SeGRA and SeGMA.....	26
Table 10: Achieved sample by county and class/form, STEM assessment .....	27
Table 11: Achieved sample by county and class/form, girl survey .....	28
Table 12: Achieved sample by county, caregiver survey.....	29
Table 13: Achieved sample by county, school survey .....	29
Table 14: Achieved sample by county, classroom observation .....	30
Table 15: Achieved sample by county, FGDs and KIIs .....	30
Table 16: Change in SeGRA scores across all rounds, intervention students .....	37
Table 17: SeGRA subtasks 1 and 2 learning bands .....	40
Table 18: SeGRA subtask 1 learning bands.....	41
Table 19: SeGRA subtask 2 learning bands.....	43
Table 20: SeGRA subtask 3 learning bands.....	44
Table 21: Change in SeGMA scores across all rounds, intervention students.....	46
Table 22: SeGMA learning bands.....	47
Table 23: SeGMA subtask 1 learning bands .....	49
Table 24: SeGMA subtask 2 learning bands .....	50
Table 25: SeGMA subtask 3 learning bands .....	52
Table 26: STEM assessment results, SAS and endline .....	53
Table 27: SeGRA scores by county, all students.....	55
Table 28: SeGMA scores by county, all students .....	56
Table 29: STEM scores by county, all intervention students .....	57
Table 30: Learning assessment results, Kwale and Marsabit .....	58
Table 31: Learning assessment scores by language, SAS and EL cohort 2 students .....	58
Table 32: Assessment results for seeing- and hearing-disabled students, endline.....	59
Table 33: Learning assessment scores by HoH gender, SAS and EL cohort 2 students .....	60
Table 34: Learning assessment scores by orphan status, SAS and EL cohort 2 students .....	61
Table 35: Learning assessment scores by household responsibilities, EL cohort 2 students.....	62
Table 36: Learning assessment scores by teacher absences, EL cohort 2 students .....	62
Table 37: Learning assessment scores by student-teacher ratio, EL cohort 2 students.....	63
Table 38: Transition rate by transition pathway .....	65
Table 39: Progression rate by grade.....	66
Table 40: Transition rates by subgroup .....	67
Table 41: Alternative transition pathways awareness.....	70
Table 42: Sustainability scores .....	74
Table 43: Community support for adolescent girls' education scores .....	76
Table 44: Caregiver support for their girl child's education.....	77
Table 45: School leadership scores.....	78
Table 46: Teaching practices score .....	79
Table 47: Attendance rates in SAS and EL, disaggregated .....	84
Table 48: Caregiver assessment of girls' attendance in school.....	85
Table 49: Caregiver assessment on frequency of girls' attendance in school.....	86

Table 50: Teaching quality score and subscores .....	88
Table 51: Teaching quality score and constitutive subscores at SAS and Endline .....	88
Table 52: Girls' view at SAS and EL on how teachers encourage students to understand and participate .....	89
Table 53: Gender subscore and constitutive questions, girl survey .....	90
Table 54: Girl survey subject subscore and constitutive questions .....	91
Table 55: Classroom observation subscores.....	92
Table 56: Classroom observation teacher attitudes subscore and constitutive questions .....	93
Table 57: Corporal punishment scores .....	94
Table 58: Summary Chart of IO2 Plausibility .....	95
Table 59: IO3 Description and Assumption .....	97
Table 60: Difference between mean LSI scores and subscores at SAS and EL.....	97
Table 61: L4L subscore by county .....	99
Table 62: Difference between L4L question scores between SAS and EL, Samburu and Tana River counties .....	100
Table 63: Summary Chart of IO3.1 Plausibility .....	101
Table 64: Household support score and subscores .....	103
Table 65: Household support score and subscores by county, endline .....	103
Table 66: Leave school subscore and constitutive questions, endline .....	104
Table 67: Future school subscore and constitutive questions, endline .....	105
Table 68: Girl's will subscore and constitutive questions, endline .....	106
Table 69: Value girl's education subscore and constitutive questions.....	107
Table 70: Summary Chart of IO4 Plausibility .....	108
Table 71: Community support for girls' education.....	109
Table 72: Community support for girls' education at endline, by county .....	109
Table 73: Willingness to support girls to continue education.....	110
Table 74: Willingness to support girls to continue education at endline, by county.....	110
Table 75: Cultural practices and girls' education .....	111
Table 76: Summary Chart of IO5 Plausibility .....	111
Table 77: Benefit levels of financial grants .....	113
Table 78: Percentage of girls aware of violent incidents or other harms against children in their communities .....	114
Table 79: Ways to prevent STDs .....	117
Table 80: SRH awareness and knowledge by county .....	118
Table 81: Confidence and self-advocacy against sexual assault .....	118
Table 82: Confidence in refusing to engage in early sexual activity .....	119
Table 83: SRH confidence and self-advocacy by county .....	119
Table 84: Willingness to use and ability to access contraceptives, by county.....	120
Table 85: SRH perceptions .....	121
Table 86: SRH Perceptions Index by county .....	121
Table 87: Rates of girls who dropped out during school closures, before re-enrolling.....	123
Table 88: Summary of attitudes towards STEM subjects .....	124
Table 89: Summary table of project-specific outcomes .....	127
Table 90: Summary of gender equity and social inclusion comments by project team and external evaluator.....	136

## List of Figures

Figure 1: Change in average EGRA and SeGRA scores for WWW girls .....	3
Figure 2: Change in average EGMA and SeGMA scores for WWW girls .....	3
Figure 3: Theory of Change .....	13
Figure 4: Map of school locations visited during fieldwork.....	25
Figure 5: Change in average EGRA and SeGRA scores across all rounds, intervention students .....	36
Figure 6: Change in overall EGRA and SeGRA scores since baseline.....	36
Figure 7: Change in SeGRA scores, panel students .....	38
Figure 8: Change in SeGRA task 1 and 2 scores, panel students .....	39
Figure 9: Change in SeGRA subtask 1 score, panel students .....	41
Figure 10: Change in SeGRA subtask 2 score, panel students .....	42
Figure 11: Change in SeGRA subtask 3 score, panel students, excluding zero scorers .....	44
Figure 12: Change in average EGMA and SeGMA scores across all rounds, intervention students.....	45
Figure 13: Change in overall SeGMA scores since baseline.....	46
Figure 14: Change in SeGMA scores, panel students.....	47
Figure 15: Change in SeGMA subtask 1 scores, panel students .....	48
Figure 16: Change in SeGMA subtask 2 scores, panel students .....	50
Figure 17: Change in SeGMA subtask 3 scores, panel students .....	51
Figure 18: Transition rates between evaluation rounds.....	66
Figure 19: Percent of endline schools reporting each number of active clubs .....	80
Figure 20: Percentage of girls aware of particular types of violence/harm against children in their communities .....	115
Figure 21: Willingness to use and ability to access contraceptives .....	120
Figure 22: Percentage of girls citing challenge to studying at home .....	122
Figure 23: Average girls' enrolment in STEM subjects, by subject and form .....	126



## Acronyms

ALP	alternative learning pathway
ASAL	Arid and Semi-Arid Lands
BL	baseline
BoM	Board of Management
CATI	Computer Assisted Telephonic Interview
CBC	competency-based curriculum
CC	community conversation
CHV	Community Health Volunteer
CSO	civil society organisation
DAC	Development Assistance Committee
EDT	Education Development Trust
EE	external evaluator
EGMA	Early Grade Mathematics Assessment
EGRA	Early Grade Reading Assessment
EL	endline
FCDO	United Kingdom Foreign, Commonwealth, and Development Office
FGD	focus group discussion
FGM	female genital mutilation
FSS	Future School Subscore
GBV	gender-based violence
GEC	Girls Education Challenge
GESI	gender equality and social inclusion
GWS	Girl's Will Subscore
HH	household
HLM	hierarchical linear modeling
HoH	head of household
HSS	Household Support Score
ICT	information and communication technology
IGA	income generating activity
IO	intermediate outcome
KCPE	Kenya Certificate of Primary Education
KCSE	Kenya Certificate of Secondary Education
KII	key informant interview
L2L	Learning to Learn
L4L	Learning for Life
LSI	Life Skills Index
LSS	Leave School Subscore
ML	midline
MoE	Ministry of Education
NGEC	National Gender Equality Commission
NGO	non-governmental organisation
OECD	Organisation for Economic Cooperation and Development
PTA	Parent-Teacher Association
RAS	Rapid Assessment Study
SAS	Special Assessment Study
SDG	Sustainable Development Goal
SeGMA	Secondary Grade Mathematics Assessment
SeGRA	Secondary Grade Reading Assessment
SIG	special interest group

SNE	special needs education
SRH	sexual and reproductive health
STD	sexually transmitted disease
STEM	science, technology, engineering, and mathematics
ToC	Theory of Change
TSC	Teachers Service Commission
TVET	technical and vocational education and training
US	urban slum
USAID	United States Agency for International Development
VfM	value for money
VGES	Value Girl's Education Subscore
WWW	<i>Wasichana Wetu Wafaulu</i> (Let Our Girls Succeed)

## Executive Summary

In this report, we present findings from the endline evaluation of the *Wasichana Wetu Wafaulu (WWW) (Let Our Girls Succeed)* programme. WWW is a GBP 29 million programme funded through the Girls' Education Challenge (GEC) Transition window. Running between May 2017 and March 2023, WWW's aim was to support 70,000 girls in marginalised communities in Kenya to achieve improved learning outcomes and to transition through key stages of education. The project operated in 521 primary schools, 90 secondary schools, and 23 TVET institutions. These locations span eight counties and two different contexts: the arid and semi-arid lands—Kilifi, Kwale, Marsabit, Samburu, Tana River, and Turkana counties—and the urban slums context in Mombasa and Nairobi counties. The project seeks to influence changes in four dimensions—the community, the home, the school, and the girl herself.

### Methodology

**The aim of our evaluation is to assess the performance of the WWW programme over time and against plan.** In particular, we examine WWW's progress in three outcome areas, learning, transition, and sustainability, as well as interrogate the relevance and wider contribution of WWW in supporting its 70,000 beneficiaries. To do this, we build on previous evaluation rounds, including the baseline (2018), midline (2020), a Rapid Assessment Survey (RAS) (April 2021) and finally a Special Assessment Survey (SAS) (December 2021) conducted in response to COVID-19.

**Our evaluation employs a mixed method, GESI-sensitive approach to analysing data that we collected between September and November 2022.** Our data collection methods and sample sizes are summarised in the table below. We collected quantitative data from surveys with girls, surveys with caregivers, school surveys, and classroom observations. We used EGRA and EGMA data (collected at both baseline and midline) as well as SeGRA and SeGMA tests to assess literacy and numeracy. We also administered a bespoke test covering biology, chemistry, and physics to assess STEM performance. To support our mixed method design and allow triangulation, we collected qualitative data through 19 group discussions with girls, 21 teachers, 11 government officials, and five programme staff. We achieved 54% of our originally intended quantitative sample size, which limits our quantitative analysis somewhat, but did achieve 100% of our qualitative sample size.

Tool	Sample – Intervention	Sample – Comparison	Total Sample
Learning assessments – SeGRA and SeGMA	1,296	159	1,455
Learning assessment – STEM	383	-	383
Girl survey	1,150	-	1,150
Caregiver survey	818	-	818
School survey	129	27	156
Classroom observations	129	-	129

**Our analysis relies primarily on trend analysis, although we do utilise a difference-in-differences approach to support the analysis of learning outcomes.** Our main analysis of relevant outcomes utilises a descriptive approach, examining the current status of outcomes and their change over time. Because the midline utilised a quasi-experimental design with learning outcomes assessed for both intervention and comparison groups, we also attempted to replicate this comparative analysis in our endline. This analysis is conducted on a longitudinal sample of 504 girls—178 from the comparison group and 326 from the intervention group. We recognise that this small sample size, which is the result of challenges in locating previously-assessed girls, combined with potential contamination issues that were identified limits the strength of the conclusions we can draw from this analysis alone. This is why our

report relies on triangulation across several data collection modes and mixed methods analysis to draw firmer conclusions about project performance.

## **Findings**

**Over its six-year history, the current iteration of WWW has supported 70,000 girls in transitioning from primary school to secondary school.** WWW has trained 6,067 teachers and 74 government officials, supported the establishment of 466 child to child clubs in primary schools and 90 health clubs in secondary schools, provided financial support to 5,307 bursary students (4,772 girls and 535 boys) and 2,443 cash transfer beneficiaries, and held 952 community forums, all with the aim of improving the life chances of its beneficiaries.

**We find that the GEC cohort made progress in literacy and numeracy between baseline and midline but experienced learning losses after the pandemic, although these losses were generally less severe than for the comparison cohort.** In other words, progress on literacy and numeracy was on a solid positive trajectory in earlier phases of the programme. However, WWW girls experienced large learning losses immediately after the COVID-19 pandemic as a result of school closures. This aligns with global trends, with studies finding substantial learning deficits globally arising early in the pandemic and persisting over time.<sup>1</sup> Learning progress resumed once girls had returned to school, with an upward trajectory on par with progress between BL and ML for the SeGRA and exceeding progress made between BL and ML for the SeGMA. At endline, our results show that girls were still recovering from these losses, as shown in the figures below. However, relative to the control group, the evidence shows that learning recovery in the intervention group was greater. Due to this, the project was unable to achieve its ambitious targets of a 60% improvement in learning over midline.

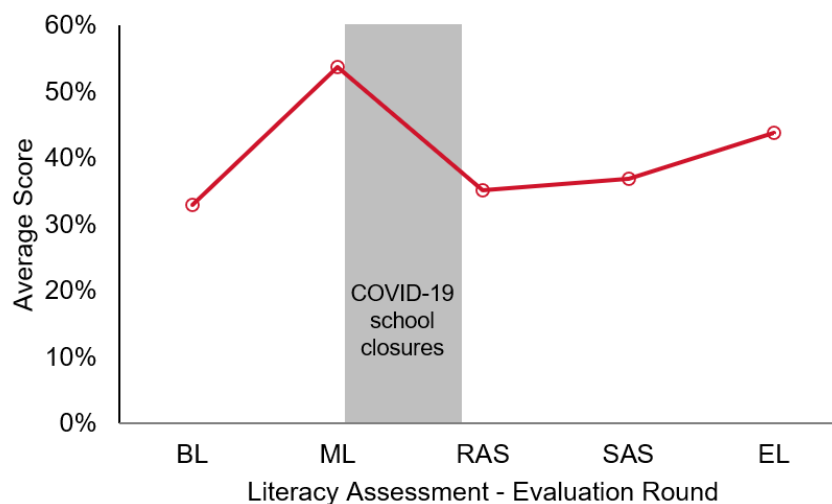
**We find indicative evidence that suggests that WWW may have had a mitigating effect on COVID-19 related learning losses.** We compared learning for WWW girls against learning witnessed in the comparison group. While we did not find a statistically significant impact for intervention students for most assessment tasks, we did find that intervention students had, on average, higher SeGRA and SeGMA scores at endline than would be expected given the results of comparison students. In other words, the WWW programme may have had a positive impact on learning for beneficiary girls, mitigating some of the worst negative effects of COVID-19.

**Project results for transition and sustainability were stronger.** It does not appear that the pandemic seriously impeded WWW girls from transitioning to higher grades, enrolling in alternative learning pathways, or re-enrolling in school after dropping out. Transition rates throughout the programme remained high, with an improvement from 93 percent at midline to 96 percent at endline. Likewise, the project had a positive impact on sustainability scores as measured along the three dimensions of community, school, and system. Using a bespoke index, we found an increase between baseline, midline, and endline, with a sustainability score of 1.8 at baseline, 2.5 at midline, and 2.7 at endline out of 4.0 possible points. Keeping in mind the severe effects the pandemic has had on government, school, and community capacity, the increase in sustainability score is a positive sign, showing progress despite substantial challenges.

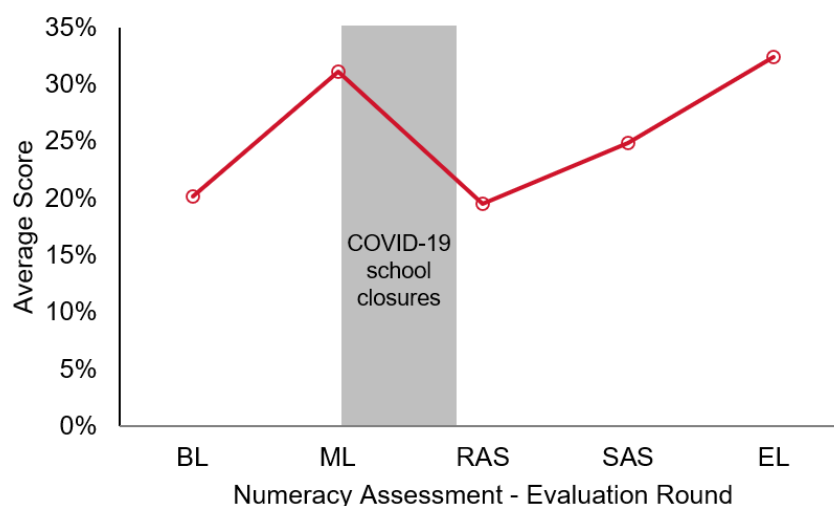
---

<sup>1</sup> Betthausen, Bastian A., Anders M. Bach-Mortensen, and Per Engzell. "A systematic review and meta-analysis of the evidence on learning during the COVID-19 pandemic." *Nature: Human Behaviour* 7 (2023): 375-85.

**Figure 1: Change in average EGRA and SeGRA scores for WWW girls**



**Figure 2: Change in average EGMA and SeGMA scores for WWW girls**



**Performance on the project's five intermediate outcomes of attendance, quality of teaching, self-efficacy, household support, and community support was generally positive**, with the most optimistic results seen in quality of teaching and self-efficacy. We summarise our headline findings below.

- **Attendance** rates are higher than baseline and midline, but lower than SAS results. Respondents noted that dips in attendance were higher for boys than those for girls.
- Our classroom observations showed very high **teaching quality** scores at endline of 83.4 out of a possible 100 points. We noted, however, that teachers often stated that boys and girls are better at different subjects, suggesting that gender stereotypes may still affect girls' learning environments. Moreover, some teachers also highlighted that their capacity to support disabled girls was limited. This suggested that while quality of teaching was strong, there may still be room for improvement.
- A slight increase in **life skills** was measured from SAS to endline; this was driven by increases in attitudes towards success in school and confidence in learning as well as attitudes towards continued pursuit of education, self-confidence, and ability to plan for the future.

- We found an extremely high **commitment to future study or training** on the part of participant girls at endline: 99.1% of girls surveyed stated that they wished to continue studying beyond the current year.
- We generally found high levels of **household support** for girls' education, with caregivers expressing especially positive responses on the value of girls' education.
- We found mixed **community views** on cultural practices hindering girls' education. While around two-thirds of caregivers stated that early marriage and pregnancy were practices that should be done away with, only 27 percent of caregivers stated that female genital mutilation should be stopped. This indicates that in spite of WWW programming, some traditional norms do persist—as should be expected, given that cultural norms are often highly resistant to change.

**The strongest contribution of the project appeared to come from its holistic, gender transformative intent.** Our analysis indicates that the programme was highly relevant to the context of Kenya, the goals of the Ministry of Education, and the parameters set by the GEC for focusing on and supporting the most marginalised girls. We found that the programme was designed with a robust gender equity and social inclusion lens and that this lens had a tangible effect on programme implementation. Evidence for example strongly suggests that gender responsive approaches were utilised in classrooms, with girls being called on as equally as boys to respond to questions. It also indicates that the programme supported girls in their understanding of sexual and reproductive health (SRH). Girls generally had good knowledge of sexually transmitted diseases (STDs) and ways of prevention at endline, and over 96 percent of girls stated that they had the right to complain if touched inappropriately by a teacher, adult, or schoolmate. Finally, we found that the programme employed an effective, holistic approach to address the life skills of girls, which according to our analysis has the greatest potential in enhancing learning outcomes.

**However, in the context of this project, we found that several barriers to girls' education do remain.** Besides the learning losses caused by the pandemic, which all students in Kenya—and globally—are still recovering from, a number of gender specific challenges remain. The endline survey with girls for instance revealed that 34.4 percent of surveyed girls were aware of instances of violence or harm against children in their communities, with this figure varying by counties. This rate has declined since midline, but increased since baseline. Girls in major urban areas were more likely to have heard of incidents of violence/harm, and girls from these areas are also the most likely to feel that children are not safe from danger and violence in their communities. In addition, we found relatively low levels of willingness to use condoms, with 63 percent of girls reporting that they would use a condom if they had sex before marriage, and even lower levels of willingness to use contraceptives in general, at only 28 percent. Unless addressed, both violence against girls and instances of early pregnancy have the potential to hinder the progress that has been supported by WWW.

### **Recommendations**

Our key recommendations for the wider sector, government, and any future programming include:

#### **To support learning**

- Invest in remedial programmes to support recovery from learning losses
- Given its importance, continue to support girls to acquire key life skills
- Continue to provide disabled girls additional support to overcome disadvantage
- Tailor teacher training support for ASALs

#### **To support transition**

- Strengthen formal and informal mechanisms of community support for girls' education
- Invest in additional SRH support for girls to improve perceptions of and access to contraceptives
- Provide flexible options for girls coming back to schools to support transition
- Address economic factors which may affect transition

## Introduction

With funding through the Girls Education Challenge (GEC), the initial *Wasichana Wote Wasome* (Let Our Girls Learn) project was implemented by Education Development Trust (EDT) between 2013 and 2017 in Kenya. This initial phase of the project focused on improving enrolment, attendance, and learning in Kenya. Based on learnings from this programme, a transition phase of the programme was developed, entitled *Wasichana Wetu Wafaulu* (WWW) (Let Our Girls Succeed). In this iteration, the project aims to support 70,000 girls in marginalised communities to achieve improved learning outcomes, transition through key stages of education, and successfully move to the next productive phase of life. This second phase started on May 1, 2017, and will close on March 31, 2023. The implementation of the WWW project was made possible by the work of consortium partners. Led by Education Development Trust (EDT), the consortium partners include Concern Worldwide in Nairobi County and Marsabit County, Kesho Kenya in Kilifi County, and the Ananda Marga Universal Relief Team (AMURT) in Samburu and Mombasa Counties.

The WWW project operated in 521 primary schools, 90 secondary schools, and 23 technical and vocational education and training (TVET) institutions, in eight counties and two different contexts: the arid and semi-arid lands—Kwale, Kilifi, Marsabit, Samburu, Tana River, and Turkana counties—and the urban slums context in Mombasa, and Nairobi counties. Through this project, EDT aims to ensure girls will have gained the skills, qualifications, and confidence required to take control of their lives to better their life chances.

This endline evaluation aims to achieve the following objectives:

- Assess the relevance and impact of WWW project interventions at the girl, school, household, community, and system levels through a gender equality and social inclusion (GESI) approach.
- Assess the delivery effectiveness and efficiency e.g., value for money (VfM), and lessons learned throughout the project implementation.
- Assess learning progress for the project beneficiaries, including girls with disabilities and girls in other GESI sub-groups.
- To assess project sustainability as per the sustainability plan, and the extent to which the net benefits of the project will continue after the project ends.
- Assess the overall gains from the interventions arising from the project implementation response and adaptations during the COVID-19.

The evaluation also seeks to address research questions guided by the Organisation for Economic Cooperation and Development (OECD) Development Assistance Committee (DAC) criteria, and better understand the project's relevance, coherence, effectiveness, efficiency, impact, and sustainability.

The target audience for the evaluation findings and recommendations is, first and foremost, EDT as the implementers of this programme. However, the evaluation will also provide useful information to the programme funder (FCDO), organizations implementing similar programmes (including those funded by GEC in Kenya and other countries), and the government of Kenya.

## Project Context

### ***Project Overview***

The United Kingdom Foreign, Commonwealth, and Development Office (FCDO) (formerly Department for International Development) is working around the world to reach the Sustainable Development Goals (SDGs) by 2030.<sup>2</sup> Progress on girls' education is critical to the achievement of these targets. Specifically,

---

<sup>2</sup> "UK and the Sustainable Development Goals", Government of the United Kingdom, accessed 1 March 2023, <https://www.gov.uk/government/topical-events/uk-voluntary-national-review-of-progress-towards-the-sustainable-development-goals>.

SDGs 4 and 5 relate to quality education and achieving gender parity, respectively. SDG 4 specifically notes ‘inclusive and quality education for all and the promotion of lifelong learning.’<sup>3</sup>

Globally, 31 million primary school aged girls have never been to school. Most of these girls come from the poorest and most marginalised communities from the most disadvantaged locations and ethnic groups. Over the last twenty years, primary school enrolments for girls have improved, along with those for boys. However, boys’ and girls’ completion rates remain equally low. At the secondary level, the differences between boys’ and girls’ participation rates start to show. Within countries, girls from the poorest households—particularly in rural areas—are subject to educational disadvantage, even at the primary school level. The GEC is helping girls most impacted by poverty to improve their lives through education and supporting better ways of enrolling girls in school, ensuring that they receive quality education with the goal of transforming their future.

Education Development Trust has supported some of the most marginalised communities across Kenya through the first Girls’ Education Challenge (GEC-1) project. From that work, the organisation developed not only a deep understanding of the highly challenging barriers that girls face, but also girls’ enormous potential, making EDT more committed than ever to help them achieve it. Through the GEC project implemented by EDT, girls gain the skills, qualifications, and confidence required to take control of their lives. Central to that vision is that most girls move from lower to upper primary school, and then into secondary school, achieving increasingly higher marks that would enable them to attend higher performing secondary schools.

In keeping with the principle of no girl left behind, the project recognizes that alternative options to secondary education are, in some cases, necessary. Therefore, GEC-T envisions that some girls’ journeys will take them from primary school into an innovative and appropriate alternative learning pathway (ALP). Those who drop out of primary school will be able to join community-based catch-up classes; with the aim of re-entering school or an ALP, ultimately preparing them for a better life.

As a step towards these goals, the Girls’ Education Challenge Transition (GEC-T) project *Wasichana Wetu Wafaulu* targeted girls with the goal of helping them complete primary school, achieve target learning outcomes, and transition successfully to a productive and positive next phase. Phase two of the programme began in 2017. At that time, there was uncertainty on the roll out of the competency-based curriculum (CBC) in Kenya; as a result, the project delayed implementation of class 4, awaiting the government’s roll-out of the CBC policy, which was delayed until 2018.<sup>45</sup> In 2017, the project profiled 52,004 girls in school and 6,183 out-of-school girls who had dropped out for various reasons. At the time, these girls were in classes five through eight. Following the 2018 release of the CBC roll-out plan and in discussions with the GEC fund manager, the project commenced implementation for grade four in 2019, targeting 12,350 girls. This brought the total number of girls in the project to 70,537.

However, due to the transition of pupils to schools outside of the project’s jurisdiction, by the first evaluation, the number of girls in school and actively supported by the project was 64,030. An additional 6,507 girls were out of school, with some enrolled in catch-up classes in readiness for re-joining the mainstream education levels.

---

<sup>3</sup> “The SDGs in Action”, United Nations Development Programme, accessed 1 March 2023, [https://www.undp.org/sustainable-development-goals?utm\\_source=EN&utm\\_medium=GSR&utm\\_content=US\\_UNDP\\_PaidSearch\\_Brand\\_English&utm\\_campaign=CENTRAL&c\\_src=CENTRAL&c\\_src2=GSR&gclid=CjwKCAiAjPyfBhBMEiwAB2CCIS3NI7QjTkBe3iiPUIhLJtxH1v3Pn tfS7Do1Z3JvQZAhbXbSyGnsRoCzykQAvD\\_BwE](https://www.undp.org/sustainable-development-goals?utm_source=EN&utm_medium=GSR&utm_content=US_UNDP_PaidSearch_Brand_English&utm_campaign=CENTRAL&c_src=CENTRAL&c_src2=GSR&gclid=CjwKCAiAjPyfBhBMEiwAB2CCIS3NI7QjTkBe3iiPUIhLJtxH1v3Pn tfS7Do1Z3JvQZAhbXbSyGnsRoCzykQAvD_BwE).

<sup>4</sup> David Njeng’ere and Lili Ji, “The Why, What and How of Competency-Based Curriculum Reforms: The Kenyan Experience”, UNESCO International Bureau of Education, June 2017.

<sup>5</sup> “Report on Competence Based Curriculum Activities”, Kenya Institute of Curriculum Development, 3 January, 2018, accessed 1 March 2023, <https://kicd.ac.ke/wp-content/uploads/2018/02/Presentation-on-CBC-Activities-Jan-2018.pdf>.



The first evaluation, the baseline, was published in May 2018 by an external evaluator (EE), the Women Educational Researchers of Kenya. In July 2019, roughly a year later, the Women Educational Researchers of Kenya conducted the first midline evaluation (ML1). Unfortunately, the COVID-19 pandemic affected project implementation. As such, after discussion with the Girls Education Challenge Transition (GEC-T) programme Fund Manager it was decided to replace ML2, which was scheduled for July 2021, with an internal Rapid Assessment Study (RAS). As such, the RAS was conducted by EDT in February 2021. A second assessment study, the Special Assessment Study (SAS) was conducted in October 2021. The purpose of the SAS was to assess the project's progress towards achieving its objectives, in turn informing adaptive programming, considering challenges brought by the COVID-19 pandemic. This endline report builds upon the work done in these previous evaluation rounds and presents an endline evaluation of the programme.

### ***Impact of Gender Inequalities and Marginalisation on Girls' Education***

The impact of gender inequality is multifaceted in the ways that gendered barriers interact with other forms of disadvantage and discrimination to negatively affect girls and women. Historically, gender inequalities entrenched unchallenged cultures of male dominance, leading to the marginalisation of women in many communities. As a result, a huge body of research evidence shows that women not only bear the brunt of poverty, but that women's empowerment through education is a central precondition for its elimination.

In the Kenya context and specifically among marginalised communities, many girls are out of school and the drop-out rate is high. In addition, girls get married early, leading to poor maternal health, high infant mortality and fertility rates, as well as increased new cases of HIV and AIDS infections. Consequently, the cycle of poverty continues to follow them and their families throughout their lives.

To address these issues, the WWW project continues to target the most marginalised girls in the most marginalised areas of Kenya: the ASALs and urban slums. The majority of girls face multiple layers of social and economic marginalisation, such as high levels of poverty, poor health, low household income, and limited access to amenities. It is therefore very difficult to group the cohort by one specific type of marginalisation. For example, some girls may live in an impoverished Arid and Semi-Arid Lands (ASAL) community where access to schools is difficult due to distance, and girls may be expected to shoulder the bulk of household chores. In another instance, a girl may be a teen mother who is also responsible for caring for a sick relative.

In general, in the ASAL areas, girls face a number of barriers to education, many of which are entrenched cultural practices linked to sex and gender, such as female genital mutilation (FGM), early marriage, and teen pregnancy. ASALs are home to pastoralist communities who face high work burdens and live in remote locations. Limited infrastructure means that girls face lengthy - and sometimes hazardous - journeys to reach distant schools/ALPs. High levels of poverty mean that many households are unable to pay school fees charged for primary or secondary education.

In urban slums, poverty is also a major barrier to girls' education, along with high levels of gender-based violence. Poor living conditions lead to poor health which can impact learners directly or indirectly. In addition, as traditional gender roles are upheld in many parts of the country, girls are often required to care for family members or the household.

Historically, inadequate investment in education means that educational resource allocation in these areas is low or not well used, and there is very little provision for special needs education (SNE). The schools in both ASALs and urban slums are characterised by untrained teachers, poor facilities, and high rates of absenteeism, leading to poor learning outcomes, high dropout rates, and low transition rates.

## **National Educational Policy Context**

In Kenya, the education system includes primary school, consisting of lower and upper primary (eight years in total), and secondary school (four years in total). The language of instruction policy is the students' mother-tongue for early grades, followed by English from upper primary through secondary school. However, in practice English is employed at the early grades, and in some counties, mother-tongues pervade well beyond early primary school.

The Kenyan Constitution (2010) stipulates the right to free and compulsory basic education for every child. The government provides free public primary education for all pupils, but parents contribute through the payment of school levies which pose a barrier to marginalised communities. At the secondary level, the government covers tuition fees for students in public day schools, while parents pay for other expenses such as uniforms and lunch. For public boarding secondary schools, households take on most of the costs, save for tuition. These expenses pose a significant barrier to transition from primary to secondary school.

National policies currently prioritise improved education quality and inclusion, as well as a wider curriculum reform. The implementation of the new 2-6-3-3-3 competency-based curriculum was envisaged to start in 2017/2018, beginning roll out in pre-primary years one and two, as well as grades one and three. The program was expected to roll out to grades four and six in 2019. This has since changed, with the roll out from grade four being carried out in individual, one-year phases; for instance, grade four rolled out in 2020, grade five in 2021, and grade 6 in 2022. The piloting of the CBC was finalised after the submission of the midline report with the grade four rollout.

Considering secondary public schools, four classifications exist: national schools, county and sub-county schools, extra-county schools (usually boarding), and private/community schools. Public school classifications are based on performance and facilities, and the quality of education varies across the classifications, in turn affecting demand on enrolment. Allocation of students to secondary schools is determined by performance in the Kenya Certificate of Primary Education (KCPE). For instance, following the 2015/2016 KCPE examination, only 3.4% of GEC-T girls enrolled in national schools. Instead, 29.4% of GEC-T girls enrolled in county schools, 41.4% in sub-county schools, and 25.8% did not transit to secondary school. The percentage of target GEC-T girls who did not transit (25.8%) was higher than the national average (19.1%).

There are also a number of TVET institutions. Girls' enrolment in TVET institutions is historically limited due to a perceived lack of the course' relevance to girls, negative social attitudes toward girls' vocational study, and a lack of awareness of the pre-tertiary qualification option. Enrolment in TVET institutions has risen steadily since 2018, with 2022 marking a record number of students enrolled in TVET programmes (265,095).<sup>6</sup> TVET management is decentralised from counties. Currently, there are very few catch-up centres with low attendance rates. The government would like to increase access and quality of TVET and community catch-up centres.

Regarding learners with disabilities, Article 54 of the Constitution provides that persons with disability have a right to access educational institutions and facilities that are integrated into society to the extent compatible with their interests and needs.<sup>7</sup> The government provides education for children with disabilities through integrated and special units in primary and secondary schools. The 2009 Special Needs Education (SNE) policy guides the provision of education for girls and boys with disabilities. The objective of the SNE policy is to "enhance gender mainstreaming in SNE programmes at all levels and ensure increased enrolment, participation, and completion rates for both girls and boys, and men and

<sup>6</sup> John Mutua, "Public university numbers dip as TVET numbers grow," Business Daily, 06 May 2022, accessed November 17, 2022. <https://www.businessdailyafrica.com/bd/economy/public-university-enrolment-dips-for-first-time-in-three-years-3805878>.

<sup>7</sup> Constitution art. 54, 2010 (Kenya).

women, with special needs and disabilities in education”. There are also other support policies: For instance, one of the functions of the National Gender Equality Commission (NGEC), spelt out in section 8(m) of the National Gender and Equality Act, is to conduct audits of the status of special interest groups (SIGs) including minorities, marginalised groups, persons with disabilities, women, youth, and children.<sup>8</sup> NGEC also acts as the principle organ of the state to ensure compliance with all treaties and conventions ratified by Kenya and relating to issues of equality and freedom from discrimination, and relating to SIGs.<sup>9</sup>

However, lack of a clear SNE implementation framework, inadequate funding, and few teachers equipped to teach children with disabilities hampers individuals’ access to SNE services. These limitations are coupled with negative attitudes, poverty, limited parent awareness, insecurity, and unsuitable institutions. Additionally, the drop-out rate for girls with special needs and disabilities is high due to teachers’ lack of sensitivity toward these kinds of learners.

In terms of early motherhood, Kenya has a relatively high rate of teen pregnancy. According to the United Nations Population Fund, 378,397 adolescent girls aged 10-19 years were pregnant between July 2016 and July 2017.<sup>10</sup> Of these girls, 28,932 were between 10 and 14 years old, while 349,465 were between the ages of 15 and 19. Eight counties with the highest number of teenage pregnancies include Narok (40%), closely followed by Homa Bay (33%), then West Pokot (29%), Nyamira (28%), Tana River (28%), and Samburu (26%), and finally Migori and Kwale (24%).<sup>11</sup> Notably, the counties where *Wasichana Wetu Wafaulu* project is being implemented include Tana River, Kwale, and Samburu, which are among the counties with the highest rates of adolescent pregnancies.

To address the barrier of teen pregnancy to girls’ education, the government of Kenya introduced re-entry guidelines in 1996 to ensure smooth readmission of adolescent mothers after delivery. The project plans to exploit this provision to support teen mothers wishing to pursue education after delivery to re-enrol in school. Such girls will also be supported by the project through catch-up centres. However, the implementation of such policies is challenging due to factors such as low levels of awareness of paths to re-entry, stigma, poverty, and lack of childcare.

### **Contextual Changes at Endline**

This section covers both the internal and external environment factors that affected the project implementation. The factors may have impacted project progress.

#### **COVID-19**

In Kenya, in response to COVID-19, schools were closed for periods of up to nine months. This hindered implementation of most in-school activities. The pandemic posed further challenges to programme activities as travel restrictions were enforced by the government. The majority of project activities in schools were affected; there was thus a need to make adjustments to the implementation and monitoring of activities.

That said, where possible, project activities were adapted to reach girls within the “new normal” of school closures and travel restrictions. For instance, as the Kenyan Institute of Curriculum Development broadcast school lessons over the radio, one such project adaptation included the provision and

---

<sup>8</sup> National Gender and Equality Commission Act (2012) Chapter 5 § 8 (Kenya).

<sup>9</sup> *Ibid.*

<sup>10</sup> Kennedy Nyamura and Benedict Ochieng, “Draft Report on Teenage Pregnancies in the Context of COVID-19,” The East African Centre for Human Rights, accessed 17 November 2022. <https://eachrights.or.ke/draft-report-on-teenage-pregnancies-in-the-context-of-covid-19/>.

Updated data is not available as of November 2022.

<sup>11</sup> *Ibid.*

distribution of solar radios.<sup>12</sup> Where girls were unable to get a signal, the project downloaded pre-recorded lessons onto flash disks and shared those with parents.<sup>13</sup>

When the project realised that students desired instant feedback, something they were missing while remote learning, the project employed some teachers to be online, answering questions in real-time over Whatsapp. In an effort to mitigate potential post-COVID dropout due to pregnancy, the project also supported reading camps during COVID. These camps were led by CHVs who facilitated tutorials created by EDT.<sup>14</sup> In some communities, the project hired a teacher to serve as a kind of camp leader.<sup>15</sup> Unfortunately, as these kinds of activities targeted the maximum number of intervention girls per community, girls with disabilities were often not reached.<sup>16</sup> Still, to better facilitate girls' continued academic engagement, the project also provided the girls with writing materials.<sup>17</sup>

In terms of the after effects of COVID-19 on the WWW project, the contract underwent an allocation cut equivalent to one percent contract value, effectively cutting money that would have gone directly to project beneficiaries, resulting in the scaling down of project interventions.<sup>18</sup> As the project is funded by the GEC which comes from a percentage of the UK's GDP, the cuts were a result of the pandemic's negative impact on the UK economy.<sup>19</sup>

In addition to financial impacts, the COVID-19 also increased the vulnerability of communities due to loss of livelihoods. This pulled communities back to a status of dependency, increased school dropouts when schools reopened, and affected learning outcomes, among other challenges.

### **Shortened School Calendars**

In the period between midline and endline, the Ministry of Education shortened both school terms and holiday breaks. As a result, holiday mentorship forums had low attendance, as the shortened holiday period meant that many beneficiaries who attended schools outside of project counties did not go home over the holidays. Additionally, schools had limited time to conduct regular club meeting sessions due to shortened terms and pressure to complete the syllabus.

### **Transfer of Trained Teachers**

Between midline and endline, some teachers within intervention schools who had been trained by the WWW programme transferred to new schools that were not targeted by the programme. This affected implementation of pedagogical skills in schools.

### **Tracking of Beneficiaries**

Prior to endline, contracts ended for Community Health Volunteers (CHVs), and coaches. These individuals played a key role in tracking project beneficiaries; as such, this hindered follow-up with beneficiaries. Additionally, tracking the learning of girls once they moved to secondary schools was challenging, as many girls joined schools outside of project implementation counties.

### **Movement Out of Project Areas During School Closures**

Movement out of project areas by nomadic pastoralists in ASAL counties as well as urban-rural movement in search of livelihoods and sustenance was a barrier to the project reaching 100 percent of

---

<sup>12</sup> KII with Implementing Staff Member, Regional Coordinator, 85.

<sup>13</sup> KII with Implementing Staff Member, Implementing Officer, 84.

<sup>14</sup> KII with Implementing Staff Member, 86.

<sup>15</sup> Ibid.

<sup>16</sup> KII with Implementing Staff Member, 83.

<sup>17</sup> KII with Implementing Staff Member, 84.

<sup>18</sup> KII with Implementing Staff Member, Financial Officer, 88.

<sup>19</sup> Ibid.

targeted learners with learning support and tutorials during COVID-19-induced school closures. Some learners also engaged in economic activities outside of their school residential areas and could not be reached for learning support during this period. The project could not control for engagement of girls in hazardous work and any likely psychological and physical impact resulting from this work.

### Security Issues

Security issues, particularly due to conflicts over issues such as cattle rustling and inter-clan clashes, have affected the project, particularly in the ASAL counties Marsabit and Turkana. Marsabit experienced ethnic clashes characterised by sporadic killings, frequent cattle raids, constant tensions, and revenge attacks in Saku Subcounty. Enduring across the last two decades, more recently the conflict in Saku subcounty peaked in July 2021 and May 2022, with seasons of more intense violence including late 2021.<sup>20</sup> This violence led to closure of schools, affecting the learning of cohort girls. In Turkana, intermittent bandit attacks throughout much of 2022 in Turkana East and parts of Turkana South made roads impassable and hindered some interventions and monitoring activities in the region.

Specifically, as the EDT team was unable to use escorts, some regions within counties like Marsabit, Samburu, and Turkana that were affected by security issues between 2019 and 2022 became unreachable. As a result, teachers were supported virtually.<sup>22</sup>

### Prolonged Drought

The drought reported on at midline continues to affect the ASAL region, including Marsabit, Kilifi, Kwale, Samburu, Tana River, and Turkana counties. Currently, 24 percent of the ASAL population—equivalent to 4.2 million people—are facing high levels of acute food insecurity.<sup>23</sup> In addition, 2.7 million people are in the crisis phase, and 785,000 people in the emergency state. These numbers mark a 10 percentage point increase from 2021, at which point 2.1 million people were recognized as living in an emergency or crisis state.<sup>24</sup>

Specifically, in Marsabit county, more than 121,000 sheep and goats, 35,000 camels, and 38,000 cattle were lost in the last few months, largely as a result of the increasing distance between water and pasture.<sup>25</sup> Reports also argue that more than 54,500 children in the ASAL counties are at risk of dropping out of school unless measures are taken to provide sufficient food and water.<sup>26</sup> This is largely due to the fact that growing food and water scarcity have driven up the prices of both. As a result, increasing proportions of income are spent on water, leaving families struggling to pay school fees.<sup>27</sup>

---

<sup>20</sup> Jacob Walter, “The Unending War, Killings in Marsabit County”, The Nation, 21 December 2021, accessed 1 March 2023, <https://nation.africa/kenya/counties/marsabit/the-unending-war-in-marsabit-3658494>.

<sup>21</sup> Victoria Amunga, “Outbreak of Ethnic Violence Kills 5 in Kenya’s Marsabit”, Voice of America, 5 May 2022, accessed 1 March 2023, <https://www.voanews.com/a/outbreak-of-ethnic-violence-kills-5-in-kenya-s-marsabit/6559550.html>.

<sup>22</sup> KII with Implementing Staff Member, Implementing Officer, 84.

<sup>23</sup> “Drought situation in the Kenya ASAL areas now at crisis level,” 5 October 2022, ASAL Humanitarian Network, accessed 17 November 2022. <https://reliefweb.int/attachments/dd9be1b9-5308-4743-aaed-a240c950876e/ASAL%20Humanitarian%20Network%20Press%20Release%20-%20October%205%202022.pdf>.

<sup>24</sup> *Ibid.*

<sup>25</sup> *Ibid.*

<sup>26</sup> *Ibid.*

<sup>27</sup> “Kenya territories tackle drought crisis,” Salvation Army, 6 October 2022, accessed 17 November 2022. <https://www.salvationarmy.org/ihq/news/inb021022-kenya-territories-tackle-drought-crisis>.

## Floods

Tana River experienced floods that affected around 2,500 households settled along riverbanks and about 2,000 hectares of planted farms. This led to displacement of communities and migration of pastoralist communities, hence affecting learning in study groups. A total of 140 project households were affected.

## Women and Girls' Rights

A September 2021 report from Human Rights Watch on gender-based violence reported an increase in violence against women and girls during the COVID-19 lockdown and the subsequent nightly curfew.<sup>28</sup> In response, President Uhuru Kenyatta pledged to direct \$50 million to ending gender-based violence (GBV) by 2026.<sup>29</sup> The prolonged drought affecting the ASAL regions has also further heightened the risk of GBV, sexual exploitation, and abuse, hindering access to education.<sup>30</sup>

## Theory of Change

The WWW Project's Theory of Change (ToC) is informed by the understanding that girls' access to education faces various complex, multi-dimensional but interrelated barriers. Fundamentally, the barriers obstructing girls' educational attainment and transition can be classified into four broad areas. These areas are the **girl** herself, at **school**, at **home**, and in the **community**. These barriers exist within each of the project pathways (outline below), and manifest differently across the arid and semi-arid lands (ASALs) and urban slum contexts.

Still, the barriers are intertwined. For example, girls' limited aspirations are closely related to the low value traditionally placed on girls' education in their communities and households. Girls' limited aspirations are also linked to their underperformance in schools whose practices and facilities are ill-equipped to equitably teach boys and girls. Some of these barriers are common across the three pathways, while others are pathway specific. For example, a lack of awareness of alternative learning pathway (ALP) options is a particular barrier affecting transition from primary school to ALPs.

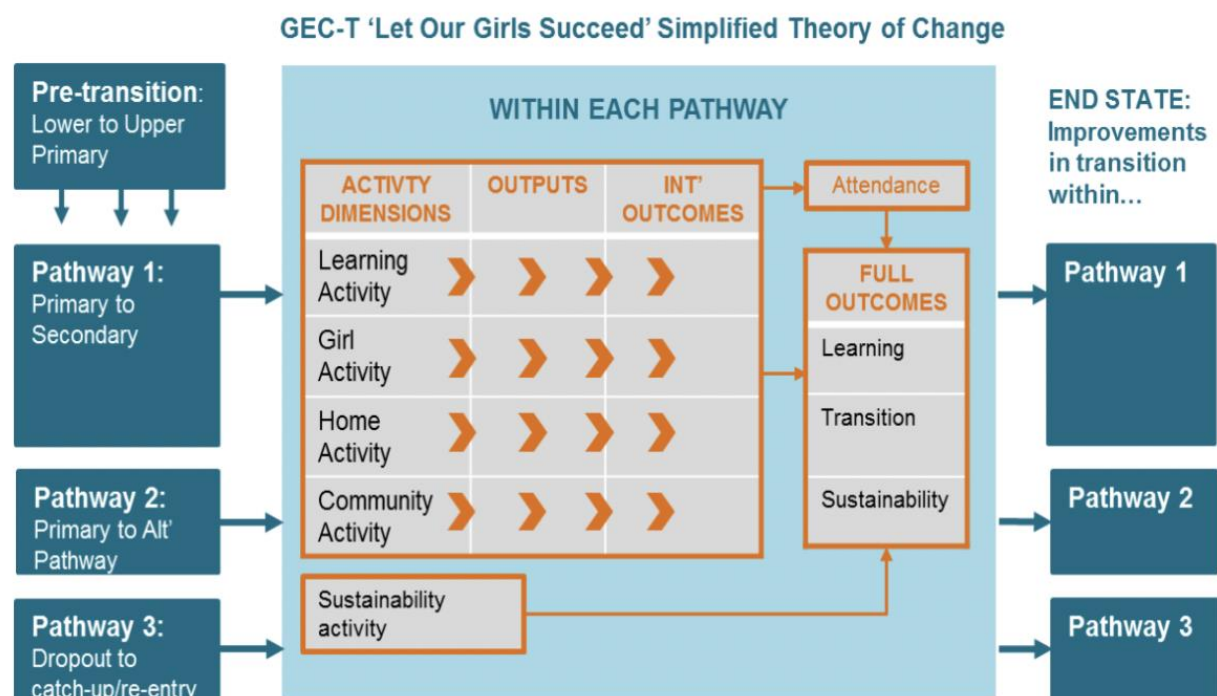
---

<sup>28</sup> "Kenya: Events of 2021," Human Rights Watch, accessed 17 November 2022. <https://www.hrw.org/world-report/2022/country-chapters/kenya>.

<sup>29</sup> *Ibid.*

<sup>30</sup> "Drought situation in the Kenya ASAL areas now at crisis level," 5 October 2022.

Figure 3: Theory of Change



## Project Outputs, Activities, and Outcomes

### Outcomes

Within the WWW project there are three primary long-term outcomes, specifically **learning**, **transition**, and **sustainability**. The learning outcome focuses on the number of girls supported by GEC who record improved learning outcomes in literacy, numeracy, and STEM. The transition outcome focuses on the number of girls who have transitioned through key stages of education, training, and employment. The sustainability outcome focuses on the programme's sustainability, specifically that the changes brought about through the project which increase learning and transition through education cycles are sustainable at the community, school, and system levels.

In addition to the three primary long-term outcomes, there are five intermediate outcomes. These focus on student attendance, quality of teaching, life skills, household support, and community-based attitudes and behaviour change. Combined with the three primary outcomes, programme outcomes are eight in total, each of which is briefly outlined below.

### Outcome 1 – Learning

Outcome one measures marginalised girls' learning outcomes across the lifecycle of the programme. The learning outcome includes literacy and numeracy improvement indicators. The literacy improvement indicator looks at target girls' aggregate literacy scores, as measured by the Secondary Grade Reading Assessment (SeGRA) (and previously, the Early Grade Reading Assessment (EGRA)) which is employed during evaluations. The numeracy improvement indicator looks at target girls' aggregate numeracy scores, as measured by the Secondary Grade Mathematics Assessment (SeGMA) (and previously, the Early Grade Mathematics Assessment (EGMA)) which is employed during evaluations. During project evaluations, Kenya Certificate of Primary Education (KCPE) and Kenya Certificate of Secondary Education (KCSE) scores were also captured in the school survey.

## Outcome 2 – Transition

Outcome two measures marginalised girls' transition outcomes across the lifecycle of the programme. Specifically, the transition outcome looks at whether girls transitioned through key stages of education, training, or fairly paid employment. The transition category is divided into three pathways. The first two pathways originate at primary school, transitioning to either secondary school (1) or an alternate learning pathway (2). The third pathway targets girls who dropped out of school, looking at their transition into catch-up classes, and ultimate school re-entry (3). The girl and caregiver surveys, which include specific questions regarding transition after primary school and re-entry into school, are used to evaluate outcome two.

## Outcome 3 – Sustainability

Outcome three measures the sustainability of the project's outcomes. Specifically, the sustainability indicator asks whether the project can demonstrate that the changes it brings about can endure beyond the lifecycle of the project. By definition, the sustainability outcome looks beyond the girl and home, focusing on effecting enduring changes at the school and community levels. While sustainability is difficult to measure while project implementation is ongoing, questions regarding sustainability are included in the girl and caregiver surveys, as well as in the key informant interviews (KIIs) conducted with select teachers, government officials, and programme staff.

## Intermediate Outcome 1 – Attendance

Intermediate outcome one looks to evaluate whether students' attendance in school has improved. To evaluate intermediate outcome one, the project relies on three main indicators. The first utilises head teacher KIIs to evaluate whether the teacher has noticed changes in student attendance between the evaluation rounds. The second measures the percentage of female students who report an improvement in school attendance. Echoing the second indicator, the third measures the percentage of female students' caregivers who report an improvement in girls' school attendance. Following the report of unchanged attendance outcomes at the midline, the attendance outcome was dropped from the project's logical framework. The attendance outcome was replaced by intermediate outcome two which concerns girls' performance in school. However, we report on both of these outcomes in this endline report.

## Intermediate Outcome 2 – Teacher Quality

Intermediate outcome three looks to evaluate whether the quality of teaching and instruction saw an improvement in schools and ALPs. Intermediate outcome three is evaluated at the girl and school levels by looking at the proportion of marginalised girls performing well academically because of accelerated remedial/catch-up lessons that promote academic aspiration and learning. As such, the learning assessments, girl surveys, and school surveys are all utilised to evaluate teacher quality.

## Intermediate Outcome 3 – Life Skills

Intermediate outcome four is concerned with girls' self-efficacy and life skills. Specifically, outcome eight looks to evaluate whether girls demonstrate improved knowledge, attitudes, and practices on issues of gender and sexual and reproductive health (SRH). The outcome also looks to quantify the percentage of adolescent girls demonstrating improved self-efficacy. To evaluate girls' improvement in self-efficacy and knowledge, attitudes, and practices on gender and SRH, we utilised two question groups within the girl surveys, one of which asks about life skills, and the other about SRH. These surveys were conducted both in person and over the phone, utilising a computer assisted telephone interviewing (CATI) scheme. The school survey also utilises data from the school survey and teacher KIIs, as both include questions regarding the status of mentoring and health clubs at the school.

## Intermediate Outcome 4 – Household Support for Transition

Intermediate outcome five looks to evaluate the degree to which households actively support the transition of girls into educational pathways, be that from primary school to secondary school or an ALP, or from a



drop-out status to a catch-up program and ultimate re-entry. To evaluate household support for girls' transitions, we utilised a survey that targeted caregivers of girls surveyed in this and past rounds. This caregiver survey was conducted over the phone, utilising a CATI scheme.

### Intermediate Outcome 5 – Community Support for Transition

Intermediate outcome six looks to evaluate the degree to which communities developed more positive attitudes toward girls' learning and transition between the baseline and endline evaluations. Specifically, intermediate outcome six is interested in measuring the percentage of communities expressing a need to do away with harmful cultural practices that hinder girls from continuing to further their education and professional training. To evaluate community support, we utilised groups of questions in the girl surveys conducted both in-person and through CATI, as well as within the caregiver survey, to ask about perceived community support for girls' education.

### Outputs

The below table shows the six main project outputs, outputs indicators, and their target populations. Changes or revisions to specific indicators from midline are also noted.

**Table 1: Project outputs**

Output	Output Indicator	Target Population	Changes/Revisions
<b>OUTPUT 1</b>	<b>Output Indicator 1.1</b>		
Teachers and school leaders in primary and secondary schools demonstrating gender-sensitive and enhanced teaching approaches (ICT and pedagogy) for improved learning.	Number of primary and secondary teachers utilising improved teaching approaches.	Teachers	Original indicator updated to include a dimension of effectiveness
	<b>Output Indicator 1.2</b>		
	Number of head teachers implementing action plans from the leadership mentor programme.	School Heads and Deputies	
	<b>Output Indicator 1.3</b>		
	Percent of primary and secondary school teachers utilising improved teaching approaches in STEM subjects.	Teachers	
<b>OUTPUT 2</b>	<b>Output Indicator 2.1</b>		
Alternative learning pathways established or expanded for girls outside or at risk of leaving school.	Number of girls enrolled and continuing or have completed education in an alternate learning pathway (catch-up program, TVET, apprenticeship, etc.). (cumulative).	Girls	
	<b>Output Indicator 2.2</b>		
	Proportion of girls with improved perceptions on the viability of the alternative learning pathways.	Girls	
<b>OUTPUT 3</b>	<b>Output Indicator 3.1</b>		
Improved self-confidence and aspirations among	Number of girls completing the mentorship programme.	Girls and Boys	Original indicator updated to include a dimension of effectiveness.

the girls in mentorship and scholarship programmes.	<b>Output Indicator 3.2</b>		
	Number of project girls and boys regularly attending girls' clubs or disability clubs.	Facilitators and Girls	New indicator.
	<b>Output Indicator 3.3</b>		
	Percent of girls with improved understanding regarding their reproductive health risks and needs.	Girls and Boys	Original indicator updated to include a dimension of effectiveness.
<b>OUTPUT 4</b>	<b>Output Indicator 4.1</b>		
Continued household support for girls' education, including in alternative learning pathways	Number of households with improved attitudes toward supporting and investing in girls' education.	Households	
	<b>Output Indicator 4.2</b>		
	Number of households reporting that financial and other material support from the WWW project has helped them keep their daughters in school ( <i>disaggregated by support package</i> )	Households	Original indicator updated to include a dimension of effectiveness.
	<b>Output Indicator 4.3</b>		
	Percentage of girls who have successfully been referred for services against the impacts of traditional initiation and other retrogressive cultural practices.	Girls	New indicator with an efficiency aspect which replaces the indicator concerning CHV activities.
<b>OUTPUT 5</b>	<b>Output Indicator 5.1</b>		
Compared to baseline, school catchment communities are more aware of the importance, benefits, and opportunities available to support girls' education	Number of community groups conducting accountability and tracking the utilisation of education funds available to schools	Groups and Individuals	Original indicator updated to include a dimension of effectiveness.
	<b>Output Indicator 5.2</b>		
	Number of groups from the catchment communities that have received funding and established functional IGAs that support girls' education	Groups	Indicator of the percentage of catchment communities that develop action plans that address barriers to girls' education has been dropped.
<b>OUTPUT 6</b>	<b>Output Indicator 6.1</b>		
The WWW project aligns to the WWW models, in turn informing emerging MoE gender and teaching approaches	Number of MoE officials trained on and conducting learning and gender responsiveness.	Ministry of Education Officials	Indicator regarding "the number of review meetings organised by MoE/TSC/County parties to address girls' education" has been dropped

Output Indicator 6.2			
	Number of MoE/TSC officials demonstrating appropriateness of the project system-led interventions as a means of improving learning and school governance structures	Ministry of Education and Teacher Service Commission (TSC) Officials	Indicator regarding “Number of MoE/TSC utilising National Leaders of Education interviews as a means of improving learning and school governance structures” has been dropped

For brevity, information on project activities is included in Annex 15.

## Project Beneficiaries

The below table shows project beneficiaries at the endline evaluation round; information on baseline and midline beneficiaries is included in Annex 15. At baseline and midline, information on the grade and ages of beneficiaries was tracked. At endline, this information was not tracked for all beneficiaries; instead, beneficiaries were tracked by county and school level (primary or secondary). As such, in Table 2 we show student beneficiaries at endline by county and school level.

**Table 2: Student beneficiaries, endline**

County	Primary School	Secondary School (Girls)	Secondary School (Boys)	Total
Kilifi	15,027	6,880	3,281	25,188
Kwale	4,860	2,331	891	8,082
Marsabit	1,924	1,677	0	3,601
Mombasa	6,352	2,230	1,673	10,255
Nairobi	23,306	4,178	1,683	29,167
Samburu	2,875	2,393	417	5,685
Tana River	9,263	2,231	2,212	13,706
Turkana	5,653	2,865	1,067	9,585
<b>EL Total</b>	<b>69,260</b>	<b>24,785</b>	<b>11,224</b>	<b>105,269</b>

Annex 15 includes additional information on target and reached populations, including girls, boys, teachers, head teachers, households, government officials, and other relevant populations.

## Barriers and Assumptions

### Barriers

Key contextual issues in urban slums and ASALs that affect education influenced both project design and delivery. These issues manifest in urban slums and ASALs in different ways and to different degrees. Below are highlights of the barriers faced by the project.

**Table 3: Overview of barriers faced in project**

Limitation	Description
<b>Poverty</b>	Households struggle to provide basic needs for education and therefore opt to take girls out of school to work.
<b>Cultural Attitudes</b>	Communities place low value on girls' education.
<b>Security</b>	Insecurity leads to high risk of sexual violence and exploitation.
<b>Workforce Capacity</b>	Low numbers of teachers, high rates of teacher attrition, and low levels of teacher skills.
<b>Infrastructure</b>	Schools in the ASALs are far apart, and the slums often have poor and unsafe infrastructure.

In response to these barriers, the project designed the Theory of Change and related activities to be delivered through three pathways:

**Pathway 1 (PW1) – Primary to Secondary Barriers:** Households' inability to pay for secondary education, girls' attendance at low quality sub-county or day schools, and consequently low aspiration to continue learning are all barriers identified in PW1.<sup>31</sup>

1. Project **activities in learning** include rolling out coaching in secondary schools and piloting information and communication technology (ICT) support for teachers and learners. Through this coaching, secondary teachers are trained to improve their knowledge of inclusive teaching practices, with the hope that such practices will lead to an increase in girls' attendance and success in school.
2. Project **activities for girls** include the establishment of girl clubs and peer mentoring in the school and community. Such activities intend to improve girls' physical and mental health, self-confidence, and academic aspirations, while supporting holistic personal/social development.
3. Project **activities in the home** include secondary school fee support which improves households' access to financial resources, allowing them to actively support girls' learning, and easing the transition from primary to secondary school.
4. Project **community activities** include forums which intend to train communities in understanding the importance of education and establishing positive attitudes/perceptions towards girls' education.

**Pathway 2 (PW2) – Primary to Alternate Learning Pathway (ALP) Barriers:** Limited alternative options for girls not transitioning from primary to secondary school, as well as negative perceptions of, and low support for alternative options, are the two main barriers identified in PW2.

1. Project **activities in learning** focus on establishing PW2 as a pathway that supports positive post-primary outcomes. Specific activities include teacher development in youth polytechnics and utilising such teacher development to expand ALP options.<sup>32</sup>
2. Project **activities for girls** include instituting educator/peer mentoring programs with the aim of growing girls' confidence and aspirations.
3. Project **activities in the home** include offering financial support to households and raising awareness of different post-primary schooling options. In raising awareness of different post-primary options, these activities intend to raise awareness for the financial resources available to

<sup>31</sup> "Low quality" is a term applied to schools that receive low marks on the Kenya Certificate of Primary Education (KCPE) Examination.

<sup>32</sup> Note, this activity was dropped as part of a scale down of the primary to ALP pathway.

girls interested in attending ALPs, in turn supporting higher rates of post-primary to ALP transition.

4. Project **activities for communities** include working with the private sector on supporting new ALPs. Activities also include those working to improve community awareness of ALPs, with the hope that through these activities, community members will assist girls' transition from primary school to ALPs.

**Pathway 3 (PW3) – School Dropouts to Catch-Up Classes/Re-Entry to Education Barriers:** Lack of catch-up opportunities for girls who drop out of school, no time or household support for extra study, and few school re-integration processes are all barriers identified in PW3.

1. Project **activities in learning** include the establishment of new ALPs through catch-up sessions. From these catch-up sessions, consequent ALPs become avenues connecting those who dropped out of school to re-entry opportunities.
2. Project **activities for girls** include promoting re-entry options and mentoring, resulting in improved mental and physical health, higher self-confidence, and an increase in academic aspirations.
3. Project **activities in the home** include Community Health Volunteers (CHVs) visits and the distribution of Back-to-School kits, both of which aim to improve access to knowledge and resources, resulting in active household support for girls' re-entry to school.

### Assumptions

The following table lists key assumptions related to programme activities, outputs, intermediate outcomes, and outcomes, as well as the reasoning for these assumptions. In cases where assumptions do not hold, the programme's Theory of Change may face validity issues.

**Table 4: Overview of project assumptions and reasoning**

Output	Assumptions	Reasoning
Activities	<ul style="list-style-type: none"> <li>- Schools and teachers are willing/have time to engage in continuing professional development activities</li> <li>- Households will engage with CHVs</li> <li>- Girls are willing to act as mentors during holiday periods</li> <li>- Partners have links with the community</li> <li>- Venues exist for Aps courses and catch-up class creation in the community</li> </ul>	<ul style="list-style-type: none"> <li>- Observed in EDT's GEC 1 programme</li> <li>- Observed in the existing WWW programme</li> <li>- Partner selection was based on this</li> <li>- Community Learning Centres in place</li> </ul>
Outputs	<ul style="list-style-type: none"> <li>- We define adequate coaches to cover primary and secondary schools (<i>engagement in the current WWW programme has been high</i>)</li> <li>- Secondary schools will engage with the programme (<i>the project has MoE support for identification of and engagement with schools, including Centres of Excellence</i>)</li> <li>- Families will send girls to day schools (<i>the WWW project supports improved performance of day/sub-county schools</i>)</li> <li>- Girls will engage with the girls' clubs, training, and mentoring (<i>observed in the existing WWW programme</i>)</li> <li>- County governments want to commit to improving TVET (<i>MoE assures that this is the case</i>)</li> </ul>	<ul style="list-style-type: none"> <li>- Engagement in the current WWW programme has been high</li> <li>- The project has MoE support for identification of and engagement with schools, including Centres of Excellence)</li> <li>- The WWW project supports improved performance of day/sub-county schools</li> <li>- Observed in the existing WWW programme</li> <li>- MoE assures that this is the case</li> </ul>

Intermediate Outcomes	<ul style="list-style-type: none"> <li>- Coaching models improve teacher practice (<i>existing model demonstrated this</i>)</li> <li>- Increased knowledge/awareness shifts girls'/families' choices around education, including alternate learning pathways (<i>this happened in our partner programmes</i>)</li> <li>- Improved understanding of girls' needs will shift schools' resources to improve girls' facilities, ie. Sanitation (evidenced by partner programmes)</li> <li>- CHVs have the time/influence to identify and prevent dropout (<i>existing programmes suggest this is the case</i>)</li> <li>- Community groups have a significant influence on the community (<i>evidence from the current WWW programme suggests this is the case</i>)</li> </ul>	<ul style="list-style-type: none"> <li>- Existing model demonstrated this</li> <li>- This happened in our partner programmes</li> <li>- Evidenced by partner programmes</li> <li>- Existing programmes suggest this is the case</li> <li>- Evidence from the current WWW programme suggests this is the case</li> </ul>
Outcomes	<ul style="list-style-type: none"> <li>- Participatory/inclusive education will achieve learning</li> <li>- Transition to high quality secondary/relevant alternative learning pathways is more attractive to girls than other life choices</li> <li>- Improved learning foundation will increase completion of primary education and transitions to secondary school</li> </ul>	<ul style="list-style-type: none"> <li>- Evidence from A Girls' Advancement Education Initiative</li> <li>- Suggested by high competition for higher-quality Kenyan secondary schools/USAID youth programmes in northeast Kenya</li> <li>- International evidence/GEC1 evidence supports this</li> </ul>

Output assumptions are further detailed in Annex 15.

## Methodology

In this section, we discuss the methodology for the WWW endline evaluation. We begin with a discussion of evaluation questions, followed by the evaluation design, data collection tools, sampling methodology and achieved sample, and limitations and assumptions. More detailed methodological analysis is included in Annex 2.

## Evaluation Questions

The endline evaluation was designed first to evaluate the programme along Organisation for Economic Co-operation and Development (OECD) Development Assistance Committee (DAC) criteria of relevance, effectiveness, efficiency, impact, and sustainability. This includes the following key evaluation questions:

- How relevant and appropriate were the WWW project interventions towards addressing the intended outcomes and intermediate outcomes as shown in the project Theory of Change and logframe?
- How effective were the WWW project interventions in reaching its intended learning, transition, and intermediate outcomes? How did the interventions address the behaviour change and perceptions around the girl's education as in the project Theory of Change?
- To what extent did WWW programme contribute to internal efficiency of the education system (reducing dropout and enhancing attendance) in the focus counties?
- What were the overall tangible and attributable short-term impacts (outcomes and intermediate outcomes) of the project interventions towards influencing behaviour change, promoting transition, and improving learning outcomes?
- How did the project cost-efficiently and cost-effectively deliver the project interventions, with special emphasis on value for money (VfM)?
- How sustainable are the project interventions at the school, system, and community towards addressing the barriers towards girl education?

Annex 2 lists specific evaluation questions in detail.

## Evaluation Design

The endline evaluation utilised a mixed-methods evaluation with quantitative and qualitative tools. For the sake of brevity, fieldwork procedures and attrition rates from the sample are discussed in Annex 2; here, we discuss the basic design of the endline evaluation, data collection tools, and the achieved sample.

Our evaluation had three distinct cohorts:

- **Cohort 1:** This cohort included girls who participated in the midline evaluation. It included both intervention and comparison schools.
  - **Intervention:** This subsample was drawn from schools that were sampled in the midline and SAS evaluations, and in which EDT works.
  - **Comparison:** This subsample included girls who were sampled at midline in schools where no EDT interventions have taken place.
- **Cohort 2:** This cohort included girls who were sampled in the SAS evaluation only and is limited to girls in target counties and schools. We generally also constrained this sample to schools with at least six girls expected to remain in attendance for logistical and budgeting purposes.<sup>33</sup>
- **Cohort 3:** This cohort included girls outside of the sample schools in three subgroups. This sample was contacted via a Computer-Assisted Telephonic Interview (CATI) methodology, as many girls were no longer present in target schools and counties and were widely spread across the country, making in-person surveys infeasible.
  - **Tracer Study:** We attempted to recontact girls who were sampled in the SAS evaluation but who were no longer in target counties or schools. This did not include girls who had completed Form 4 and successfully graduated.
  - **Alternative Pathways:** This subsample included girls from TVET, apprenticeships, or in-school catch-up programmes.
  - **Kwale and Marsabit:** This subsample included girls reached at midline in Kwale and Marsabit counties, who could not be surveyed in person by Consilient for logistical and budgetary reasons. Additionally, EDT conducted an internal, in-person learning survey of girls in Marsabit and Kwale.

These three cohorts contribute to analysis in different ways. Cohort 1 data is critical for analysis related to programme impact, as it is the only cohort with comparison data. Difference in difference analysis around learning outcomes relies on midline data and Cohort 1 endline data. This design allows inferences to be drawn regarding changes in girls' assessment scores over time, by comparing the relative change among girls in intervention schools to the relative change among girls in comparison schools. However, Cohort 1 is limited in its small sample size, an issue discussed further below.

Cohort 2, as the largest sample, is required for subgroup analysis and disaggregation between groups. We also rely on Cohort 2 data to understand predictors (at the girl, household, school, and community levels) of key outcomes of interest. This is supplemented by endline data from Cohorts 1 and 3.

Cohorts 2 and 3 are the primary data contributors to transition analysis. Analysis of Cohort 2 learners will provide insights into predictors of transition and re-enrollment. The Cohort 3 tracer subsample was intended to be used to track various pathways (and their predictors) of learners who have left their schools since the SAS evaluation; however, issues recontacting these girls, discussed further below, means that we do not conduct analysis for this subsample.

---

<sup>33</sup> This eliminated 25 schools and 180 girls from the sample.

### ***Cohort 1 Analysis Methodology***

For analysis using Cohort 1, the evaluation utilises a quasi-experimental difference-in-differences design. This design incorporates an explicit comparison group of schools/communities that did not receive WWW interventions; school-age girls sampled from these comparison schools constitute the comparison group for assessment of learning outcomes. Data collection in comparison schools and intervention schools occurred at midline, and we tracked girls using information from this evaluation round. This design allows for inferences to be drawn regarding changes in girls' assessment scores over time by comparing the relative change among girls in intervention schools to the relative change among girls in comparison schools.

Use of a difference-in-differences approach mitigates for two important sources of bias when drawing inferences regarding programme impact. First, by comparing the same respondents across evaluation rounds, the design accounts for differences between treatment and comparison groups. For instance, if programme schools already perform at a higher level than non-intervention schools regardless of programme intervention, this would bias cross-sectional comparisons of outcomes, and would lead us to overestimate programme impact. By adjusting for gaps that existed prior to project implementation, the difference-in-differences approach ensures that pre-existing differences do not bias the conclusions drawn about project impact. Second, the design controls for systematic changes in outcomes over time that are not attributable to the programme itself. For instance, in the context of a widespread shock that reduced learning outcomes across most of the region (such as COVID-19), a simple longitudinal comparison of learning outcomes in programme schools over time would suggest that the programme reduced student performance. A difference-in-differences approach, in contrast, may show that the programme helped mitigate negative external effects.

### ***Cohort 2 and 3 Analysis Methodology***

For cohorts 2 and 3, due to the methods employed in previous evaluation rounds, we had no ability to assess comparison groups. As such, we utilised these cohorts to conduct trend analysis of changes in outcomes of interest over time, provide descriptive analysis of the status of outcomes of interest, test the theory of change, and conduct a predictive analysis of the relationship between various student- and school-level characteristics and outcomes of interest.

### ***Qualitative Data Analysis Methodology***

Our analysis of qualitative data from key informant interviews (KIIs) and focus group discussions (FGDs) focuses on validating results from the quantitative survey and providing information on programme outcomes that are difficult to measure through quantitative data, such as sustainability and efficiency. Qualitative interviews provide insight into the 'how' and 'why' of trends observed in the quantitative data and gleaned through secondary sources. Interviews allow for nuanced and open-ended responses to difficult and/or complex topics, thereby eliciting more information on attitudes, perceptions, and experiences that otherwise cannot be obtained by a structured survey.

For the qualitative data, we focus on thematic analysis to understand reasons underlying changes in outcomes of interest, how the WWW programme contributed to results, and what other factors may have influenced results. Overall, the goal of the qualitative analysis is to assess and validate the programme's impact, better understand whether impact varies across relevant subgroups, and understand the mechanisms of change.

### ***Use of Secondary Sources***

Analysis of primary data was supplemented using secondary sources such as previous evaluation and monitoring reports, data from past evaluation rounds, programme documents (such as reports on activities undertaken and challenges faced), and literature (including academic and "grey" literature) on education in Kenya. During surveys, some additional secondary information was collected from schools,



such as exam results and club activities. These secondary sources were primarily used to answer research questions around value for money and, to a lesser extent, sustainability. Another key use (particularly of external academic and “grey” literature sources) was to further understand questions around COVID-19-specific challenges faced by students and schools in Kenya, as well as issues surrounding education of marginalized groups such as girls with disabilities.

## Data Collection Tools

The design of this endline evaluation closely matches that of previous evaluation rounds in the tools used, with the exception of an additional classroom observation tool added to better understand use of positive and negative teaching practices. Full survey scripts and qualitative interview guides are annexed to this report. The tools included the following:

- Quantitative tools
  - Learning assessments: Early Grade Reading Assessment (EGRA), Early Grade Mathematics Assessment (EGMA), Secondary Grade Maths Assessment (SeGMA), Secondary Grade Reading Assessment (SeGRA), and science, technology, engineering, and mathematics (STEM) assessment
  - Girl survey
  - Caregiver survey
  - School (head teacher) survey
  - Classroom observation
- Qualitative tools
  - FGDs with girl students
  - KIIs with teachers
  - KIIs with government officials and programme staff

Tools are described in detail in Annex 2.

## Target and Achieved Sample

As this evaluation builds off of and refers to four previous evaluation rounds—the baseline, midline, RAS, and SAS—the below table shows data collected in each of these rounds. The endline sample is discussed in more detail below.

**Table 5: Data collected within each evaluation round**

	Baseline	Midline	RAS	SAS	Endline
<b>Learning Assessments</b>					
Intervention	5,168	11,788	1,285	1,861	1,278
Comparison	1,700	2,666	-	-	178
<b>Girl Survey</b>					
Intervention	<i>See footnote<sup>34</sup></i>	6,777	1,285	1,861	933
Comparison		1,707	-	-	-
<b>Caregiver Survey</b>					
Intervention	<i>See footnote</i>	5,345	-	498	818
Comparison		1,358	-	-	-

<sup>34</sup> Sample size for the girl survey and caregiver survey was not recorded at baseline, and is thus not included here. These assessments were conducted with intervention and comparison groups, however.

School Survey					
Intervention	197	204	123	166	129
Comparison	50	63	-	-	27
Classroom Observation					
Intervention	-	-	-	-	129
Comparison	-	-	-	-	-

Table 6 below shows the target sample by tool, Table 7 shows the target sample for learning assessments by county, and Table 8 shows the target sample for schools by county.<sup>35</sup>

**Table 6: Target sample by tool**

Tool	Cohort 1 Sample		Cohort 2 Sample	Cohort 3 Sample	Total Target
	Comparison	Intervention			
SeGRA/SeGMA	572	779	1,346	0	2,697
STEM assessment	0	0	551	0	551
Girl survey	0	0	1,346	900	2,246
Caregiver survey	0	0	1,346 <sup>36</sup>	0	1,346
School survey	25	133	133 <sup>37</sup>	0	158
Classroom observation	0	133	133 <sup>38</sup>	0	133
Teacher KIIs	0	20	20 <sup>39</sup>	0	20
Student FGDs	0	20	20 <sup>40</sup>	0	20

**Table 7: Target sample for learning assessments by county**

County	Cohort 1 Sample		Cohort 2 Sample	Total Target
	Comparison	Intervention		
Kilifi	56	108	243	407
Mombasa	59	67	79	205
Nairobi	201	279	470	950
Samburu	39	48	138	225
Tana River	176	221	219	616
Turkana	41	56	197	294
<b>Total</b>	<b>572</b>	<b>779</b>	<b>1,346</b>	<b>2,697</b>

<sup>35</sup> The sampling design is discussed in more detail in Annex 2.

<sup>36</sup> Surveyed via CATI, rather than in person.

<sup>37</sup> The same schools as included in the Cohort 1 Intervention group.

<sup>38</sup> The same classrooms as included in the Cohort 1 Intervention group.

<sup>39</sup> Teachers were identified within the intervention schools included in the Cohort 1 sample.

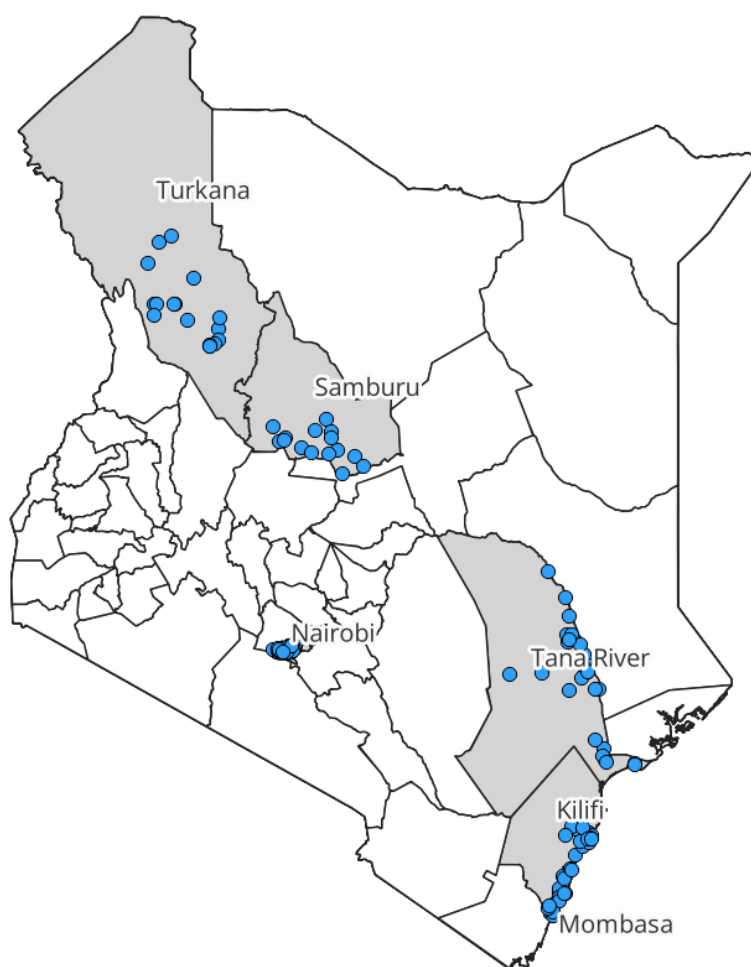
<sup>40</sup> Students for the FGDs were drawn from the same group of students targeted in the Cohort 1 sample for learning assessments in intervention schools.

Table 8: Target sample for schools by county

County	Cohort 1 Sample		Total Target
	Comparison	Intervention	
Kilifi	2	26	28
Mombasa	2	5	7
Nairobi	12	51	63
Samburu	3	12	15
Tana River	5	23	28
Turkana	1	16	17
<b>Total</b>	<b>25</b>	<b>133</b>	<b>158</b>

The below map shows the locations of schools and communities visited for the endline evaluation.

Figure 4: Map of school locations visited during fieldwork



We now discuss the achieved sample for each tool.

### **Learning Assessments**

For the SeGRA and SeGMA, a total of 1,456 girls were recontacted and consented to take the assessments. This represents a recontact rate of 54 percent, with 46 percent attrition.<sup>41</sup> By cohort, recontact rates were as follows:

- 504 girls were assessed within Cohort 1 (37 percent recontact rate, 63 percent attrition)
  - 326 girls were assessed in intervention schools (42 percent recontact rate, 58 percent attrition)
  - 178 girls were assessed within comparison schools (31 percent recontact rate, 69 percent attrition)
- 952 girls were assessed within Cohort 2 (71 percent recontact rate, 29 percent attrition)

It is worth noting the significantly lower recontact rates for Cohort 1 compared to Cohort 2. Cohort 1 included girls assessed two years prior at midline; during fieldwork, it was found that information on the locations and contacts for these girls was frequently out-of-date or incomplete, resulting in lower recontact rates (see Annex 2 – During Data Collection for a more extensive discussion of recontact procedures). In contrast, for Cohort 2—girls contacted during the SAS assessment—recontact rates were around 71 percent, on par with previous GEC evaluations conducted by Consilient. These dynamics are discussed further below under Challenges and Limitations.

The below tables show the achieved sample for learning assessments by county, cohort, and class/form. We note that we do not calculate recontact rates by class/form as much of our data on expected class/form was found to be incorrect.<sup>42</sup> However, comparing to data on class/form of all beneficiaries tracked at midline (and extrapolated to expected grades at endline), we note that our sample underrepresents girls in Class 7 and Form 1 and overrepresents girls in Forms 2, 3, and 4.

**Table 9: Achieved sample by county and class/form, SeGRA and SeGMA**

County	Cohort 1 Sample		Cohort 2 Sample	Total	Target	Recontact Rate
	Comparison	Intervention				
Kilifi	39	50	192	281	407	69.0%
Mombasa	21	22	47	90	205	43.9%
Nairobi	22	97	302	421	950	44.3%
Samburu	9	25	122	156	225	69.3%
Tana River	64	101	152	317	616	51.4%
Turkana	23	31	137	191	294	65.0%
<b>Total</b>	<b>178</b>	<b>326</b>	<b>952</b>	<b>1,456</b>	<b>2,697</b>	<b>54.0%</b>

<sup>41</sup> An additional 32 girls were recontacted but did not consent to take the learning assessments.

<sup>42</sup> We found substantial discrepancies in previously collected beneficiary data related to class/form; indeed, these discrepancies contributed in part to low recontact rate, as many girls who were expected to still be in primary schools, given their grade recorded at midline, had since moved to new secondary schools. As a result, we do not calculate recontact rate by class/form.

Class/Form						
Class 7	1	1	9	11	-	-
Class 8	3	10	450	463	-	-
Form 1	1	5	91	97	-	-
Form 2	16	11	131	158	-	-
Form 3	34	10	145	189	-	-
Form 4	271	141	126	538	-	-
<b>Total</b>	<b>178</b>	<b>326</b>	<b>952</b>	<b>1,456</b>	<b>2,697</b>	<b>54.0%</b>

For the STEM assessment, 383 assessments were administered. The target sample for STEM assessments agreed upon in the inception phase was 551; as such, the achieved sample represents a contact rate of 69.5 percent, with 30.5 percent attrition. However, in order to increase the likelihood of achieving the target sample, a total of 620 girls were targeted during fieldwork to take STEM assessments, giving us a contact rate of 61.8 percent, with 38.2 percent attrition. The below tables disaggregate by county and class/form for the total sample of 620 targeted during fieldwork; the tables do not disaggregate by cohort, as all STEM assessments were administered as intended with Cohort 2 girls.

**Table 10: Achieved sample by county and class/form, STEM assessment**

County	Achieved Sample	Target	Recontact Rate
Kilifi	105	166	63.3%
Mombasa	39	84	46.4%
Nairobi	65	104	62.5%
Samburu	55	63	87.3%
Tana River	62	107	57.9%
Turkana	57	96	59.4%
<b>Total</b>	<b>383</b>	<b>620</b>	<b>61.8%</b>
Class/Form			
Class 7	0	-	-
Class 8	0	-	-
Form 1	3	-	-
Form 2	126	-	-
Form 3	132	-	-
Form 4	113	-	-
Out of school	0	-	-
<b>Total</b>	<b>383</b>	<b>620</b>	<b>61.8%</b>

## Girl Survey

For the girl survey, 933 girls were recontacted and consented to the survey, a recontact rate of 69 percent with attrition of 31 percent.<sup>43</sup> The below tables show recontact rates for the girl survey by county and class/form. As above, we do not provide recontact rates by class/form as data on expected class/form was frequently incorrect.

**Table 11: Achieved sample by county and class/form, girl survey**

County	Achieved Sample	Target	Recontact Rate
Kilifi	189	243	77.8%
Mombasa	33	79	41.8%
Nairobi	310	470	66.0%
Samburu	124	138	90.0%
Tana River	141	219	64.4%
Turkana	136	197	69.0%
<b>Total</b>	<b>933</b>	<b>1,346</b>	<b>69.3%</b>
Class/Form			
Class 7	8	-	-
Class 8	431	-	-
Form 1	83	-	-
Form 2	133	-	-
Form 3	148	-	-
Form 4	125	-	-
Out of school	5	-	-
<b>Total</b>	<b>933</b>	<b>1,346</b>	<b>69.3%</b>

For girls that were not found, enumerators were instructed to submit data on where the girl is now. While this data is incomplete, it provides some indication of why girls were unable to be recontacted. Enumerators were able to find families of 55 girls within communities surrounding target schools; among this group, 76 percent of girls now attended boarding school or had transferred to a new school. Enumerators were able to find information about an additional 52 girls; among this group, 46 percent were confirmed to be attending a new school in a different location or to have graduated from Form 4, while 33 percent were confirmed to have moved to a new location, but their schooling status could not be confirmed.

## Caregiver Survey

Caregiver phone numbers were provided by girls during the girl survey. Using these numbers, 818 caregivers were contacted via CATI procedure, or 60.8 percent of the target sample. The below table shows contact rates for caregivers by county. We note that some caregivers of Marsabit/Kwale girls (from

<sup>43</sup> An additional 92 girls were recontacted but did not consent to take the girl survey, for a total recontact rate of 76 percent.

cohort 3) were contacted during the CATI using phone numbers obtained during the CATI with cohort 3 girls.

**Table 12: Achieved sample by county, caregiver survey**

County	Achieved Sample	Target	Contact Rate
Kilifi	163	243	67.1%
Kwale	32	0	-
Marsabit	9	0	-
Mombasa	28	79	35.4%
Nairobi	361	470	76.8%
Samburu	59	138	42.8%
Tana River	95	219	43.4%
Turkana	71	197	36.0%
<b>Total</b>	<b>818</b>	<b>1,346</b>	<b>60.8%</b>

This table shows, notably, that contact rates for caregivers in Nairobi and Kilifi were substantially higher than those for other counties. This may represent a limitation to the CATI procedure, whereby we are only able to survey caregivers who have access to a phone and a cellular network. This limitation is discussed more below.

### **School Survey**

All 158 target schools were visited by teams, but one school in Nairobi did not consent to a school survey and one school in Nairobi was closed. As a result, 156 school surveys were conducted. We note that during fieldwork, the EDT team provided information that several targeted schools were comparison, rather than intervention, schools. As such, the achieved sample was 129 intervention schools and 27 comparison schools. This corresponds to 108 primary schools and 48 secondary schools. The table below shows the achieved sample by county and intervention status.

**Table 13: Achieved sample by county, school survey**

County	Comparison	Intervention	Total	Target	Recontact Rate
Kilifi	3	25	28	28	100.0%
Mombasa	2	5	7	7	100.0%
Nairobi	13	48	61	63	96.8%
Samburu	2	13	15	15	100.0%
Tana River	5	23	28	28	100.0%
Turkana	2	15	17	17	100.0%
<b>Total</b>	<b>27</b>	<b>129</b>	<b>156</b>	<b>158</b>	<b>98.7%</b>

### ***Classroom Observation***

Classroom observations were conducted in 127 intervention schools;<sup>44</sup> two schools in Nairobi did not consent to the classroom observation. This corresponds to 106 primary schools and 48 secondary schools. The following table shows the achieved sample by county.

**Table 14: Achieved sample by county, classroom observation**

<b>County</b>	<b>Total</b>	<b>Target</b>	<b>Recontact Rate</b>
Kilifi	25	24	104.2%
Mombasa	5	5	100.0%
Nairobi	47	48	97.9%
Samburu	13	13	100.0%
Tana River	23	22	104.5%
Turkana	16	16	100.0%
<b>Total</b>	<b>129</b>	<b>128</b>	<b>100.8%</b>

### ***Qualitative***

In total, 19 focus group discussions (FGDs) with girl students and 21 key informant interviews (KIIs) with teachers were conducted, mostly aligned with the target sample. The following table shows the sample by county.

**Table 15: Achieved sample by county, FGDs and KIIs**

<b>County</b>	<b>FGDs</b>	<b>KIIs</b>	<b>Total</b>
Kilifi	3	4	7
Mombasa	2	2	3
Nairobi	7	7	14
Samburu	2	2	4
Tana River	3	3	6
Turkana	2	3	5
<b>Total</b>	<b>19</b>	<b>21</b>	<b>40</b>

Sixteen additional KIIs were conducted with government officials and programme staff. In total, 11 KIIs were conducted with government officials (one in Kilifi and two in each of the remaining five counties), including County Directors and Sub-directors of Education, Curriculum Support Officers, and Quality Assurance and Standards Officers, and five KIIs were conducted with programme staff.

<sup>44</sup> Two classroom observations were also conducted in comparison schools.



## Limitations and Assumptions

A number of key limitations apply to our methodology and achieved sample.

**Difference-in-differences assumptions and limitations:** Inferences drawn via difference-in-differences rely on two key assumptions. The first assumption is that under the counterfactual condition—i.e., in the absence of intervention—the change in outcomes over time in treatment and comparison schools will be similar. This assumption is often referred to as the “parallel trends assumption” because it assumes that, in the absence of treatment, trends in outcomes in the treatment group would evolve in parallel to those in the comparison group. This first assumption is generally untestable under a difference-in-differences framework. In the absence of randomisation, the parallel trends assumption may or may not hold, though the selection process leaves no reason to expect divergent trends under the counterfactual.

The second assumption is that comparison group respondents are not exposed to the treatment or to similar interventions affecting learning and transition. Informally, this assumption is often stated as the “no spillover” assumption. The second assumption is testable *ex post*, by directly measuring whether comparison group respondents received any of the project’s interventions.

If both assumptions are satisfied, inferences drawn using difference-in-differences are expected to be unbiased. However, we note that, in the context of this endline evaluation, there is expected to be some contamination among comparison schools from other programme (i.e., not WWW) interventions.

Our school survey tool included questions on whether comparison (and intervention) schools have received support from other NGOs in order for us to better understand the extent of spillover effects. We find that among all endline schools, 94 out of 156 (60.3 percent) reported receiving investments from NGOs. This includes 5 out of 27 comparison schools, or 18.5 percent, who reported benefitting from interventions targeting teacher training, provision of learning materials and school supplies, and strengthening reading and mathematics—similar to those interventions provided by the WWW programme. This poses a limitation to the validity of the difference in differences approach, although only a minority of comparison schools were affected.

We also note that 38 intervention schools (29.5 percent of all endline intervention schools) reported receiving assistance from NGOs other than EDT (or Concern Worldwide, an implementing partner). This included interventions from Asante Africa, Kesho Kenya, UNICEF, the Safaricom Foundation, and the School Based Teacher Support System, among others. Given that interventions related to learning, teaching, and school quality were not solely implemented by the WWW programme in target schools, this may cause us to overestimate the impact of the WWW programme, some of which may be attributable to other programmes.

**Consistent assessment tools:** one of the difficulties related to the inability to use the same assessment instruments throughout the duration of the programme due to girls requiring different levels of difficulty in assessments as they progressed through school. For baseline, EGRA and EGMA were both used as they were the most grade-appropriate assessments for the girls at that time. By midline, EGMA and EGRA were determined to be the wrong difficulty level for girls, meaning the programme needed to use SeGRA and SeGMA.

**Tracking across evaluation rounds:** No tracking data on girls sampled at baseline was available at endline. This means that our analysis is not able to measure programme impact on girls from the full lifecycle of the programme. Similarly, because our past data extends at most to the midline evaluation (two years prior to the endline) and often only to the SAS evaluation (one year before endline), our analysis will only show change over a limited period of time. For some expected programme impacts, such as learning outcomes, we may still expect a measurable change. However, for some expected impacts that tend to be gradual—such as impacts on confidence and leadership—we may measure only

a limited impact. This would not necessarily be due to the efficacy (or lack thereof) of the programme, but rather due to the limited timeframe for analysis.

**Sample size of cohort 1:** The sample size of cohort 1 is small due to limited tracking data on girls sampled in the midline evaluation. Because attrition rates of girls from our sample were high, the statistical power of the evaluation is low, reducing our ability to draw statistically significant conclusions using a difference-in-differences methodology for analysis.

Our approach included substantial efforts to maximize sample size within constraints on timeline, budget, and tracking information from past rounds. However, our overall sample size remained low. Low recontact rates, particularly for Cohort 1 (midline) girls, were primarily due to incorrect information on the current location of the girl. Many girls were no longer present in the school recorded at midline; anecdotally, team leaders and enumerators often reported from asking the head teacher and community members that these girls had since graduated or moved to new schools.

The low recontact rates for Cohort 1 present a limitation for the difference-in-differences analysis of learning outcomes, as the sample size for this cohort is under-powered. This means that the chance of a “Type II” error in statistical analysis—accepting the null hypothesis of “no treatment effect” when there is, in fact, a treatment effect—is higher. However, it is important to note that the small sample size does not increase the risk of a “Type I” error—finding a treatment effect when there is no actual treatment effect. In other words, if our analysis shows no effect of the programme on learning outcomes, we cannot confidently say that this is accurate due to low sample size. However, if our analysis shows a significant effect of the programme on learning outcomes, the small sample size does not detract from this finding.

**Downstream effects of girl survey sample size:** Recontact rates for the girl survey had downstream effects on the sample size of the caregiver survey. Among the 69 percent of girls who were recontacted for the girl survey, only around 70 percent provided a phone number for their caregiver. This means that the total sample of caregiver phone numbers collected during fieldwork for the CATI is only around 650 numbers. Given that response rates to the CATI were inevitably not 100 percent, these dynamics mean that the caregiver sample was substantially smaller than the target sample.

**Recontact biases:** There may be some biases in our sample from girls that could not be recontacted. For example, girls who dropped out of school were more difficult to recontact, and may also have relatively lower learning, transition, or life outcomes leading them to drop out. As such, since these girls are excluded from our sample (because they could not be recontacted), we may overestimate our measures of programme impact. In contrast, girls who have since graduated or moved to new schools (such as boarding schools) were also difficult to recontact, but may have relatively higher learning, transition, or life outcomes which led them to successfully graduate or continue in school; this would lead us to underestimate programme impact. Since we do not have information on girls who were not recontacted, we cannot definitively say the direction or extent of this bias.

The difference-in-differences approach to analysis of cohort 1 is designed to mitigate some of this impact, as we would expect similar types of girls to drop out of both comparison and intervention schools. However, this bias is cumulative from previous rounds, and is thus an important limitation.

Annex 2 includes analysis on attrition characteristics and predictors of recontact. While we find that differential attrition by demographic characteristics is relatively small, there do appear to be some differences in girls who were recontacted compared to those who were not, including relatively higher recontact rates in Samburu, lower rates in Mombasa and Tana River, and higher recontact rates for disabled girls. We also find some differences in household responsibilities between recontacted and attrition girls, although small sample size limits our ability to draw conclusions about these differences. Lastly, we find that recontacted students had relatively higher SeGRA scores and (for intervention students) SeGMA scores. All of these findings suggest potential bias in our difference in differences analysis.

The CATI procedure may also introduce some biases, as we were unable to contact caregivers or girls who do not have phone access (who are also more likely to belong to marginalized groups and have high levels of poverty). Indeed, the above analysis of the achieved sample finds much higher contact rates in Nairobi and Kilifi than in other counties. This may limit the applicability of our conclusions to other counties with lower contact rates.

**Marking of learning assessments:** Due to time and logistical constraints, enumerators marked the SeGRA and SeGMA assessments themselves during fieldwork. This could potentially affect our measures of programme impact if enumerators did not accurately mark assessments, particularly if assessments were consistently marked lower (or higher) than appropriate.

To address this potential issue, first, enumerators and team leaders were extensively trained on marking procedures during training. For the SeGMA, where answers are objective (i.e., either correct or incorrect), this involved a review of all questions and the marking scheme. For the SeGRA, some answers were subjective (i.e., have multiple possible correct answers, or are marked on a scale, rather than simply marked as correct/incorrect). As such, for the SeGRA, questions and the marking scheme were reviewed, but enumerators and team leaders also underwent extensive practice marking example answers to subjective questions. All examples were discussed at length to ensure that enumerators understood not only the correct mark, but also the rationale for the mark.

Second, quality control conducted during fieldwork identified enumerators whose marks were consistently below or above the average for all students (see Annex 2 for more detail). For these enumerators, we further investigated potential reasons for the deviations, such as the age and form of the students whose assessments were marked. If the deviations were not explainable by such factors, we requested that team leaders review the assessments in order to ensure that marking was accurate.

Third, after fieldwork concluded, all enumerators returned copies of their assessments to our offices. Experienced maths and English teachers were hired to review the assessments at random, ensuring that at least three SeGRA and three SeGMA assessments were reviewed for each enumerator. If a substantial deviation from the appropriate marking was found, the enumerator's assessments would be re-marked. Using these procedures, one enumerator was identified whose assessments were then re-marked.

Given the technicality of the STEM assessment, copies of this assessment were kept by team leaders, and then marked later by individuals with education in science. As such, the above limitations do not apply to the STEM assessment.

**Format of learning assessments:** The format of learning assessments was adjusted slightly from midline/SAS. At midline/SAS, girls were given learning assessments which included both questions and spaces for answers on the same sheet. In contrast, at endline, girls were provided with separate question and answer sheets. This may have slightly increased the difficulty of learning assessments as girls may have made mistakes matching question and answer numbers, copying numbers for maths problems, or coming back to questions previously skipped (although this issue was not noted by enumerators during fieldwork). However, we note that the same learning assessments were administered to both comparison and intervention girls, meaning that any change in difficulty is equivalent across groups and will not affect analysis. Furthermore, several additional steps were taken to reduce the increase in difficulty, including by providing every girl with a question sheet (so that she did not have to copy from one "master sheet" used by multiple girls) and instructing enumerators to mark answers correctly if they had been consistently copied to the wrong question numbers of the answer sheet.<sup>45</sup> Given these steps and the fact

---

<sup>45</sup> For example, if a girl accidentally answered questions 3 through 12 in the answer boxes for questions 4 through 13, these were marked as if they were written in the correct answer boxes.

that girls are in upper primary or secondary schools, we expect a minimal, at most, impact of these changes in format.

**Administration of SeGRA:** At endline, many students scored 0 on subtask 3 (short story writing) of the SeGRA. Given that scores for subtasks 1 and 2 were relatively high, the percent of zero scorers likely does not reflect students' inability to complete this subtask, but rather reflects time management or survey administration issues. At the beginning of fieldwork, for example, enumerators did not prompt students with the time remaining in the assessment, which may have led students to not attempt subtask 3. While this issue was noted and enumerators were instructed to begin prompting students with time remaining, the issue may have led to lower-than-expected subtask 3 results. This issue is discussed in more detail in the section on Learning Outcomes.

**Floor effects for learning assessments:** Subtask 3 of the SeGMA and all subtasks of the STEM demonstrated substantial "floor effects", as evidenced by a large percentage of students scoring 0 on these subtasks. These floor effects had the most substantial effect for the STEM assessment; in essence, this assessment was too difficult for most students at both SAS and endline. This effect resulted in limited change in results from SAS to endline and makes it difficult to meaningfully compare average scores across groups.

**Benchmarking of learning assessments:** In line with wider contextual constraints with learning assessments in Kenya, there was a lack of benchmarking for the SeGRA, SeGMA, and STEM assessments on what should be considered a "pass" or "fail" and what score demonstrates grade level proficiency. While our analysis includes benchmarks for proficiency, these are not tied to any standardized Kenya-wide benchmarking system, but are retained from past evaluation rounds for the sake of comparability.

**Fieldwork challenges:** Outside of these limitations to the analysis, the main challenges faced during fieldwork were logistical. In Turkana, Samburu, and Tana River, long distances between many schools made transportation challenging and costly. This issue was resolved in part through the provision of transportation by EDT teams.

**Soliciting information through CATI:** In general, it is more difficult to establish rapport over the phone than in person, which may reduce the quality and depth of the information solicited through CATI approaches. Our CATI training addressed these issues to mitigate their impact on data. We note that enumerators did not report substantial difficulties establishing rapport during the CATI.

**Higher recontact rates for minority language speakers and girls with disabilities:** As discussed in Annex 2, we were significantly more likely to recontact minority language speakers and girls with disabilities than individuals not belonging to these groups, controlling for other factors. Our analysis suggests that at endline, girls belonging to these groups generally had lower assessment scores than majority language speakers and girls without disabilities, likely due to the greater systematic difficulties faced by marginalized girls to pursue education as well as, for girls with disabilities, insufficient accommodations during assessments. This could theoretically reduce our measures of learning in comparison to midline.

## Learning Outcomes

Literacy, numeracy, and STEM outcomes were measured using a Early Grade Reading Assessment (EGRA) and Early Grade Mathematics Assessment (EGMA) at baseline, and Secondary Grade Reading Assessment (SeGRA), Secondary Grade Mathematics Assessment (SeGMA) from midline onwards, with a bespoke test for biology, chemistry, and physics. The change in assessments from baseline to midline was due to the need for a test of higher levels of difficulty, given the girls were two years older at midline compared to baseline. Overall, we find that WWW girls made progress in literacy and numeracy between baseline and midline. However, in line with global trends, WWW girls experienced large learning losses immediately after the COVID-19 pandemic as a result of school closures. Learning progress resumed once girls had returned to school. At endline, our results show that girls were still recovering from these losses; as such, the ambitious endline target of a 60% improvement above midline levels was not met.

However, in this section, we find indicative evidence that suggests that WWW may have had a mitigating effect on learning losses. While we do not find a statistically significant impact for intervention students for most assessment tasks, we find that intervention students had, on average, higher SeGRA and SeGMA scores at endline than would be expected given the results of comparison students. In other words, the WWW programme may have had a positive impact on learning for beneficiary girls, mitigating some of the worst negative effects of COVID-19.

### Summary of Findings

1. Targets of a 60% improvement over midline for literacy and numeracy were not achieved at endline, largely due to widespread learning losses in Kenya resulting from the COVID-19 pandemic. It may have been appropriate to revise this target downward in response to the pandemic.
2. WWW girls made progress in literacy and numeracy between baseline and midline, before experiencing large learning losses immediately after the onset of the COVID-19 pandemic as a result of school closures. At endline, learning progress has resumed when compared to assessment results from immediately after school re-openings, and girls are beginning to recover from learning losses.
3. Indicative evidence suggests that WWW girls may have experienced a more rapid recovery of learning than girls in the control group, suggesting that the programme may have mitigated some of the negative impacts of school closures.
4. Learning outcomes were generally worse for girls from ASAL counties and minority backgrounds, indicating that greater support may still be needed for these groups, who may also have faced more acute effects from COVID and school closures.
5. In focus group discussions across all counties, girls believed that they were progressing well in school and attributed this to effective teaching practices such as group discussions, allowing students to ask questions, and explaining topics in new ways for students who are struggling.

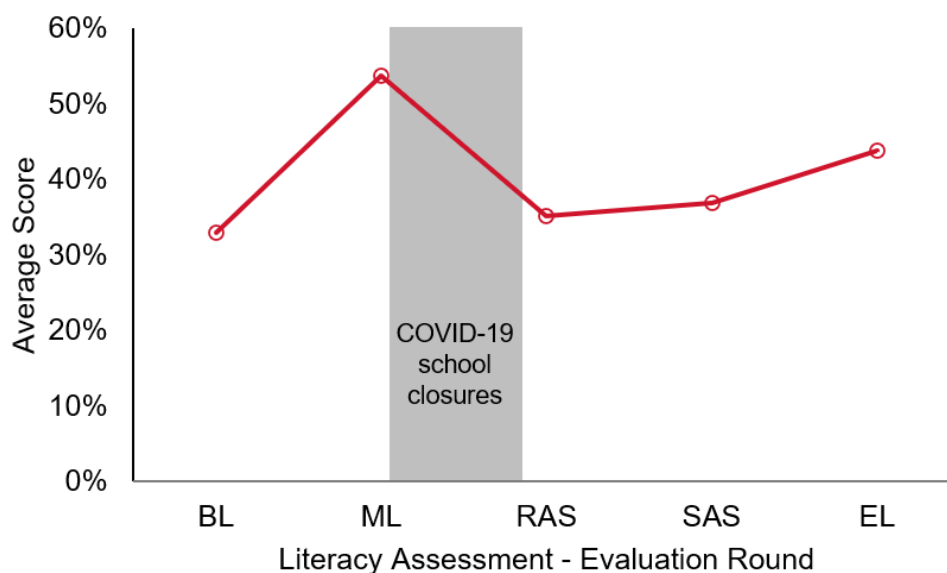
## Literacy Outcomes

Students who participated in the SeGRA were assessed with three subtasks. The first involved reading a short passage and answering basic reading comprehension questions with answers that could be drawn directly from the text. The second involved reading a short passage and answered reading comprehension questions that required interpretation or analysis of the text. The third involved writing a short story based on a half-sentence prompt. Each task was scored as a percent of items correctly answered out of the total possible score – six for the first subtask, 19 for the second, and 20 for the third. An overall literacy score was also calculated by dividing the total number of points achieved over the total possible score of 45. This means that individual test items have different weights, since individual

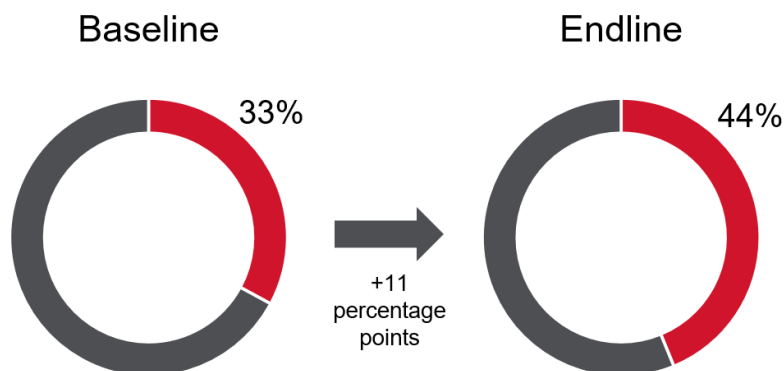
questions could be scored out of different numbers of maximum points. We note that while weighting of test items varied across assessment rounds, assessment difficulty in each round was calibrated against the previous round to allow for equivalence, and total scores have been calibrated to account for differences in number of items.<sup>46</sup>

The below figure shows change in average EGRA and SeGRA scores from all evaluation rounds for intervention students. In the two-year period between baseline and midline, average reading assessment scores increased by 21 percentage points, with an average SeGRA score of 53.7% at midline. The rapid assessment (RAS) results, which were collected when schools reopened to assess learning loss, showed average scores had dropped close to baseline performance. In the one year-period between the RAS and endline, girls were showing positive signs of learning recovery on a similar rate of progress that was noted between baseline and midline. There was an overall increase in average score on respective tests of 11 percentage points from baseline to endline.

**Figure 5: Change in average EGRA and SeGRA scores across all rounds, intervention students**



**Figure 6: Change in overall EGRA and SeGRA scores since baseline**



<sup>46</sup> For more details on assessment calibration, see Annex 12.

**Table 16: Change in SeGRA scores across all rounds, intervention students**

Task	BL (EGRA)	ML (SeGRA)	RAS (SeGRA)	SAS (SeGRA)	EL (SeGRA)
n	2,521	917	953	1,161	1,265
EGRA/SeGRA overall score	32.9	53.7	35.2	36.9	43.8
Subtask 1	-	73.2	42.1	51.5	69.7
Subtask 2	-	45.7	30.3	28.9	40.1
Subtask 3	-	42.1	33.4	30.5	35.6

A large portion of the decline in scores since midline is likely attributable to the COVID-19 pandemic and related school closures, which had a detrimental effect on learning worldwide.<sup>47</sup> Indeed, research has shown substantial learning losses due to COVID-19: Patrinos et al. (2022), for example, finds that across a range of countries, learning losses averaged around half of a year's worth of learning. Other studies have found learning losses ranging between half a year to a full year's worth of learning.<sup>48</sup>

To better understand the impact of the WWW programme on literacy scores, we now use a difference-in-differences model to compare the change in scores among intervention and comparison groups from midline to endline. This allows us to understand the impact of the programme separate from the more general negative impact of COVID-19 on all students. This model is limited to students who were surveyed at both midline and endline – i.e., “panel” students.<sup>49</sup> Because this dataset includes the same students assessed at two different times, we can be reasonably confident that any findings are due to programme impact, rather than differences between midline and endline student groups.<sup>50</sup>

Using this model, we find a positive but not significant effect for the intervention group compared to the comparison group. While average scores for both groups declined at endline compared to midline, average scores for intervention students exposed to the programme declined less than those for comparison students. Figure 9 shows that at endline, average calibrated scores for the comparison group had declined by 10.6 percentage points, from 50.1 percent at midline to 39.5 percent at endline, while average scores for the intervention group had declined by 6.7 percentage points, from 54.8 percent at midline to 48.2 percent at endline. This suggests that the WWW programme may have had a positive effect on mitigating the negative impacts of COVID-19 on learning.

<sup>47</sup> Some of the change in scores may also be due to differences in student characteristics or assessment difficulty across rounds. These limitations are discussed more in Annex 2 and Annex 16. Annex 16 also includes additional analysis of the cross-sectional sample, as a check for robustness of the difference-in-differences findings.

<sup>48</sup> These findings are validated by findings from Angrist et al. (2021), in which Early Grade Reading Assessments were conducted in Ethiopia, Kenya, Liberia, Tanzania, and Uganda and the authors found between half a year to more than a year's worth of learning loss. They are also validated by findings from a study conducted by Whizz Education (2021) in Kenya, which similarly found learning losses equivalent to slightly over one year. Sources: Angrist et al., 2021. “Building back better to avert a learning catastrophe: Estimating learning loss from COVID-19 school shutdowns in Africa and facilitating short-term and long-term learning recovery”. *IJED-UNICEF*. Patrinos, Harry Anthony, Emiliana Vegas, and Rohan Carter-Rau, 2022. “An Analysis of COVID-19 Student Learning Loss”. *World Bank Policy Research Working Paper* 10033.

Whizz Education, 2021. “Measuring the Impact of COVID-19 on Learning in Rural Kenya”.

<sup>49</sup> The sample for this model includes 366 intervention students and 159 comparison students assessed at both midline and endline.

<sup>50</sup> This model also calibrates student's scores based on a comparison of the relative difficulty of assessments between midline and endline, discussed more in Annex 12.

**Figure 7: Change in SeGRA scores, panel students**



However, as will be discussed further below, at endline, many students scored 0 percent on subtask 3. Follow-ups with these students revealed that this was not due to inability to write a short story, but rather because students ran out of time either because they were not prompted with the time remaining in the assessment or because of time management issues. As such, inclusion of subtask 3 in the overall results biases our analysis, and we thus now analyse results for just subtask 1 and 2 scores.<sup>51</sup>

The below figure shows that at endline, calibrated subtask 1 and 2 scores increased for intervention students, while decreasing for comparison students. At endline, calibrated intervention scores increased by an average of 2.1 percentage points, while comparison scores decreased by an average of 3.4 percentage points. While these findings remain insignificant in the regression analysis, they are suggestive of positive programme impact.

<sup>51</sup> These scores are similarly calibrated by excluding subtask 3 scores from the assessments used for calibration.



**Figure 8: Change in SeGRA task 1 and 2 scores, panel students**



Further robustness checks are included in Annex 16, including controlling for the age of students and school characteristics and analysis of cross-sectional data. These analyses generally show similar positive results as those reported above, although results remain insignificant in all regressions.

Qualitative data from teachers further suggests that there may have been a link between WWW interventions and positive learning outcomes. All interviewed teachers stated that they felt more confident teaching English now than three years ago, and many stated that they had learned new teaching practices through the WWW programme that had helped them improve student learning. For example, one teacher in Turkana stated that he now used simple English which allowed learners to understand better,<sup>52</sup> while another teacher in Nairobi stated that she had seen improvements in student learning after she changed teaching practices to focus on vocabulary first.<sup>53</sup>

Lastly, we now conduct a foundational skill diagnosis by categorising learners into four achievement bands, as at midline. These bands include non-learners, categorised as those who score 0 percent; emergent learners, for those who score 1 percent to 40 percent; established learners, who score 41 percent to 80 percent; and proficient learners, who score 81 percent or higher. Given challenges with subtask 3, this table shows learning bands for calibrated scores on subtasks 1 and 2; analysis of overall scores including subtask 3 is included in Annex 16.

The below table shows that at endline, among the panel sample, there was an increase in proficient and established learners in the intervention group. In contrast, there was a decrease in the percent of established learners in the comparison group. This suggests, as above, that the WWW programme may have helped improve literacy scores among intervention students, representing a rebound from learning loss during the pandemic.

<sup>52</sup> KII with teacher, Turkana, Int. 5

<sup>53</sup> KII with teacher, Nairobi, Int. 11

**Table 17: SeGRA subtasks 1 and 2 learning bands**

Panel Students	Intervention		Comparison	
	ML	EL	ML	EL
Non-learner (0%)	0.3%	0.3%	0.6%	0.0%
Emergent learner (1% - 40%)	19.4%	11.2%	18.9%	27.7%
Established learner (41% - 80%)	59.6%	61.2%	71.1%	59.8%
Proficient learner (81% - 100%)	20.8%	27.3%	9.4%	12.6%

To further investigate these impacts, we now analyse changes in scores on specific subtasks. We focus this discussion on changes in the panel sample from midline to endline, as this allows us to draw some inferences regarding programme impact.

### **Subtask 1 Outcomes**

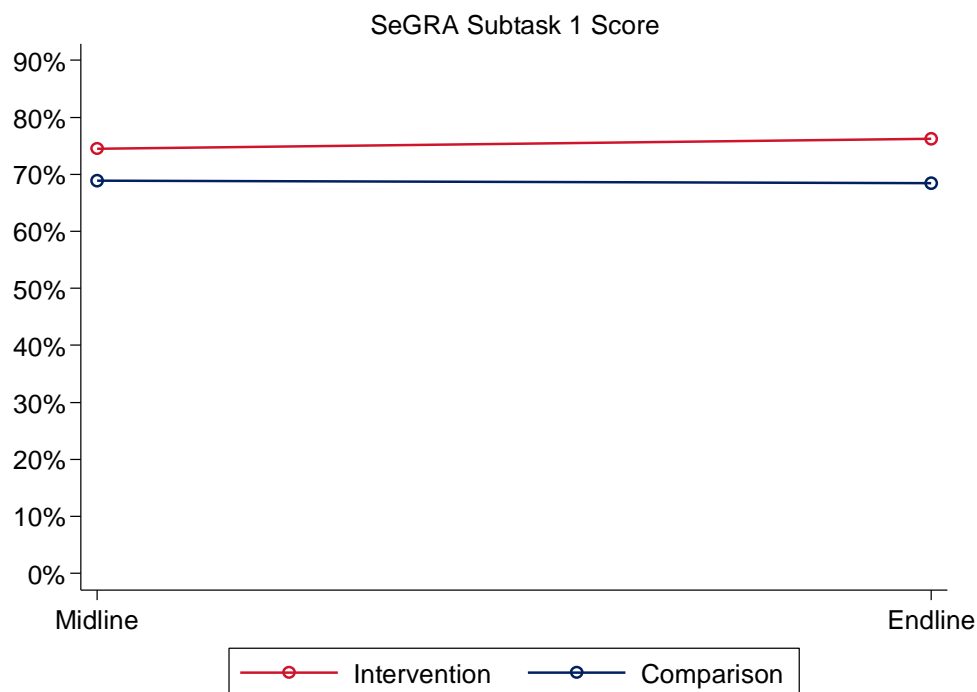
For subtask 1, students were asked to read a passage and answer simple reading comprehension questions. For all intervention students at endline,<sup>54</sup> the mean subtask 1 score was 69.7 percent, compared to a mean subtask 1 score of 73.2 percent for all midline intervention students (see Table 16).<sup>55</sup>

To better understand programme impact, as with the overall literacy results, we now utilise difference-in-differences analysis.<sup>56</sup> The below figure compares calibrated subtask 1 results. We find that average scores increased by 1.8 percentage points among intervention students, while remaining stagnant among comparison students. While these results are not significant, they do represent a slight positive change among students who benefitted from the WWW programme, particularly notable given the impact of COVID between midline and endline.

<sup>54</sup> i.e., the cross-sectional sample.

<sup>55</sup> This score is uncalibrated. However, subtask 1 was well-calibrated across midline and endline, with a calibration factor of 0.999. In other words, a student who took midline subtask 1 followed immediately by endline subtask 1 was likely to, on average, receive almost exactly the same score on both subtasks.

<sup>56</sup> Annex 16 includes additional robustness checks, with are substantively similar to the results shown here.

**Figure 9: Change in SeGRA subtask 1 score, panel students**

Looking now at learning bands, for panel students, the below table shows that more students were categorised as proficient and less as non-learners at endline than at midline. Increases in the percent of proficient learners were similar across groups, with an increase of 3.9 percentage points for intervention students and 5.7 percentage points for comparison students.<sup>57</sup>

**Table 18: SeGRA subtask 1 learning bands**

Panel Students	Intervention		Comparison	
	ML	EL	ML	EL
Non-learner (0%)	1.9%	0.3%	2.5%	0.6%
Emergent learner (1% - 40%)	6.8%	4.6%	10.7%	10.1%
Established learner (41% - 80%)	51.9%	51.9%	64.2%	61.0%
Proficient learner (81% - 100%)	39.3%	43.2%	22.6%	28.3%

As with overall literacy scores, the results of the difference-in-differences analysis suggest a potential positive impact of the WWW programme on literacy levels. The results found for all assessed students show that the vast majority of students have at least basic reading abilities, with very few “non-learners”; indeed, only 1.7 percent of students—51 students total—scored 0 percent on this subtask, including 2.3 percent of comparison students and 1.6 percent of intervention students.

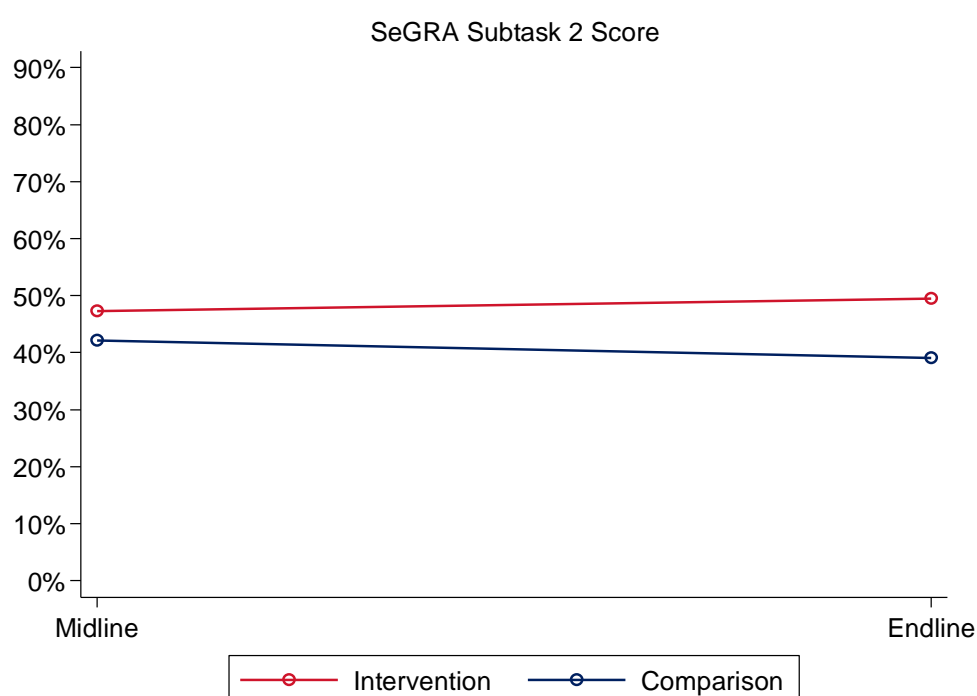
<sup>57</sup> The distribution of scores suggests that this subtask may have had a moderate ceiling effect, with over 10 percent of students scoring 100 percent on the subtask, including 13.8 percent of students at midline and 11.6 percent of students at endline.

## Subtask 2 Outcomes

For subtask 2, students were asked somewhat more advanced reading comprehension questions which generally involved interpretation or analysis of the text or which required the student to draw on information learned in school. For example, at endline, one question asked students to pick out a simile in the story; another asked them to describe the moral lesson learned from the story.

On average, scores for this subtask were lower than for subtask 1, showing its increased difficulty. The below figure shows the difference in calibrated subtask 2 results between rounds for the panel sample. Among these students assessed at both midline and endline, we find that endline scores were 2.3 percentage points higher, on average, among intervention students, and 3.0 percentage points lower among comparison students. While these findings were not significant, they again point to potentially positive programme impact.<sup>58</sup>

**Figure 10: Change in SeGRA subtask 2 score, panel students**



Examining learning bands in Table 19 below, we see that the number of non-learners consistently decreased at endline among all groups. Indeed, there were no non-learners among comparison students at endline. However, there was also a decline in the number of proficient learners. Among intervention students, more panel students were categorized as established learners for this subtask at endline compared to midline. This includes some learners who were previously categorized as proficient, as well as learners who improved from weaker learning bands. In contrast, comparison group students showed modest improvement in the proportion of proficient learners (1.3 percentage point increase), but also had a greater proportion of students classified in the lower “emergent learners” category.

<sup>58</sup> Further analysis and robustness checks are included in Annex 16.

**Table 19: SeGRA subtask 2 learning bands**

Panel Students	Intervention		Comparison	
	ML	EL	ML	EL
Non-learner (0%)	6.1	0.8	3.8	0.0
Emergent learner (1% - 40%)	42.9	36.1	52.8	57.9
Established learner (41% - 80%)	39.9	54.9	40.9	38.4
Proficient learner (81% - 100%)	11.2	8.2	2.5	3.8

Overall, consistent with results thus far, these findings, though not significant, show potentially positive programme impact. Unlike with subtask 1, for which many students were proficient, the vast majority of students are classified as either emergent or established learners with regards to the more advanced reading comprehension skills required for this subtask. This suggests that there may be a need to strengthen students' interpretation and analysis skills, building off of high levels of basic literacy already achieved.

### **Subtask 3 Outcomes**

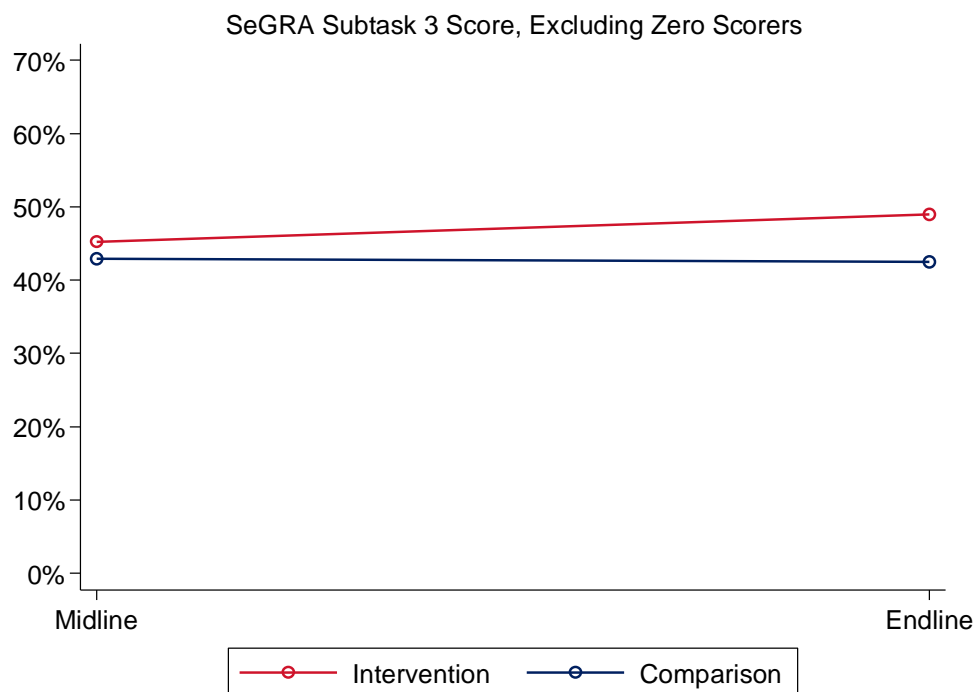
For subtask 3, students were given the following prompt: "At long last the long-awaited wedding day arrived..." They were then asked to write a 250- to 350-word short story based on this prompt. For all intervention students at endline, the mean subtask 3 score was 35.6 percent, compared to a mean subtask 3 score of 42.1 percent for midline intervention students (see Table 16).<sup>59</sup> Underlying this difference in average scores was a substantially different score distribution; among all students, at midline, only 1.3 percent scored 0 on this subtask, while at endline, 21.2 percent scored 0. This suggests that at endline, many students did not attempt this subtask, as students who attempted to write a story were generally given a score of at least 5 percent (1 out of 20 points). Follow-ups with girls who scored zero on this subtask reinforced this conclusion, and suggest that the zero scores are not representative of students' actual writing abilities; this is discussed more in Annex 16. In order to better understand programme results, we thus analyse results for non-zero scorers below.<sup>60</sup>

The below figure shows the difference-in-differences results for the panel sample excluding zero scorers.<sup>61</sup> We find an increase of 3.8 percentage points for intervention students at endline and slight decreases in average scores for comparison students, suggesting potential positive programme impact (although results are not significant).

<sup>59</sup> We do not calibrate subtask 3 results as there is no underlying reason to do so, as students were simply asked to write a short story, and thus there is no inherent difference in difficulty of the tasks.

<sup>60</sup> Results for all students, including zero scorers, are also included in Annex 16.

<sup>61</sup> This sample includes 307 intervention students and 123 comparison students.

**Figure 11: Change in SeGRA subtask 3 score, panel students, excluding zero scorers**

In our analysis of learning bands shown in Table 20, we first note that, as described above, substantially more students were classified as non-learners at endline than at midline. However, because very few students were non-learners for subtasks 1 and 2—which also involve reading and writing—it is unlikely that the substantial number of students who scored 0 percent for this subtask did so because they did not know how to write a story. Rather, these students did not reserve sufficient time to reach subtask 3, either due to time management or assessment administration issues (see Annex 16), and thus did not attempt this subtask.

Among the panel sample, a higher percentage of comparison students (though not significantly so) were classified as non-learners (22 percent) than intervention students (15 percent). This could occur if intervention students have relatively stronger reading comprehension skills than comparison students; if so, they may have been able to complete subtasks 1 and 2 more quickly, thus allowing them more time to attempt subtask 3. Second, if intervention students have learned better time management and examination skills, they may have been more likely to reserve time to attempt subtask 3.

**Table 20: SeGRA subtask 3 learning bands**

Panel Students	Intervention		Comparison	
	ML	EL	ML	EL
Non-learner (0%)	1.1	15.0	1.9	22.0
Emergent learner (1% - 40%)	45.6	34.4	47.2	42.8
Established learner (41% - 80%)	51.4	42.9	50.9	30.2
Proficient learner (81% - 100%)	1.9	7.7	0.0	5.0

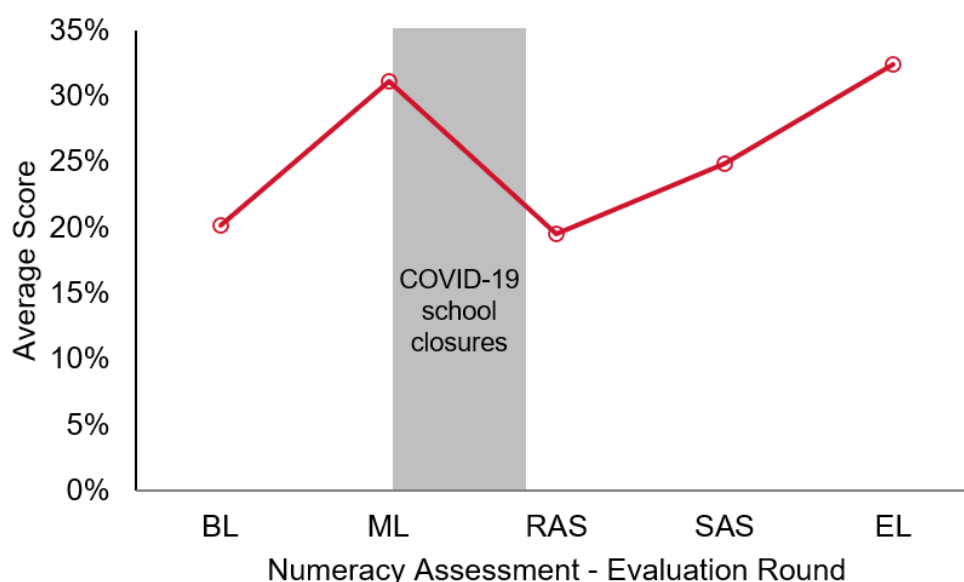
Outside of these patterns for non-learners, we also note that at endline, the percent of proficient learners increased, while the percent of emergent and established learners decreased. These patterns may suggest a divergence in learning outcomes, with some students performing very well and others falling behind. However, overall, these results continue to show a potentially positive programme impact on the most difficult literacy task assessed—the ability to write a creative, original, and grammatically sound short story.

## Numeracy Outcomes

Students who were assessed with the Secondary Grade Mathematics Assessment (SeGMA) were asked to attempt three subtasks. The first involved questions which mostly required multiplication, division, and knowledge of basic geometry. The second involved solving algebraic expressions. The third involved solving more complex algebraic expressions that needed to be derived from word problems. As with the SeGRA, each task was scored as a percent of items correctly answered out of the total possible score – 16 for the first subtask, 14 for the second, and 15 for the third, for a total of 45 possible points. The overall mathematics score was then calculated as the number of points scored out of the total possible points.<sup>62</sup>

The figures and table below show changes in overall EGMA and SeGMA and subtask scores for all intervention students assessed in the five evaluation rounds.<sup>63</sup> As with the SeGRA, this table shows a clear impact of COVID-19 on mathematics learning. Average scores increased substantially from baseline to midline, and then dropped to a lower-than-baseline average at RAS in conjunction with COVID-19 school closures. Mathematics learning levels have since gradually recovered to now reach a slightly higher level at endline than that at midline, with a 12.3 percentage point increase since baseline.

**Figure 12: Change in average EGMA and SeGMA scores across all rounds, intervention students**



<sup>62</sup> As with the SeGRA, we note that the midline and baseline assessments had different total numbers of points and different subtask weighting. This is addressed by reporting scores only as percentages and by calibrating results. However, scores are not calibrated in the initial two figures and tables showing results for all intervention students across all rounds, as not all assessments have been calibrated against each other (e.g., the endline assessment was not calibrated against the baseline or RAS assessments).

<sup>63</sup> i.e., the cross-sectional sample.

Figure 13: Change in overall SeGMA scores since baseline

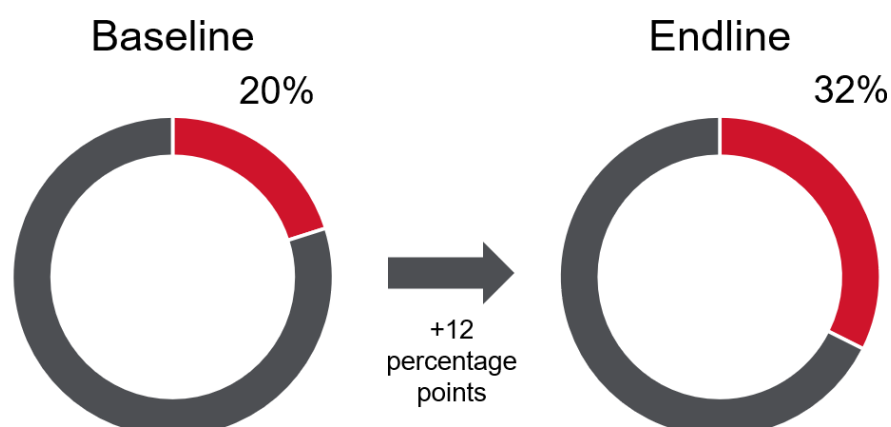


Table 21: Change in SeGMA scores across all rounds, intervention students

Task	BL (EGMA)	ML (SeGMA)	RAS (SeGMA)	SAS (SeGMA)	EL (SeGMA)
n	2,014	903	973	1,168	1,265
EGMA/SeGMA overall score	20.1	31.1	19.5	24.8	32.4
Subtask 1	-	48.2	34.5	43.7	51.8
Subtask 2	-	32.1	15.6	23.2	32.1
Subtask 3	-	13.0	8.4	7.5	12.0

We now use a difference-in-differences model to compare changes in SeGMA scores between intervention and comparison students to provide a better indication of programme impact. Using this model and calibrated overall scores,<sup>64</sup> we find a positive but not significant effect for the intervention group compared to the comparison group. Figure 14 shows that at endline, average calibrated scores for the comparison group had increased by around 4.8 percentage points, from 28.2 percent at midline to 33.0 percent at endline, while average scores for the intervention group had increased by around 7.0 percentage points, from 34.2 percent at midline to 41.1 percent at endline. This suggests that, as a result of the WWW programme, intervention students may have scored an average of 2.2 percentage points higher on the SeGMA than they would have in the absence of programme intervention.

<sup>64</sup> The endline SeGMA was found to be somewhat more difficult than the midline SeGMA. As such, calibration boosts endline scores.



**Figure 14: Change in SeGMA scores, panel students**

Robustness tests and further analysis of the cross-sectional sample are included in Annex 16. Results from this analysis broadly conform to the findings above.

Lastly, to further understand potential programme impact and mathematics learning levels, we now analyse learning bands for the SeGMA. Table 22 shows that for the calibrated SeGMA, there was an increase in the percent of students classified as proficient learners and a decrease in the percent of emergent learners among both intervention and comparison students. Overall, these results, combined with those discussed above, suggest a recovery in mathematics learning and slight improvement over pre-COVID levels.

**Table 22: SeGMA learning bands**

Panel Students	Intervention		Comparison	
	ML	EL	ML	EL
Non-learner (0%)	0.8%	1.1%	1.3%	1.9%
Emergent learner (1% - 40%)	57.7%	50.8%	67.9%	62.9%
Established learner (41% - 80%)	38.8%	39.3%	28.3%	30.2%
Proficient learner (81% - 100%)	2.7%	8.7%	2.5%	5.0%

We note that mathematics skills, as measured by SeGMA scores, were substantially weaker than literacy skills. Some of this gap may be due to examination fatigue, as students completed the SeGMA after the SeGRA. However, this gap also indicates a general need to continue strengthening mathematics skills, as most students seem to face more challenges to master these skills. Indeed, in the qualitative data, the vast majority of students stated that they struggle most with technical subjects including mathematics,

physics, and chemistry. However, in a positive note, most interviewed teachers expressed that girls' attitudes towards studying mathematics had begun to change in recent years. One teacher in Nairobi, for example, stated that many girls had built confidence in these subjects and now competed with boys for the top marks.<sup>65</sup> Another teacher in Mombasa noted that while girls' enrolment in maths tended to lag behind boys, the girls who did study the subject often excelled.<sup>66</sup> Overall, these findings suggest that the WWW programme may have had some impact on improving girls' confidence in studying mathematics.

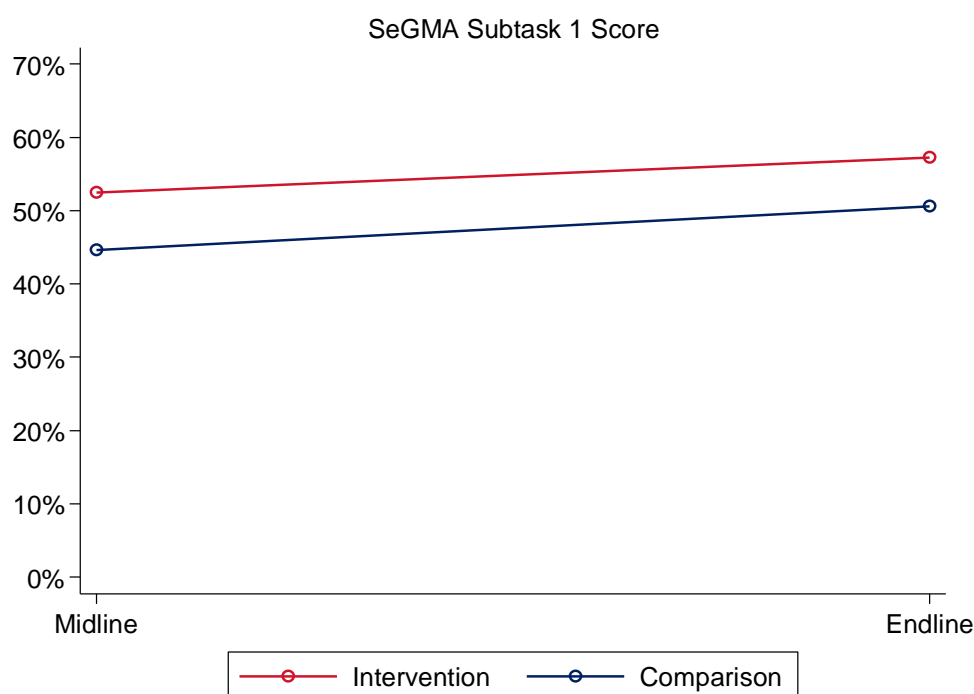
To better understand these dynamics, we now analyse results by subtask.

### Subtask 1 Outcomes

For subtask 1, students were asked questions involving multi-digit multiplication, division, fractions, or basic geometry. For all intervention students at endline, the mean uncalibrated subtask 1 score was 51.8 percent, compared to a mean subtask 1 score of 48.2 percent for midline intervention students, an increase of 3.6 percentage points (see Table 21). Calibrating results to account for a slight increase in subtask 1 difficulty at endline, we find a mean endline score of 53.3 percent, suggesting that among the cross-sectional sample, overall scores increased somewhat for intervention students at endline.

To better understand whether this is the case, we utilise the panel dataset for a difference-in-differences analysis. The below figure shows the results of this analysis; comparison students saw a slightly greater increase in average subtask 1 scores at endline compared to intervention students, although both groups saw increases. The improvement of comparison students relative to intervention students is not significant; the general trend, however, suggests modest improvements in basic mathematics skills post-COVID.

**Figure 15: Change in SeGMA subtask 1 scores, panel students**



<sup>65</sup> KII with teacher, Nairobi, Int. 29

<sup>66</sup> KII with teacher, Mombasa, Int. 30

We now analyse results by learning bands. For panel students, the below table shows that there were very few non-learners at both midline and endline, but also little change in the percent of non-learners across evaluation rounds. Notably, the percent of proficient learners declined among both intervention and comparison students from midline to endline, suggesting that at endline, fewer students were highly capable of completing the types of maths assessed in this subtask. There was, however, also an increase in the percent of established learners and a decrease in the percent of emergent learners. Given the widespread negative effects of COVID-19 on learning, this pattern is notable. Many learners risked being left behind due to the COVID-19 pandemic and related school closures; the data suggests that, at least for this subtask, this may not have occurred, showing resilience of students and past learning.

**Table 23: SeGMA subtask 1 learning bands**

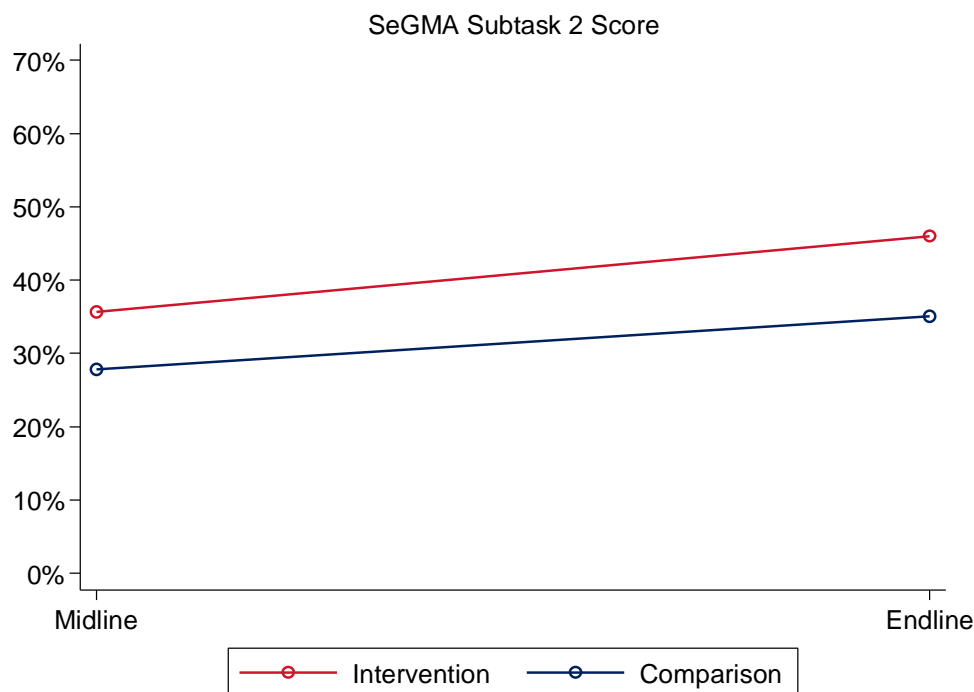
Panel Students	Intervention		Comparison	
	ML	EL	ML	EL
Non-learner (0%)	1.6%	1.6%	2.5%	1.9%
Emergent learner (1% - 40%)	34.4%	27.6%	41.5%	39.0%
Established learner (41% - 80%)	40.0%	49.2%	39.0%	43.4%
Proficient learner (81% - 100%)	24.0%	21.6%	17.0%	15.7%

Overall, these findings suggest that mastery of the more basic mathematics skills assessed in this subtask increased fairly evenly across intervention and comparison groups, potentially as a result of increased years of schooling and widespread recovery from COVID-induced learning losses.

### **Subtask 2 Outcomes**

For subtask 2, students were asked to simplify or solve algebraic expressions/equations. The figure below shows that while average scores were lower for subtask 2 than subtask 1, scores improved more substantially for this subtask. Indeed, calibrated scores for intervention students improved, on average, by 10.4 percentage points from midline to endline, while calibrated scores for comparison students improved by 7.1 percentage points. This shows not only an improvement in ability to solve more complex mathematics problems, but also a slight differential improvement in students affected by the WWW programme compared to those in non-intervention schools. While this difference is not significant, it is suggestive of potential positive programme impact.

**Figure 16: Change in SeGMA subtask 2 scores, panel students**



Looking at learning bands in the table below, we first note that there were a substantially larger number of non-learners for subtask 2 than subtask 1 in both evaluation rounds. The lowest number of non-learners, however, occurred for the intervention panel sample at endline; for this group, non-learners decreased by 10.1 percentage points at endline, while the comparison group had 8.8 percentage points fewer non-learners at endline. While this difference is not significant, it shows potential programme impact on learning amongst those students who were furthest behind at midline, a positive sign that the programme was able to target all learners, not just those who had already mastered some mathematics skills.

**Table 24: SeGMA subtask 2 learning bands**

Panel Students	Intervention		Comparison	
	ML	EL	ML	EL
Non-learner (0%)	19.7%	9.6%	26.4%	17.6%
Emergent learner (1% - 40%)	40.7%	34.4%	41.5%	42.1%
Established learner (41% - 80%)	29.5%	41.5%	26.4%	32.1%
Proficient learner (81% - 100%)	10.1%	14.5%	5.7%	8.2%

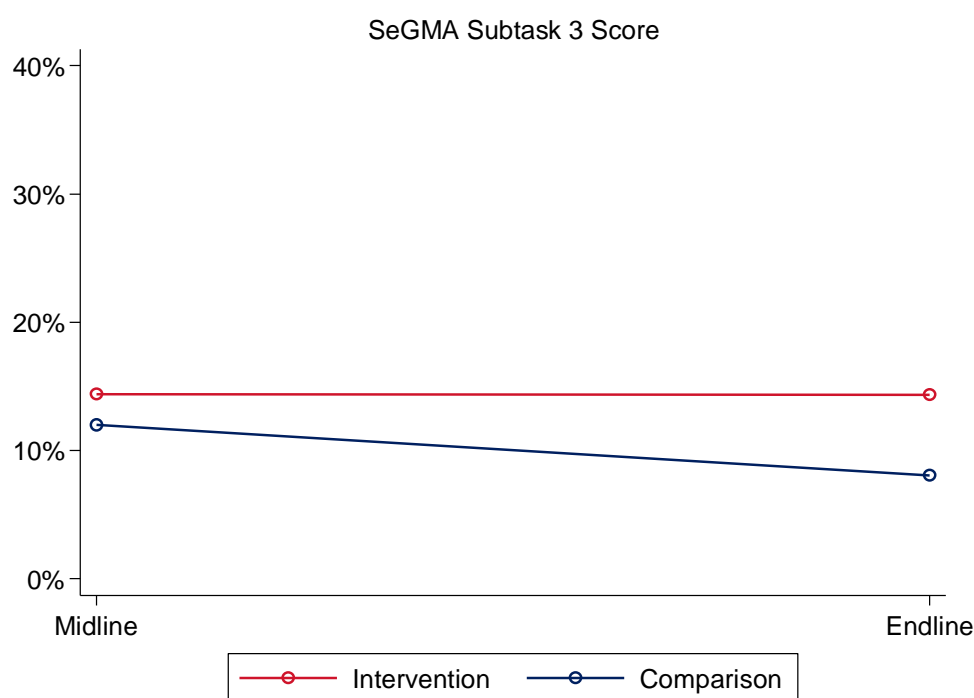
Additionally, unlike with subtask 1, we also find an increase in the percent of proficient learners at endline, with a slightly larger increase among the intervention group than the comparison group. Along with an increase in established learners, this suggests a general improvement in mathematics skills, especially among intervention students affected by the programme.

### Subtask 3 Outcomes

For subtask 3, students were asked to solve fairly complex mathematics word problems that required the students to develop their own algebraic equations and solve those equations.<sup>67</sup> Given that this subtask was the most difficult of the three, average scores were unsurprisingly lower than those for subtask 1 or 2. At endline, the average score for all intervention students was only 12 percent, a small decrease from the average midline score of 13 percent. For all comparison students, average scores similarly decreased by one percentage point, from 9 percent at midline to 8 percent at endline (see Table 21).

Among our panel sample, we still find that average scores decreased from midline to endline. However, notably, we find a substantially smaller decline among intervention students—for whom average scores decreased by just 0.1 percentage points—compared to comparison students, for whom average scores declined by 3.9 percentage points. This difference is significant, suggesting positive programme impact on mitigating COVID-induced declines in advanced mathematics skills for intervention students.

**Figure 17: Change in SeGMA subtask 3 scores, panel students**



When combined with the results from subtask 1 and 2, these results suggest that the WWW programme may have had relatively greater impact on higher-level mathematics skills as compared to “simpler” skills such as multiplication and division. We note that all questions in subtask 3 required students to read a word problem; in contrast, many questions in subtask 1 and 2 required only very limited reading of instructions for students to solve a problem. As such, it is possible that the relatively stronger subtask 3 outcomes of intervention students are related to improved literacy skills, including strengthened abilities both to read and to pull out the important information in a written passage.

<sup>67</sup> Scores for this subtask are not calibrated. During the calibration exercise, an insufficient number of students scored above 0 on this task to allow for an appropriate sample size for calibration.

Analysing learning bands further shows the relatively greater improvement of subtask 3 proficiency among intervention students; these students were relatively more likely to be classified as proficient or established at endline compared to comparison students, and less likely to be categorized as non-learners. We note, however, that among both intervention students, we see increases in both the percent of proficient learners and the percent of non-learners compared to midline. Indeed, we find that at endline, many students who were categorized as emergent learners at midline fell back into the non-learner category. These results may suggest a widening of the learning gap between students with relatively weaker mathematics skills and those with stronger mathematics skills.

**Table 25: SeGMA subtask 3 learning bands**

Panel Students	Intervention		Comparison	
	ML	EL	ML	EL
Non-learner (0%)	38.5%	53.0%	41.5%	64.2%
Emergent learner (1% - 40%)	51.6%	31.2%	51.6%	28.3%
Established learner (41% - 80%)	7.9%	12.3%	4.4%	6.9%
Proficient learner (81% - 100%)	1.9%	3.6%	2.5%	0.6%

Overall, however, the difference-in-differences results suggest a positive effect of the WWW programme on subtask 3 scores.

## STEM Outcomes

We now provide a brief analysis of STEM learning outcomes. The STEM assessment was administered for the first time during the SAS evaluation. As such, we have no comparison data for which a difference-in-differences analysis can be conducted to understand programme impact on STEM learning. Instead, below we provide a descriptive analysis of STEM results across the SAS and endline rounds. We note that subgroup STEM results are discussed in the subsequent section, and attitudes towards and enrolment in STEM subjects are discussed under a below section titled *Attitudes Towards STEM*.

The STEM assessment included three subtasks, one each for the topics chemistry, physics, and biology. Within each subtask, students were asked a range of questions about the topic which generally required the student to write a short answer, conduct basic maths, or draw a diagram. Each subtask was marked out of the total possible points, and the overall assessment was then marked as a percent score of the points obtained over the total points possible across all three subtasks.

The below table shows average STEM assessment results for all students assessed at SAS and endline. In both rounds, we find very low overall assessment scores of 11.7 percent at SAS and 11.6 percent at endline. These results suggest very limited mastery of STEM subjects that has not changed over the past year.

**Table 26: STEM assessment results, SAS and endline**

	Intervention		
	SAS	EL	Diff.
Number of respondents	410	383	-
<b>Overall score</b>	<b>SAS</b>	<b>EL</b>	<b>Diff.</b>
Mean score	11.7	11.6	-0.1
Standard deviation	6.2	8.4	-
<b>Subtask 1 (Chemistry)</b>	<b>SAS</b>	<b>EL</b>	<b>Diff.</b>
Mean score	15.2	12.8	-2.4
Standard deviation	10.8	10.3	-
<b>Subtask 2 (Physics)</b>	<b>SAS</b>	<b>EL</b>	<b>Diff.</b>
Mean score	10.1	7.3	-2.8
Standard deviation	7.3	8.8	-
<b>Subtask 3 (Biology)</b>	<b>SAS</b>	<b>EL</b>	<b>Diff.</b>
Mean score	9.9	14.7	4.8
Standard deviation	7.0	13.7	-

Examining results by subtask, at SAS, we find the highest scores for subtask 1, chemistry, with progressively decreasing scores for subtask 2, physics, and subtask 3, biology. At endline, we find the highest scores for biology (subtask 3), at 14.7 percent. Indeed, scores for subtask 3 increased from SAS to endline, while scores for subtasks 1 and 2 decreased. While these results may suggest some improvement in biology, there remains a clear need to further strengthen girl students' STEM skills.

#### **Box 1: Girl Students' Experiences with STEM Learning and Teaching**

In focus group discussions across all counties, girls stated that they were learning in school. This, in and of itself, is a positive finding after the widespread negative impact of COVID-19 on learning both in Kenya and worldwide. In this box, we highlight girls' learning experiences—especially of STEM subjects—and the impact of positive teaching practices encouraged by the WWW programme on learning.

In Tana River, girl students stated that they were learning in school because teachers were using effective practices such as group discussions, allowing students to ask questions, and explaining topics in new ways for students who were struggling. While the girls highlighted that chemistry, mathematics, and physics were the most difficult subjects for students, they also stated that they had become more interested in sciences in recent years because they enjoyed practicals, which they also found helpful for learning.<sup>68</sup>

In Turkana, girl students stated that while many students enjoyed mathematics and sciences, these were also the subjects where students performed the worst. When probed further, the students highlighted the importance of teacher retention and having quality teachers at all grade levels for learning, stating that there are “teachers that you understand when teaching some topics, but after

<sup>68</sup> FGD with girls, Tana River, Int. 41

some time you end up being taught by a different teacher who you aren't used to". Despite these challenges, the girls still stated that they had developed a positive attitude towards STEM subjects over the last year, and appreciated teachers' attempts to help struggling students.<sup>69</sup>

Lastly, in Nairobi, girls highlighted the importance of understanding subjects' real-life applicability in increasing their interest and engagement in the subject. For example, two girls stated that they enjoy biology and chemistry because they directly relate to their desired careers in clinical medicine and pharmacy.<sup>70</sup>

Overall, these findings suggest that while students may struggle with STEM subjects, they remain interested in pursuing those subjects. Furthermore, there is indication that STEM teaching may be improving through the use of practicals and other positive teaching practices. These findings are discussed further in subsequent sections on teaching practices and attitudes towards STEM.

## Subgroup Learning Outcomes

We now analyse learning outcomes by subgroup in order to begin to understand the interactions between marginalisation and learning (further explored in Annex 16). For this analysis, we primarily focus on the full sample of endline students who completed both learning assessments and girl surveys, and then analyse results for students for whom we have both a learning assessment and caregiver survey (a smaller sample). We also utilise the SAS sample to check for robustness of results and to provide an indication of changes over time.<sup>71</sup> Because we did not conduct girl surveys or caregiver surveys with midline students or with comparison students, we generally do not include results from this evaluation round or the comparison group in the below analysis, except where data exists from the learning assessments (e.g., for the analysis by county and school characteristics).

### Outcomes by County

First, we analyse change across evaluation rounds in learning assessment scores by county. We cannot utilise a difference-in-differences methodology for this analysis due to limited sample size. However, the analysis still provides information about how learning varies by county.

Table 27 shows uncalibrated SeGRA scores<sup>72</sup> by county, round, and intervention/comparison group.<sup>73</sup> We find that while learning scores generally declined between midline and endline in all counties, scores tended to decline more in the ASAL counties, Kilifi, Samburu, Tana River, and Turkana. In contrast, in Nairobi, average SeGRA scores improved for both intervention and comparison students at endline, after dropping at SAS. For Nairobi, this pattern meant that intervention students' average scores improved from being relatively on par with ASAL students at midline to being substantially higher than ASAL students at endline—the average score among Nairobi intervention students was 18.6 percentage points higher at endline than the average score among all ASAL intervention students, resulting in substantially higher average scores in urban slums than ASALs (56.0 percent in urban slums compared to 36.8 percent in ASALs).

<sup>69</sup> FGD with girls, Turkana, Int. 16

<sup>70</sup> FGD with girls, Nairobi, Int. 15

<sup>71</sup> We note that there were a large number of duplicate UIDs—679 in total—in the SAS learning assessment data. These observations were dropped from the analysis as we could not determine the reason for the duplication.

<sup>72</sup> Uncalibrated scores are used because the SAS was not calibrated to either ML or EL assessments.

<sup>73</sup> We note that sample size is very low for some groups, particularly in Mombasa, Samburu, and Turkana for midline and endline.



**Table 27: SeGRA scores by county, all students<sup>74</sup>**

ASALs	Intervention			Comparison		
	ML	SAS	EL	ML	SAS	EL
Kilifi	51.8 (n = 91)	27.3 (287)	41.1 (n = 242)	61.3 (n = 56)	-	45.1 (n = 39)
Samburu	57.0 (n = 48)	34.8 (148)	34.9 (n = 147)	25.4 (n = 26)	-	33.6 (n = 9)
Tana River	51.7 (n = 235)	29.2 (205)	37.7 (n = 249)	45.3 (n = 170)	-	35.7 (n = 68)
Turkana	50.6 (n = 71)	31.0 (100)	30.6 (n = 163)	62.1 (n = 41)	-	30.3 (n = 27)
All ASALs	52.1 (n = 445)	29.8 (n = 740)	36.8 (n = 801)	49.0 (n = 293)	-	37.1 (n = 143)
Urban Slums	ML	SAS	EL	ML	SAS	EL
Mombasa	66.8 (n = 54)	-	59.1 (n = 69)	48.6 (n = 58)	-	45.8 (n = 21)
Nairobi	54.9 (n = 360)	49.4 (421)	55.4 (n = 395)	46.1 (n = 186)	-	58.3 (n = 26)
All urban slums	56.4 (n = 414)	-	56.0 (n = 464)	46.7 (n = 244)	-	52.7 (n = 47)

This pattern of ASAL students falling behind students in urban slums holds, on average, when all counties are taken into account. Among all ASALs, we see a decline in SeGRA scores of 22.3 percentage points among intervention students; in contrast, in urban slums, scores declined by only 0.4 percentage points among intervention students. While this pattern is predominantly driven by students in Nairobi—for whom scores increased at endline—scores in Mombasa also decreased by fewer percentage points than scores for ASAL students. Overall, while these results are not indicative of programme impact, they are suggestive of a substantial and increasing level of literacy learning deprivation in ASALs compared to urban slums. This dynamic may have been influenced by the severe drought and rising levels of conflict experienced in ASALs over the past few years which has led to more frequent school closures in ASALs than in urban slums. Overall, this may suggest a need for expanded interventions in ASALs.

Looking at SeGMA scores in the table below, we notice a similar but less distinct pattern for ASALs as compared to urban slums. Among all ASAL counties, the average score for an intervention students decreased by 1.1 percentage points from midline to endline, while the average score increased by 4.4 percentage points for intervention students in urban slums. These results for urban slums, however, are driven entirely by increases in SeGMA scores in Nairobi. In fact, SeGMA scores declined in Mombasa by more than scores in any of the ASAL counties.

<sup>74</sup> We note that 76 midline students had missing or incorrectly labeled school names and no data was available on the county; these students are dropped from the sample.

**Table 28: SeGMA scores by county, all students**

ASALs	Intervention			Comparison		
	ML	SAS	EL	ML	SAS	EL
Kilifi	36.2 (n = 90)	18.4 (n = 287)	30.3 (n = 242)	39.3 (n = 56)	-	36.4 (n = 39)
Samburu	23.5 (n = 48)	25.2 (n = 149)	29.0 (n = 147)	16.1 (n = 24)	-	28.6 (n = 9)
Tana River	29.8 (n = 235)	21.1 (n = 214)	31.9 (n = 249)	18.8 (n = 170)	-	23.0 (n = 68)
Turkana	28.6 (n = 71)	27.8 (n = 100)	23.0 (n = 163)	33.3 (n = 41)	-	22.5 (n = 27)
All ASALs	30.2 (n = 444)	21.8 (n = 750)	29.1 (n = 801)	24.5 (n = 291)	-	26.9 (n = 143)
Urban Slums	ML	SAS	EL	ML	SAS	EL
Mombasa	47.8 (n = 54)	-	36.3 (n = 69)	15.1 (n = 58)	-	21.5 (n = 21)
Nairobi	31.5 (n = 349)	30.2 (n = 418)	38.4 (n = 395)	27.4 (n = 182)	-	39.8 (n = 26)
All urban slums	33.7 (n = 403)	-	38.1 (n = 464)	24.5 (n = 240)	-	31.6 (n = 47)

In Tana River, Kilifi, and Nairobi, we observe a pattern whereby average SeGMA scores declined at SAS and recovered somewhat at endline, showing the negative impact of COVID-19 and subsequent progress towards recovery. However, the pattern in Turkana notably deviates from this, with a continued learning loss at endline compared to both midline and SAS. It is possible that this continued deterioration in SeGMA learning scores could be due to the negative effects of drought and conflict on the county, leading to school closures. While we cannot draw definitive conclusions from this due to the lack of comparability of samples, overall, these findings suggest a need to further strengthen mathematics learning in Turkana county.

Lastly, for STEM assessments, we do not have data from midline and thus cannot conduct a similar analysis of change over time. As such, the below table disaggregates SAS and endline STEM results by county. We find that average scores declined in Tana River, Turkana, and Mombasa and increased in Samburu (although this county had an extremely low sample size at SAS) and Nairobi. Overall, at SAS, scores were approximately equivalent across the ASALs and urban slums. In contrast, at endline, average scores were significantly higher in urban slums than in ASALs, with scores in urban slums an average of 3.1 percentage points higher than those in ASALs.

**Table 29: STEM scores by county, all intervention students**

ASALs	Intervention		
	SAS	EL	Difference
Kilifi	-	11.8 (n = 105)	-
Samburu	8.0 (n = 11)	11.0 (n = 55)	3.0
Tana River	13.5 (n = 55)	11.2 (n = 62)	-2.3
Turkana	11.6 (n = 124)	7.3 (n = 57)	-4.3
All ASALs	11.9 <sup>75</sup> (n = 190)	10.6 (n = 279)	-1.3
Urban Slums	SAS	EL	Difference
Mombasa	10.7 (n = 71)	9.9 (n = 39)	-0.8
Nairobi	12.0 (n = 149)	16.7 (n = 65)	4.7
All urban slums	11.6 (n = 220)	14.2 (n = 104)	2.6

Overall, these disaggregated results for the SeGRA, SeGMA, and STEM assessments suggest relatively higher levels of learning deprivation in ASALs compared to urban slums. However, most of the relatively higher performance of urban slums is driven by results in Nairobi county, rather than those in Mombasa, which are often on par with results in the ASAL counties.

In addition to this analysis, the table below also shows learning assessment results for Kwale and Marsabit counties (as well as for intervention students in the four ASALs and two urban slums evaluated by the external evaluator). Due to the small number of tracked girls in these counties and logistical limitations, the external evaluator did not target these counties for the endline evaluation. However, in order to better understand learning outcomes, an internal EDT team conducted learning assessments with girls in Kwale and Marsabit following the fieldwork for the external endline evaluation. Because the assessments were not conducted by the evaluator, we do not include them in our analysis of aggregate results above or in other disaggregated results below, but address them here.

<sup>75</sup> Excluding Kilifi

**Table 30: Learning assessment results, Kwale and Marsabit**

County	SeGRA	SeGMA	STEM
Kwale	26.6 (n = 78)	25.0 (n = 78)	11.7 (n = 48)
Marsabit	48.0 (n = 42)	39.4 (n = 42)	24.4 (n = 35)
ASALs	36.8 (n = 801)	29.1 (n = 801)	10.6 (n = 279)
Urban slums	56.0 (n = 464)	38.1 (n = 464)	14.2 (n = 104)

Table 30 shows that SeGRA and SeGMA scores in Kwale fell substantially below the average for the six counties analysed above. Comparing to the urban slums, Kwale had an average SeGRA score almost 30 percentage points lower than the average for this region, an average SeGMA score 13.1 percentage points lower, and an average STEM score 2.5 percentage points lower. In contrast, results in Marsabit were higher than average results for the six counties above, with results exceeding average scores in ASAL counties. Compared to other ASALs, the average SeGRA score in Marsabit was 11.2 percentage points higher, the average SeGMA score was 10.3 percentage points higher, and the average STEM score was 13.8 percentage points higher.

### **Outcomes by Language**

We now analyse learning assessment results by whether students speak a majority or minority language. We define majority language as students who reported speaking English or Swahili at home, or who speak Samburu in Samburu or Turkana in Turkana. These students may have relative advantages while learning, as they either speak the language of instruction or the (likely) primary language of the instructor.

The below table shows that at both SAS and endline, SeGRA and SeGMA scores were significantly lower for students who speak a minority language compared to those who speak a majority language. The differences were particularly substantial for students at SAS, when minority language speakers scored, on average, 16.8 percentage points lower than majority students on the SeGRA and 9.5 percentage points lower on the SeGMA. There were no significant differences in STEM scores across majority and minority language speakers in either round.

**Table 31: Learning assessment scores by language, SAS and EL cohort 2 students**

SeGRA Score	Majority language	Minority language	Difference	P-value
SAS	45.4	28.6	-16.8	<0.001
EL	43.0	38.0	-5.0	0.001
SeGMA Score				
SAS	32.6	23.1	-9.5	<0.001
EL	32.1	28.4	-3.7	0.01
STEM Score				
SAS	11.3	12.3	1.0	0.16
EL	10.5	10.2	-0.3	0.78

While these patterns point to potentially substantial learning difficulties faced by minority language speakers, one promising finding comes from comparing SAS and endline results. At endline, majority language speakers' SeGRA and SeGMA scores had stagnated, with slight declines in both. In contrast, at endline, minority language speakers' SeGRA and SeGMA scores had increased substantially since SAS. While their scores remained lower than those of majority language speakers, this suggests that WWW programming (or some other effect) may have had relatively more benefits for minority language speakers, helping these marginalised students catch up to their peers.

### **Outcomes by Disability**

Next, we analyse learning outcomes by disability status. Before analysing results, we make two notes. First, the sample of disabled girls at SAS was very small, limiting our ability to draw conclusions about disabled girls' performance on learning assessments in this round. Second, at endline, girls in special schools (i.e., schools established to serve girls who are deaf and girls who are blind) were given less total time to take assessments than in previous rounds, and were not provided with adequate accommodations to be able to demonstrate their knowledge.<sup>76</sup> As such, below, we do not compare performance between girls with disabilities and girls without disabilities, as results from endline are likely to somewhat underestimate disabled girls' actual abilities.

In general, sample size was too small for most specific types of disability to allow for further disaggregation. However, at endline, 16 students reported having a vision-related disability and 46 students reported having a hearing-related disability. SeGRA, SeGMA, and STEM results for these students are reported below.

**Table 32: Assessment results for seeing- and hearing-disabled students, endline**

	SeGRA Score	SeGMA Score	STEM Score
Vision disability (n = 16)	42.2	26.3	.77
Hearing disability (n = 46)	18.8	14.6	2.7 <sup>78</sup>

Students with hearing disabilities had noticeably low results for all assessments, while students with vision-related disabilities tended to perform approximately on par with able-bodied students. While we again note that assessment administration issues could have driven some of these results, these results still suggest a particular need to focus on strengthening learning for students who have difficulty hearing.

#### **Box 2: Attitudes Towards Students with Disabilities**

To better understand how girls with disabilities are perceived and treated by students and teachers, in focus group discussions, girls were presented with a vignette about Zuleikha, a girl who needs a crutch to walk and who enjoys learning but feels shy at school and started school late. FGD participants were asked how Zuleikha's classmates and teachers might treat her and whether Zuleikha would face any problems at school. The vignette structure allowed girls to more openly express their opinions without fearing judgment for any negative statements.

<sup>76</sup> The expedited research design and data collection time periods limited our ability to verify the accommodations used in previous evaluation rounds and implement these accommodations. This shortcoming was noted during field quality assurance, and while corrective action was taken immediately, not all girls with disabilities were able to benefit from the appropriate accommodations at endline.

<sup>77</sup> Only eight students with a vision disability took the STEM assessment.

<sup>78</sup> N = 23

Across counties, many girls mentioned that Zuleikha would likely face difficulties due to her disability. In Nairobi, for example, girls in all seven FGDs stated that Zuleikha might face mockery from classmates, teachers, or the community due to her use of crutches. Girls in Tana River, Samburu, and Nairobi also stated that Zuleikha would be likely to drop out of school due to discrimination, as well as the challenges she would face from starting school late.

However, many girls also held positive attitudes towards Zuleikha and her potential to succeed in school. In Turkana, for example, girls in one FGD stated that the community would help Zuleikha acquire a wheelchair and medical care, and might help her find a sponsor for her education.<sup>79</sup> In Tana River, girls suggested that teachers might treat Zuleikha better than other students and encourage her to work hard because of her situation.<sup>80</sup> Furthermore, one girl in Nairobi stated that if Zuleikha was smart, despite her disability and her late start to schooling, she would be able to catch up in class. This girl stated that she herself knew a girl who had started very late in school, but now led the class.<sup>81</sup> Overall, while these responses to the vignette suggest that girls with disabilities still face many challenges to learning, the challenges are neither universal nor insurmountable. While it remains important for programming to support students with disabilities, these students do have opportunities to learn and, at least in some cases, are supported by classmates, teachers, and communities.

### **Outcomes by Household Characteristics**

We now examine learning outcomes across three household characteristics: whether the student belongs to a female headed household, whether the student is an orphan (neither parent is alive, or only one parent is alive),<sup>82</sup> and the extent to which the student is responsible for housework and other non-school duties, such as agricultural work or helping with a business. All three of these aspects may be expected to be associated with lower learning outcomes, the first two because of their association with marginalisation and the last because it reduces the amount of time the student can dedicate to learning.

Table 33 shows learning assessment score by the gender of the head of household (HoH). At SAS, there were no significant differences in assessments scores by the HoH gender. However, at endline, scores were significantly lower for the SeGRA and SeGMA for girls in households with female heads compared to those with male heads. While the difference in SeGRA and SeGMA scores was not very large—only 4 percentage points for the SeGRA and 3 percentage points for the SeGMA—this finding still suggests that girls with female HoHs may face relatively more marginalisation and challenges to learning.

**Table 33: Learning assessment scores by HoH gender, SAS and EL cohort 2 students**

SeGRA Score	Male HoH	Female HoH	Difference	P-value
SAS	35.5	36.5	1.0	0.61
EL	42.1	37.9	-4.2	0.03
SeGMA Score				
SAS	27.4	26.2	-1.2	0.55
EL	31.4	28.4	-3.0	0.08

<sup>79</sup> FGD with girls, Turkana, Int. 16

<sup>80</sup> FGD with girls, Tana River, Int. 41

<sup>81</sup> FGD with girls, Nairobi, Int. 15

<sup>82</sup> We choose to define orphan in this manner for the purposes of this analysis as the sample size of both-parent orphans was very small at endline.

STEM Score				
SAS	11.7	11.8	0.1	0.97
EL	10.4	10.1	-0.3	0.77

Examining now by orphan status, we similarly find that at endline, SeGRA and SeGMA scores were significantly lower for students who had only one or no parents alive compared to students who had both parents alive. At SAS, however, there were no significant differences in assessment scores, nor are there significant differences in STEM scores at endline (although this may be because overall scores for both subgroups are extremely low). This finding again suggests that household factors have a tangible effect on learning outcomes, although we note that the relationships found in this section may be due to confounding factors: For example, students who are orphans may also have lower socioeconomic status and thus be more likely to work in their free time, reducing the amount of time they have to study. These dynamics are further explored in the section on Predictors of Learning Outcomes in Annex 16.

**Table 34: Learning assessment scores by orphan status, SAS and EL cohort 2 students**

SeGRA Score	Both parents alive	Orphan	Difference	P-value
SAS	35.8	35.3	-0.5	0.76
EL	42.4	38.3	-4.1	0.02
SeGMA Score				
SAS	27.8	25.7	-2.1	0.20
EL	32.1	26.9	-5.2	0.001
STEM Score				
SAS	12.1	11.4	-0.7	0.28
EL	10.7	9.4	-1.3	0.14

Lastly, we look at learning assessment scores by level of household responsibility. We do not have comparable data from SAS for this dynamic, and thus only examine results for endline students. At endline, students were asked how much time they typically spend on a normal school day to do tasks unrelated to school and learning—caring for family members, doing housework, fetching water, doing agricultural work, or helping with a family business. In the table below, we compare learning assessment results for girls who reported spending less than 2 hours per day on these tasks versus those who reported spending more than 2 hours per day.

The table below shows that at endline, students who reported higher levels of household responsibilities had significantly lower SeGRA and SeGMA scores, and lower—although not significantly so—STEM scores. The difference was particularly acute for the SeGRA, for which students who reported doing more than two hours a day of chores scored, on average, almost 7 percentage points lower than students with smaller chore burdens. As expected, this shows that students who have greater household responsibilities, and thus less time to dedicate to studying and learning, face worse learning outcomes.

**Table 35: Learning assessment scores by household responsibilities, EL cohort 2 students**

	Less than two hours per day	More than two hours per day	Difference	P-value
SeGRA Score	42.6	35.8	-6.8	<0.001
SeGMA Score	31.6	27.7	-3.9	0.02
STEM Score	10.5	9.9	-0.6	0.55

Overall, this analysis shows that the characteristics of a student's household have a significant impact on that student's learning. It is important to again emphasise, however, that the findings above may be due to confounding factors, rather than the specific dynamic analysed. For example, students with greater household responsibilities, students who are orphans, and students who have a female HoH may also come from families with less income. If this is the case, it may be the deprivation associated with poverty, rather than the factors directly observed above, that truly has an impact on learning.

### **Outcomes by School Characteristics**

Lastly, we examine learning outcomes by school characteristics. We focus on two characteristics: teacher attendance and student-teacher ratio. We expect learning outcomes to be better in schools where teachers attend more frequently and in schools where the student-teacher ratio is lower (i.e., class sizes are smaller).

Examining first by teacher attendance, we utilise a variable from the girl survey in which respondents were asked to state their level of agreement or disagreement with the statement "My teachers are often absent from class/lessons".<sup>83</sup> We note that at SAS, almost every girl surveyed—over 98 percent—stated that they agreed or strongly agreed that their teachers were often absent. While this is an important finding in and of itself, it does not lend itself to subgroup analysis, as there is no "infrequently absent" subgroup to analyse. As such, the table below reports results only for endline students.

For endline students, as expected, we find significantly lower SeGRA and SeGMA scores for girls who report that their teachers are often absent. STEM scores are also lower, although not significantly so. These findings emphasise the importance of teacher attendance on learning, a dynamic which will be further explored in subsequent sections of the report.

**Table 36: Learning assessment scores by teacher absences, EL cohort 2 students**

	Not often absent	Often absent	Difference	P-value
SeGRA Score	42.2	34.9	-7.3	0.003
SeGMA Score	31.4	25.9	-5.5	0.01
STEM Score	10.6	9.2	-1.4	0.20

Looking now at student-teacher ratio, we run a regression analysis to understand the relationship between this variable and learning outcomes. The below table shows that for students at SAS, we find a significant and negative relationship between SeGRA and SeGMA scores and student-teacher ratio. In

<sup>83</sup> As teachers and head teachers have incentives to overreport their attendance, this variable provides the most reliable measurement of teacher attendance.



other words, as student-teacher ratio increases—i.e., as class sizes grow larger—SeGRA and SeGMA scores tend to decrease.

**Table 37: Learning assessment scores by student-teacher ratio, EL cohort 2 students**

	EL		SAS	
	Coefficient	P-value	Coefficient	P-value
SeGRA Score	0.08	0.26	-0.8	<0.001
SeGMA Score	0.02	0.72	-0.4	<0.001
STEM Score	-0.04	0.34	-0.03	0.42

Overall, these findings emphasise that school characteristics do, unsurprisingly, have an impact on students' learning outcomes. Annex 16 further examines these findings in a predictive analysis which suggests that the number of school clubs is a significant predictor of learning outcomes. These dynamics are investigated further in subsequent sections of the report which analyse aspects such as school governance and management, attendance, and the quality of teaching.

## Reflections on Learning Outcomes

In this section, we have first found substantial learning losses—likely due to COVID-19, as has occurred globally—which assessed girl students are still recovering from. These learning losses have resulted in lower SeGRA scores at endline than at midline among both intervention and comparison groups, and similar SeGMA scores at endline compared to midline. However, endline scores for both learning assessments have still improved over baseline values and have now returned to an upward trajectory following COVID.

Despite COVID-induced learning losses, we found significant positive programme impact on SeGMA subtask 3 scores (complex mathematics word problems). We did not find a significant impact of the programme on other learning outcomes. However, we did find that intervention students had, on average, higher SeGRA and SeGMA scores at endline than would be expected given the results of comparison students. These findings were robust to the inclusion of control variables for age and school-level factors within subsequent regression models. In other words, while the evidence is only indicative, it is likely that the WWW programme had positive impact at mitigating the negative effects of COVID-19 on learning, with overall SeGRA results<sup>84</sup> 5.5 percentage points higher among intervention students than expected given results of comparison students, and overall SeGMA results 2.2 percentage points higher than expected.

For STEM outcomes, we did not have comparison data with which to analyse programme impact. However, we find very low STEM assessment results at both SAS and endline, with no overall improvement in scores across rounds, potentially due to the limited time (one year) between these evaluation rounds. At endline, scores were highest for biology and lowest for physics.

Analysing learning outcomes by subgroup, we found that at endline, ASAL students had begun to fall behind in comparison to students in urban slums. However, this result was predominantly driven by the relatively strong performance of students in Nairobi. We additionally found that minority language speakers, students with a female HoH, orphans, and students with greater household responses had significantly lower learning assessment results. Examining school-level characteristics, we found that

<sup>84</sup> Excluding subtask 3 due to high levels of zero scorers.

teacher absences were significantly correlated with lower SeGRA and SeGMA results, as was large class sizes.

Our predictive analysis, however, found no significant correlations between individual demographic characteristics, such as language and disability, and learning outcomes. Instead, we found a positive and significant relationship between life skills and learning outcomes. While these variables are somewhat cross-correlated—for example, disabled students may have lower self-confidence, and thus score lower on our life skills index—this suggests that interventions to strengthen students’ confidence, time management, and other skills have important effects on the student in school and beyond.

At the school level, we found significant relationships between the number of school clubs and learning outcomes. This may be a proxy for the level of resources available to schools and students, or school clubs may benefit students’ learning by increasing life skills. In part, this finding suggests the important of co-curricular activities for learning, as well as interventions increasing school resources.

## Transition Outcomes

### Summary of Findings

1. Transition rates increased to 96 percent at endline, from 94 percent at baseline and 93 percent at midline. While this narrowly misses the programme’s ambitious transition goal of 99 percent, it is still a notably high result, particularly considering widespread school closures during COVID-19. These high results despite COVID may have been due to both programme and government interventions.
2. Given the context of Kenya and the impact of COVID-19, a goal of 99 percent transition was likely overly ambitious.
3. Contributing to high transition rates, girls have very high aspiration levels regarding education, with over 95 percent stating that they wish to continue in education until they meet their goals. There has also been an increase in WWWW girls’ awareness of alternative learning paths (ALPs), such as TVET and apprenticeships, as a viable transition option.
4. Repetition rates, however, were relatively high, with 17.6 percent of girls repeating a grade, likely due in part to the impact of COVID on learning.
5. Economic barriers remain the biggest challenge to transition, as many families lacked the ability to pay for their girl’s education. This highlights the importance that WWWW scholarships and cash transfers have played in strengthening transition.
6. Transition rates were very similar across all subgroups. Girls in ASALs, however, had higher grade repetition rates than girls in urban slums, as did girls in female-headed households.

Transition in WWWW is understood as progression into and through successive grades of formal and non-formal education or vocational training or into safe, fairly paid employment or self-employment. The goal of this project was not only to improve the learning outcomes of girls but also for girls to transition through the education system to go on to have productive and positive lives. Therefore, the project considers three possible pathways of successful transitions:

- **Pathway 1:** Transition from primary to secondary school (this is the preferred pathway)
- **Pathway 2:** After having dropped out of school enrolling into an alternative learning pathway
- **Pathway 3:** After having dropped out of school enrolling back into primary or secondary school

Table 38 presents the transition rates per sample and pathway, as well as the aggregated overall transition rate. Each column in the table presents the transition rate of our sample to each pathway. The final two columns present the total transition rate per sample and the total sample size. In the survey, girls were asked which type of learning facility they currently attended; this included a response for out-of-school girls.

**Table 38: Transition rate by transition pathway**

Sample	Pathway 1 & 3	Pathway 2	Unsuccessful transitions	Total	Sample Size
SAS sample	99.7%	0.0%	0.3%	99.7%	933
Kwale and Marsabit sample	61.0%	5.0%	26.6%	66.0%	94
Alternative Pathways sample	0.0%	100.0%	0.0%	100.0%	49
Total	91.8%	5.1%	3.9%	96.1%	1,076

The aggregate transition rate across all the sample and transition pathways is 96.1 percent. The highest transition rates by far are that of the SAS sample (i.e., girls surveyed at both SAS and endline), at 99.7 percent. The difference between these samples is expected as the samples targeted girls who had followed different transition pathways.

Noticeably, the Marsabit and Kwale girls had a much lower transition rate than that of the SAS sample at 66 percent. This may be due to several reasons. Firstly, the Marsabit and Kwale sample was based on the midline sample (as the SAS did not include girls from these two counties). The midline took place a year and a half before the SAS (and two and a half years before the endline). This is a longer period of time, where more girls may have dropped out than is the case for the SAS sample. Secondly, the data collection methods may have also influenced these results. The Marsabit and Kwale girls were surveyed over the phone whereas the SAS sample girls were surveyed in person. Since this sample was not surveyed by going in person to the schools, we were more likely to contact out-of-school girls and girls who had enrolled in an ALP course.

Overall, when looking at the SAS sample as our primary measure of transition, the transition outcome target for the endline was met.<sup>85</sup> In addition to WWW interventions, another factor which may have contributed to the high transition rate is the Kenyan Government's push for 100 percent transition from primary to secondary since 2020. This has been continued and expanded as explained by one government official. She explained that a policy is currently being implemented which targets students who have finished Class 8 to help them transition<sup>86</sup> into secondary school. These enrolment drives are carried out by stakeholders who find and talk to children who have not reported to secondary school to try to address the issues these children are facing. The alignment between this policy and WWW interventions is positive, as it suggests that progress made through WWW will be continued through government initiatives.

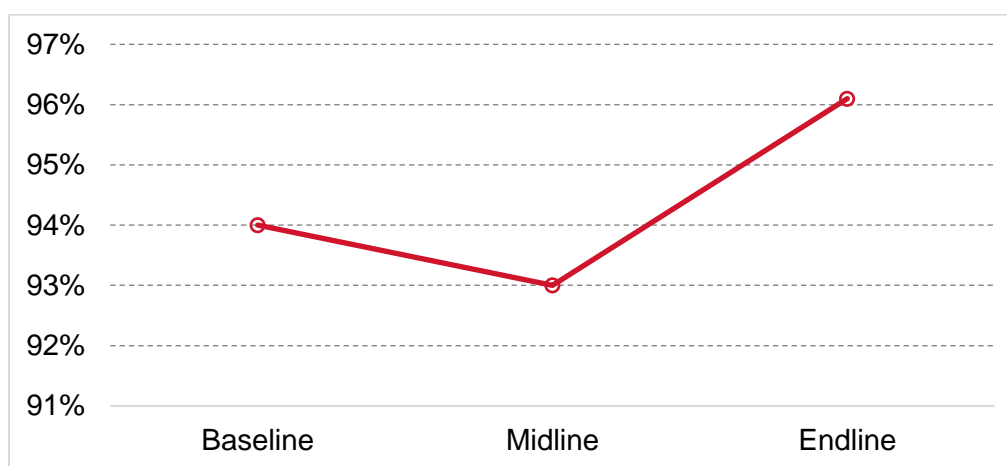
Figure 18 below shows the transition rate from the baseline to endline evaluations. We find that the transition rate has increased between the midline and endline evaluation from 93 percent to 96 percent.

<sup>85</sup> The SAS sample may overestimate transition, as only one year passed between the SAS and endline surveys and because the SAS sample is mostly comprised of girls that we were able to track in schools, and who had thus successfully transitioned. These limitations are discussed more in Annex 17.

<sup>86</sup> KII with Government official, Nairobi, Int 87.

While the endline result did not meet the programme's target, a transition rate of 96 percent is still notably high and represents an improvement over both baseline and midline.

**Figure 18: Transition rates between evaluation rounds**



*Note: SAS study sample contained only in school girls and, therefore, no transition rate is available for that evaluation round*

A second indicator of the successful transition of girls is to analyse the grade-to-grade progression rate of girls. Since the SAS took place one year before the endline evaluation we assume that students in the cohort sample should have progressed through one grade at school.<sup>87</sup> Table 39 presents the progression rate of girls by grade using the Cohort 2 sample.

**Table 39: Progression rate by grade**

Grade	Sample Size	Number of Girls who repeated	Progression rate
Class 7	142	4	2.8%
Class 8	348	120	34.5%
Form 1	77	5	6.5%
Form 2	101	5	5.0%
Form 3	98	10	10.2%
Form 4	68	3	4.4%
Total	834	147	17.6%

This table shows that there were a notable number of girls, 17.6 percent (147 out of 842), who repeated the grade they were in the previous year. The large majority of the girls who repeated grades were those girls in Class 8. This could be attributed to national examinations before joining secondary school; girls may not have been able to get a high enough grade in these examinations to graduate and move on to

<sup>87</sup> It is important to note that the data contained some inaccuracies as some girls had progressed more grades than expected or had moved back grades in ways which did not make sense. Progression of more than one grade could in part be explained by the compressed school calendar, which meant that some students covered more than one academic year in the course of one calendar year; however, this does not explain all anomalies in the data. Therefore, this limits our ability to make claims on the grade-to-grade progression of girls in the project, but the data gives an indication nonetheless.

secondary school. Furthermore, this repetition rate is higher than that reported at midline (9.2 percent) which could be due to learning loss during COVID. However, it is also notably positive that although the data suggest that a high number of girls repeated grades, this did not encourage them to drop out of school. This could be for a number of reasons including family support and encouragement to stay in school, or girls' self-confidence and ambition to continue their education. As presented later on in this section, girls reported very high ambitions regarding their education. Of the girls in our sample, 97 percent of them reported they would transition through the education system until they meet their goals and, therefore, girls may see repetition as an obstacle to overcome rather than a reason to leave school altogether.

## Subgroup Transition Rates

The previous subsection discussed transition in terms of the different transition pathways. However, in this section we investigate how transition rates differ across demographic groups in the sample. In Annex 17, we also describe the results of a predictive analysis of the relationship between student- and school-level characteristics and transition rates.

The table below represents the transition rates by subgroups using the SAS-ML sample. Due to the extremely high transition rate, there is little information on transition rates when split into subgroups. The only noticeable characteristic is that all three girls who did not transition in the SAS sample had no father. Notably, all 65 girls with disabilities in the SAS samples had successfully transitioned.

**Table 40: Transition rates by subgroup**

Characteristics	Successful Transition		Unsuccessful Transition	
	N	%	N	%
Overall	834	99.7%	3	0.3%
<b>Regions</b>				
ASAL	449	99.6%	2	0.4%
Urban Slums	385	99.7%	1	0.3%
<b>County</b>				
Kilifi	175	100.0%	0	0.0%
Mombasa	30	100.0%	0	0.0%
Nairobi	263	99.6%	1	0.4%
Samburu	115	99.1%	1	0.9%
Tana River	125	100.0%	0	0.0%
Turkana	126	99.2%	1	0.8%
<b>Girls Characteristics</b>				
Girl has any disability	65	100.0%	0	0.0%
Girl is orphaned (no father)	150	98.0%	3	2.0%
Girl who is orphaned (no mother)	28	100.0%	0	0.0%
Girl is double orphaned (no parents)	17	100.0%	0	0.0%
Minority language speakers	274	100.0%	0	0.0%

Girl who spends more than a quarter of a day doing HH chores	47	98.0%	1	2.0%
Caretaker Characteristics				
HH less than secondary education	122	99.0%	1	1.0%
Caretaker unemployed	85	99.0%	1	1.0%
Female headed households	150	98.0%	3	2.0%

When analysing the grade-to-grade progression by subgroups we find that girls in ASALs had a higher repetition grade than those in urban slums (20 percent and 15 percent respectively) and that girls in Nairobi had the highest repetition rate by a 10-percentage point margin at 30 percent. Furthermore, we find that women headed households also have a higher repetition rate than male headed households by 5 percentage points (21 percent and 16 percent respectively). All other sub-groups are below the average repetition rate. This includes the subgroup of disabled girls, only 5 of which had repeated grades (7.7 percent).

It is useful to note that caregivers highlighted that one of the biggest challenges for girls to transition are economic issues within households. This is highlighted by the fact that 83.6 percent of caregivers find it difficult to afford girls education. This was further discussed by respondents, including teachers and government officials, who highlighted one of the biggest challenges to girls transitioning remains the high levels of poverty in communities which limits parents' ability to afford school materials or to pay school fees.<sup>88</sup> This is despite the government implementing a number of initiatives including capitation grants and provision of school materials in government owned schools.<sup>89</sup> The following quote summarises this challenge:

*...The greatest barriers or challenges in enrolling or staying in school is just because of the current economy and most of our parents, most of here where we are, are really struggling to meet their daily needs, so meeting their daily needs and then [paying] school fees becomes a challenge, so most of these girls are really struggling in staying in school.<sup>90</sup>*

This also highlighted the importance of project activities such as scholarships, provision of school materials, and cash transfers for the transition of girls in these communities. For example, we asked girls if they thought they would be able to continue their education without support from the programme and 34 percent of the girls said no. Within this group, the most common reason for not being able to continue their education was not having the money to support their education in the next academic year. The importance of these WWW activities for the transition of girls was also mentioned in qualitative discussion with girls and programme staff, who claimed that the programme had helped girls remain in school by paying for their school fees and providing books and other learning materials.<sup>91</sup>

Respondents also highlighted that another of the biggest challenges to transition in communities remain cultural practices and negative attitudes to education. Teachers and government officials highlighted that it is sometimes difficult to engage with caregivers to support their daughter's education as they do not understand the importance of education and prefer that their daughters begin working at a young age to

<sup>88</sup> KII with Teacher, Nairobi, Int 1. KII with Teacher, Nairobi, Int 2. KII with MOE Official, Kilifi, Int 3. KII with Teacher, Tana River, Int 6. KII with Teacher, Nairobi, Int 7. KII with Teacher, Mombasa, Int 8.

<sup>89</sup> KII with Government official, Nairobi, Int 87.

<sup>90</sup> KII with Teacher, Nairobi, Int. 11.

<sup>91</sup> FGD with girls, Mombasa, Int 9. FGD with girls, Nairobi, Int 10. FGD with girls, Kilifi, Int 13. FGD with girls, Nairobi, Int 14. FGD with programme staff, Int 88.

help bring money to the household.<sup>92</sup> Furthermore, girls becoming pregnant at an early age was also cited as another issue that can hinder the transition of girls.<sup>93</sup> Early pregnancy was often cited as a barrier to transition as girls would not only be expected to drop out to look after their children, but they would also face stigmatization from their peers, community, and family.<sup>94</sup> This means that these children would then not only not have the resources to re-join school, but would also not want to due to the discrimination they would face at school. However, in FGDs with girls, respondents suggested that some progress has been made in this regard, as girls shared a number of stories where their friend or classmates had become pregnant but were convinced to come back to school and had been able to transition through education.<sup>95</sup>

Overall, this section suggests that there no specific sub-groups which had particularly high transition or repetition rates. It is important to reiterate that this is to be expected when the sample has such a high transition rate. This sub-section also notes that, as in previous evaluations, the high levels of poverty in communities remain the biggest challenge for girls to transition through education, largely because of the difficulties parents face to be able to pay school fees. This not only highlights the importance of the bursaries supplied by the programme but also the difficulties which girls may face now that the programme is coming to a close. Therefore, it would be recommendable that the government continue bursaries or financial help to those families which struggle to pay school fees and that EDT offer communities a list or database of organizations and programmes where families can go for help when facing financial difficulties.

Lastly, we note that while the subgroup analysis finds few differences in transition or repetition rates by student characteristic, our predictive analysis, discussed in detail in Annex 17, finds a significant relationship between some school characteristics and repetition rates. Higher teacher attendance, higher numbers of trained teachers, and lower number of lessons taught by teachers per week are all significantly correlated with lower repetition rates. This finding emphasizes the importance of a strong school environment for learning.

## Alternative Transition Pathways

One of the goals of the WWW programme was to not only improve the transition of girls through formal learning pathways but also through alternative learning pathways such as TVET, catch up centres, and apprenticeships. Girls often face limited alternative options and low support for these options. The WWW programme aimed to improve access to these ALPs through activities such as raising awareness of the options and working with the private sector on alternative pathways.

In the girl's survey with the SAS cohort, a number of statements were read to girls to gauge their levels of aspirations as well as their awareness on alternative learning pathways. The table below displays the items that were used to question the girls' aspirations levels and alternative pathway knowledge, showing results for the RAS, SAS, and endline. The percentages in this table represent the percentage of girls who 'agree' or 'highly agree' with the statements.

---

<sup>92</sup> KII with Teacher, Turkana, Int 4. KII with government official, Nairobi, Int 87.

<sup>93</sup> KII with Teacher, Turkana, Int 5. KII with Teacher, Mombasa, Int 8. KII with Teacher, Samburu, Int 12.

<sup>94</sup> FGD with girls, Nairobi, Int 21. FGD with girls, Turkana, Int 22. FGD with girls, Nairobi, Int 15. FGD with girls, Tana River, Int 23.

<sup>95</sup> FGD with girls, Nairobi, Int 21. FGD with girls, Tana River, Int 17. FGD with girls, Tana River, Int 24. FGD with girls, Samburu, Int 25. FGD with girls, Mombasa, Int 20.

**Table 41: Alternative transition pathways awareness**

Statement	Region	Wave			Change (SAS to EL)
		RAS	SAS	Endline	
I would like to continue studying/attending school after this year	Urban Slums	96.9%	99.3%	99.0%	-0.3%
	ASALs	97.4%	98.9%	99.1%	0.3%
	Total	97.2%	99.1%	99.1%	0.0%
I am aware of the different transition pathways in education system	Urban Slums	80.2%	81.7%	90.2%	8.5%
	ASALs	86.6%	91.9%	91.8%	-0.1%
	Total	83.8%	87.8%	90.9%	3.1%
I am willing to transition throughout the education pathways until I meet my goal	Urban Slums	93.6%	94.2%	96.8%	2.6%
	ASALs	95.3%	96.1%	95.8%	-0.3%
	Total	94.6%	95.3%	96.4%	1.1%
I am aware about the alternative pathways of education (TVETs, Apprenticeships, etc)	Urban Slums	59.5%	66.4%	77.9%	11.5%
	ASALs	77.8%	82.8%	75.9%	-6.9%
	Total	69.8%	76.2%	77.0%	0.8%
I would like to continue learning a vocation or trade	Urban Slums	76.2%	65.7%	64.7%	-1.0%
	ASALs	85.2%	65.8%	63.4%	-2.4%
	Total	81.3%	65.8%	64.1%	-1.7%

The table shows that findings have remained positive across evaluation rounds, especially regarding desire to continue attending school and willingness to reach educational goals. The largest changes can be observed in the transition pathways awareness statements. Firstly, we see an 8.5 percentage point increase in the awareness of the different transition pathways in urban slums and an increase in the awareness of alternative pathways in urban slums of 11.5 percentage points. Although in previous rounds ASAL girls seemed to be more aware of their transition options, this has evened out and is now almost equal in ASALs and urban slums. This suggests that among all WWW girls, awareness of transition pathways has increased.

The table also highlights, however, the continued decrease in the girls who would like to continue learning a vocation or trade. Although the percentage of girls who agreed they would like to continue a vocation or trade was quite high at RAS, this sharply declined at SAS and has shown a further decline at the endline by 1.7 percentage points. Although these results may be affected by social desirability bias,<sup>96</sup> this trend suggests that although awareness of ALPs has increased, girls find these alternative pathways less appealing than continuing in traditional education or high-skilled employment. Overall, this suggests that

<sup>96</sup> Social desirability bias refers to respondents' tendency to underreport socially undesirable attitudes and behaviours. In this case, pursuing a trade or vocation may be viewed as socially undesirable—a pathway for "failures" who are unable to continue to university. As such, girls may be likely to underreport their desire to pursue these pathways. Despite this limitation, we note that social desirability bias is likely to affect both SAS and endline results in similar ways; as such, the underlying trend of a decrease in desire to pursue a vocation/trade at endline remains valid.



while increasing awareness of ALPs is useful, it is not necessarily linked to desirability of ALPs, and that more work may be needed to improve perceptions of ALPs.

During interviews with a number of respondents, the increase in awareness of TVET was also noted. Government officials highlighted that there had been an increase in TVET enrolment of girls and that families had become more supportive of these pathways as they understood that this could lead to girls having an income of their own to sustain themselves.<sup>97</sup> When girls in ALP programmes were asked about where they had learned about the programme the most common response amongst beneficiaries was the WWW project (50 percent). This was attributed by programme staff and government officials to the work of the programme in raising awareness of this option and trying to remove the stigma which these courses carried, as well as a renewed emphasis by county ministries to improve the quality of centres by equipping them appropriately and coaching the instructors in these facilities.<sup>98</sup> Programme staff mentioned that the biggest challenge they faced was the stigma that TVET courses had of being for failures and those who were not smart enough to be in school.<sup>99</sup> However, this was addressed by having girls who had made a successful living after TVET courses talk to the cohort girls and encourage them to join these programmes. During qualitative interviews with girls there also seemed to be a general awareness of TVET programmes and what TVET could provide for them, and there did not seem to be stigma from these girls towards TVET programmes. Furthermore, when we surveyed ALP girls and asked why they had joined a TVET programme the most common response was that they had joined in order to learn a specific trade, rather than because they could not continue to secondary school.

*I am targeting to finish high school with a mean grade of B- or B. Thereafter, I will consider what to do. Depending on the outcome, we can also go for vocational training and upgrade to the field we want to be in.<sup>100</sup>*

Another important statistic that stands out from Table 59 is the high aspiration of girls in regard to their education, as the agreement response rate for both “I would like to continue studying/attending school after this year” and “I am willing to transition throughout the education pathways until I meet my goal” have remained at over 90 percent throughout all three evaluations. For the statement assessing girls’ aspirations for next year, the agreement rate has remained the same as at SAS, at 99 percent. This may also help explain the extremely high transition rates we have found in our analysis, as the girls who were surveyed at SAS had intentions to remain at school for the next year and our own results suggest that they have done so.

As part of this evaluation a small sample of girls who were attending, or had attended, an alternative learning pathway were surveyed on their general thoughts on their courses. The full results are discussed in Annex 17. We note that overall, girls had high perceptions of the quality of TVET programmes and believed programmes were preparing them for work.

Another alternative learning pathway which was promoted during the project but has not been discussed so far are catch-up centres for those students who had dropped out of school, often due to early marriage or pregnancies. Programme staff highlighted that at first, it was difficult to encourage girls to join these programmes as they had children to take care of at home; however, this was addressed by offering bursaries to girls who joined the programme.<sup>101</sup>

While catch-up centres offer a useful flexible option, girls in these programmes still faced challenges in transitioning to secondary school. For girls that were able to complete this course, they expected that there would be a similar flexible system in secondary school where girls could choose their own hours.

<sup>97</sup> KII with MOE Official, Kilifi, Int 3.

<sup>98</sup> KII with MOE Official, Kilifi, Int 3. KII with Programme Staff, Tana River, Int 86.

<sup>99</sup> KII with Programme Staff, Tana River, Int 86. KII with Programme Staff, Nairobi, Int 85.

<sup>100</sup> FGD with girls, Nairobi, Int 15.

<sup>101</sup> KII with Programme Staff, Tana River, Int 86.

Programme staff explained that this system was not currently in place and, therefore, once girls had completed the catch-up centres, they were unable to attend secondary schooling due to the lack of flexibility in school scheduling.<sup>102</sup> This meant that those girls who had re-enrolled and completed the courses at catch-up centres once again had to drop out.<sup>103</sup> It would be wise to encourage the government to implement a similar catch-up centre system in secondary schools, so that girls in these circumstances can continue their education past primary school.

### Box 3: Experience with Catch-up Centres

One girl in Nairobi, “Mercy”,<sup>104</sup> reported participating in a catch-up centre in the girl survey. Mercy stated that she had left school because she needed to help at home and had to get a job to support her family. When she later sought to return to school, she was informed about catch-up centres by a teacher. Mercy benefitted substantially from the catch-up centre and other support provided to her while she was involved in the programme; she stated that while enrolled in the programme, she had also received support from Community Health Volunteer (CHV) activities, distribution of back-to-school kits, peer mentoring, and educator mentoring. As a result, Mercy had very positive views of the catch-up centre programme: She stated that she would recommend the programme to other girls and that the programme was preparing her well for work and would be helpful for her to later secure a job. After completing the programme, Mercy wants to start a business.

Another girl in Marsabit, “Fatuma”, also reported participating in a catch-up centre after dropping out of school when her parents were unable to pay school fees. She was informed about the programme by a friend who had also enrolled in a catch-up centre. Like Mercy, Fatuma benefitted from many support activities while participating in the catch-up centre, including support from CHV activities, back-to-school kits, apprenticeship, and peer mentoring. These support activities had given her improved interest in what she was learning, as well as strengthened her financial knowledge. Fatuma stated that she would recommend the programme and felt it was preparing her well for work; she would like to start a business after completing the programme.

## Reflections on Transition Outcomes

Our sample had an extremely high transition rate of 96.1 percent; although this does not meet the programme’s target of 99 percent, this should still be viewed as a positive result, especially in light of the impact of COVID-19 on school attendance. This high transition rate is likely attributable both to the WWW programme as well as government efforts pushing for a 100 percent transition rate from primary to secondary school.

In terms of ALP, our findings suggest that there has been a general increase in the awareness levels of girls for these transition options. It seems that girls and parents understand that TVET is a viable option for girls to gain skills for the labour market which will allow them to make a living, rather than being seen as something for “failures”. The general quality of TVET courses was reported to be high as girls agreed that participating in these courses was giving them the tools to be successful in the workforce.

Overall, the largest challenge to transition remains the ability of families to pay for the girl’s education. This highlights the importance that scholarships and cash transfer have played in keeping girls at schools and maintaining support for the transition of girls. However, it also suggests that the programme may leave a gap here once it comes to an end as the girls who had been beneficiaries of the programme may find themselves struggling to pay for their school fees again.

<sup>102</sup> KII with Programme Staff, Nairobi, Int 85.

<sup>103</sup> KII with Programme Staff, Nairobi, Int 85.

<sup>104</sup> Names have been anonymized.

To conclude, two recommendations are formed as a result of the analysis in this section:

1. As just stated, the provision of cash transfers and scholarships was noted as being one of the most important activities within this programme. Therefore, it is possible that once project activities come to an end, a number of cohort girls will be forced to drop out as families are no longer able to afford their education without the support of the programme. It would be recommendable for the government to keep focusing on the provision of free education and to continue some of the activities which have been monetarily supporting those families who are not able to afford their daughters education.
2. It was also noted that catch up centres have proved a viable option for girls to come back to school under a flexible schedule which allows them to study while also attending to their household responsibilities. However, this project can only have a limited impact as long as there is not a similar flexible option at secondary school level to which these girls can transition once they have completed the primary level catch up courses. It would be recommendable for the government to study possible options for instituting catch up centres at secondary school level for marginalised girls to transition through education.

## Sustainability Outcomes

### Summary of Findings

1. The project performed well on sustainability scores as measured along the three dimensions of community, school, and system. Using a bespoke index, we found an increase between baseline, midline, and endline, with a score of 1.8 at baseline, 2.5 at midline, and 2.7 at endline (out of a total possible score of 4).
2. Sustainability is strongest at the school level, driven by strong teaching practices and extensive extracurricular activities that seek to improve girls' life skills, health, and other aspects of relevance to the programme.
3. Government involvement in learning and gender responsiveness was also very strong; the MoE and its departments are actively involved in providing training to teachers and other school personnel, which will likely reinforce gains made from WWW after the close of the programme.
4. Community support for girls' education and adoption of community action plans to strengthen girls' education are still emerging areas, but have improved since baseline and midline.
5. Keeping in mind the severe effects the pandemic has had on government, school, and community capacity, the increase in sustainability score is a positive sign, showing progress despite substantial challenges.

As at midline, sustainability for the project is measured along three dimensions: the community, school, and system. The community dimension is measured through community participation in community action plans and household support for adolescent girls' education. The school dimension is measured through the strength of school leadership, teaching practices, and co-curricular activities. The system dimension is measured through the status of gender analysis and reporting and support for TVET.<sup>105</sup>

---

<sup>105</sup> At midline, an indicator was also included measuring support for National Leaders of Education interventions. This indicator was dropped from the WWW programme framework at endline (see *Introduction*). As such, we do not include it in measurements.

At the community level, participation in community action plans is expected to increase programme sustainability by increasing community support for education and training of marginalised girls, which may continue after programme support is removed. Similarly, improved attitudes towards girls' education will increase sustainability by increasing the likelihood that girls remain in school due to family support, even after the close of the WWW programme.

At the school level, strengthened school leadership may improve sustainability by enabling school leaders to better manage and train teachers and manage school finances, thus improving quality of education and allowing schools to be more self-sufficient. By strengthening teaching practices, the WWW programme may have a long-term impact on learning outcomes, as more effective, better-trained, and gender- and disability-sensitive teachers will likely continue using these positive practices throughout their careers. Lastly, increased uptake of extracurricular activities—especially if continued after the end of the WWW programme—will help strengthen students' life skills and confidence, skills that they will utilise throughout their lives.

At the system level, improved government monitoring of teaching and learning activities and gender responsiveness, if continued after the end of the programme, will help inform emerging MoE gender and teaching approaches. This will allow for future government- or school-led interventions to further strengthen girls' education. Increased support for TVET, meanwhile, will help expand alternative learning pathways for girls, thus improving transition outcomes and girls' abilities to provide for themselves. We note that the former indicator—learning and gender responsiveness—has been changed since midline, in which the indicator measured the number of MoE officials trained on and conducting gender analysis and reporting. As such, we do not report baseline or midline values for this indicator. We also do not include this indicator in our calculation of the final sustainability score at endline in order to increase comparability across rounds.

To analyse these dimensions of sustainability, we utilise quantitative and qualitative data to calculate sustainability scores.<sup>106</sup> The methods of calculating scores are discussed more in Annex 17. Sustainability scores for each indicator can range from 0 to 4, where 0 is “negligible”, 1 is “latent”, 2 is “emerging”, 3 is “becoming established”, and 4 is “established”. Scores for indicators at the same level (community, school, or system) are aggregated into a single level score. These scores are then average to produce an overall sustainability score. The below table shows indicator values and scores for baseline, midline, and endline.<sup>107</sup>

**Table 42: Sustainability scores**

Rating	BL	ML	EL
<b>Community level</b>			
Community action plans	1.0	2.0	2.3
Support for girls' education	2.0	2.3	2.4
Average community score	1.5	2.2	2.4

<sup>106</sup> While rationale for indicator scores is provided in the narrative sections below, it is important to note that some scores are subjective, although we have relied on objective data where possible. In some cases, scoring is based on limited qualitative interviews, while in other cases scores include quantitative data from multiple respondent groups and qualitative data.

<sup>107</sup> We additionally report SAS values for specific indicators, but do not focus on them here as SAS data is not available for all indicators.

School level			
School leadership	2.0	2.7	2.7
Teaching practices	2.0	3.2	3.7
Extracurricular activities in schools	2.0	2.9	3.2
Average school score	2.0	2.9	3.2
System level			
Learning and gender responsiveness	-	-	3.3 <sup>108</sup>
TVET support	2.0	2.0	2.0
Average system score	2.0	2.0	2.0
Total sustainability score			
Total score	<b>1.8</b>	<b>2.5</b>	<b>2.7</b>

Before discussing specific scores, it is worth noting the potential impact of the COVID-19 pandemic on sustainability indicators at all three levels. As will be discussed below, COVID-19 affected individuals' and communities' economic stability, thus affecting their ability to support girls' (and boys') education. At the school level, school closures led to disruptions in programme activities which may have limited the programme's ability to conduct some activities meant to bolster sustainability.<sup>109</sup> A programme staff member also noted high turnover in schools, in part due to the pandemic, limited sustainability.<sup>110</sup> At the system level, lastly, the myriad challenges brought on by the pandemic may have reduced government capacity to prioritise programme activities or support schools or TVET institutions, although these challenges were not specifically mentioned in interviews. With this in mind, the increase in sustainability score shown above is a positive sign, showing progress despite substantial challenges.

## Community Level

### **Community Action Plans**

At endline, within the caregiver survey, caregivers were asked questions about community conversations (CCs), including their awareness of and participation in these groups, topics discussed during CCs, and the benefits of these groups. This data is used to better understand community plans to support girls' education. Additionally, data on community initiatives undertaken to support girls' education provides further information of relevance to this indicator.

At endline, we find that 38 percent of surveyed caregivers were aware of CCs sponsored by the WWW programme. The vast majority of respondents—90 percent—stated that CCs were somewhat or very beneficial to the community.

Community support for girls' education is discussed in more detail in the section *Community Behaviour and Support*. Here, we note that at endline, 64 percent of caregivers reported that a girl in their

<sup>108</sup> Not included in calculations of average and total scores in order to improve comparability of scores across rounds.

<sup>109</sup> For example, if the programme was unable to carry out planned teacher training activities during the pandemic, this may have inevitably reduced the reach of teacher training activities and thus reduced sustainability. In qualitative interviews, a programme staff member indeed noted this as a challenge, stating that due to COVID, the programme had a very limited timeline to train teachers (KII with programme staff, Int. 85).

<sup>110</sup> KII with programme staff, Int. 83

community had received support from the community to further her education, showing a fairly high prevalence of support for girls' education.

Overall, while CC groups do seem to benefit communities to form plans to support girls' education, the limited prevalence of these groups—with only 38 percent of caregivers reporting awareness—means that this indicator remains emerging. However, compared to midline, a much higher percentage of caregivers reported community (or personal) support for a girl to pursue her education. As such, we rate this indicator at a value of 2.3 for endline, indicating that while much work is still needed to establish community action plans, an improvement has been observed over midline.

### **Support for Girls' Education**

The support for girls' education indicator is calculated using one quantitative variable, which asked caregivers to rate the support for adolescent girls' education within the surrounding community. We then validate our findings for this indicator using other quantitative and qualitative data on related topics.

As reported in Table 43, we calculate an average score of 2.4 for this indicator at endline.<sup>111</sup> We note that the baseline score is based on the score reported in this evaluation report, and is not directly comparable to other values.

**Table 43: Community support for adolescent girls' education scores**

*Respondents were asked: "How do you rate the support for adolescent girls' education within the surrounding community?"*

Rating	BL	ML	SAS	EL
n	-	1,454	294	814
Very poor	-	2.3%	9.5%	3.0%
Poor	-	9.8%	18.4%	11.3%
Average	-	47.0%	50.0%	34.8%
Good	-	33.0%	19.1%	41.0%
Excellent	-	7.9%	3.1%	10.0%
Average score	2.0	2.3	1.9	2.4

We find that support for girls' education dropped between midline and SAS, from 2.3 to 1.9, before recovering to a slightly higher value of 2.4 at endline. This pattern may be due, in part, to the negative impact of COVID-19 upon households' and communities' livelihoods. If communities' livelihoods and socioeconomic stability were threatened due to COVID-19, then households' abilities to support education in general may have decreased if caregivers were no longer able to pay for schooling costs or if children were needed to support their families through work.

Triangulating these findings with the qualitative data, we find that almost all interviewed teachers stated that the community in which they worked supported girls' education. One teacher from Kilifi, for example, stated the following:

<sup>111</sup> The variable used to calculate this value included five answer options; we thus normalised the average value for this indicator to a maximum score of four by averaging the overall rating so that a rating of "very poor" was equivalent to a score of 0, while a rating of "excellent" was equivalent to a score of 4.

*The community is now enlightened on the importance of girl-child education and they are now educating their girls not like before. We have lady teachers in this school who come from this community and this shows how positively the community have embraced girl education.<sup>112</sup>*

Along these lines, in FGDs, girls also often expressed that their parents supported them to attend school, but there were limitations to this support. The greatest limitation was a lack of school fees; many girls stated that if their parents did not have money, they would be unable to attend school.<sup>113</sup>

Looking at additional quantitative data, the table below shows the level of education that caregivers were willing and able to support their girl child to achieve at endline. While the majority of caregivers expressed that they would like their girl child to attend university and would be willing to support her to attend university, only around one-third of caregivers stated that they would be able to support their girl to attend university. Around one-quarter of caregivers, in contrast, stated that they were only able to support their girl child through secondary school, or that they didn't know the level of education they would be able to support.

**Table 44: Caregiver support for their girl child's education**

Level of Education	Would Like Girl to Achieve	Willing to Support Girl to Achieve	Able to Support Girl to Achieve
Primary	0.0%	0.5%	3.8%
Secondary	1.5%	4.8%	24.0%
College/TVET	12.7%	13.5%	12.3%
University	80.8%	71.9%	34.0%
Don't know	5.0%	9.3%	25.9%

Overall, these findings suggest that while there are high theoretical levels of support for girls' education, barriers remain to these girls actually completing higher levels of education. Data suggests that financial difficulties are a dominant barrier; at endline, 83.6 percent of caregivers reported that it was difficult to afford for their girl child to go to school.

A positive finding, however, is that most caregivers expressed similar levels of value for girls' education as boys': At endline, 95 percent of caregivers agreed that a girl is just as likely to use her education as a boy. Caregivers were also likely to report that their household's support for girls' education had increased in the past year, with 69.2 percent of respondents stating that support had improved at endline. This suggests that while barriers remain, caregiver and community support and value for girls' education is high, potentially increasing the sustainability of programme interventions.

Furthermore, caregivers expressed relatively high levels of support for girls who had dropped out or not transitioned to secondary school to continue their education. At endline, 92.3 percent of caregivers stated that they would be willing to support such a girl, while 54.8 percent stated that they had supported such a girl. This suggests that not only are caregivers willing to support girls who have stayed in school, they are also willing to support those who have struggled with school or had to drop out due to life circumstances. This is a positive finding for the sustainability of the programme's impact on marginalised girls, as these girls are more likely to drop out and need future community support.

<sup>112</sup> KII with teacher, Kilifi, Int. 26

<sup>113</sup> E.g., FGD with girls, Tana River, Int. 17; FGD with girls, Samburu, Int. 25; FGD with girls, Nairobi, Int. 18

## School Level

### School Leadership

School governance and management are discussed in Annex 19; as such, in this section, we only discuss the school leadership score used to calculate the sustainability outcome. We calculate a score using data from the caregiver survey, in which caregivers were asked to rate the quality of school management, its improvement over time, and the quality of the head teacher/principal and Board of Management (BoM).

Our score is calculated based on caregiver responses to the questions “How well is the school [*your girl child*] attends managed?” and “How would you rate the performance of the school head teacher or principal of the school that [*your girl child*] attends?” At both midline and endline, we find that the vast majority of caregivers—89 percent and 90 percent respectively—stated that their child’s school was managed well or extremely well. We also find that caregivers were similarly unlikely to rate head teacher performance as poor at both midline and endline.

**Table 45: School leadership scores**

Score	BL	ML	EL
<b>School management</b>			
Not managed well at all	-	5.5%	4.8%
Well managed	-	73.2%	71.4%
Extremely well managed	-	15.9%	18.9%
Don't know <sup>114</sup>	-	5.4%	4.9%
Score	-	2.5	2.6
<b>Head teacher performance<sup>115</sup></b>			
Very poor	-	-	1.1%
Poor	-	3.3%	1.1%
Average	-	52.6%	24.5%
Good	-	36.9%	53.8%
Excellent	-	-	19.5%
Don't know	-	7.2%	-
Score	-	2.9	2.9
<b>Overall score</b>			
Average score	2.0	2.7	2.7

Overall, we find little change in the average school management score from midline to endline, with a score of around 2.7 for both rounds. This is a substantial improvement over baseline, however, where

<sup>114</sup> Values excluded from calculation of indicator score.

<sup>115</sup> At midline, only three choice options were provided for head teacher performance. In addition, “don’t know” was not included as a choice option at endline. “Don’t know” answers are excluded from the average score calculation.



school management was scored as a 2 (emerging). Positive school management practices appear to still be in the process of established, rather than being fully established. This finding is not overly surprising due to setbacks from COVID-19 and the WWW programme’s targeting of marginalised schools with fewer resources.

The qualitative data supports this finding, showing mixed support from school leadership for teacher development and strengthening of teaching practices. About half of teachers stated that they had received mentoring from head teachers or principals, or that head teachers had helped facilitate mentoring or training sessions held by others. Most teachers had also received mentoring/training from another party, such as EDT or government officials.

### **Teaching Practices**

Teaching practices are discussed at length in a subsequent section under *Intermediate Outcomes*; as such, here, we only present the calculation and findings for the sustainability score. In order to allow for comparability with past data, we calculate a score for teaching practices based on data from the girl survey, in which girls were asked several questions about the quality of their teachers.

Scores presented below in Table 46 are calculated as the percent of “positive” answers to five statements in the girl survey.<sup>116</sup> These statements are as follows:

1. My teachers make me feel welcome in the classroom.
2. My teachers explain the lessons well in class.
3. My teachers respond to learner’s questions well in class.
4. My teachers treat boys and girls differently in the classroom.
5. My teachers are often absent from class/lessons.

The below table shows that at endline, the average teaching practices score was 3.7 out of 4, representing generally good quality of teaching as rated by girls. This was a substantial improvement over the score of 3.0 out of 4 for SAS.<sup>117</sup> Comparing to values reported in previous reports, we find an improvement in teaching practices at endline compared to previous rounds.

**Table 46: Teaching practices score**

Score	BL	ML	SAS	EL
0	-	-	0.0%	0.1%
1	-	-	0.0%	0.6%
2	-	-	0.4%	0.8%
3	-	-	30.3%	8.2%
4	-	-	69.1%	34.8%
5	-	-	0.3%	55.5%
Average score	2.0	3.2	3.0	3.7

<sup>116</sup> For statements 1 through 3, positive answers were “agree” or “strongly agree”. For statements 4 and 5, positive answers were “disagree” or “strongly disagree”. Using this method, a total teaching practice score was calculated first as a total out of 5, then normalised to an average score from 0 to 4 points.

<sup>117</sup> This lower score was mainly driven by reported high levels of teacher absenteeism.

Qualitative data on teaching practices is also analysed in more detail in the subsequent section on teaching quality. In summary, we find that girls generally expressed that their teachers help them when they are struggling with a topic and do not punish struggling students.<sup>118</sup> In FGDs, many girls stated that if they did not understand a topic, they would feel comfortable raising their hand or following up with the teacher after the lesson, and that the teacher would then explain the topic in another way.

However, in an important area for future improvement, many girls mentioned that teachers still used negative teaching practices such as corporal punishment or other harsh punishments. In FGDs, girls described the use of caning, physical labour, and exercise as punishments, as well as assignment of extra schoolwork and suspension from school. While most girls stated that students did not feel embarrassed by punishment, this is indicative of the commonality of negative punishments.

Regarding teacher attendance, in FGDs, many girls reported that teacher attendance had significantly improved over the past year. A girl from Tana River, for example, stated that teachers “always attend and don’t miss lessons” and that “their attendance has gotten better since COVID”.<sup>119</sup> Overall, these findings validate the quantitative findings above.

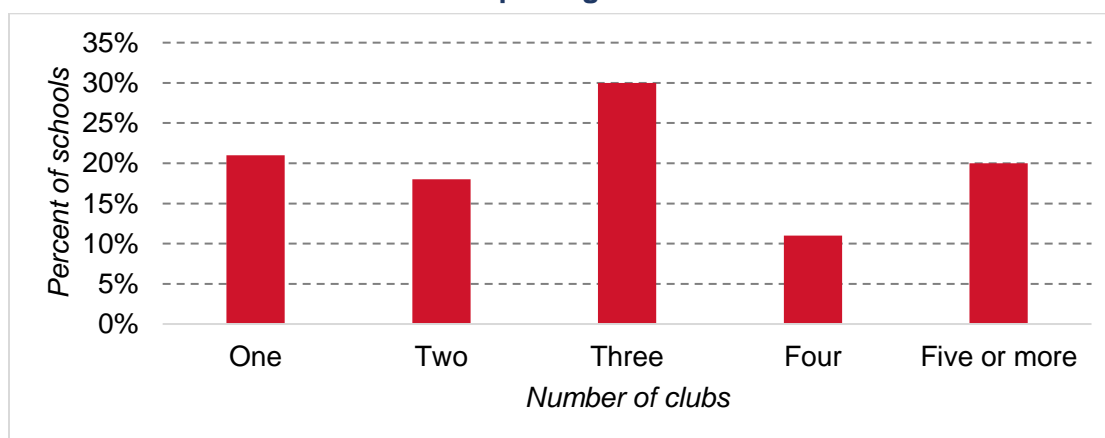
### **Extracurricular Activities in Schools**

The extracurricular activities indicator is calculated based on the number of clubs offered at schools and on analysis of the qualitative data from girls, in which they were asked about participation and usefulness of extracurricular activities.

At midline, within qualitative data, students reported membership in school clubs which enhanced their career choices, knowledge of SRH, and life goals. Other respondents also described a strengthening of school activities promoting girls’ self-confidence and knowledge of SRH which were run by schools without project support. As such, this indicator was given a score of 2.9, representing an indicator that is becoming established. This was a substantial improvement over the score of 2.0 at baseline.

At endline, we find that schools had, on average, around three school clubs. As shown in the below figure, schools also most frequently reported having three clubs, with 30 percent of schools reporting three active clubs. These clubs addressed a wide range of topics, including health, sports, maths, science, the environment, agriculture, journalism, music, guidance and counselling, debate, and drama, among others.

**Figure 19: Percent of endline schools reporting each number of active clubs**



<sup>118</sup> E.g., FGD with girls, Tana River, Int. 17; FGD with girls, Turkana, Int. 22; FGD with girls, Samburu, Int. 12; FGD with girls, Nairobi, Int. 21

<sup>119</sup> FGD with girls, Tana River, Int. 23

Respondents to the school survey reported a positive effect of these school clubs and extracurricular activities: Clubs were reported to have improved a number of outcomes of relevance to the programme. The first of these outcomes can be broadly categorised as “life skills”, and includes reported improvements in students’ and girls’ confidence, self-esteem, and attitudes towards education. Clubs were also reported to have improved learning outcomes, reduced the number of students dropping out, and taught students important skills, such as health and hygiene topics and agriculture and animal rearing. By teaching students these skills, some head teachers also reported that clubs had reduced the prevalence of teenage pregnancy and reduced illnesses among students.

Overall, extracurricular activities appear relatively well-established at many schools. As such, at endline, we give this indicator a score of 3.2. This represents our judgment that the status of extracurricular activities is beyond becoming established in many schools, but still needs more work to be fully and sustainably established within all schools.

## System Level

### *Learning and Gender Responsiveness*

In order to strengthen sustainability at the system level, Ministry of Education (MoE) officials were trained on how to undertake monitoring of teaching and learning activities as a way of ensuring sustainability. One of the areas to be monitored was teachers’ use of gender responsive pedagogy. This indicator was revised from baseline/midline; as such, we only include endline results in this section. Data for this indicator primarily comes from qualitative interviews with government officials, who were asked about the types of training, school monitoring, and support for teaching and learning that the government provided.

All government officials stated that the MoE and its departments were actively involved in providing training to teachers and other school personnel. This included training on the curriculum, including preparedness for the competency-based curriculum (CBC), curriculum checks, and support for curriculum delivery to ensure teachers were teaching appropriate topics and using appropriate teaching methods. One government official from Nairobi, for example, stated that “a lot of trainings have happened in curriculum development; this is not limited to the CBC but also to the 8-4-4 curriculum development, and it is subject based”.<sup>120</sup> Teacher training sessions were also organized to train subject teachers on how best to deliver the curriculum and to improve pedagogy.<sup>121</sup> Some officials also stated that training on school management was provided for principals, their deputies, and senior teachers.<sup>122</sup> Additionally, an official in Mombasa stated that there was occasional training on gender responsiveness.<sup>123</sup> Trainings were described as effective.<sup>124</sup>

In KIIs, many teachers also stated that they had received training from the MoE or other relevant government bodies, validating the above findings. For example, one teacher from Tana River stated that he had received a training from the MoE on the CBC, and that this training was very beneficial.<sup>125</sup> We note, however, that a minority of teachers stated that they had not received mentorship or training, suggesting a need to continue expanding the reach of government trainings.<sup>126</sup> Furthermore, no teachers explicitly mentioned receiving training on gender responsiveness or gender-sensitive teaching practices.

While these findings point to a generally very active role for the MoE in training and learning, we note one challenge described by an official from Nairobi. She noted that government support activities—including

---

<sup>120</sup> KII with government official, Nairobi, Int. 93

<sup>121</sup> See e.g., KII with government official, Nairobi, Int. 93

<sup>122</sup> KII with government official, Samburu, Int. 92

<sup>123</sup> KII with government official, Mombasa, Int. 94

<sup>124</sup> KII with government official, Kilifi, Int. 3

<sup>125</sup> KII with teacher, Tana River, Int. 27

<sup>126</sup> KII with teacher, Nairobi, Int. 29; KII with teacher, Nairobi, Int. 7; KII with teacher, Mombasa, Int. 30

training and other forms of support—were generally limited to government institutions. However, in some counties, there are far more informal, private, or non-governmental schools than public schools. In Nairobi, for example, she stated that there were over 300 government primary and secondary schools, but more than 2,000 non-governmental schools. Government support to these schools was limited, in part because the management of these schools was not transparent and did not accept support.<sup>127</sup>

Despite this challenge, overall, MoE involvement in monitoring of teaching and learning activities and teacher training appears high. Given these findings, we rate this aspect of sustainability at a 3.3, as these activities are well on their way to becoming established.

### **Support for TVET**

Information on support for TVET institutions at the system level comes from qualitative interviews conducted with government officials. Some officials described a high level of government support for TVET programmes. For example, an official in Mombasa stated that “TVET programmes are a good alternative and the government is keen on reviving them... there is generally a positive attitude”.<sup>128</sup> Another official from Nairobi stated similarly that “the ministry has been vibrant in trying to promote [TVET institutions]”.<sup>129</sup>

However, some respondents also expressed challenges regarding support for TVET institutions at both the system and the individual levels. A government official from Samburu, for example, stated the following:

*We have a big challenge with regards to TVET institutions, this is so because the uptake has been very, very slow and the negative perception from the past still lingers. The mentality of white-collar jobs is still alive. TVET courses are classified as “kazi ya mkono”(manual labor) which carries along with it the stigma of school dropout or exam failures.<sup>130</sup>*

A government official from Turkana similarly stated that regarding support for TVET, “cultural attitudes” were a constraining factor and TVET programmes were still undervalued and underrated,<sup>131</sup> while another official from Tana River stated that TVET programmes were “generally shunned” as people held the perception that TVET was a below-par transition pathway.<sup>132</sup>

Overall, these interviews show that support for TVET programmes remains under development. While government officials appeared to support this pathway, negative perspectives from communities are a major constraint. At midline, findings were very similar, with the midline report noting that “relatively negative perspectives by communities – terming TVET as inferior [sic], made it harder for officials to encourage communities to embrace this option”. As such, we score this aspect of sustainability the same at endline as at midline, a 2.0.

## **Reflections on Sustainability Outcomes**

The above analysis shows an increase in sustainability from baseline to endline and a small increase between midline and endline. Sustainability is strongest at the school level, particularly driven by strong teaching practices and co-curricular activities, leading to a sustainability score of 3.2 at endline (or somewhat above “becoming established”). While the system score was lowest of the three component scores at endline, this is primarily driven by still-emerging support for TVET programmes. In contrast,

<sup>127</sup> KII with government official, Nairobi, Int. 87

<sup>128</sup> KII with government official, Mombasa, Int. 94

<sup>129</sup> KII with government official, Nairobi, Int. 93

<sup>130</sup> KII with government official, Samburu, Int. 92

<sup>131</sup> KII with government official, Turkana, Int. 90

<sup>132</sup> KII with government official, Tana River, Int. 97

government involvement in learning and gender responsiveness—a new indicator at endline, and thus not included in calculations of the total score—is strong. If this indicator is included in calculations of the total score, the endline sustainability score increases to 2.8 (from 2.7).

Overall, these findings suggest generally positive results for sustainability. These dynamics are discussed more under *Project-Specific Outcomes*.

## Intermediate Outcomes

In addition to the three primary long-term outcomes which the WWW project aims to achieve, the project also aims to create improvements in several intermediate outcomes, namely: 1) girls' attendance in school, 2) quality of instruction, 3) girls' life skills, 4) household support for girl's education, and 5) community-based attitudes and behavioural changes. The extent to which changes in these intermediate outcomes can be attributed to WWW is significantly limited by the evaluation design, as girls for whom intermediate outcome data was collected during the SAS and EL lack comparison groups. As such, analyses of intermediate outcomes are focused on two steps. The first is a descriptive analysis of changes in the panel sample who were sampled at both the SAS and EL. The second is a predictive analysis of various school- and individual-level factors, and how well they predict intermediate outcomes.

### Attendance

#### Summary of Findings

1. Attendance rates have improved from baseline and midline, but decreased since SAS.
2. Compared to SAS, attendance has declined almost twice as much for boys compared to girls. Given the WWW programme's focus on girls, this suggests that the programme may have had a mitigating effect on attendance decreases, resulting in improved outcomes for girls relative to boys.
3. While attendance measured through school records decreased, caregivers were more likely to state that their girl's attendance had improved at endline. This was reiterated in qualitative interviews, which suggested that girls' attendance had improved over the past year. These findings may represent caregivers' and girls' perceptions of improving attendance post-COVID-19.
4. The presence of school clubs was positively and significantly associated with attendance rates. Supporting the establishment of school clubs, especially those focusing on life skills and confidence, may be a useful intervention to further improve attendance.

The primary method for measuring attendance used in the WWW evaluations is to cross-reference enrolment records for relevant streams and class/forms with a headcount of the number of students physically present in class during the school survey. Attendance rates are then taken by dividing the number of students present by the number of students who are formally enrolled; in cases where more students were present in class than those formally enrolled, the attendance may exceed 100 percent, which would suggest an overcrowding of classes in relation to initial planning of class sizes.

Using this measure, the BL data indicates that attendance on class observation days measured at around 88.0 percent for intervention schools, and lower for non-intervention schools, at around 77.0 percent. The ML data highlighted a 3.0 percentage point drop in attendance rates for intervention schools between BL and ML, before this figure increased by 6.2 percentage points to reach 91.2 percent for all students during the SAS round. Girls' attendance rates follow a similar trajectory, dropping from BL to ML, before rising

substantially leading up the SAS round.<sup>133</sup> Table 47 below highlights changes from the SAS to EL rounds, as well as disaggregating the data by gender and by county.<sup>134</sup>

**Table 47: Attendance rates in SAS and EL, disaggregated**

County	Girls			Boys			All		
	SAS	EL	Diff.	SAS	EL	Diff.	SAS	EL	Diff.
Nairobi	93.2%	87.6%	-5.6	90.6%	83.5%	-7.1	92.2%	85.5%	-6.7
Kilifi	89.9%	91.9%	2.0	87.0%	88.0%	1.0	88.6%	89.6%	1.0
Samburu	95.8%	96.2%	0.4	98.2%	82.0%	-16.1	93.3%	90.0%	-3.3
Turkana	96.4%	94.7%	-1.7	95.9%	82.9%	-12.9	96.1%	88.1%	-8.0
Tana River	87.7%	92.3%	4.6	83.9%	88.8%	4.9	86.1%	90.5%	4.4
All	92.4%	90.2%	-2.2	90.2%	84.5%	-5.7	91.2%	87.2%	-3.9

A general trend across genders and counties is the falling attendance rates from SAS to EL, with attendance rates dropping by 3.9 percentage points on aggregate and by 8.0 and 6.7 points in Turkana and Nairobi, respectively, though some of these rates remain higher than at both BL and ML. The decline is more pronounced among boys than girls, with the former registering a decline in attendance rate that is more than twice of the decline seen among girls. Given the WWW programme's focus on girls, this suggests that the programme may have had a mitigating effect on attendance decreases, resulting in improved outcomes for girls relative to boys.

A number of factors may explain these trends. The COVID-19 lockdowns, for instance, may have led some learners who had not attended school to now find it difficult to return, either because they have taken on new activities during lockdown that they are now continuing, or because the lockdown period severely hindered their learning routine, causing them to fall behind and become maladjusted to keep up with schoolwork once the regular pace of in-person schooling returned. For example, some girls may have become more involved in household chores during the lockdowns, and an increased dependence on their contributions may mean that some parents would prefer them to stay at home more frequently and thus miss class.<sup>135</sup> Alternatively, some students might have started a small business or otherwise become involved in other income-generating activities that would take time away from attending school.<sup>136,137</sup>

Indeed, as the COVID-19 pandemic waned in severity, and as public health measures around physical distancing and opening hours for businesses are relaxed, the ensuing rebound in business and economic activities<sup>138</sup> may mean that learners once again increasingly have incentives to undertake income-

<sup>133</sup> Charity Limboro et al. (2020). "Washichana Wetu Wafaulu GEC- T Midline 1 Report," Education Development Trust, 131.

<sup>134</sup> All intervention schools surveyed at both SAS and endline, 123 schools, are included in this table. Comparison data is not included.

<sup>135</sup> No data is available on girls' time spent in household chores during SAS, which limits empirical analysis of these changes between rounds.

<sup>136</sup> FGD with Girls, Nairobi, Int.10

<sup>137</sup> The above measure of attendance relies on cross-referencing enrolment records with headcounts, but it is unclear how up-to-date enrolment records are for students who had already dropped out entirely. Nor is it clear whether pursuing economic opportunities means that learners are wholly, or periodically, absent from schools. Thus, in this instance, we are unable to fully distinguish the effects of these dynamics from attendance and retention separately, and primarily discuss findings in relation to absences from school, whether they be *ad hoc* absences or a more permanent drop-out from school. Still, we find it plausible that issues such as the need to pursue economic opportunities can negatively impact both attendance and retention.

<sup>138</sup> World Bank Group. (2021). "Kenya's Economy is Showing Resilience as Output Rises Above Pre-Pandemic Levels Driven by Rebound in Services Sector," *World Bank*, December 14, <https://www.worldbank.org/en/news/press-release/2021/12/14/kenya-s-economy-is-showing-resilience-as-output-rises-above-pre-pandemic-levels-driven-by-a-rebound-in-the-services-sect>.

generating activities, either on their own accord or by request of their family. That one of the sharpest declines in attendance rates was observed in Nairobi – the heart of Kenya’s economy – adds credence to this explanation.

*I know as I started talking about the locality, the area where we are in, it is a bit challenging. The pupils want to start up businesses at an early age, there are so many businesses around ... [these young learners] are easily wooed [by economic opportunities].<sup>139</sup>*

Gendered differences in attendance rates can only be explained more speculatively, as the bulk of this study’s data collection instruments and analysis centred around the experiences of girl learners. However, to the extent that the above explanation holds true, it is possible that male learners are more involved in a wide variety of income-generating activities due to there being more opportunities afforded to them than to their female counterparts, including lesser expectations that male household members are to help with unpaid domestic household work compared to female members.<sup>140</sup> In this case, the trade-off between time spent in education and time spent helping generate income may be steeper for boys than girls, leading the former’s attendance rates to drop more precipitously as they spend more time, on average, pursuing income-generating activities.

The above method of measuring attendance has its limitation in that older age cohorts of learners did not have an attendance measure taken. We explain the implications of this more fully in in Annex 18. As such, it is useful to triangulate the findings with information from other measures. In this case, we use data obtained from asking caregivers of intervention students to estimate the attendance of girl learners in their schools, which encompasses girls of all age groups covered in this study. Caregiver perceptions are more subjective and thus prone to recall and/or social desirability biases than the previous method of measuring attendance; they also do not give us information about the attendance of male learners. Nonetheless these additional measures provide a useful robustness check against the school survey data.

**Table 48: Caregiver assessment of girls' attendance in school**

Response	Has girl attended school most days since opening?			Has girl's attendance improved over the past year?		
	SAS	EL	Change	SAS	EL	Difference
No	6.5%	7.7%	1.1	13.4%	15.2%	1.8
Yes	92.3%	92.2%	-0.1	83.5%	80.8%	-2.7
Do Not Know	1.1%	0.2%	-1.0	3.1%	4.1%	1.0

<sup>139</sup> KII with Teachers, Nairobi, Int. 11

<sup>140</sup> Amanda Ellis et al. (2007). "Gender and Economic Growth in Kenya: Unleashing the Power of Women," The World Bank, 52.

**Table 49: Caregiver assessment on frequency of girls' attendance in school**

Response	SAS	EL	Difference
More than half of the time <sup>141</sup>	45.0%	89.6%	44.6
About half of the time	25.0%	7.1%	-17.9
Less than half of the time	5.0%	2.6%	-2.4
Not enrolled	0.0%	0.2%	0.2
Do not know	25.0%	0.6%	-24.4

Data from Table 48 and Table 49 above, both gathered from caregivers via CATI,<sup>142</sup> paint a different picture than the data from the school survey headcounts. Table 48 shows a high level of caregivers reporting that attendance was good (“attended most days of school”) or improving, at more than 90 percent and 80 percent, respectively. A sizable increase of 44.6 percentage points was also observed in the proportion of caregivers who claimed that the girls under their care were in school for more than half the time in which they were supposed to be.<sup>143</sup>

The perception of improving attendance rates expressed by caregivers is generally in line with many of the responses from qualitative KIIs, many of whom indicated that girls’ attendance had in fact seen improvements over the past year. One teacher believed that assistance provided by various organisations has helped improve attendance, particularly among girls from disadvantaged backgrounds,<sup>144</sup> with another affirming that subsidised or free education programs promoted by the WWW have helped increase attendance.<sup>145</sup> Overall, these quantitative and qualitative findings may represent caregivers’ and girls’ perceptions of improving attendance following COVID-19 school closures.

Taken together, the general picture is that even accounting for decreases in attendance rates between SAS and EL, as is suggested by the headcount school survey data, the attendance rates across most counties remain higher than the rates observed during both the BL and ML rounds of data collection.<sup>146</sup> Thus, the overarching, albeit somewhat small, improvements from BL to EL are clear.

Lastly, we note that the presence of school clubs was positively and significantly associated with improved attendance rates (see Annex 18). This suggests that supporting the establishment of school clubs, especially those focusing on life skills and confidence, may be a useful intervention to further bolster student attendance.

---

<sup>141</sup> Looking at the mean age of girls whose respondents provided this answer indeed suggests some truth to the aforementioned possibility of headcounts missing important trends among learners who are, on average, older. Specifically, the mean age for girls whose caregivers claimed as much was 15.8, compared to 13.7 and 14.4 for Class 7 and 8, respectively. This might suggest that the more positive responses from caregivers may be, in part, driven by the inclusion of older girls compared to the headcounts in Class 7 and 8.

<sup>142</sup> This includes caregivers of intervention students only within schools that were surveyed at both SAS and endline.

<sup>143</sup> The large decrease in respondents claiming to not know how often their girls attended school suggests that some caregivers may have simply become more aware of their dependent’s attendance, rather than their dependents actually increasing attendance. Caution is thus warranted in using this measure to determine attendance rates. Still, one might nonetheless interpret caregivers being more attentive to their dependent’s education as being a positive trend, albeit not as significant as a realised increase in girl learners’ attendance.

<sup>144</sup> KII with Teacher, Tana River, Int. 6

<sup>145</sup> KII with Teacher, Mombasa, Int. 30

<sup>146</sup> Limboro, *Midline 1*, 131.



## Teaching Quality

### Summary of Findings

1. Teaching quality was found to be very high at endline, with an overall score of 83.4 out of 100. Qualitative data suggests that many improvements in teaching quality may be due to WWW programme interventions and trainings.
2. A key contributor to strong teaching quality was teachers' use of methods to engage slower learners, motivate students, and use practical demonstrations. Furthermore, girls had strong perceptions of gender equity in the classroom.
3. However, teachers often stated that boys and girls are better at different subjects, suggesting that gender stereotypes may still affect girls' learning experiences. Additionally, around one-quarter of teachers expressed acceptance of corporal punishment.
4. Several teaching quality metrics were lower in ASALs than urban slums. This may be due to higher teacher attrition in ASALs than urban slums, which reduced ASAL teachers' exposure to WWW project interventions.

This section analyses differences in teaching quality between the SAS and EL rounds, a point of interest described by Intermediate Outcome 2 (IO2), measuring whether the “quality of teaching/instruction improved in schools and alternative pathways”. As teaching quality is difficult to measure, we established a number of “scores” to serve as proxies for teaching quality. Reported by girls via the girl survey, three of these ten scores include girls' perceptions of teacher attitudes and behaviours vis-à-vis gender; where applicable, data regarding the subjects and topics that girls believe their teachers need to refocus; and judgements vis-a-vis girls' classroom learning environments.

In addition to scores generated from the girl survey, data from classroom observations includes a report on corporal punishment; an evaluation of the teacher's socioemotional skills; a review of the way the teacher delivered instruction to the students in the classroom, evaluated against the girl-first pedagogy; a measure of the teacher's attitudes self-reported by the teacher following the class period; and three measures of observed classroom culture, including a general measure, a gender-specific measure, and a disability-specific measure.

In order to evaluate these measures, which are comprised of relevant questions from either the girl survey or classroom observation survey, we developed a ranking scheme for each question that was then used to create a series of scores. These scores include the cumulative teaching quality score, the girl survey subscore which includes three sub-subscores, and the classroom observation subscore, which includes seven sub-subscores.<sup>147</sup>

Within this section, we begin by discussing changes between scores generated from the girl survey between SAS and EL. Moving to classroom observations, we do not have classroom observation data evaluating the same teachers at SAS, so we are unable to discuss any observed changes between the two rounds. Instead, we discuss observed differences in the scores and subscores across different school characteristics at EL, such as region and county. The below table shows teaching quality scores for intervention schools, including both primary and secondary schools.

<sup>147</sup> An explanation of the methodology used to calculate the scores and subscores is included in the Annex.

**Table 50: Teaching quality score and subscores**

Score Type	Score at SAS	Score at EL	Difference
<b>Teaching Quality Score</b>	<b>83.4</b>		
<i>Girl Survey Subscores</i>	77.2	79.6	2.4
Gender	88.2	88.0	-0.2
Subject	81.5	78.9	-2.6
Learning Environment	67.3	76.0	8.7
<i>Classroom Observation Subscores</i>	84.0		
Corporal Punishment	73.0		
Socioemotional Skills	81.3		
Instruction	82.4		
Teacher Attitudes	81.3		
Classroom Culture	97.0		
Classroom Culture, Disabilities	90.1		
Classroom Culture, Gender	85.8		

### **Girl Survey Subscores**

Considering the overall Girl Survey Subscore, scores increased by 2.4 percentage points between SAS and EL, indicating an improvement in the quality of teaching/instruction as reported by intervention girls. Overall, we may say that between SAS and EL, the quality of teaching and instruction delivered to girls improved across the board, especially that delivered to girls who report having a disability (see Annex 18).

**Table 51: Teaching quality score and constitutive subscores at SAS and Endline**

Score Type	Score at SAS	Score at EL	Difference
<b>Teaching Quality Score</b>	<b>83.4</b>		
<i>Girl Survey Subscores</i>	77.2	79.6	2.4
Learning Environment	67.3	76.0	8.7
Subject	81.5	78.9	-2.6
Gender	88.2	88.0	-0.2

Evaluating the three subscores from the girl survey, one recorded a large positive difference between the SAS and EL evaluations: the learning environment subscore. Breaking down the learning environment subscore into its six constitutive questions, we see that one question drove the learning environment score increase more than any other, which asked for students' agreement with the statement "my teachers are often absent from class/lessons". Considering the nature of the question and timing of the SAS round, responses to this question were most likely influenced by the COVID-19 pandemic, with schools struggling to repopulate their classes with teachers as schools reopened.

Moving to questions considering encouragement in the classroom, intervention girls were asked “If you don’t understand something, do teachers use a different language to help you understand?” Responses of “often” and “sometimes” correlate with a negative score, while “rarely” or “never” correlate with a positive score. This is because to increase students’ knowledge of and confidence in English, teachers are instructed to use only English, in order to create an environment where girls can best learn the language and in turn be prepared for the national exams. That said, as girls are also enrolled in a Kiswahili class, teachers are also instructed to use only Kiswahili in class, and not translate or explain concepts in dialect.

**Table 52: Girls’ view at SAS and EL on how teachers encourage students to understand and participate**

Question	Region	Options	SAS (%)	Endline (%)	Difference (%)
If you don’t understand something, do your teachers use a different language to help you understand?	ASALs	Often/ Sometimes	87.6	91.5	3.9
		Rarely/Never	12.5	6.6	-5.9
		Don’t know	0.4	1.9	1.5
	Urban Slums (US)	Often/ Sometimes	97.8	94.9	-2.9
		Rarely/Never	12.0	4.9	-7.1
		Don’t know	0.2	0.2	-0.1
	Total	Often/ Sometimes	87.6	92.9	5.2
		Rarely/Never	12.0	5.9	-6.1
		Don’t know	0.3	1.2	0.9

Overall, we find that teachers frequently used a different language to help students understand lessons at both SAS and endline, and that this rate increased by 5.2 percentage points at endline. As girls were home for much of 2019 and 2020 due to the COVID-19 pandemic, if girls were communicating with families and community members in languages other than English or Kiswahili, the rate at which the girls may struggle to fully understand in English or Kiswahili may have increased, leading to this result.

Disaggregating by region, the rate at which teachers used a different language in class increased in ASALs while decreasing in urban slums. This difference between the ASALs and urban slums may be the result of a higher presence of minority languages and dialects in ASALs, which may mean that girls are less likely to fully understand a lesson taught in English or Kiswahili in ASALs. As a result, future programmes or interventions may work with teachers in ASAL counties to emphasise the need to instruct the girls in only English or Kiswahili or develop language-specific programs to help fill gaps in ASAL girls’ English and Kiswahili proficiency.

Moving to the gender subscore, though the mean difference between the subscore at SAS and EL is nearly zero, breaking the subscore down by question reveals more variation. For the first question comprising gender subscore (G1)—responses to “My teacher treats boys and girls differently”—at SAS,

73.2 percent of intervention girls either “disagree[d]” or “strongly disagree[d]” with the statement, compared to 82.5 percent of intervention girls at endline, resulting in a positive difference of 9.4 percentage points between the two rounds. In sum, on average, teachers’ equal treatment of boy and girl students improved. This increase in the number of girls disagreeing with “My teacher treats boys and girls differently” is a positive indicator vis-a-vis IO2.

**Table 53: Gender subscore and constitutive questions, girl survey**

Score Type	Score at SAS	Score at EL	Difference
<i>Girl Survey Subscore</i>	77.2	79.6	2.4
Gender	88.2	88.0	-0.2
G1: “My teacher treats boys and girls differently” (% strongly disagree or disagree)	73.2	82.5	9.4
G2: “My teacher asks more questions to...” (% girls or equal number of girls and boys)	96.4	91.1	-5.3
G3: “My teacher asks harder questions to...” (% girls or equal number of boys and girls)	95.0	90.5	-4.6

Disaggregating by region, at SAS, 71.2 percent of girls enrolled in ASAL counties “disagreed” or “strongly disagreed” with the statement “my teacher treats boys and girls differently”, while 77.1 percent of girls from urban slums “disagreed” or “strongly disagreed” with the same. At endline, the number of girls from ASAL counties disagreeing with the statement increased by 7.3 percentage points, while that of girls in urban slums increased by 11.4 percentage points. With these changes in mind, though fewer girls from schools in ASAL and urban slum counties reported teachers treating boys and girls differently at endline, girls from schools in urban slum counties saw a larger positive change.

Considering the fact that we followed girls between rounds and not teachers, the difference in the change in girls’ responses to G1 in urban slum versus ASAL counties may be a result of higher rates of teacher attrition in ASAL counties versus urban slums, as noted by an implementing staff member.<sup>148</sup> With a higher rate of teacher turnover in schools located within the ASAL counties, girls in these schools may be less likely to have been taught by teachers who consistently received WWW project support and training over the past three years, compared to their peers in urban schools. Ultimately, this could be indicative of positive endorsement of WWW programme activities vis-a-vis teacher trainings.

Reviewing questions G2 and G3, we see that the rate at which girls responded with “girls” or “an equal number of boys and girls” decreased by 5.3 and 4.6 percentage points between SAS and EL, respectively. Looking at driving factors in the decline in G2’s score, disaggregating by region we find that the rate at which girls from urban slums selected “girls” or “an equal number of boys and girls” declined by 5.7 percentage points, while that of girls enrolled in schools in the ASAL counties decline by 19.5 percentage points. Though the difference in G3’s change was less pronounced when disaggregated by region, the rate decreased by 7.9 percentage points in ASAL counties, compared to a more modest 1.0 percentage point in urban slums. As hypothesised above, as these questions evaluate teachers’ behaviours, or at least girls’ perception of such behaviours, it follows that attrition of teachers likely drove the difference in the scores. As such, we suggest that a higher rate of teacher attrition resulting from the

<sup>148</sup> KII with Implementing Staff Member, Int. 88.

COVID-19 pandemic in ASAL counties likely contributed to the decline in students' rating of teachers' behaviours vis-à-vis boy and girl students. With this in mind, it's important to note that the effects of COVID-19 were largely beyond the control of the WWW project, and negative impacts or lack of improvements caused by the COVID-19 disruption are not indicative of programme impact.

Moving to the subject subscore, the mean subject subscore was 81.5 percent at SAS and declined by 2.6 percentage points, resulting in a subscore of 78.9 percent at endline. Situating the subscore's evaluation, a score of 100 percent was awarded if the girl said that there were no topics that the teacher needed to refocus on for the girl to better understand. In answering that there *were* such topics, the girl was given four subsequent questions about English, Kiswahili, mathematics, and science, and asked to identify the topics within each subject that the teacher needed to refocus on.<sup>149</sup>

**Table 54: Girl survey subject subscore and constitutive questions**

Score Type	Score at SAS	Score at EL	Difference
<b>Teaching Quality Score</b>	<b>83.4</b>		
<i>Girl Survey Subscores</i>	77.2	79.6	2.4
Subject	81.5	78.9	-2.6
Question	% of girls who said "yes" at SAS	% of girls who said "yes" at EL	Difference
Are there topics that you think the teacher needs to refocus for you to better understand?	78.6	81.9	3.3
English	13.4	12.0	-1.4
Kiswahili	8.0	6.7	-1.3
Math	57.4	55.1	-2.3
Science	44.0	48.5	4.5

Breaking the subscore down, we see that the percentage of girls responding that there are topics that the teacher needs to refocus on in order for the student to better understand increased by 3.3 percentage points. While the percentage of girls responding that the teacher needed to refocus topics in English, Kiswahili, or maths decreased by 1.4, 1.3, and 2.3 percentage points, respectively, the number of students reporting that teachers needed to refocus on topics in science increased by 4.5 percentage points.

In some ways, this is a positive finding; while it may suggest lower levels of understanding of science subjects, it also suggests high levels of interest in those subjects, and a desire to successfully learn. Indeed, throughout the focus group discussions, girls repeatedly cited science as the subject in which they were most interested.<sup>150</sup> For more on self-efficacy in learning attitudes toward STEM, please consult the *Attitudes Towards STEM* section.

<sup>149</sup> For each subject, the girl was given a set number of topics from which to choose from, as well as the option to add an "other" topic. As a result, there was the possibility for the girl to name four English topics, four Kiswahili topics, nine mathematics topics, and 41 science topics.

<sup>150</sup> FGD with Girls, Nairobi, Int. 38; FGD with girls, Nairobi, Int. 37; FGD with Girls, Nairobi, Int. 21; FGD with Girls, Nairobi, Int. 15; FGD with Girls, Samburu, Int. 25; FGD with Girls, Kilifi, Int. 34; FGD with Girls, Tana River, Int. 17.

Beyond science, though a reported need for a teacher to refocus any topic is likely associated with a lack of understanding, the identification that a topic needs refocusing can be indicative of a higher commitment to the subject. As girls' Learning to Learn (L2L) Life Skills scores increased across the board, it's likely that this recognition of a need for the teacher to refocus certain topics may come from students' increased commitment to the topic and learning in general. If this is the case, such an effect may be at least partially attributable to the success of programme mentorship programmes, as well as teacher training, as mentioned above.

### ***Classroom Observation Subscores***

Moving to the data collected through classroom observations in intervention schools, we do not compare results to prior rounds as observations were conducted for the first time at EL. Instead, we identify current strengths and weaknesses, which can in turn be used to inform future interventions.

The classroom observation subscore was comprised of seven components measuring use of corporal punishment, positive teaching practices, and classroom culture, among other dynamics. These scores range from a minimum of 0 to a maximum of 100 percent. Looking at the scores in the table below, we find the corporal punishment score is the lowest of the seven, at 73.0 percent, suggesting the greatest need for improvement in this area. We also notice that the teacher attitudes score, which was self-reported by teachers, is the second lowest score at 81.3 percent. We note that the corporal punishment score was also partially calculated from teacher self-responses, rather than solely from observations. As such, it seems that teacher attitudes may paint a more negative picture of realities within the classroom than measurements taken during isolated classroom observations.<sup>151</sup>

**Table 55. Classroom observation subscores**

Score Type	Score at EL (%)
<b>Teaching Quality Score</b>	<b>83.4</b>
<i>Classroom Observation Subscores</i>	84.4
Corporal Punishment	73.0
Socioemotional Skills	81.3
Instruction	82.4
Teacher Attitudes	81.3
Classroom Culture	97.0
Classroom Culture, Disabilities	90.1
Classroom Culture, Gender	85.8

Beginning with the teacher attitudes subscore, teachers were asked thirteen questions, included below. Looking at the question-level scores, low levels of disagreement with the statement “boys and girls are better at different subjects” had a large influence over the overall subscore. Disaggregating by region, 31.3 percent of teachers in ASAL counties disagreed with the statement, compared to 26.2 percent of teachers from urban slum counties. Looking at the teacher KIs, no teacher said that boys and girls are

<sup>151</sup> Classroom observations may also be somewhat biased, as teachers are more likely to utilize positive practices and avoid negative practices when under observation.

inherently better at different subjects, but the refrain that the two are succeeding in different subjects was common across all counties.<sup>152</sup>

**Table 56: Classroom observation teacher attitudes subscore and constitutive questions**

Question	Score at EL (%)
<b>Teaching Quality Score</b>	<b>83.4</b>
<i>Classroom Observation Subscores</i>	84.4
Teacher Attitudes	81.3
Question	Score (%)
Coaching is an additional burden to my schedule	9.2
I desire frequent coaching support and professional feedback from the CSO	98.7
Schools should be supported to build their internal capacities for teacher coaching.	98.5
I have the same expectations for boys and girls in my class	84.2
Boys and girls are better at different subjects	71.3
It is important to design lesson and teaching strategies that are gender sensitive	86.9
When I see a student is struggling, I adjust my lesson to help them	94.9
When a student is struggling, they just need to work harder	39.0
I can identify when a student is struggling before they take an exam	95.6
I use the same strategies to teach the students are performing well and the students who are struggling	26.3

Moving to the corporal punishment score, enumerators reported witnessing students being “physically disciplined” in 19.1 percent of the classrooms where observations were conducted. This was corroborated in FGDs, in which girls in some schools in all counties reported physical punishment, mostly through caning (on the hand or buttocks), as well as assignment of physical labour as a punishment.

While girls were observed to be physically disciplined more than boys, eight of these schools in which corporal punishment was observed were all girl schools, and enumerators did not observe any classroom where boys were punished and not girls. This suggests that there may be little difference in the use of corporal punishment with boys as opposed to girls.

<sup>152</sup> KII with Teacher, Kilifi, Int. 28; KII with Teacher, Mombasa, Int. 30; KII with Teacher, Nairobi, Int. QUESTIONS W TRANSLATIONS; KII with Teacher, Nairobi, Int. QUESTIONS W TRANSLATIONS (1); KII with Teacher, Nairobi, Int. 35; KII with Teacher, Nairobi, Int. 7; Samburu, Int. 11; KII with Teacher, Int. 29; KII with Teacher, Turkana, Int. 42; KII with Teacher, Turkana, Int. 4; KII with Teacher, Turkana Int. 5; KII with Teacher, Tana River, Int. 12; KII with Teacher, Tana River, Int. 41; KII with Teacher, Tana River, Int. 27.

**Table 57: Corporal punishment scores**

Observed	Boys were physically disciplined	Girls were physically disciplined	Students were physically disciplined
Percent of classroom observations where students were physically disciplined in class.	14.5%	19.1%	19.1%
Self-reported Attitudes (Teachers)	Response		% of classrooms
Corporal punishment is sometimes necessary in the classroom	Agree/Agree strongly		24.5%
	Disagree/Disagree strongly		75.5%
Corporal punishment is never acceptable	Agree/Agree strongly		71.9%
	Disagree/Disagree strongly		28.1%
Corporal punishment slows down learning	Agree/Agree strongly		69.7%
	Disagree/Disagree strongly		30.3%

Moving to teachers' self-reported attitudes on punishment, across all three questions regarding teachers' attitudes vis-à-vis corporal punishment, the majority of teachers disagreed that corporal punishment was sometimes necessary and agreed that corporal punishment is unacceptable and slows down learning.<sup>153</sup> This suggests mostly positive attitudes toward this negative practice.

Disaggregating by county, in response to the latter two corporal punishment statements, "corporal punishment is never acceptable" and "corporal punishment slows down learning", teachers from Turkana recorded the highest rate of negative responses across the six counties, with 47.7 percent of Turkana teachers disagreeing with the second statement, and 44.7 percent of Turkana teachers disagreeing with the third. While the two focus group discussions conducted with girls from Turkana revealed that girls in both schools experience corporal punishment, Turkana girls were not alone as girls in *all* FGDs reported experiencing corporal punishment. That said, girls from three different schools across Kilifi, Nairobi, and Mombasa reported some kind of recourse when subjected to corporal punishment. Specifically, an FGD with girls in Kilifi revealed that students can refuse corporal punishment, but in doing so they are given a letter to take home to their parents, and the parents can choose to return it with a note allowing for the punishment.

In sum, it is clear that corporal punishment is still occasionally employed in schools across all counties of interest. With future interventions in mind, we suggest that emphasis on the absolute inappropriateness of corporal punishment of any kind be centred in teacher training.

#### **Box 4: Improvement in Teaching Practices**

The endline evaluation suggests that strengthened teaching practices have been a key outcome of the WWW programme. This is an important outcome as positive teaching practices not only have an impact on learning, but also on the sustainability of outcomes as teachers are likely to continue using positive practices beyond the close of the programme. In this box, we highlight several teachers' positive experiences using new teaching practices.

In Turkana, a young biology and chemistry teacher stated that he had substantially changed his mode of teaching over the past three years. When he began teaching, he had focused primarily on syllabus

<sup>153</sup> A chart displaying teacher attitudes vis-a-vis corporal punishment by county is included in the annex.



coverage, but came to realize that moving quickly through the syllabus left many learners behind. Now, he uses a learner-centred approach to try to engage slower learners, in part through the use of simpler terminologies and more in-depth explanations. He attributed this change to both trainings received from the WWW programme and government-held annual trainings.<sup>154</sup>

In Nairobi, an experienced mathematics teacher similarly described substantial changes to his teaching practices due to training he had received. One key change was in the way he encouraged and motivated his students to study mathematics and related mathematics lessons to real life. Due to these changes, he stated that, “most children tend to think that mathematics is a very difficult subject, but through the new methods of teaching, they have improved their attitude... [and] have now started seeing that mathematics is just a normal subject like any other subject”. Additionally, this teacher stated that he had found trainings on gender sensitivity particularly helpful as a male teacher; these trainings allowed him to improve his interactions with female students so that they would feel included in the class.<sup>155</sup>

Lastly, in Mombasa, an experienced physics teacher stated that he now uses hands-on activities to engage students, explaining that because his subject area could be very abstract, demonstrations allowed students to understand things more clearly than using the conventional method of lecturing. For example, he described a project where students produced bio-gas, which gave the students a tangible demonstration of and experience with alternative sources of energy. He stated that these projects not only helped students understand the subject, but also helped them understand its applicability to real life.<sup>156</sup>

### Summarizing IO2 Plausibility vis-a-vis the Girl Survey and Classroom Observations

**Table 58: Summary Chart of IO2 Plausibility**

Criteria	Assessment
Plausibility: Is the theory of change plausible?	Moderate. The theory behind IO2 focuses on girls' demonstration of improved performance vis-a-vis the assessments. While this is perhaps the most concrete way to measure improved teaching and instruction, provision to include girls' perceptions of teaching quality and instruction, and provision to re-contact and monitor changes in these girls' responses, should ultimately inform measurement of IO2 as well.
Evidentiary confirmation of key elements: To what extent are the key elements of the theory of change confirmed by new or existing evidence.	Moderate. Qualitative data suggests some link between IO2 and improvement in learning assessment scores.
Identification and examination of other influencing factors: To what extent have other influencing factors been identified.	Moderate. Through KIIs and FGDs, girls and teachers alike note the presence of other programmes and interventions within their schools. Still, as correlation between teacher attrition in ASAL counties seems to correlate with girls reporting lower levels of teacher quality, it seems that, through an inverse logic, where teachers have remained in school, the WWW programme teacher training has had a positive impact.

<sup>154</sup> KII with teacher, Turkana, Int. 42

<sup>155</sup> KII with teacher, Nairobi, Int. 29

<sup>156</sup> KII with teacher, Mombasa, Int. 30

## Recommendations

In light of the findings from the Girl Survey and Classroom Observations, we present three recommendations. The first, considering the noted differences between ASAL and urban slum girls' reported comfort with English instruction, future programmes or interventions may work with teachers in ASAL counties to emphasise the need to instruct the girls in only English or Kiswahili or develop language-specific programs to help fill gaps in ASAL girls' English and Kiswahili proficiency.

In addition, during the COVID-19 pandemic, teacher attrition proved higher in ASAL counties than those of urban slums. As such, the number of trained teachers working in ASAL schools at EL was lower than that of those working in urban slum schools. With this differential in mind, future interventions may take care to diversify intervention focus by county, accounting for the difference in the level of training completed by teachers after this intervention round. For instance, teacher training that includes gender sensitization may be emphasised in ASAL counties, specifically counties like Samburu.

What's more, it is clear that corporal punishment is still employed in schools across all counties of interest. With future interventions in mind, we suggest that emphasis on the absolute inappropriateness of corporal punishment of any kind be centred in teacher training. In addition, rates of observed corporal punishment varied by the gender make-up of the school. This is something to note for future interventions.

## Life Skills

### Summary of Findings

1. At endline, life skills, as measured by the standard GEC Life Skills Index, were very high with an overall score of 87.9 out of 100. This exceeds the endline target of 81.0.
2. Life skills increased slightly from SAS to endline, driven by improved attitudes towards success in school, continued pursuit of education, and self-confidence.
3. However, we find a significant and negative relationship between whether a girl is disabled and some life skills. This suggests a need to continue strengthening the confidence and learning-related attitudes of girls with disabilities.

This section analyses differences in girls' decision-making and life skills between the SAS and EL rounds. These categories are mapped to the WWW project's third intermediate outcome (IO3). IO3 is described as investigating whether "marginalised girls have increased self-efficacy and life skills". From this description, the outcome's indicators are twofold. The first—and the concern of this section—is "the percent of adolescent girls demonstrating improved-self efficacy" (IO3.1). The second indicator analyses girls' demonstration of "improved knowledge, attitudes, and practices on gender and sexual and reproductive health (SRH)" (IO3.2). As this report includes specific sections on SRH and child protection, to avoid redundancy, only the first indicator (IO3.1) is evaluated in this section.

**Table 59: IO3 Description and Assumption**

Description	Assumption
Intermediate Outcome 3	Marginalised adolescent girls have increased self-efficacy and life skills. Measured by 1) % of adolescent girls demonstrating improved self-efficacy (IO3.1) and 2) girls demonstrating improved knowledge, attitudes and practices on gender and SRHR (IO3.2).
IO3 Assumption	<ul style="list-style-type: none"> <li>- Community engagement and mobilisation efforts are sufficient to mitigate resistance, stimulate demand for education, and enable girls' participation in all components of the programme.</li> <li>- Project content is engaging and meaningful to the girls</li> <li>- The content within the mentorship policy will transform the learners and teachers to better understand gender equity and social inclusion.</li> <li>- The girls will be open to discuss their knowledge on issues SRH during the evaluation sessions (IO3.2).</li> </ul>

To evaluate IO3.1 we look at girls' reported Life Skills scores. At both SAS and EL, decision-making and life skills were measured by girls' perceptions of self, as reported through the girl survey. Within the girl survey is a section of questions specifically relating to life skills and decision-making. This section has thirty-one questions, twenty-four of which are adapted from the Life Skills Index (LSI).

### **Life Skills Index**

The LSI is an index that is standardised across most Girls' Education Challenge programmes. The complete LSI is split into three indices as defined by the UK FCDO. The "learning to learn" (L2L) index consists of seven questions about self-confidence in the classroom. The "learning for life" (L4L) index consists of ten questions about education and sources of motivation. The third index, the "agency" index, is comprised of eight questions about the girl's role in decision-making and general agency. Higher scores indicate a high level of overall life skill.

Twenty-four LSI questions were asked at both the SAS and endline evaluations to intervention students. In terms of respondents, 837 girls were asked the twenty-four LSI questions at both the SAS and EL round, comprising this section's population of concern.<sup>157</sup>

**Table 60: Difference between mean LSI scores and subscores at SAS and EL**

Standardised Scores / Subscores	SAS	EL	Difference
<b>Life Skills Index</b>	86.4	87.9	1.4
Learning to Learn	83.7	88.7	4.0
Learning for Life	87.5	90.2	2.8
Agency	87.5	83.9	-3.6

<sup>157</sup> A breakdown of the constitutive LSI questions, along with the methodology used to calculate the scores and subscores, are included in the annex.

As evidenced in Table 60 above, the overall difference in girls' LSI scores was insignificant but positive, with a positive difference of 1.4 percentage points between the SAS and EL evaluation rounds. This suggests that between SAS and EL, the goal of adolescent girls demonstrating improved-self efficacy (IO3.1) was met.

To better understand this increase in self-efficacy, we look at the three constitutive subscores. We find increases in the learning to learn (L2L) and learning for life (L4L) subscores and a decrease in the agency subscore. Between the SAS and EL evaluation rounds, the L2L subscore recorded the greatest increase of the constitutive life skills subscores, increasing by 4.0 percentage points. In terms of program activities, we can cautiously assume that this increase may be correlated with a general improvement in intervention schools' environments, aided through teacher training and access to school materials. In addition, as many girls expressed the positive impact of having their school fees partially or wholly covered by the WWW project, it is possible that the removal of stress generated by needing to meet school fees allowed girls to become more excited about learning for learning itself.

Disaggregating by county, the increase in the L2L subscore was spread across all six counties, but especially amplified in Mombasa County where the subscore recorded a 11.2 percentage point increase. In contrast to Mombasa, the lowest county difference between the subscore at SAS and EL was recorded in Kilifi, with a positive difference of 1.0 percentage points.<sup>158</sup>

Breaking the L2L subscore down to the question level, both across counties and in Mombasa it appears that the question driving the increase in the L2L subscore is that asking for agreement vis-à-vis the question "When I succeed at school it is because I worked hard". Across characteristics, the overall rate at which girls responded to this question with either "agree" or "strongly agree" increased by 8.5 percentage points. In Mombasa, the increase observed across characteristics nearly doubled, increasing by 16.7 percentage points between the two rounds. The rate of agreement also increased by a large margin in Tana River. Considering potentially impactful programming, the implementation of mentorship programs may play a leading role in making girls believe that when they succeed at school it is because they worked hard.<sup>159</sup>

The statement "I can stay focused on a goal despite things getting in the way" also recorded a modest (but second largest) score increase between SAS and EL. Specifically, the percent of girls agreeing with the statement increased by 1.2 percentage points. Along with the above results, we hypothesize that improvements in attitudes towards hard work and focus on aspirations may have been affected by the opportunity to engage with older girls or women who have achieved "success at school" and are able to share the connection between their experience at school and success. As such, it seems likely that the implementation of mentorship programmes may have played a positive and potentially significant role in approaching and ultimately meeting IO3.1.

Looking now at the "learning for life" (L4L) subscore, we find that this subscore also increased between the SAS and EL rounds, recording a positive difference of 2.8 percentage points. Disaggregating by county, this increase was largely driven by increases in Samburu and Tana River counties. Having the two lowest aggregate L4L subscores at SAS, the increases seen in Samburu and Tana River between the two rounds is not only remarkable for the size of their increase, but for effectively moving the counties from the two counties with the lowest L4L subscore to the two counties with the highest.

---

<sup>158</sup> Chart depicting the L2L subscore by county is included in the Annex.

<sup>159</sup> Chart depicting the Difference between L2L question scores between SAS and EL, Mombasa County and overall is included in the Annex.

**Table 61: L4L subscore by county**

<i>L4L Subscore</i>			
Characteristic	SAS	EL	Dif.
Counties			
Kilifi	89.5	89.8	0.3
Mombasa	91.0	89.2	-1.8
Nairobi	85.9	87.9	2.0
Samburu	86.3	92.9	6.6
Tana River	83.5	91.7	8.2
Turkana	91.9	92.2	0.2

In trying to understand what drove these notable increases in Samburu and Tana River's scores we break the subscore down to the question level. In doing so, we see that both Samburu and Tana River recorded notable increases in girls' responses to "I ask the teacher if I don't understand something", with the rate of positive responses from respondents increasing by 11.6 and 9.6 percentage points, respectively, between SAS and EL. Along these lines, teachers from Samburu and Tana River shared that their teaching practices have changed over the past year to include more integrated teaching methods.<sup>160</sup> Specifically, one teacher from Tana River shared:

*I feel more confident [in my teaching] because the students are more open to me and ask questions and some even volunteer to answer questions in class. I feel more confident because I have more strategies on how to teach. The change has been brought by the training that I got from EDT and the Ministry.<sup>161</sup>*

In addition, teachers from Samburu and Tana River noted integrating "peer teaching" practices whereby they allow students to help each other as the students "understand each other more than teachers".<sup>162</sup> Creating an environment where students are free to collaborate and problem-solve amongst themselves likely contributes to the creation of an environment whereby students generally feel more comfortable and confident in working through problems. As such, it makes sense that if a girl can pursue learning and solutions amongst her peers, when the group is unable to figure out a question or concept, they are more confident when asking the teacher for help. Further, gaining confidence in one's ability to problem solve either individually or with a group of peers is a valuable skill to have beyond the classroom, and clearly correlates with the idea of improved self-efficacy (IO3.1).

<sup>160</sup> KII with Teacher, Samburu, Int. 12; KII with teacher, Samburu, Int. 25; KII with Teacher, Tana River, Int. 6; KII with Teacher, Tana River, Int. 41.

<sup>161</sup> KII with Teacher, Tana River, Int. 41.

<sup>162</sup> KII with Teacher, Samburu, Int. 12.

**Table 62: Difference between L4L question scores between SAS and EL, Samburu and Tana River counties**

Question	Samburu	Tana River
I would like to continue studying / attending school after this year.	1.3	0.8
I am aware of the different transition pathways in the education system.	6.5	7.2
I am willing to transition throughout the education pathways until I meet my goal.	1.3	-0.4
I am aware of the alternative pathways of education (TVETs, apprenticeships, etc.)	2.2	10.0
I would like to continue learning a vocation or trade	10.8	6.4
I can put a plan in place and stick with it.	-0.4	2.0
I recognise that choices I make today about my studies can affect my life in the future.	-5.6	2.0
I study well in a group with other peers during the non-school hours.	-2.2	5.2
When I have/had the opportunity, I can / could organise my peers or friends to do an activity.	-1.3	-6.4
I ask the teacher if I don't understand something.	11.6	9.6

In Samburu, the overall increase in the L4L subscore was also led by a 10.8 percentage point increase in positive responses to the question “I would like to continue learning a vocation or trade”, while the L4L question that saw the largest increase in positive responses from Tana River was “I am aware of the alternative pathways of education (TVETs, apprenticeships, etc.)”. It is worth noting that Samburu and Tana River were the *only* counties to record an increase in response to either question.

Additionally, we note that disability status was significantly correlated with negative L4L and L2L subscores. In other words, girls with disabilities tended to have lower L4L and L2L scores than girls without disabilities. This is not entirely surprising, given the substantial barriers that girls with disabilities may face to achieve their goals in both education and life, and suggests a need to continue focusing on girls with disabilities in interventions targeting life skills.

Looking now at the agency subscore, the 3.6 percentage point decline in this subscore was largely driven by declines in response to two questions. The first, “Who decides whether or not you will go back to school?” declined by 4.3 percentage points between the two rounds. The second, “Who decides whether or not you will continue in school past this year?” declined by 6.0 percentage points between the two rounds, more than any other question constituting the agency subscore.

A decline in the agency subscore questions means that the share of girls responding to one of the seven subscore questions with “my family decides for me” increased while those reporting “I decide” or “I decide jointly with my family” decreased. It appears that this negative change corresponds with the girls’ grade in school. For instance, the mean change in question scores between the two evaluation rounds for girls from Class 7 was positive. In contrast, girls in Form 3 saw the greatest negative change between the two rounds, declining by 13.9 and 11.1 percentage points, respectively.

It appears that parents of girls in primary school support their daughters’ aspirations to transition to secondary school. However, once girls reach secondary school, no initiatives exist on the scale of the

WWW and MoE transition initiatives to urge parents to continue to support their girls' aspirations in school. As such, as girls get older and their parents' wish for them to leave school and marry or work contradicts the girls' hope to remain in school, the differing will of the parent vis-à-vis the daughter becomes clear, and the prevalence of girls reporting "my family decides for me" grows.

Disaggregating the agency subscore by county, Kilifi, Mombasa, and Nairobi saw very little changes to their aggregate agency subscore between the two rounds.<sup>163</sup> In contrast, Samburu, Tana River, and Turkana saw notable negative changes, with those of Samburu and Turkana declining by 9.9 and 12.1 points, respectively. This may relate to cultural practices and attitudes in these ASAL counties.

If these hypotheses regarding the decline in the agency subscore between the SAS and EL rounds are correct, they are not purely negative. Though the assumption is that parents are in disagreement with the girl, deciding that she should leave school or finish school at an earlier age, such an assumption also means that the girl wishes to stay in school. With this in mind, it's important to note that given the questions we asked the girls, we don't know if the parents are disagreeing with the girl about her education timeline and more actively voicing their plan for the girl to leave school as the girl matures, or if the girl extended her education timeline between SAS and EL and now hopes to remain in school longer, even transitioning to tertiary-level education. If the FGDs with girls enrolled in school in Turkana are any indication, all of them would like to complete secondary school and transition to university.<sup>164</sup>

Considering the notable decline across the two agency subscore questions noted above, a decline generally correlated with girls' increasing ages, we recommend that future projects consider programmes to increase household support for girls' education beyond the initial transition from primary to secondary school. In addition, considering the stark contrast between girls' responses from Samburu and Turkana to those of other girls, future projects could consider programs aimed at educating communities on post-secondary education opportunities for girls and establishing clear pathways for access, e.g., transportation routes and accommodation, where applicable.

### **Summarizing IO3.1 Plausibility**

**Table 63: Summary Chart of IO3.1 Plausibility**

Criteria	Assessment
Plausibility: Is the theory of change plausible?	High. The underlying assumptions behind improving girls' reported self-efficacy follows conventional thinking, as evidenced by the use of the Life Skills Index to gather and measure responses to IO3.
Evidentiary confirmation of key elements: To what extent are the key elements of the theory of change confirmed by new or existing evidence.	Moderate. There appears to have been a small increase in LSI scores from SAS to EL. Results varied by county.
Identification and examination of other influencing factors: To what extent have other influencing factors been identified.	Moderate. Through KIIs and FGDs, girls and teachers alike noted the presence of other programmes and interventions within their schools.

<sup>163</sup> Girls from Marsabit and Kwale were not asked the 24 LSI questions at SAS and as such their responses given at EL are not considered in this section.

<sup>164</sup> FGD with Girls, Turkana, Int. 16; FGD with Girls, Turkana, Int. 21.

## Recommendations

The Life Skills Index Score improved by 1.4 percentage points between SAS and EL, indicating an increase in “the percent of adolescent girls demonstrating improved-self efficacy” (IO3.1). In order to continue this positive trend in future intervention rounds, we recommend that future projects consider programmes to increase household support for girls’ education beyond the initial transition from primary to secondary school. Specifically, we hypothesise that the implementation of programmes designed to advocate for girls’ education beyond secondary school, at the community level, may help to increase girls’ agency subscores and as a result, girls self-efficacy scores (IO3.1).

## Household Support

### Summary of Findings

1. At endline, we calculate an overall Household Support Score of 82.1. This score falls slightly below the programme’s target of 88.0, but is still notably high.
2. While household support may have been somewhat affected by COVID-19 due to the stresses placed on households’ economic standings, the overall target was still a realistic, though ambitious, goal for endline.
3. Overall, we find high levels of household support for girls to remain in education and pursue future schooling. Caregivers also highly value girls’ education and actively involve girls in discussions and decisions around their education and future career.
4. While household support was strong across all counties, it was highest in Turkana and Tana River and lowest in Kilifi and Mombasa. In all four of these counties, the variation in household support was mostly driven by differing perceptions on when it is acceptable for a girl not to attend school; in Turkana, for example, caregivers stated that there are almost no acceptable reasons for a girl not to attend school.
5. Over 90 percent of caregivers want their girl to achieve college/TVET or university. However, only around half of caregivers said they would be able to support this level of education, reiterating the major challenge of economic barriers to continuing education and again highlighting the relevance of WWW scholarship and bursary activities.

This section analyses caregivers’ support for the transition of girls into educational pathways, as described by the WWW project’s fourth intermediate outcome (IO4). In order to best understand caregivers’ attitudes at endline, we analyse caregivers’ responses to eighteen questions at endline with the goal of understanding caregivers’ support for girls’ education, using that understanding to inform recommendations for future interventions.<sup>165</sup>

Considering means of endline analysis, the Household Support Score is our overall score, composed of eighteen questions. To better understand household support, the overall Household Support Score is broken down into four subscores: the Leave School Subscore (LSS), Future School Subscore (FSS), Girl’s Will Subscore (GWS), and Value Girl’s Education Subscore (VGES). The four subscores are composed of eleven, three, two, and two questions, respectively, and were balanced when feeding into the Household Support Score.

<sup>165</sup> Scores are for caregivers of intervention students only.



**Table 64: Household support score and subscores**

Score Type	Score at EL
<b>Household Support Score (HSS)</b>	<b>82.1</b>
Leave School (LSS)	76.8
Future School (FSS)	88.0
Girl's Will (GWS)	89.5
Value Girl's Education (VGES)	95.1

Looking at the table above, the overall Household Support Score was 82.1 percent at endline. Looking at the subscores, we see that at 76.8, the Leave School Subscore drove the Household Support Score down, countered by the Value Girl's Education Subscore at 95.1 percent, which drove the overall score up. Between LSS and VGES, the Future School and Girl's Will Subscores middled out the overall score, coming in at 88.0 and 89.5 percent, respectively.

**Table 65: Household support score and subscores by county, endline**

County	Household Support Score	LSS	FSS	GWS	VGES
<b>Kilifi</b>	76.2	70.8	91.1	85.5	95.4
<b>Kwale</b>	83.7	79.8	90.8	89.1	92.2
<b>Mombasa</b>	77.1	68.5	90.7	90.4	96.2
<b>Nairobi</b>	82.3	75.3	86.6	93.5	95.5
<b>Samburu</b>	84.0	83.3	86.0	87.7	86.8
<b>Tana River</b>	86.0	81.2	82.4	90.7	96.7
<b>Turkana</b>	89.5	87.1	94.3	84.5	98.3

\* Marsabit not included due to low sample size.

Disaggregating by county, at endline, we see that the Household Support Score varied by 13.3 percentage points across the eight counties, peaking in Turkana with a score of 89.5 percent, and bottoming in Kilifi with a score of 76.2 percent. In all counties, the Value Girl's Education Subscore pulls the score up, in line with pre-disaggregate trends. The Leave School Subscore was also lowest in almost every county, with the exception of Turkana. This is discussed further below.

### **Leave School Subscore**

The Leave School Subscore is characterised by eleven possible points, giving it the heaviest weight within the Household Support Score. Specifically, the caregiver was asked "under which of the following conditions do you think it is acceptable for a child to not attend school?" and then given ten scenarios, plus the opportunity to "think of any other condition where you think it is acceptable for a girl to not attend school". The ten stated scenarios are included in the table below.

**Table 66: Leave school subscore and constitutive questions, endline**

Acceptable for a girl to not attend school	Overall	Kilifi	Kwale	Mombasa	Nairobi	Samburu	Tana River	Turkana
LS1: Child may be physically harmed or teased on way to school	18.6%	24.8%	31.3%	23.1%	20.1%	9.4%	18.7%	0.0%
LS2: Child may be physically harmed or teased at school	14.6%	22.7%	25.0%	19.2%	13.0%	9.4%	14.7%	0.0%
LS3: Child needs to work	14.4%	25.5%	3.1%	19.2%	15.6%	7.6%	4.0%	8.6%
LS4: Child needs to help at home	16.3%	30.5%	3.1%	30.8%	16.1%	7.6%	5.3%	8.6%
LS5: Child is married	11.2%	20.6%	6.3%	11.5%	11.6%	7.6%	5.3%	0.0%
LS6: Child is too old	14.6%	28.4%	6.3%	26.9%	15.6%	7.6%	2.7%	0.0%
LS7: Child has physical or learning needs	16.8%	29.1%	6.3%	26.9%	18.3%	5.7%	6.7%	8.6%
LS8: Child is unable to learn	11.8%	19.9%	9.4%	19.2%	12.5%	11.3%	4.0%	0.0%
LS9: Education is too costly	17.6%	31.2%	9.4%	34.6%	15.2%	18.9%	5.3%	8.6%
LS10: Child is a mother	19.7%	35.5%	21.9%	34.6%	18.3%	11.3%	9.3%	3.5%
LS11: Other	17.5%	17.0%	9.4%	7.7%	23.2%	13.2%	25.3%	1.7%
Overall LSS	76.8	70.8	79.8	68.5	75.3	83.3	81.2	87.1

\* Marsabit not included due to low sample size.

Looking at the eleven situations, the two questions with the highest levels of agreement (i.e., lowest scores) are “The child may be physically harmed or teased at school or on the way to/from school” and “The child is a mother”. For both questions, respondents from Kwale drove the questions’ low scores, with 31.3 percent and 21.9 percent of respondents from Kwale, respectively, believing that the two reasons were “acceptable” reasons for the girl to leave school.

While the two question scores from Kwale stand out, those from Kilifi are also low, with 24.8 percent and 35.5 percent of caregivers from Kilifi believing that the threat of physical harm or teasing at or on the way to/from school is an acceptable reason for a child to leave school, as well as if the child is a mother. While Kilis with teachers don’t speak to caregivers’ feelings vis-à-vis a child who is a mother and her attendance in school, a teacher from Kilifi noted that “many young people [in Kilifi] indulge in early sexual activities” and that if a student gets pregnant, the parent may “not...want their child to continue learning”.<sup>166</sup> In turn, the teacher added that the situation requires a specific intervention because “the girls feel that education is not very important to them” which “affects our work here as teachers”.<sup>167</sup>

We might expect caregivers in Turkana to believe that the possibility of the child being harmed “on the way to/from school” would be considered unacceptable at a high rate, given the insecurity in Turkana and distances that students must often travel in order to reach school. Instead, 0 percent of caregivers from

<sup>166</sup> KII with Teacher, Kilifi, Int. 26.

<sup>167</sup> *Ibid.*

Turkana that considered the threat of physical harm an “acceptable” reason to miss school. This may lead us to believe that the more routine the threat of physical violence on the way to/from school, the less it presents as a threat, instead being accepted as part of daily life. In addition, this example shows that the threat of physical violence is not an absolute reason for a child to miss school, and that in some places, the value of education may overpower the threat of violence.

### Future School Subscore

With an overall subscore of 88.0 at endline, the future school subscore is composed of three questions.<sup>168</sup> With a question score of 92.9 percent, the question asking caregivers “What level of schooling *would you like [girl]* to achieve?” recorded the highest future school question score. On the other end of the spectrum, at 46.8 percent the question asking “What level of schooling *are you able* to support your *[girl]* in achieving” received the lowest question score.

**Table 67: Future school subscore and constitutive questions, endline**

	Overall	Kilifi	Kwale	Mombasa	Nairobi	Samburu	Tana River	Turkana
FS1: Level of school would like girl to achieve (% college/TVET or university)	92.9% <sup>169</sup>	96.5%	87.5%	92.3%	94.7%	90.6%	88.0%	100.0%
FS2: Level of school willing to support girl to achieve (% college/TVET or university)	85.4%	95.8%	90.6%	88.5%	88.4%	50.9%	77.3%	96.6%
FS3: Level of schooling able to support girl to achieve (% college/TVET or university)	46.8%	61.7%	34.4%	73.1%	51.3%	22.6%	41.3%	19.0%
Overall FSS	88.0	91.1	90.8	90.7	86.6	86.0	82.4	94.3

\* Marsabit not included due to low sample size.

Disaggregating by county, excluding Marsabit which had a very low sample size, at 88.0 percent Tana River had the lowest FS1 score. As Tana River has been heavily impacted by drought over the past couple of years, resulting in a wave of extreme poverty, the relatively low level of agreement with the FS1 statement regarding the level of schooling the caregiver would *like* their girl to achieve may be the result of wanting the girl to begin working in order to help sustain and support herself and the household. This potential relationship between insecurity and support is important to note. Though the Theory of Change recognises that the girls of concern come from marginalised communities, differences in type and degree of insecurity continue to affect caregivers’ abilities to voice support for girls’ education.

Looking at FS2 which asks, “What level of schooling would you be willing to support your *[girl]* to achieve?” at 50.9 percent Samburu had the lowest FS2 score. During a focus group discussion with girls from Samburu, one girl stated, “In our community we still have a culture where parents don’t take young kids to school, they go for grazing”.<sup>170</sup> Commenting on the difficulties of the culture in Samburu vis-à-vis girls’ education, one teacher said, “I understand much is hard for the girls in this community to change

<sup>168</sup> These questions are included in the table below, as well as the Annex under “Calculating the Future School Subscore”.

<sup>169</sup> In 2018, the Baseline Report recorded that 93.6 percent of surveyed caregivers whose girls were enrolled in an intervention school said they would like their daughter to achieve “College / University”.

“GEC - T Baseline Report,” Education Development Trust, May 2018, page 114.

<sup>170</sup> FGD with Girls, Samburu, Int. 40.

their culture”.<sup>171</sup> With these thoughts in mind, it seems that a mix of the labour demands of pastoralism, plus general cultural attitudes towards (or against) girls’ education may be driving the relatively low willingness to support girls in pursuing college, a TVET program, or university.

Considering FS3 which asks, “What level of schooling are you able to support your [girl] to achieve?” at 34.4 percent, caregivers from Kwale comprised the lowest question score. On the other side of the spectrum, at 73.1 percent and 61.7 percent respectively, caregivers from Mombasa and Kilifi responded to FS3 with “Collage/TVET” or “University” at a higher rate than caregivers from any other county. As a high percentage of caregivers who are *willing* to support their child in achieving college, TVET, or university are also *able* to do so, it appears that compared to other counties, the economic position of caregivers who are willing to support their daughter in achieving higher levels of education is relatively strong.

### **Girl’s Will Subscore**

The two questions comprising the Girl’s Will Subscore are “Do you discuss with [girl] what she wants to become in the future?” and “Do you listen to the views of [girl] when you make decisions about her education or are those decisions made by adult members of the family only?”<sup>172</sup>

**Table 68: Girl’s will subscore and constitutive questions, endline**

	Overall	Kilifi	Kwale	Mombasa	Nairobi	Samburu	Tana River	Turkana
Overall GWS	89.5	85.5	89.1	90.4	93.5	87.7	90.7	84.5
GW1: Discuss girl’s future goals	86.9%	83.7%	81.3%	92.3%	93.8%	84.9%	88.0%	70.7%
GW2: Listen to girl’s opinion	91.4%	87.2%	96.9%	84.6%	92.9%	90.6%	92.0%	96.6%

\* Marsabit is not included due to low sample size.

Looking at the Girl’s Will Subscore’s constitutive questions disaggregated by county we see that Turkana drove down GW1 with a question score of 70.7 percent, while Mombasa and Kilifi reported the lowest GW2 question scores, 84.6 percent and 87.2 percent respectively.

Considering GW1 which asks, “Do you discuss with [girl] what she wants to become in the future?” Turkana’s low score may be due to the fact that much of the county is made up of pastoralist families. As a result, one teacher from Turkana noted that parents are often not home. As a result, the low GW1 score may be the result of parents not being home to discuss the girl’s future with the girl.<sup>173</sup> With this in mind, future interventions may include an intervention at the household level, priming heads of households to discuss their girl’s future with the girl and creating a kind of framework or agreement vis-à-vis her goals and how the household will commit to supporting her in achieving them.

### **Value Girl’s Education Subscore**

At 95.1 percent, the Value Girl’s Education Subscore is the highest of the four subscores. The VGES is composed of two questions, “Even when funds are limited it is worth investing in [girl’s] education” and “A girl is just as likely to use her education as a boy”. The average VGES was 95.1 at endline, a very high score.

<sup>171</sup> KII with Teachers, Samburu, 39.

<sup>172</sup> The specifics behind calculating the subscore are included in the Annex under “Calculating the Girl’s Will Subscore”.

<sup>173</sup> KII with Teacher, Turkana, Int. 5.

**Table 69: Value girl's education subscore and constitutive questions**

	Overall	Kilifi	Kwale	Mombasa	Nairobi	Samburu	Tana River	Turkana
VGE1: Worth investing in girl's education	95.5%	95.8%	87.5%	96.2%	96.4%	88.7%	97.3%	98.3%
VGE2: Girl as likely as boy to use education	94.7%	95.0%	96.9%	96.2%	94.6%	84.9%	96.0%	98.3%
Overall VGES	95.1%	95.4%	92.2%	96.2%	95.5%	86.8%	96.7%	98.3%

Disaggregating by county, with a VGE1 score of 87.5 percent, caregivers in Kwale county were more likely than caregivers from any of the other seven counties to disagree with the statement “Even when funds are limited it is worth investing in [girl's] education”. In contrast, at 97.3 percent, caregivers from Tana River were most likely to agree. The relatively low level of agreement from caregivers in Kwale may have to do with the high level of poverty found across the county, interplaying with generally large household sizes. Specifically, in Kwale, adolescents make up 78.0 percent of the population,<sup>174</sup> and the average household size in Kwale is 1.3 times that of the national average.<sup>175</sup> With such a young population, the economic burden imposed on families by way of school fees and incidentals for multiple children is high. In turn, some caregivers in Kwale may feel that it is better for girls, and perhaps boys as well, to gain employment and work to aid the family's financial burdens, not afforded the luxury of considering the long-term benefits – financial and otherwise – of an education.

Looking at VGE2 which tests caregivers' agreement with the statement “A girl is just as likely to use her education as a boy”, at 84.9 percent caregivers from Samburu county were more likely than caregivers from any of the other seven counties to disagree with the statement. This relatively high level of disbelief in the equality of boys' and girls' educational acumen may be somewhat related to beading. In Samburu, beading is a practice whereby a young man in the community, named a “Moran” upon completion of a ceremony, selects a young girl to be his partner.<sup>176</sup> In doing so, he selects a young girl with whom he can have sex, but not marry.<sup>177</sup> Considering that beading engages young girls in non-consensual sexual activity, at the selection of a young man, exemplifies a situation where a girl's individuality, will, and future is not as valued and regarded as that of a man's. If this line of logic exists within the community, we can use it to think through and better understand why a relatively high percentage of caregivers from Samburu disagree that a girl is just as likely to use her education as a boy.

Though Future School Subscore question 3 (FS3) is the household support question with the lowest score at endline, the Value Girl's Education Subscore Question 1 (VGE1) leaves us on a hopeful note. Specifically, while the mean FS3 score is 46.8 percent, the mean VGE1 score is 95.5 percent. This leads us to believe that though many households are unfortunately struggling to *be able* to send their girls to school, the vast majority still see the importance of doing so.

<sup>174</sup> “The Need,” Kenya Kesho School for Girls, accessed 12 January 2022, <https://www.kenyakeshoschoolforgirls.org/our-projects/>; Share of Adolescent Population in Kenya 2010 – 2020,” Statista, accessed 12 January 2022, <https://www.statista.com/statistics/1266462/share-of-adolescent-population-in-kenya/>.

<sup>175</sup> Share of Adolescent Population in Kenya 2010 – 2020,” Statista, accessed 12 January 2022, <https://www.statista.com/statistics/1266462/share-of-adolescent-population-in-kenya/>.

<sup>176</sup> “Silent Sacrifice: Girl-child beading in the Samburu Community of Kenya,” The International Work Group for Indigenous Affairs, accessed 12 January 2022, [https://www.iwgia.org/images/publications/0607\\_SEEDO\\_research\\_report.pdf](https://www.iwgia.org/images/publications/0607_SEEDO_research_report.pdf).

<sup>177</sup> *Ibid.*

### Summarizing IO4 Plausibility

**Table 70: Summary Chart of IO4 Plausibility**

Criteria	Assessment
Plausibility: Is the theory of change plausible?	Moderate. The theory behind IO4 is concerned with households' active support of girls into educational pathways. Though the caregiver survey asks caregivers about the degree to which they support their girl's transition into educational pathways, evaluating the outcome is somewhat limited by an inability to verify their actual actions (e.g., expenditure on girls' education relative to boys').
Evidentiary confirmation of key elements: To what extent are the key elements of the theory of change confirmed by new or existing evidence.	High. Household support for girls' education was generally strong at endline, and in qualitative interviews, it was reported that bursaries and similar WWW interventions helped households further support their girls to continue in school.
Identification and examination of other influencing factors: To what extent have other influencing factors been identified.	Moderate. Through KIIs and FGDs, girls and teachers alike note the presence of other programmes and interventions within their schools and communities.

## Community Behaviour and Support

### Summary of Findings

1. At endline, 77 percent of caregivers reported a need to do away with at least one harmful practice affecting girls' education. This falls slightly below the programme's target of 85 percent.
2. The programme's target was likely overambitious, given that traditional and cultural norms (including support for harmful practices) tend to change very slowly over time.
3. More communities reported undertaking formal initiatives to support girls' education in the past year at endline than at SAS. Furthermore, almost two-thirds of caregivers reported that girls had received community support for education, suggesting that informal initiatives to support girls' education may be robust.
4. Community views on cultural practices hindering girls' education were mixed. While around two-thirds of caregivers stated that early marriage and pregnancy were practices that should be done away with, only 27 percent of caregivers stated that female genital mutilation should be stopped.

This section analyses community behaviour toward and support for girls' education, collected through caregiver surveys and described by the WWW project's Intermediate Outcome 5 (IO5). Specifically, IO5 is "communities develop more positive attitudes to enable girls' learning and transition" and is measured by the "percent of communities expressing need to do away with harmful cultural practices that hinder girls from continuing to further their education and training".

To evaluate IO5, we analyse caregiver responses to questions revolving around community initiatives, community support, cultural practices, household support, surrounding family support, and will to support girls' education. In order to best understand caregivers' attitudes at endline, we analyse caregivers' responses to the eleven community behaviour and support questions with the goal of understanding

communities' behaviour towards and support for girls' education, using that understanding to inform recommendations for future interventions.<sup>178</sup>

First, we examine community initiatives and support for girls' education. The below table shows a low rate of formal community initiatives for girls' education. However, the rate at which girls received community support was substantially higher, suggesting that informal mechanisms to support girls' education are more common than formal "community initiatives". There was also a positive increase in reported community initiatives to support girls' education over the past 12 months of 8.8 percentage points.

**Table 71: Community support for girls' education**

	SAS	EL	Difference
Community initiatives undertaken in past 12 months	31.3%	40.1%	8.8
Any girl has received community support	64.4%	64.4%	0.0

Disaggregating by county for endline, we find the highest level of reported community initiatives to support girls' education in Samburu, where over 56 percent of caregivers reported that there had been community initiatives in the past 12 months. In general, rates of community initiatives were higher in ASAL counties than in urban slums, perhaps reflecting stronger community systems in these counties. Notably, however, we find that caregivers in Kilifi, Turkana, and Nairobi reported the highest rates of girls receiving community support, running counter to the pattern from the first question. This suggests that while urban slums may not have robust community systems by which members of the community can come together to pursue priorities, they do still indeed support girls' education.

**Table 72: Community support for girls' education at endline, by county**

	Kilifi	Mombasa	Nairobi	Samburu	Tana River	Turkana
Community initiatives	28.8%	30.8%	37.9%	56.6%	46.7%	53.7%
Girl received community support	69.4%	50.0%	68.0%	56.7%	57.1%	69.0%

The qualitative data sheds some light on results from Turkana. Reviewing the key informant interviews conducted with teachers from Turkana, high rates of financial support and mentoring in Turkana may be the result of county bursaries and the National Government Constituencies Development Fund paying many Turkana girls' school fees.<sup>179</sup> If community members lobbied for such aid, this may be considered as a community initiative. What's more, in terms of mentoring, another teacher from Turkana noted that their school had made efforts to "motivational speakers" into the school every week to talk to the students about different subjects.<sup>180</sup>

We now examine caregivers' reported willingness and ability to support girls through education. The below table reports results for relevant questions at SAS and endline. We find very high levels of willingness to support girls who have not been selected for secondary school, or who have dropped out,

<sup>178</sup> Data is for caregivers of intervention students only.

<sup>179</sup> KII with Teacher, Turkana, Int. 42.

<sup>180</sup> KII with Teacher, Turkana, Int. 4.

to continue their education or training. While rates were high at both SAS and endline, there was a 6 percentage point increase at endline, a positive sign.

**Table 73: Willingness to support girls to continue education**

	SAS	EL	Difference
Willing to support girl who has not been selected for secondary school or has dropped out to continue education	86.2%	92.2%	6.0
Have supported such a girl	34.2%	54.2%	20.0

The percent of caregivers who reported supporting such a girl was substantially lower, perhaps reflecting economic difficulties reported elsewhere in this document. Only around one-third of caregivers at SAS and slightly over half of caregivers at endline reported providing this support. However, the increase of 20 percentage points at endline is large, and reflects a promising increase in household support.

Disaggregating by county in the below table, we find the lowest level of willingness to support girls in Kilifi, although it is still fairly high in absolute terms, at 87 percent. Levels were highest in Turkana, where every caregiver stated that they would be willing to support a girl to continue her education, and in Samburu, where 98 percent of caregivers stated this.

Caregivers in Turkana also frequently reported that they had supported such girls, with 87 percent stating that this had happened. However, rates of support were second highest in Mombasa, at 65 percent, followed then by Samburu. We note that there is no clear pattern by urban slum/ASAL counties, with Tana River reporting relatively low levels of willingness to support and actual support compared to other counties.

**Table 74: Willingness to support girls to continue education at endline, by county**

	Kilifi	Mombasa	Nairobi	Samburu	Tana River	Turkana
Willing to support girl	87.2%	92.3%	91.8%	98.1%	92.0%	100.0%
Have supported girl	43.2%	65.4%	50.5%	62.3%	49.3%	87.0%

Caregivers were also asked how their household's support for adolescent girls' education has changed over the past year. At SAS, 60.8 percent of caregivers stated that it had improved, while at endline, 68.8 percent—or 8 percentage points more—of caregivers said it had improved. This is again a promising sign, showing strengthening support for girls' education.

Lastly, at endline, caregivers were asked about the main cultural practices that hinder girls from pursuing education, and whether they supported steps to reduce such cultural practices. Results for endline are reported in the table below.<sup>181</sup> We note that 77 percent of caregivers reported a need to do away with at least one harmful cultural practice.

<sup>181</sup> Caregivers were allowed to select more than one option; as such, response rates may not sum to 100 percent.



**Table 75: Cultural practices and girls' education**

	Practice hinders	Practice should be done away with
Early marriage/pregnancy	65.7%	69.8%
Girl beading	2.0%	9.1%
Female genital mutilation (FGM)	16.7%	27.3%
Moranism	1.5%	8.0%
Disco matangas	8.2%	19.7%
Herding	2.2%	7.1%
Wedding ceremonies	2.2%	6.1%
Preference of boys	9.5%	17.4%
Other	36.6%	-

By far the most frequently selected practice that hinders girls' education was early marriage and pregnancy, with almost two-thirds of caregivers stating this option. Indeed, in qualitative interviews, pregnancy was frequently discussed as a substantial barrier to education, with girls who became pregnant facing logistical difficulties to continue their education as well as shame and stigma.

The second most frequently mentioned practice was "other". Looking into specific responses, we find that many caregivers stated that students' discipline, performance, and school performance were main barriers to girls' education. This suggests that many caregivers may believe that factors inherent to the girl, or the strength of the girl's school, are more important than cultural practices.

The third most frequently selected option was FGM. Disaggregating by county, we find the highest rates of selection of this option in Nairobi, Tana River, and Samburu, suggesting that such practices may be more prevalent in these counties. In contrast, no caregivers stated that FGM was a barrier in Turkana.

Looking now at which practices caregivers believed should be done away with, we find that almost 70 percent of caregivers stated that early marriage and pregnancy should no longer be a practice. This is a promising sign, as it suggests that caregivers—and perhaps communities more broadly—recognize the harm done by this practice and the significant impact it has on a girl's ability to succeed in education.

The next-highest level of support was to do away with the practice of FGM. However, it is notable that only around one-quarter of caregivers said this practice should be done away with, a low value in absolute terms. Given the major negative effects of FGM on girls—both in terms of their education and life outcomes—there is clearly a need to continue sensitizing caregivers to the need to do away with this practice.

### **Summarizing IO5 Plausibility**

**Table 76: Summary Chart of IO5 Plausibility**

Criteria	Assessment
Plausibility: Is the theory of change plausible?	Moderate. The underlying assumption is sound: that community support for doing away with harmful cultural practices that affect girls' education logically correlates with more positive attitudes enabling girls' learning and transition. However, the supposed relationship is neither exclusive nor absolute.

Evidentiary confirmation of key elements: To what extent are the key elements of the theory of change confirmed by new or existing evidence.	Moderate. Willingness to support girls' education was high and improved at endline. However, barriers to community support appear to still remain.
Identification and examination of other influencing factors: To what extent have other influencing factors been identified.	Moderate. Through KIIs and FGDs, girls and teachers alike note the presence of other programmes and interventions within their schools. In addition, expressions of a need to do away with "harmful" cultural practices does not automatically correlate with the development of more positive attitudes vis-a-vis girl's learning and transition.

## Auxiliary Outcomes of Interest

Auxiliary to the five main outcomes of interest, as identified by EDT's logical framework, economic empowerment, child protection, sexual and reproductive health (SRH), and the impact of the COVID-19 pandemic are additional outcomes of interest.<sup>182</sup>

## Economic Empowerment

### Summary of Findings

1. At endline, 22 percent of caregivers were aware of financial grants provided by WWW to support girls' education. The vast majority of caregivers who had experience with these grants stated that the grants were beneficial.

One of the activities of the WWW programme included the disbursement of financial grants to community groups, with the aim of empowering communities to use such grants to directly or indirectly support girls' education. As of the ML data collection round, for instance, WWW had disbursed cash transfers for up to 1,906 households, bursaries to 1,143 girls, and travel grants for a further 843 girls.<sup>183</sup> Among recipient households, it was shown that 89.4 percent of household heads believed that such disbursements helped girls attend school more regularly than before.

Perhaps due to the lesser intensity of disbursements between SAS and EL, as output targets had been exceeded in previous rounds, a relatively small percentage of contacted households reported to have heard of WWW financial grants. Specifically, 181 of the 814 caregivers contacted via CATI<sup>184</sup> (or 22.2 percent) reported to be aware of such grants, with the highest number – 80 caregivers – located in Nairobi due to the concentration of the evaluation sample there. However, Marsabit and Kwale had the highest proportion of caregivers who have heard of such grants, at 60.0 and 35.7 percent, respectively.

Of the 181 who had heard of financial grants, 43.1 percent belonged to community groups who had received financial grants. The vast majority of them, 84.6 percent, further claimed that such grants were "very beneficial" to their community groups, with an additional 9.0 percent claiming that grants were "somewhat beneficial".

<sup>182</sup> School management and governance is additionally discussed in Annex 19.

<sup>183</sup> Limboro *Midline 1*, 198

<sup>184</sup> Data is for caregivers of intervention students only.

**Table 77: Benefit levels of financial grants**

Response	Percent of respondents at EL
Very beneficial	84.6%
Somewhat beneficial	9.0%
Slightly beneficial	3.8%
Not at all beneficial	2.6%

The majority of the 78 recipients surveyed had used grants partly or wholly to support girls' education. This primarily occurred through using grants to pay for school fees, while others also used grants to purchase items necessary for school, such as schoolbooks, stationaries, and transportation money. In addition, several recipients also used grant money as investments for their business or trade, such as by purchasing goats, with an aim of strengthening income sources to further support girls' education.

These findings are consistent with ML findings on the effectiveness and beneficial nature of financial grants provided to communities. The qualitative data also indicates that bursaries can help with school attendance of girl learners, with one teacher claiming that bursaries helped increase attendance in their area.<sup>185</sup>

Moreover, a different measure used in our caregiver survey suggests that support provided to households might be one major reason for improved school attendance of girls: caregivers who claimed that the school attendance rate of their dependents had improved over the preceding year were most likely to cite financial and material support provided to their households as one of the major reasons behind the improvements, with 38.8 percent of caregivers citing this reason. The idea that support provided to households can help improve girls' attendance in school is consistent findings from the ML and the qualitative data.

## Child Protection

### Summary of Findings

1. Around one-third of surveyed girls were aware of instances of violence or harm against children in their communities. Girls in major urban areas were more likely to have heard of incidents of violence/harm, and girls from these areas were also the most likely to feel that children are not safe from danger and violence in their communities.
2. However, almost no girls stated that violence against children occurs in schools. Furthermore, less than 1 percent of girls stated that teachers were perpetrators of violence. While more work may be needed to reduce violence against children in communities and households, girls appear to view schools as safe spaces.

In this subsection, we examine the extent to which girls are aware of various types of threats against children's welfare in the community, including various types of violent incidents, as well as non-violent harm more generally. As these questions were not included in the SAS round of data collection, this subsection aims to provide a contextual overview at a single point in time for intervention girls, as well as top-level comparisons to baseline and midline.

As seen in Table 78, the major urban centres of Nairobi and had the highest levels of girls' awareness of incidents of violence or harm to children. This may be because the density of major urban areas

<sup>185</sup> KII with Teacher, Tana River, Int.27

corresponds to a higher concentration of crime/criminality in less secure areas, which in turn affects girls' and children's safety on their way to and from school. One teacher in Nairobi cites the example of young male drug users who may threaten the safety of girls in densely populated areas.<sup>186</sup> Indeed, one of the major emerging findings from qualitative respondents is that the early hours by which some girls need to leave their homes in order to arrive to school on time puts them at heightened risks of violence, as they would often travel before sunrise. To accommodate this, some girls choose to come to school at a later time, which in turn might impact learning due to girls potentially spending less time in school per day.<sup>187</sup>

**Table 78: Percentage of girls aware of violent incidents or other harms against children in their communities<sup>188</sup>**

	Nairobi	Kilifi	Samburu	Turkana	Tana River	Mombasa	Marsabit	Kwale	All
Yes	38.4%	37.0%	37.7%	31.7%	26.8%	52.9%	3.1%	28.6%	34.4%
No	60.1%	50.8%	45.4%	62.6%	66.5%	44.1%	96.9%	70.1%	59.0%

Examining results from past evaluation rounds, at baseline, we find that only 12 percent of intervention caregivers in ASALs and 20 percent in urban slums were aware of violent incidents against children. However, at midline, this rate had increased substantially, with 40 percent of intervention girls claiming awareness of violent incidents against children. Overall, while results are not entirely comparable (with caregivers surveyed at baseline and girls surveyed at midline/endline), this suggests that violent incidents against children may have declined over the past two years, but may remain higher than at baseline. The underlying reasons for this pattern, however, are not entirely clear.

On aggregate, defilement of children is the type of incident of which the highest percentage of girls were aware, followed by child labour and physical violence done onto children, including both homicide and non-fatal physical acts of violence. Between counties, some variation in the kinds of harm/violence done onto children can be observed. For example, girls from Nairobi often explain that they are afraid of making their early morning commute to schools due to the presence of safety threats posed by potential kidnappers and predatory young men.<sup>189</sup> Conversely, the girls from the more rural Tana River county instead cite natural threats as the major safety concern for school aged children on their way to school, including leopards, elephants, and rivers which some students cross without bridges.<sup>190</sup>

When asked specifically about acts of violence, rather than harm in general, respondents were most likely to cite children's parents or caregivers as the main perpetrators of violence, with 35.5 percent of girls claiming as much. This is followed at some distance by relatives in general, whom 21.3 percent of respondents claimed to be the main perpetrators. Teachers, on the other hand, were the category of actors least cited by respondents as perpetrators of violence, with fewer than 1.0 percent of respondents claiming as much, which suggests a positive level of comfort among girls about the security and safety of their learning environments.<sup>191</sup>

Indeed, when asked where violence has tended to occur to children in their respective communities, schools are near the bottom of the list, with only 0.5 percent of respondents claiming so. Instead, the

<sup>186</sup> KII with Teachers, Nairobi, Int. 29

<sup>187</sup> KII with Teachers, Nairobi, Int. 21

<sup>188</sup> An additional "don't know" category was included in the answer choices to this question.

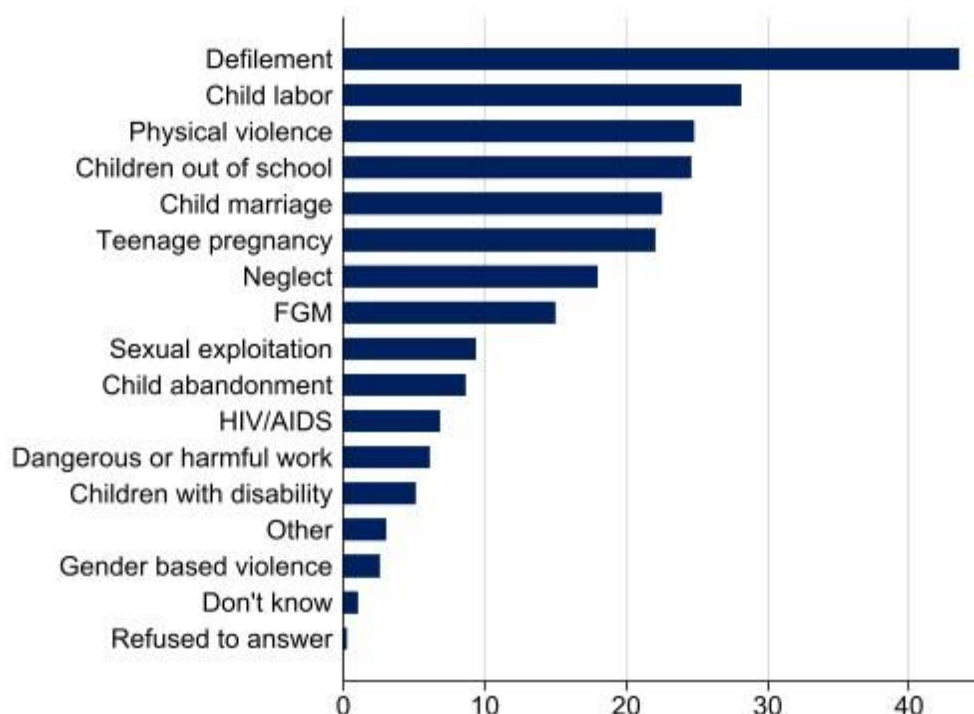
<sup>189</sup> FGD with Girls, Nairobi, Int.21; FGD with Girls, Nairobi, Int. 37

<sup>190</sup> FGD with Girls, Tana River, Int. 17; FGD with Girls, Tana River, Int. 24

<sup>191</sup> Interestingly, this finding contrasts somewhat with earlier findings that corporal punishment is still used in schools. It is possible that students do not view corporal punishment as "violence", but rather as a standard method of punishment on par with punishments like cleaning or tidying school grounds. It is also possible that the structure of the questioning, whereby students were asked about severe methods of violence such as defilement and FGM, led students to downplay the extent to which corporal punishment qualifies as violence.

community and the household are the two most cited locations in which violence occurs, with 95.1 percent of respondents selecting one of these two options as the primary location where violence is perpetrated against children.<sup>192</sup> For the communities, this is unsurprising, as multiple qualitative respondents have explained that children's walk to school from their homes, especially for girls, is amongst the most dangerous times for children when done very early in the morning or after dark. Qualitative respondents did not discuss household physical violence against girls, but cited neglect of or discrimination against girls by their parents as a significant harm done upon them, especially in cases where parents choose to neglect their daughters who had accidentally become pregnant.<sup>193</sup>

**Figure 20: Percentage of girls aware of particular types of violence/harm against children in their communities**



Several community structures may be available to help support children who have encountered one or more of incidents of violence or harm, but the extent to which girls are aware of such structures, as well as how to engage with them, remains an open question. When asked about whether they felt that community structures to support children existed, just under half – 45.8 percent – of respondents answered in the affirmative. This is an important finding, as it implies that more than half of girls in communities might not be willing to, or know how to, seek support if they encountered incidents of violence or harm.

Among those who did acknowledge the existence of community structures, girls most frequently claimed that NGOs are a potential source of support for children, with 25.1 percent of surveyed girls citing NGOs as community support structures. This is followed by children's homes (19.1 percent), churches (18.1

<sup>192</sup> 53.4 percent of respondents felt that it will most likely happen somewhere in the community, while 41.7 felt that violence would more likely occur inside the house.

<sup>193</sup> For example, FGD with Girls, Int. 14

percent), and children’s departments in government (15.3 percent). It is also worth noting that among girls who claimed to know of community support structures, 61.0 percent were only able to cite one such structure. Once again, future programs related to girls’ education and empowerment might consider activities that both increase the capacity of more structures to address children’s protection issues, as well as ensuring that girls are made aware of the variety of options for support available to them.

Lastly, a predictive analysis discussed in Annex 19 provides one particularly important finding: use of corporal punishment by teachers is associated with worse child protection outcomes and less perceived safety in communities. This further reiterates the need to eliminate this negative practice.

## Sexual and Reproductive Health

### Summary of Findings

1. At both endline and SAS, girls overwhelmingly believed sexual education was important and expressed generally strong knowledge of sexual and reproductive health, including issues related to sexually transmitted diseases.
2. Over 96 percent of girls were confident in their right to complain if touched inappropriately by a teacher, adult, or classmate. Girls were also highly confident in their ability to refuse unwanted sexual intercourse.
3. However, we find relatively low levels of willingness to use condoms, with 63 percent of girls reporting that they would use a condom if they had sex before marriage, and even lower levels of willingness to use contraceptives in general, at only 28 percent. Access to contraceptives was also low. Further strengthening these areas will help bolster other positive SRH outcomes achieved during WWW implementation.

A goal of the WWW project was to improve girls’ knowledge of and attitudes towards sexual and reproductive health. As described in previous sections, in the qualitative data, girls commonly mentioned that pregnancy was a frequent reason for girls to drop out of school. Early pregnancy is also associated with poorer overall life outcomes; the World Bank, for example, finds that girls who give birth before adulthood are “likely to bear increased health risks, social stigma, and adverse economic impacts for the rest of their lives”.<sup>194</sup> Unsafe sexual practices also put girls at risk for long-term health issues, including HIV/AIDs and other sexually transmitted diseases (STDs) and infections.

To analyse these dynamics, within the girl survey, intervention girls were asked about their awareness of SRH topics, their self-confidence in advocating for themselves regarding sexual health and consent, their willingness to use contraceptives, their perceptions of SRH, and their knowledge of the consequences of unsafe SRH practices. In this section, we analyse this data from the girl survey and data from FGDs with girls to provide a descriptive analysis of SRH practices and perceptions at SAS and endline. We also analyse predictors of girls’ confidence in advocating for positive SRH practices. More detailed analysis is provided in Annex 19.

### **SRH Awareness and Knowledge**

Girls were first asked about the importance of sexual education. At SAS, 91.5 percent of girls stated that sexual education is important; a similar proportion of girls, 90.3 percent, stated the same at endline. This shows a high level of interest in and prioritisation of education about SRH, a positive sign.

<sup>194</sup> World Bank, 2022. “The Social and Educational Consequences of Adolescent Childbearing”. <https://genderdata.worldbank.org/data-stories/adolescent-fertility/>.

Girls were then asked about their knowledge of specific SRH topics; positively, only a small percentage of girls, 0 percent at SAS and 4 percent at endline, stated that they did not know the consequences of early sexual activity. Girls most frequently stated that pregnancy was a potential consequence of early sex, with 82 percent of girls listing this as a consequence at endline. Additionally, nearly sixty percent of girls stated that dropping out of school could be a consequence of early sexual activity. While this is a high percentage, the even higher percentage of girls stating that pregnancy is a consequence may mean that girls do not perceive pregnancy as an insurmountable barrier to continuing their education. Indeed, in response to a vignette about a pregnant girl presented during the FGDs, some girls stated that it could be possible for a pregnant girl to return to school. Two respondents in an FGD, for example, stated the following:

*Girl 1: It will depend on how [her] parents will react to the pregnancy. If [her] parents are understanding they will forgive her and let her continue with her education, but we have some parents who might even chase her out of the home.*

*Girl 2: If [she] was determined to continue with her education and got the support of her parents she would go back to school and continue with her education.<sup>195</sup>*

We note, however, that many girls stated that a pregnant girl would face substantial stigma which would make her very uncomfortable in school. Further work may be needed to destigmatize education for teenage mothers.

In addition to questions about the consequences of early sex, girls were also asked about ways to prevent STDs. The below table shows that girls were by far most likely to state that abstinence was a way to prevent STDs, with over 85 percent of girls stating this option at both SAS and endline. In contrast, only one-third of girls at SAS and 41 percent of girls at endline stated that using condoms could prevent STDs, and only 16 percent of girls at SAS and 24 percent of girls at endline stated that faithfulness could prevent STDs. While this is technically true—neither condom use nor faithfulness provides complete protection against STDs—it is important to note that both of these are still useful steps to reduce the risk of STDs. As such, it may be useful to present sexual education about STDs in a more nuanced manner, emphasizing that while condoms and faithfulness are not perfect steps, they are still helpful SRH practices.

**Table 79: Ways to prevent STDs**

Method	SAS	EL
Abstinence	85.6%	86.5%
Use of condoms	33.2%	40.9%
Being faithful to one partner	16.4%	23.7%
Other	0.0%	7.7%

Below, we report endline results on importance of SRH by county; further disaggregation is included in Annex 19. Looking across all questions, we generally find the highest levels of prioritisation for sexual education and knowledge of SRH in Turkana. In this county, almost 99 percent of girls stated that sexual education is important and 97 percent stated they had good or some knowledge of STDs.

<sup>195</sup> FGD with girls, Tana River, Int. TRG1

**Table 80: SRH awareness and knowledge by county**

	Kilifi	Kwale	Marsabit	Mombasa	Nairobi	Samburu	Tana River	Turkana
Sexual education is important	83.6%	81.8%	71.9%	76.5%	95.8%	88.5%	93.3%	98.6%
Good or some knowledge of STDs	88.7%	93.5%	84.4%	94.1%	96.7%	86.9%	89.6%	97.1%

In contrast, we find relatively lower levels of awareness and knowledge in Marsabit, Mombasa, and Kwale. In Marsabit and Mombasa, less than 80 percent of girls stated that sexual education was important. Marsabit also had the lowest level of girls reporting to have knowledge of STDs. In Kwale and Kilifi, girls also placed relatively less importance on sexual education. These findings point to a need to continue strengthening sexual education in these counties.

### ***SRH Confidence and Self-advocacy***

We now analyse girls' confidence and self-advocacy regarding sexual health. First, we analyse responses to questions that assess a girls' confidence in complaining if a teacher, adult, or peer touches her inappropriately. Second, we analyse girls' confidence in saying no to unwanted sex and resisting peer pressure to have sex.

Table 81 shows girls' responses to questions about whether they feel they have the right to complain if touched inappropriately by a teacher, an adult, and a schoolmate. We find that almost all girls stated that they did feel they had this right in all three situations at both SAS and endline.

**Table 81: Confidence and self-advocacy against sexual assault**

I have the right to complain if...	SAS	EL	Difference
A teacher touches me inappropriately	99.2%	97.6%	-1.6
An adult touches me inappropriately	99.2%	98.5%	-0.7
A schoolmate touches me inappropriately	98.7%	96.3%	-2.4

The table below further investigates this dynamic. We find very high levels of confidence that girls could refuse to engage in unwanted sex and resist peer pressure to have pre-marital sex. While around one-third of girls at endline stated that their friends would laugh at them if they refused to have sex, the high levels of confidence in resisting peer pressure may mitigate any negative impacts of peer pressure.



**Table 82: Confidence in refusing to engage in early sexual activity**

	SAS	EL
"I am confident that I could refuse to engage in sexual intercourse if I did not want it" (% agree)	97.1%	97.0%
"I am confident that I could resist peer pressure to participate in risky behaviours such as pre-marital sex" (% agree)	93.9%	95.1%
"My friends would laugh at me for refusing to have sex" (% disagree)	69.6%	67.8%

Lastly, the table below shows endline results by county. In general, we find higher average levels of SRH confidence and self-advocacy in the urban slums compared to the ASALs. In Mombasa, for example, 100 percent of girls stated that they would be able to complain if they were inappropriately touched by any party; in Nairobi, this statement was echoed by 99 percent of girls. In contrast, in Marsabit, Samburu, and Tana River, only around 94 percent of girls stated that they would be able to complain if they were inappropriately touched by a teacher. Furthermore, in Turkana, only 86 percent of girls stated that they would be able to complain if inappropriately touched by a schoolmate.

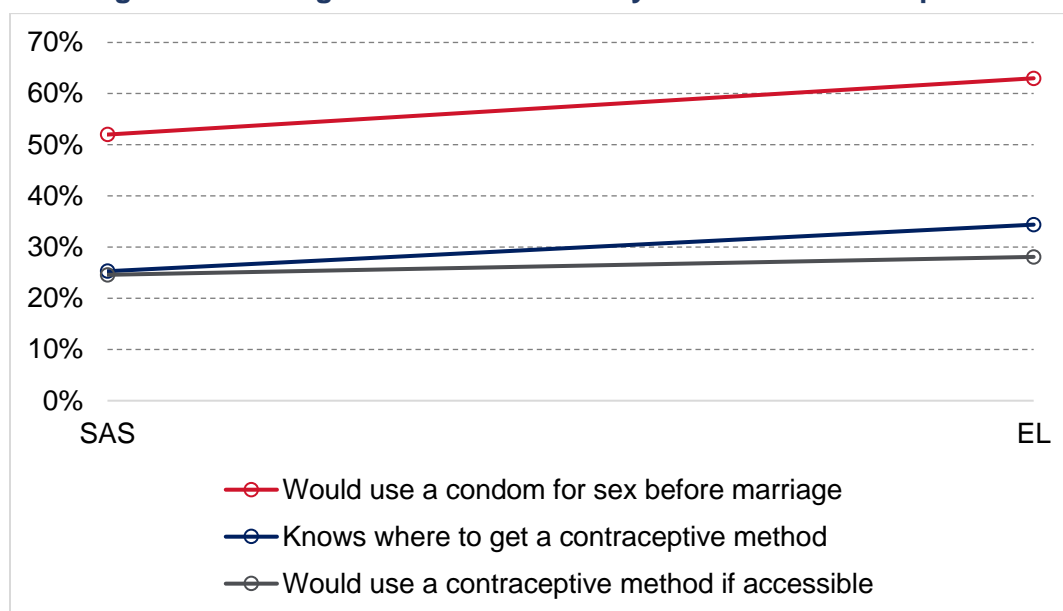
**Table 83: SRH confidence and self-advocacy by county**

	Kilifi	Kwale	Marsabit	Mombasa	Nairobi	Samburu	Tana River	Turkana
<b>Ability to complain if facing inappropriate touching from...</b>								
Teacher	99.2%	97.4%	93.8%	100.0%	99.4%	93.1%	98.2%	94.2%
Adult	99.6%	97.4%	93.8%	100.0%	99.4%	95.4%	98.2%	99.3%
Schoolmate	97.9%	97.4%	93.8%	100.0%	99.1%	93.9%	98.2%	85.6%
<b>Confidence to refuse early sexual activity</b>								
Refuse unwanted sex	97.1%	93.5%	90.6%	100.0%	98.8%	93.1%	96.3%	99.3%
Resist peer pressure	94.1%	84.4%	87.5%	100.0%	97.6%	90.8%	97.0%	99.3%
Friends would (not) laugh	60.9%	57.1%	53.1%	73.5%	72.0%	70.0%	80.5%	60.4%

These patterns generally continue when looking at questions around girls' confidence in refusing sexual activity. Confidence remains high in Mombasa and Nairobi in particular, and is notably low in Marsabit. Girls in Kwale also reported relatively low levels of confidence in refusing sexual activity compared to other counties, with only 84 percent stating that they would be able to resist peer pressure to participate in risky sexual behaviours.

### **Willingness to Use Contraceptives**

We now analyse girls' willingness to use contraceptives, including condoms, as well as their ability to access contraceptives. The figure below presents results for questions regarding contraceptive use. At SAS, only around half of girls stated that they would use a condom if they were to have sex before marriage; at endline, this had increased to 63 percent of girls. Similarly, we find a higher percentage of girls stated that they knew where to get a contraceptive method at endline.

**Figure 21: Willingness to use and ability to access contraceptives**

It is important to note that overall, the percent of girls who know where to access contraceptives remains small, at only one-third of respondents at endline. There is clearly a need to continue improving access to contraceptives for adolescents.

Furthermore, only around 28 percent of girls stated that they would use a contraceptive method if they knew how to access it at endline. These findings run somewhat counter to the results for the first question about condom use; however, girls may have interpreted this question to ask about contraceptives other than condoms. If so, this suggests that stigma around contraceptives remains, and that girls feel uncomfortable using contraceptives, especially non-condom contraceptives.

Looking now at endline results by county, in Table 84, we first note very low levels of willingness to use and access to contraceptives in Tana River. Only around 40 percent of girls in this county stated that they would use a condom if they had sex before marriage and only 18 percent knew how to access contraceptives and stated that they would use contraceptives if accessible. This suggests a need to improve attitudes towards contraceptives and availability of contraceptives in Tana River.

**Table 84: Willingness to use and ability to access contraceptives, by county**

	Kilifi	Kwale	Marsabit	Mombasa	Nairobi	Samburu	Tana River	Turkana
Willing to use condom	66.0%	53.3%	56.3%	64.7%	62.8%	84.6%	39.6%	71.9%
Knows how to access contraceptives	39.5%	28.6%	40.6%	44.1%	39.6%	43.9%	18.3%	23.0%
Would use contraceptive if accessible	30.7%	22.1%	37.5%	50.0%	30.1%	36.2%	17.7%	19.4%

Second, in contrast to the previous subsection, for willingness to use a condom, we note relatively more positive results in Samburu and Turkana relative to other counties. In Samburu especially, almost 85 percent of girls stated that they would use a condom for premarital sex.

Lastly, we note that in Mombasa county, attitudes towards contraceptive use were significantly better than attitudes in other counties, with 50 percent of girls stating that they would use a contraceptive if accessible. It may be worth further examining sexual education practices in Mombasa regarding contraceptives—especially non-condom contraceptives—to see if any practices may be replicable in other counties in order to improve girls’ attitudes.

### **SRH Perceptions**

Within the girl survey, intervention girls were also asked a series of 12 questions about their perceptions towards SRH. This includes perceptions about girls who become pregnant, sexual education, sexual health and hygiene, and level of comfort speaking about SRH topics. In Annex 19, we present the percentage of positive responses to each of the 12 questions. Here, we calculate an overall “SRH Perceptions Index” as a score based on the number of positive answers to each question. Scores are normalised to a range from 0 to 100, where 100 would represent a positive answer to all 12 questions (or every question that the girl did not refuse to answer), while 0 would represent a negative answer to all 12 questions. Index values are reported for SAS and endline, and by county for endline respondents.

The below table shows overall SRH Perceptions Index scores. We find fairly high overall scores at endline. This suggests that girl students’ perceptions of the importance of SRH education may be increasing and stigma against pregnant girls may be decreasing.

**Table 85: SRH perceptions**

	SAS	EL
Average SRH Perceptions Index score	77.8	80.7

The table below additionally shows endline SRH Perceptions Index scores by county. Variance across counties is relatively limited, ranging from a low average score of 74.2 in Marsabit to a high of 85.7 in Turkana. While this difference is not extremely large, it is significant. There are no clear patterns, however, by ASALs and urban slums, with two ASAL counties—Turkana and Marsabit—having the highest and lowest scores, while urban slum counties’ scores were generally clustered towards the middle of the distribution.

**Table 86: SRH Perceptions Index by county**

	Kilifi	Kwale	Marsabit	Mombasa	Nairobi	Samburu	Tana River	Turkana
Average SRH Perceptions Index score	76.5	77.3	74.2	80.7	81.7	83.7	81.2	85.7

In Annex 19, we present a detailed analysis of predictors of SRH confidence and perceptions. Here, we emphasise one finding: a significant and negative relationship between disability status and SRH confidence. Although girls with disabilities had high SRH knowledge/perceptions, they remain less confident about their ability to refuse unwanted sexual activity than similar girls without disabilities. Indeed, 13 percent of girls with disabilities stated that they would not have the right to complain about a schoolmate if touched inappropriately. These findings emphasise the importance of continuing to address the unique issues and challenges faced by girls with disabilities.

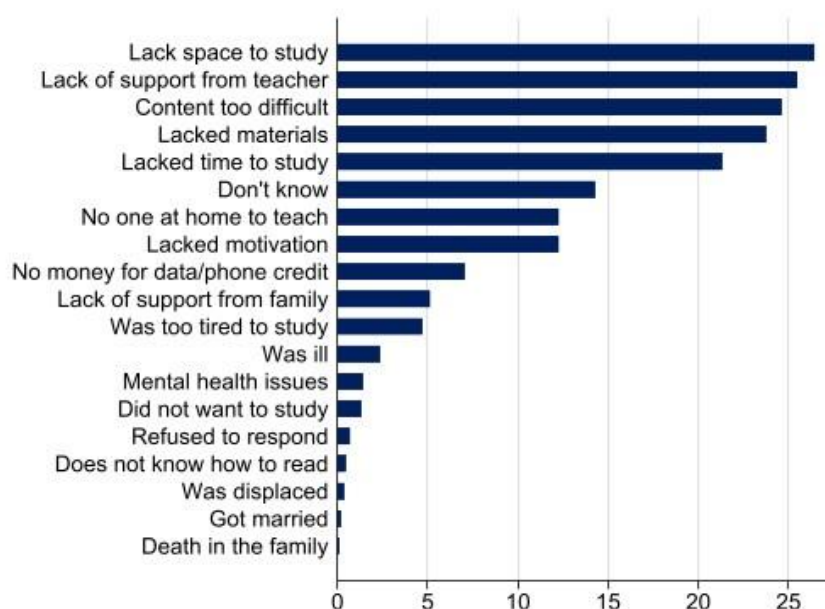
## COVID-19 Impact

### Summary of Findings

1. Learning at home posed many challenges for Kenyan girls, primarily because their home environments were not suitable for learning and they lacked learning materials.
2. COVID-19 also acutely affected girls' households, especially through economic losses and heightened levels of stress.
3. However, training teachers on the use of ICT had a mitigating effect on negative learning and transition outcomes; girls from schools where teachers were trained in ICT had a lower likelihood of dropping out.
4. Girls in urban slums, especially Mombasa, were less likely to drop out during COVID, potentially due to their relatively greater access to telecommunications infrastructure to facilitate at-home learning.

The sudden transition to remote schooling, and the significant disruption it caused for students lacking the necessary resources to adapt, likely ranks among the most consequential side effects of public health measures to curb the spread of COVID-19.<sup>196</sup> In Kenya, school closures caused many students to return home and attempt to continue their education under highly challenging circumstances. In this sub-section, we briefly examine how the COVID-19 pandemic and ensuing public health measures impacted the intervention girls surveyed during EL, as well as their families.

**Figure 22: Percentage of girls citing challenge to studying at home**



From Table 87 below, it is clear that learning from home posed myriad challenges for Kenyan girls. Chief among these is the fact that girls' home environments were often not suitable for learning, either because

<sup>196</sup> Mohamed Ihsan Ajwad and Simon Bilu. (2022). "The effects of COVID-19 school disruptions will last decades," *World Bank Blogs*, June 20. <https://blogs.worldbank.org/developmenttalk/effects-covid-19-school-disruptions-will-last-decades>

girls' households are overcrowded and girls did not have adequate space to study, or because there is lack of materials, such as books and stationaries, which girls can use at home to continue studying effectively. Indeed, over 30.0 percent of girls surveyed claimed that they could only rely on their own materials to continue their studies during school closures, or that they had no such materials at all, which severely hindered their learning. The general lack of support from teaching staff, coupled with school class content perceived by some girls are being too challenging, compound the challenges which a sudden shift to home-based learning caused. As one girl succinctly put, "if it was time for her [a girl learner] to study she would say 'no', because she could not study the subject alone. This is because she lacked teachers' support," though the respondent admitted that siblings may try to help her.<sup>197</sup><sup>198</sup>

More broadly, 67.0 percent of all girls claimed that they or their families were affected in some way by the pandemic. These include, among other things, economic losses and heightened levels of stress. At times, the impact of the pandemic and challenges to home-based learning had led some girls to drop out of school, before eventually re-enrolling prior to the EL round. Dropping out and then resuming one's studies is possibly amongst the most disruptive and consequential impacts on girls' education. The overall percentage of surveyed girls who dropped out of school and re-entered remains relatively small, with only 8.4 percent of girls claiming as much in our sample. However, there are geographic differences in these proportions, as demonstrated in Table 87 below. Major urban areas such as Nairobi and Mombasa had among the lowest rates of dropouts in our sample, with our Mombasa sub-sample not having a single girl who dropped out of school. This is not an entirely surprising finding, as major urban areas may also have more developed telecommunications infrastructures and other resources that better facilitate girls' home-based learning. While the small sub-sample sizes outside Nairobi warrant caution, respondents in KIIs and FGDs also highlighted the connectivity challenges in more rural areas.<sup>199</sup>

**Table 87: Rates of girls who dropped out during school closures, before re-enrolling**

	Nairobi	Kilifi	Samburu	Turkana	Tana River	Mombasa	Marsabit	Kwale
Yes	6.3%	10.9%	6.9%	3.6%	8.5%	0.0%	9.4%	24.7%
No	93.8%	89.1%	93.1%	96.4%	91.5%	100.0%	90.6%	75.3%

Lastly, we note that the predictive analysis included in Annex 19 found that girls with disabilities were slightly more likely than their peers to have dropped out at some point during school closures, suggesting that these girls may have faced even more difficulties adapting to remote or hybrid learning. In a positive finding, however, we find that training teachers on ICT had a mitigating effect: Girls from schools where teachers were trained on ICT use had a lower likelihood of dropping out and re-enrolling.

<sup>197</sup> FGD with Girls, Mombasa, Int. 9

<sup>198</sup> At the same time, it is worth noting that the COVID-19 lockdowns may have, in fact, created an opportunity for increased learnings for some girls if their original school environment posed challenges for their learning. For example, one girl claimed that "during lockdown there was an improvement [in learning] as she [a hypothetical girl] was comfortable at home but when the school opened she feared [her teacher]." See FGD with Girls, Mombasa, Int.19. However, the general trend remains that COVID-19 was overall detrimental to girls' learning.

<sup>199</sup> KII with Teachers, Turkana, Int.42; KII with Teachers, Kilifi, Int.28

## Attitudes Towards STEM

### Summary of Findings

1. Girls had very positive attitudes about whether girls, as well as boys, should study STEM subjects.
2. In qualitative discussions, many girls mentioned that they were highly interested in STEM subjects, especially if they wished to pursue a career in STEM.
3. At endline, there was a significant decrease in the percent of girls who stated that they disliked STEM classes because they were difficult. This may be reflective of an improvement in girls' life skills and confidence, including their willingness to tackle challenging subjects in pursuit of their educational and career goals.

In this section, we briefly analyse intervention girls' attitudes towards and uptake of STEM subjects, focusing on areas with substantial change between SAS and endline. More detailed analysis is included in Annex 19. Success in STEM subjects enables students to pursue rewarding and stable careers in fields such as medicine, engineering, and information and communications technology (ICT). However, both worldwide and in Kenya, girls are less likely to take STEM subjects and pursue STEM-related careers, in part because of lack of self-confidence, parents' and teachers' expectations, stereotypes, family obligations, and barriers to academic performance.<sup>200</sup> As such, a goal of the programme is to improve girls' attitudes towards and enrolment in STEM subjects to help provide them with the academic background to succeed in their careers.

The below table provides a highlight of results on attitudes towards STEM and reasons why students like and dislike STEM classes. We find that girl students had very positive attitudes about whether girls, as well as boys, should study STEM subjects.

**Table 88: Summary of attitudes towards STEM subjects**

	SAS	EL
"I believe science and mathematics subjects are not meant for girls but for boys" (% disagree)	87.0%	89.1%
"Girls perform better in mathematics and science subjects than boys" (% agree)	35.3%	30.1%
"Girls understand mathematics and science subjects better than boys" (% agree)	34.1%	28.1%

<sup>200</sup> Amunga, James, and Amadalo Maurice Musasia, 2020. "The Gender STEM Gap and Its Impact on Sustainable Development Goals and the Big Four Agenda in Kenya: A Synthesis of Literature." *International Journal of Contemporary Education* 4(1): 1.

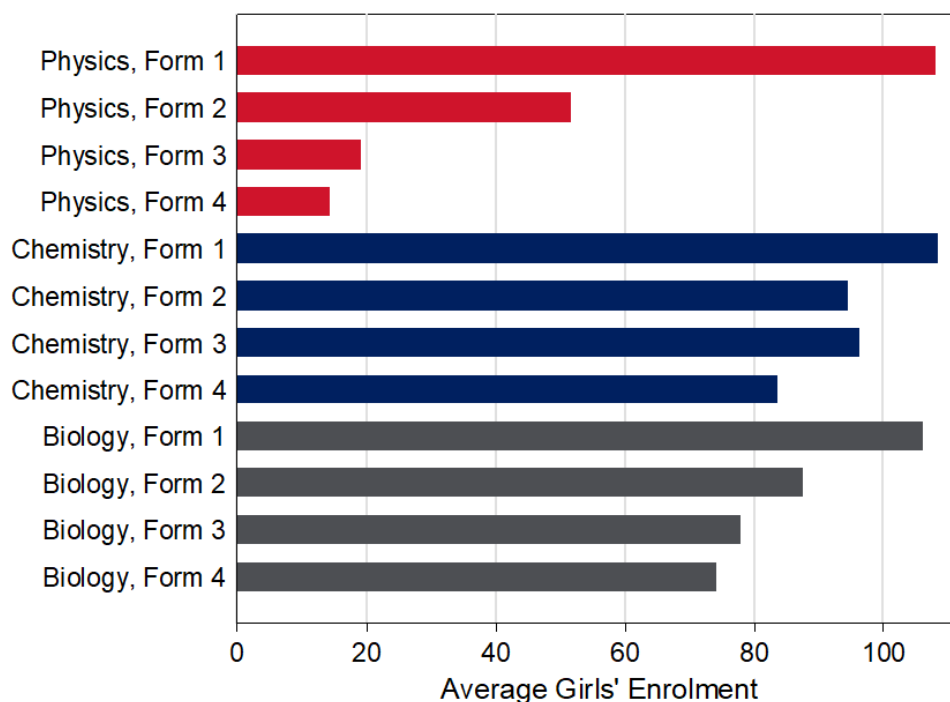
<b>Reasons for liking STEM classes</b>		
"I found science classes easy"	48.5%	50.4%
"It is easy to see how science would be useful for the future"	41.6%	36.6%
"I would like to pursue a career in the sciences"	-	32.0%
<b>Reasons for disliking STEM classes</b>		
"Science classes are hard"	62.8%	22.7%

Interestingly, in contrast to these findings, we find that almost half of girls at endline said that they like science classes because they find them easy. Similarly—and positively—at endline, a drastically lower percentage of girls stated that they disliked science classes because they were hard, compared to SAS. This suggests that while girls may still believe that boys perform better at STEM subjects, they may have relatively high self-confidence in pursuing sciences. Furthermore, it is useful to note that almost one-third of girls stated that they wanted to pursue careers in the sciences.

We now analyse enrolment in STEM subjects to further understand girls' involvement with these subjects. We note that at endline, 10 schools reported higher girls' enrolment in at least one STEM class than total girls' enrolment in Forms 1 or 2, nine schools higher STEM than total Form 3 enrolment, and 11 schools higher STEM than Form 4 enrolment. We drop these schools from the analysis as we cannot explain this discrepancy.

The figure below shows that at endline, in Form 1, average girls' enrolment in physics, chemistry, and biology across all schools exceeded 100 students. This suggests a high level of interest in and desire to pursue STEM subjects. Enrolment in all STEM subjects tends to decline in higher forms, but this mirrors a decline in total enrolment in higher forms, with average girls' enrolment in school declining from around 113 students in Form 1, to 104 students in Form 2, 98 students in Form 3, and 86 students in Form 4. This partially coincides with STEM subjects becoming elective in Form 3, with only two sciences required at this grade.

**Figure 23: Average girls' enrolment in STEM subjects, by subject and form**



However, it is notable that enrolment in physics declines substantially more than general enrolment or enrolment in chemistry and biology. While roughly equivalent numbers of girl students are enrolled in physics as in chemistry and biology in Form 1, by Form 4, average girls' enrolment in physics had dropped to only 16 students, around one-fifth of the enrolment in chemistry and biology. This suggests, as above, that girls may struggle substantially more with physics courses, which require higher levels of mathematics abilities.

## Project-specific Outcomes

In this section, we discuss project-specific outcomes broadly defined along OECD DAC criteria of relevance/appropriateness, effectiveness, short-term impact, efficiency and value for money (VfM), sustainability, and gender equity and social inclusion (GESI). Relevant questions for the evaluation were defined in the inception phase, and are answered within each subsection below.

The below table summarises the findings and evidence base for each analysed criterion. For each criterion, we provide a summary of the evidence used in analysis and rate the quality of evidence. We find that evidence is strong for all criteria except efficiency/VfM and GESI, for which project documents would have provided useful secondary information. We also summarise results for each project-specific outcome and rate the result on a five-point scale, including very weak, weak, mixed, strong, and very strong.



Table 89: Summary table of project-specific outcomes

Outcome	Definition	Evidence	Result
Relevance/ appropriateness	Extent to which intervention responds to beneficiaries and stakeholder priorities	Quantitative and qualitative data  <b>Quality of evidence: Strong</b>	Intervention was highly relevant to context, including beneficiary (student, teacher, etc.) needs and Kenyan government priorities.  <b>Rating: Very strong</b>
Effectiveness	Extent to which intervention achieved objectives	Quantitative and qualitative data  <b>Quality of evidence: Strong</b>	Achievement of objectives was mixed, but it is important to note that the COVID-19 pandemic presented many challenges for programming.  <b>Rating: Strong</b>
Short-term impact	Extent to which intervention has generated significant effects	Quantitative and qualitative data  <b>Quality of evidence: Strong</b>	Some interventions, such as those to improve teaching practices or provide financial assistance to girls to continue education, appear to have had a strong impact. However, barriers to girls' education, especially for the most marginalised girls, remain.  <b>Rating: Strong</b>
Efficiency and VfM	Extent to which intervention delivers results in an economic and timely way	Interviews with programme staff; interviews with government officials; quantitative data  <b>Quality of evidence: Moderate</b>	The programme was generally implemented efficiently and utilised adaptive management practices. However, several interventions were noted as less cost effective.  <b>Rating: Strong</b>
Sustainability	Extent to which benefits of intervention are likely to continue	Quantitative and qualitative data; interviews with government officials  <b>Quality of evidence: Strong</b>	The programme worked closely with government officials to strengthen sustainability, and many outcomes are expected to have long-term impacts. However, some interventions, such as bursaries, were noted as unsustainable.  <b>Rating: Strong</b>
Gender equity and social inclusion	Incorporation of gender equity and social inclusion concerns into programming	Interviews with programme staff and government officials; quantitative data  <b>Quality of evidence: Moderate</b>	In line with GEC priorities, the programme had a strong focus on the most marginalised girl students.  <b>Rating: Very strong</b>

## Relevance/Appropriateness

Activities and changes made that contributed to learning improvements from BL and ML, including the use of positive and/or innovative teaching methods, were sufficiently “sticky” that learners still benefitted from the learning done in that period, which enabled them to be more resilient against shocks (especially COVID-19) in subsequent periods. Furthermore, it is likely that programme activities during and after COVID-19 school closures contributed to preventing severe losses in learning. This seems particularly true given that only 36.6 percent of girls claimed that they were supported by their teachers while schools were closed. Some of the specific activities undertaken by the WWW-T team were also incredibly relevant to the changing context. By EL, for example, over 77.0 percent of teachers were provided training that included ICT training. As our section on COVID-19 impacts highlighted, this was one of the school-level factors that reduced the likelihood that girls would drop out from and re-enrol in school, which would have undoubtedly led to even more substantial learning losses. Overall, while the programme was clearly unable to fully mitigate the effects of the COVID-19 pandemic on learning outcomes, its interventions were timely and relevant for girls’ learning, such that the programme was able to prevent even steeper losses of learning between ML and EL and to maintain some of the gains made from BL to ML.

Regarding transition, it is worth noting in that the programme implementation also coincided with the Kenyan Government’s major push for 100 percent transition, showing high relevancy to government priorities. Several programme interventions were noted as particularly relevant to increase transition, such as bursaries and the provision of other economic support measures.

The extent to which improvements in learning and transitions are made between BL and EL can likely be attributed to the use of positive participatory learning practices and their adoption in classroom settings, showing the high relevance of teaching-oriented project activities. Furthermore, the integration and teaching of new learning methods also likely impact positively how well prepared, and how confident, instructors felt to deliver lessons, which in and of itself may have positive effects on students’ learning. One teacher’s observation, worth quoting at length, demonstrates the benefits which introducing the DIM method specifically had both on their technical abilities to teach and their confidence to navigate the classroom setting:

*For the last 3 years after going through this training, we’ve been taught a lot, to stop thinking in terms of ‘I do’ ‘we do’ ‘you do’, to the DIM method in which we now act like facilitators. We’ve been engaging our learners, and the teaching has been teacher/pupil centred, unlike before, and this one has also helped me a lot for the last 3 years. Using this method encouraged my learners, they’ve really been enjoying the lessons more when it comes to literacy, they’ve become very, very active. During the training, we were also taught on gender, so we’ve been engaging both boys and girls about the numeracy/literacy and other learning areas equally.*

*I have more confidence now because the last training that we went into, we were how to deal with different areas when the DIM method was being introduced, emphasising grammar teaching, writing, and also reading for comprehension ... After the training, things were now better and confidence is high; now when you are going to teach a lesson, you have confidence because you know the way to deliver.<sup>201</sup>*

Additional modes of learning not previously designed by the program also emerged as a result of the pandemic. Specifically, online learning methods using applications such as WhatsApp were – by necessity – quickly adopted by many schools and teachers in response to the closures of schools. Our surveys did not specifically measure how effective students felt these online instruction methods were for their learning. However, overall, they likely helped prevent even more severe learning losses by girls,

<sup>201</sup> KII with Teacher, Nairobi, Int. 7

which, given the severity and pressing circumstances of COVID-19, can be considered a commendable achievement.

Improvements made to the teaching methods used in the classroom were also supplemented by the programme's emphasis on garnering increased community support for girls' education. Specifically, through community conversations, it is hoped that communities can become involved in schools and girls' education. While the prevalence and level of involvement of these groups appears to remain limited, the majority of caregivers who were aware of these groups found them very beneficial to the community, suggesting a high level of relevance.

Outside of these more structured mechanisms for community involvement in girls' education, however, we find that communities, largely independently from the program, are already willing to support, and have been supporting, girls' education. This includes providing material support, such as school supplies, financial support, or mentoring of caregivers to better care for their dependents. Thus, while the WWW-T component on community engagement and the formalisation of their plans to support education remain works in progress, the fundamental logic of leveraging community resources to support girls' education is sound. Future iterations of the program might consider a more careful mapping of existing community support structures for girls' education and leveraging their already ongoing activities, rather than building wholly new structures where existing ones that duplicate or interfere with the functioning of existing ones.

The one area that would likely require closer intervention from programme implementers is likely on community support towards girls with disabilities and pregnant girls, as there still remain cases where communities failed to provide support for such girls, or worse, community members such as fellow students actively shame, stigmatise, or mock them. In these cases, relying on existing community structures may perpetuate existing norms that do not support the learning outcome improvements of particularly vulnerable girls.

Finally, in regard to how project beneficiaries see their roles as agents of change, many have expressed a desire to continually improve outcomes in the education system. This is true regarding their own achievements and ambitions, such as teachers striving to improve their own skills and learners striving to do as well as they can in their studies and reach higher levels of education. One girl, for example, adamantly claims that she wants to continue working hard to study to become a doctor because her mother is diabetic, and she would like to be able to treat the mother herself.<sup>202</sup> This is also true in terms of broader contributions to girls' education. For example, more than 80.0 percent of caregivers during both the SAS and EL claim that they would like their girls to obtain university-level education. This demonstrates that the key stakeholders engaged via this programme have high potential to contribute to future improvements. Teachers interviewed for this study also highlighted the value of previous training programs and their desire to continually help girls achieve better education outcomes.

What can be gleaned from this is that the willingness to become change agents is present in many of the key stakeholders relevant to the project. Larger obstacles that prevent them from fully enacting change include resource constraints, especially in light of the pandemic, and cultural beliefs and norms that prevent some girls from fully taking advantage of their educational opportunities. Creating sustainable channels for resources, as well as addressing cultural norms in some segments of communities should likely be the main focus of future programming.

## Effectiveness

We now discuss the WWW programme's effectiveness. The OECD defines this as the extent to which the intervention achieved its objectives and its results, including any differential results across groups. We note that the majority of this report is dedicated to discussing the programme's effectiveness, as well as

---

<sup>202</sup> FGD with Girls 1, Nairobi

its short-term impact; as such, this section predominantly provides a summary of findings discussed in previous sections.

As discussed in previous sections, learning outcomes increased slightly for the SeGMA and declined slightly for the SeGRA and STEM assessment at endline, likely as a result of the COVID-19 pandemic. The WWW programme likely contributed to mitigating learning losses in intervention schools, which showed smaller learning losses than comparison schools, but results were inconclusive, potentially due to the small sample size used for analysis. For transition outcomes, we found high transition rates at endline of around 96 percent among all surveyed girls.

For Intermediate Outcome 1, attendance, we find that attendance rates in most counties remain higher than the rates found at baseline and midline. For Intermediate Outcome 2, girls' performance, as discussed above we find declining learning outcomes, but with possible mitigation of learning loss due to WWW programming. For Intermediate Outcome 3, teacher quality, and Output 1, use of gender sensitive teaching approaches, we find generally high overall teaching quality scores, including measures of gender equity in teaching. Quality of teaching was one of the intermediate outcomes with the highest scores at endline; this may drive related improvements in learning outcomes and transition. For Intermediate Outcome 4, life skills, and Output 3, improved confidence, we find a slight overall increase in life skills from SAS to endline, driven by improvement in attitudes towards success in school, confidence in learning, desire to continue pursuing education and training, and general self-confidence. For Intermediate Outcome 5 and Output 4, both related to household support, we find generally high levels of support for girls to pursue school and remain in school. Relatedly, for Intermediate Outcome 6 and Output 5, both related to community support, we find increasing levels of community support from SAS to endline. We note that many caregivers perceive early marriage and pregnancy as a major hindrance towards girls' education, and a majority supported doing away with these practices. However, support for doing away with other harmful practices, such as FGM, was relatively low. Furthermore, we note that many economic barriers remain to support for girls' education, a point emphasized in both quantitative and qualitative data, suggesting that economic factors are also important determinants of community support. For Output 2, establishment of ALPs, there was slight progress towards increasing support for TVET, as well as investment in programmes such as catch-up centres. Lastly, for Output 6, related to strengthening MoE approaches, progress (as discussed in *Sustainability Outcomes*) was strong.

Overall, this summary points to mixed achievement of planned results, but with generally positive trends for most results. It is again important to note the major detrimental effects of the COVID-19 pandemic on outcomes of interest, as well as on project implementation. As such, these findings should not be taken as a decisive statement of project success or lack thereof; indeed, the project's ability to adapt to the challenges of COVID-19 is laudable, despite the pandemic's impact on the effectiveness of specific interventions.

In addition to this analysis of effectiveness, the below questions were presented as areas for investigation in the Inception Report, and are thus addressed here:

**Did the project outputs significantly contribute to the achievement of the intermediate outcomes?**

Project outputs were closely tied to intermediate outcomes, as can be seen from the summary of results above. As such, achievement of outputs was directly linked to intermediate outcome results.

**To what extent were services and items delivered in a timely manner, and to what degree was service provision adequately supported to achieve objectives on schedule?**

In interviews, project staff stated that implementation had matched the vision of the project, with one staff member stating that "fidelity was very high".<sup>203</sup> However, it is again worth noting that the COVID-19

---

<sup>203</sup> KII with project staff, Int. 83

pandemic posed major challenges to programme delivery, albeit challenges that the implementers strove to adapt to and address.

### **Did the project monitoring and evaluation system function satisfactorily?**

While project monitoring helped inform activities and implementation, the lack of tracking data on girls who were evaluated at baseline and midline, as well as limitations to past data, made it impossible to use a quasi-experimental difference-in-differences methodology for the endline evaluation. This issue limits our ability to draw any definitive conclusions about programme impact.

### **Were the indicators and targets used sufficient and adequate to evaluate the impact of the programme?**

Indicators and targets were sufficient to evaluate the impact of the programme, and addressed all areas of relevance, including learning, transition, sustainability, teaching quality, community and household support, life skills, and sexual and reproductive health, among others discussed in this report.

### **How well did the program effectively link with the other GEC-T programs and other stakeholders in the county/sub-county? Were there consistent overlaps in programming and support?**

As described further below under *Efficiency and Value for Money*, many schools reported benefitting from interventions run by other NGOs/groups, and there were overlaps in programming in at least some schools.

### **What are the other hindrances to girls' education that may have been affected by the interventions?**

In interviews, project staff did not mention any unintended consequences of the project on other barriers to girls' education; in part, this is because the GEC Theory of Change very comprehensively addresses barriers to girls' education. However, project staff noted that the programme had downstream effects on boy students because of the impact on teacher training.

## **Short-term Impact**

The main body of this report is dedicated to discussing impact, which is also summarised in *Relevance/Appropriateness* and *Effectiveness* above. As such, we do not repeat that analysis here. However, we wish to emphasise several notable positive findings and barriers to short-term impact in this section.

In general, the bursaries, scholarships, and other financial grants provided to learners and their families have been cited as useful initiatives to help girls obtain an education. Similarly, the increasing use of positive teaching practices, such as instructions methods that emphasise participatory class sessions with learners, have also been cited as key changes that help make both girl learners feel more supported in their learning, and help teachers feel more confident in their work.

However, it is worth repeating that several obstacles remain that prevent further progress, especially for marginalised girls. Ongoing stigma against girls with disabilities and girls who are pregnant continue to create uncomfortable learning environments for some girls, to the point that some may prefer to drop out altogether. Travelling to school and home continues to raise concerns for child protection, as threats to girls' safety continue to be relevant, especially when schools are unable to provide transportation services for girls.

## Efficiency and Value for Money

We now discuss the programme's efficiency and value for money, focusing on answering six questions set out in the Inception Report. According to the OECD, efficiency is defined as the extent to which the intervention delivers results in an economic and timely way. For this analysis, we examine the relative impact of various interventions compared to the resources used for implementing those interventions; ideally, this would include a review of project budgets, but we were not able to review these documents at the time of the evaluation. We then identify which interventions were more cost effective based on a combination of overall impact and resource-intensity. This data is validated through KIIs with project staff which queried aspects of efficiency such as the project's flexibility in using resources, adherence to budgets, and cost reduction measures. KIIs with government officials and quantitative data were used to understand overlaps between the WWWW programme and other programmes within Kenya, which may suggest low levels of overall efficiency within the sector and a need for improved coordination across education implementing agencies.

### **How did the project apply adaptive management practices towards ensuring the best use of resources?**

Project staff described holding "very frequent" meetings with implementation teams to discuss the intended design and implementation of the programme, which enabled adaptive management and the adjustment of programming based on the success (or lack thereof) of implementation.<sup>204</sup>

Adaptations were particularly necessary in light of the COVID-19 pandemic and resultant school closures. To address this issue, the WWWW programme recruited remedial teachers who helped identify students lagging behind and supported these students. While this shows effective adaptation and reassignment of resources to new areas of focus, it is worth noting that many post-COVID activities were rushed due to a revised, shorter school calendar.

### **To what extent did the WWWW program contribute to internal efficiency of the overall education improvement in the eight counties through reduction of dropout and attendance enhancement?**

As shown in the *Transition Outcomes* section, while transition rates were already high at baseline, these rates increased across evaluation rounds. This includes through a reduction of dropout as well as through transition from primary to secondary school, into TVET, or into gainful employment (if of legal age). As shown in the intermediate outcome *Attendance*, however, we found generally declining rates of attendance from SAS to endline, and at endline, attendance rates as measured by headcounts fell below 90 percent in almost all counties.

Overall, these findings suggest that while the programme contributed to enhancing the efficiency of the Kenyan education system by supporting improved transition rates—in line with national policies—the effect on student attendance was limited. More work may need to be done to further strengthen this component of efficiency.

### **How cost-efficient were the implemented project interventions at the school, household, community, system, and girl levels?**

As we could not review project budgets, we cannot provide definitive information on the cost-effectiveness of programming. However, we make several relevant notes on this aspect of efficiency and VfM. First, we note that project staff's activities included ensuring that the project was implemented according to the workplan budget and the defined scope in order to strengthen cost effectiveness and VfM.

---

<sup>204</sup> KII with project staff, Int. 83

Second, we note that interventions seeking to provide bursaries, strengthen income generating activities, or otherwise improve girls' financial access to school were generally noted as less effective (from cost and implementation perspectives) and likely unsustainable. Despite this issue, these activities were continued throughout the life of the project, in part because they were one of the only project interventions to address a substantial barrier to girls' education – economic factors. While bursaries were shown to be useful interventions, this represents a limitation on cost effectiveness.

Third, a project staff member also highlighted that the cost per unit of catch up programmes was high, suggesting that the cost-effectiveness of this intervention may have been low. However, this staff member noted that the output was delivered successfully.<sup>205</sup>

Lastly, turnover of teachers, as highlighted in prior sections of the report, represented a limit on cost efficiency, as money spent to train teachers no longer contributed to improved teaching and learning in project schools. However, it is important to note that this money is not "lost"—teachers, and students in the schools to which they have transferred, still benefit from these interventions, although the benefit is not measurable through this evaluation.

### **Was the project implemented in the most efficient way compared to other alternatives?**

While there were some limitations to the efficiency of program interventions as highlighted above, in general, the programme was implemented in an efficient way. The inclusion of interventions at a wide range of levels—the girl, the community, the school, and the system—serves to strengthen efficiency relative to projects that only address one issue. This is because this project structure addresses the multiple, intersecting issues that limit girls' access to and success in education.

### **What aspects of the project didn't seem cost-efficient and why? How else could the project have implemented them to remain relevant and cost efficient?**

Aspects with limited cost efficiency are noted above, and include bursaries and income-generating activities, catch-up centres, and (to a lesser extent) teacher training. The main area for improvement highlighted by government officials and project staff was bursaries, which were noted as a project activity that would likely not continue beyond the life of the project (discussed further below). While the focus on improving girls' and households' ability to afford education is justifiable—particularly given qualitative data which suggests that this remains a major constraint to school attendance—this appears to be a larger issue than can be effectively addressed by just one programme. As such, in order to improve cost efficiency of these and other similar interventions, it may have been more effective for WWW to partner with other NGOs whose primary focus is on economic growth and livelihoods.

### **Did the evaluation reveal consistent overlaps between the WWW project and other concurrent projects?**

In the school survey, many head teachers noted that their school had benefitted from interventions held by other NGOs to address teaching and learning. As noted in the section on evaluation limitations, 60 percent of all endline schools, including 38 total intervention schools, reported receiving support from NGOs outside of EDT/Concern/WWW. This suggests that there are, indeed, overlaps between the WWW project and other concurrent projects. Improved coordination between the WWW project and NGOs working in similar areas may have been beneficial to improve cost effectiveness and VfM, as well as to strengthen the overall impact of programming.

---

<sup>205</sup> KII with project staff, Int. 83

## Sustainability

We now discuss sustainability, or the extent to which the benefits of the WWW interventions are likely to continue. We note that sustainability is a main outcome of the programme; we do not replicate that analysis here. Instead, we discuss other aspects of sustainability utilising data not included in the *Sustainability Outcomes* section.

The WWW programme worked closely with communities, county authorities, and the MoE to achieve and sustain education outcomes. At the community level, the project worked with community members on income generating activities to improve the sustainability of bursaries, and worked with community groups to challenge cultural barriers to education and to address action points to strengthen girls' education. The project also worked with community health volunteers (CHVs) to provide well being support to girls and households, improve the provision of mentoring when schools were closed and follow up with girls who were not attending school. As these CHVs are permanent staff employed by the Ministry of Health, their work continues after the close of the project.

At the county and MoE levels, the programme worked closely with relevant authorities. For example, in order to strengthen sustainability at the system level, Ministry of Education (MoE) officials were trained on how to undertake monitoring of teaching and learning activities as a way of ensuring sustainability. The project also strengthened linkages between curriculum support officers, quality assurance officers, and schools to help strengthen teaching and learning. The programme is now intended to be handed over to the MoE as an exit strategy.

Sustainability mechanisms are discussed in detail in the previous *Sustainability Outcomes* section. In summary, the project has focused on the community, school, and system levels to strengthen sustainability. This includes interventions focusing on increasing community participation in the creation and implementation of community action plans, improving community support for girls' education, strengthening school leadership and teaching practices, expanding access to co-curricular activities, putting systems in place to enable the MoE to support learning and gender responsiveness, and increasing support for TVET.

As shown in the relevant outcomes section, sustainability had not been fully embedded by the endline evaluation, with particularly limited success at the community level. However, progress across all dimensions of sustainability did increase from baseline to endline.

We additionally note that, outside of indicators tied to the sustainability outcome, the project's interventions to strengthen girls' learning, life skills, and other outcomes are expected to have long-term sustainable impacts. The United Nations, for example, states that education and reproductive health for girls are key to sustainable development.<sup>206</sup>

As described above, county education officials were highly involved in the project, and were targeted with some project activities. Given that the project's exit strategy is handover to the MoE, strong input from relevant county and national government authorities is expected beyond the life of the project. However, both government officials and project staff expressed some reservations about the sustainability of the project, and specifically some sub-components. The most frequently highlighted issue revolved around monetary support for project beneficiaries, including through bursaries. One project staff member, for example, stated that income-generating activities were ineffective, limiting sustainability as communities would no longer be able to monetarily support girls to attend school once programme support is

---

<sup>206</sup> United Nations, 2016. "Education and reproductive health for girls key to sustainable development". [www.un.org/youthenvoy/2016/03/education-reproductive-health-girls-key-sustainable-development/](http://www.un.org/youthenvoy/2016/03/education-reproductive-health-girls-key-sustainable-development/).



removed.<sup>207</sup> Another staff member similarly stated that the bursary program is likely to go under because “other donors and partners don’t see it as necessary or successful”.<sup>208</sup>

The project has been generally viewed positively by local communities and schools, and is highly credible. Teachers who were familiar with the project, for example, stated that its interventions for teaching and learning were useful; one teacher at a special school in Mombasa provided the following example:

*The EDT [training], in fact we have been going there I think three or four times. The last time we were there, we benefitted as a special needs school because it touched our learners. I remember when we had been going to the other seminars... most of the time we realised that the organizers were not considering [the learning environment of disabled learners], until the time we went to [the EDT training] and I was introduced to somebody who really helped us.*<sup>209</sup>

In general, the project’s focus on a variety of avenues affecting girls’ success in school and beyond—including the girl herself, households, communities, schools, and the system—has enabled the project to engage closely with local populations, strengthening this aspect of sustainability.

Furthermore, in KIIs with government officials, officials consistently stated that the WWW programme aligned with the priorities of the MoE and with existing education policies and strategies. An official from Turkana, for example, stated that “the objectives have been consistent and none of them have been out of policy; they’ve value added to the priorities of the Ministry”.<sup>210</sup> An official from Samburu similarly stated that “the [WWW objectives are] very consistent with the MoE objective to have every child have basic education and to be self-sufficient”.<sup>211</sup>

County level government officials did not mention any policy changes that had come about as a result of WWW implementation; however, at national level, the project is informing the formulation of policies such as the teacher education policy, guidance and counselling policy, school health policy, guidelines on accelerated education and development of a new education strategic plan .

## Gender Equity and Social Inclusion

In this section, we provide a quick overview of gender equity and social inclusion (GESI) considerations of the programme and build off the comments provided during the ML analysis, during which GESI considerations were last elaborated on. As we have not been provided with an opportunity to engage with programme implementers directly to replicate the WWW Project GESI Assessment tool seen in the ML, the table below highlights the external evaluators’ comments and observations during the EL, in relation to comments previously provided by the project team at BL and ML.

---

<sup>207</sup> KII with project staff, Int. 88

<sup>208</sup> KII with project staff, Int. 84

<sup>209</sup> KII with teacher, Mombasa, Int. 30

<sup>210</sup> KII with government official, Turkana, Int. 91

<sup>211</sup> KII with government official, Samburu, Int. 96

**Table 90: Summary of gender equity and social inclusion comments by project team and external evaluator**

Outputs	Baseline (Project Comments)	Midline (Project Comments)	Endline (EE Assessment)
Output 1: Teachers and school leaders in primary and secondary schools demonstrating gender sensitive and enhanced teaching approaches (ICT and pedagogy) for improved learning	Sensitive to the needs of special needs children. Teacher training includes adaptive content for teaching children with special needs.	Adaptive content on SNE is integrated in the teacher training manuals, teacher training and delivery recognizes the diverse needs of learners, e.g. over age girls; young mothers and children with disabilities. Instructional coaches have been trained on how to support school teachers to address these differences and marginalization. However, SNE assessments are yet to be completed to allow for individualized and targeted learning.  The project has provided SNE equipment to learners and increased textbook accessibility for learners with disability through the use of orbit readers which converts print material into Braille.	Evidence from the girl surveys and classroom observations suggest that gender sensitive and enhanced teaching approaches are increasingly utilised in classroom settings. These include equal, if not higher, frequency of calling on girls to answer questions, and use of facilitative teaching approaches.  Special needs education, while somewhat implemented, remains scarce, as a number of teachers explained that they do not have the capacity or resources to fulfil SNE requirements. Having a disability is one of the significant predictors of having dropped out and re-enrolled during school closures, after controlling for other factors.
Output 2: Alternative learning pathways established or expanded for girls outside or at risk of leaving school		The project is acknowledging that there are missing drop-out girls who are not accounted for, therefore we are blind to their exclusion factors, and no current interventions seem to be attracting them.	Awareness of alternative learning pathways has increased among learners, though the percentage willing to go through this route continues to decrease, from 81.3 to 64.1 percent of surveyed girls between the RAS and the EL. What this does suggest is that traditional pathways have remained more prioritised by learners, to the point that TVET and ALP programmes are considered the same as being a drop out. This portends potential difficulties in getting out of school or a-

			risk girls to take up ALP options.
Output 3: Improved self-confidence and aspirations among the girls in mentorship and scholarship programmes	The selection criteria for scholarship was transparent and mindful of special needs children, orphans, and child headed households.	Local community resource persons are engaged in identifying marginalized girls and then serve as mentors during community theme days and holiday mentorship forums; this ensures that the positive values and community resources are utilised, and their positive interests are safeguarded. The project is yet to establish if there are any social groups underserved by the local resource persons.	EE not provided data on recipients of such programmes, and did not inquire specifically about it during data collection. However, some improvement in all girls' life skills was measured at endline compared to SAS.
Output 4: Households continued support for girls' education including in alternative pathways	CHVs are part of the community and locality thus speaks the native language. Information collected at HH level also on social wellbeing to continued support of girls and boys.	The project has adopted a whole community approach for engaging households to support girls' education. The project is yet to determine if the community approach is socially inclusive as the project might be blind sighted if there are community exclusions for example if some tribal and/or social group is not included in the community groups.	EE not provided data on the composition of community groups, though our finding that only 38.0 percent of caregivers had heard of community conversations does strongly suggest that already marginalised social groups are likely to not be engaged by such initiatives.
Output 5: School catchment communities more aware of the importance, benefits and opportunities available to support girls for productive education	CFs selected by the community based on those who know how to read and write. No data on the number of people with disabilities involved in groups, or number of CFs with disability or how inclusion is covered in conversations.	The project facilitates community dialogue but is unaware if there are any social groups not participating in the process. The project believes all groups are involved but needs to confirm this.	Awareness of formal community structures to support girls' education, such as the community conversations and community action plans, remains low. This strongly suggests that some marginalised social groups are likely to be excluded from these formal opportunities.  However, informal community support mechanisms exist, whereby community members support girls' abilities to obtain an education. Seventy-nine

			percent of surveyed caregivers reported benefitting from financial support from communities, 48 percent from material support, and 39 percent from mentoring. It is possible that their inclusion therein is higher.
Output 6: WWW project aligned to WWW models inform emerging MoE gender and teaching approaches	Stakeholder meetings done in some counties. Stakeholders' forums do not have an agenda for gender dimension nor social inclusion. The forums are mainly used for coordination and information sharing	The project is also implementing dormant policies such as non-formal education and back to school policy for teen mothers, community of practice among teachers as well as school-community linkages programmes. Through this, the project is evidencing modalities of implementing these policies and lessons for policymakers' learning. With continuous implementation, documentation and research, the project stands to have evidence that can be used in the revision of these same policies. Through this, the project can then stand as a thought leader in the evolving formulation and roll out of these policies that focus on the marginalized.	Government officials indicate desire to continue using teaching methods, including gender sensitive ones, after the end of the WWW-T programme. They have expressed concerns around whether these will even have a change to be applied in practice, as the end of cash transfers and bursaries, which government officials have claimed will not be continued, means that some of the previously overcome barriers preventing regular attendance might become relevant again after the project ends.

Based on this assessment, several key findings emerge:

- Bursaries, scholarships, and financial support are seen as among the most effective initiatives for helping girls in their education. However, government officials have claimed that while they are appreciative of the WWW-T's contributions and would like to continue best practices into the future, they will face difficulties in continuing financial support for learners and their families, which in turn might roll back the gains made over the course of the project. Priorities for future programmes should include emphasising sustainable channels of funding and support for girls, including by tapping into existing community resources and support structures.
- In the same vein, one way in which community support mobilisation may be improved to increase resources for learning may be to emphasise that some of the investments made to school facilities and staff may have carry-over effects for boy learners as well. These benefits which others can experience could be communicated via mechanisms such as the community conversations, while taking due care to emphasise that girls face uniquely difficult challenges compared to their male peers.

- As in the ML, support for girls with disabilities could still be strengthened, including through the mainstreaming of SNE and other accommodative initiatives, including those to positive changes in community attitudes towards girls with disabilities.
- Alternative pathways of education and TVET programs are increasingly seen as non-viable options for education, with the percentage of girls willing to go through such programs dropping by nearly 20 percentage points between the RAS and EL. It is possible that our data, collected from girls enrolled in schools, leads to bias against ALPs, and that girls who are already dropped out and struggle to re-enrol may in fact be more willing to take enrol in ALP programmes. Nonetheless, the frequent insinuation that ALPs and TVET programmes are for those who have “failed” the formal education system and are as good as not enrolling in any education programmes suggests that additional initiatives to highlight the value of ALPs may be beneficial for the most at-risk girls, especially those who may shy away from such programmes for fear of social stigma.

## Conclusion

### Primary Outcomes

This endline evaluation has built upon previous evaluation rounds, particularly the midline conducted in 2020 and the SAS conducted in 2021. Overall, we find that students have begun to recover from learning losses incurred during the COVID-19 pandemic. In general, this recovery has occurred at a greater rate for intervention students than for comparison students—although not significantly so—suggesting a potential impact of the WWW programme on learning outcomes.

When further examining learning outcomes by subgroup, we find that ASAL students had begun to fall behind at endline in comparison to students in urban slums. This dynamic, however, was predominantly driven by the relatively strong performance of students in Nairobi, suggesting that girl students in Nairobi may face fewer barriers to learning than other marginalized girls.

We further note that our predictive analysis of learning outcomes by demographic characteristics found no significant correlations between individual demographic factors and learning, but a positive and significant relationship between life skills and learning outcomes. This finding validates the programme’s holistic approach towards strengthening life skills, decision-making, confidence, and other related skills.

Looking at transition, we find very high transition rates—although somewhat below programme goals—at endline, a positive sign of potential programme impact. Because transition rates are extremely high, we find no significant differences in transition by subgroup, nor any significant predictors of transition. However, when analysing grade-to-grade progression rates, we find that ASAL girls have higher repetition rates, on average, than girls in urban slums, although girls in Nairobi had the highest repetition rate.

Girls were also considered to have successfully transferred if they pursued an alternative learning programme, such as a TVET programme. Our findings suggest that there has been a general increase in the awareness levels of girls for these transition options. Although TVET courses remain a secondary option in case girls are not able to get the grades to go to university, it seems that girls and parents understand that it is a viable option for girls to gain skills for the labour market.

Overall, the largest challenge to transition is the ability of families to pay for a girl’s education, a dynamic also highlighted in analysis of household and community support for education. This highlights the importance that scholarships and cash transfer have played in keeping girls at schools. However, it also suggests that the programme may leave a gap here once it comes to an end as the girls who had been beneficiaries of the programme may find themselves struggling to pay for their school fees again.

Regarding sustainability, we find an improvement in sustainability at endline compared to midline and baseline, with find higher levels of sustainability at the school level than at the community or system levels. High levels of school sustainability were driven first and foremost by very a strong teaching practices score, which improved significantly and substantially between baseline and endline. However, while there has been an increase in overall sustainability from baseline to endline, the increase between midline and endline was limited—perhaps due to the effects of COVID-19 on project implementation. We note also that in KIIs, respondents from the government and project staff similarly suggested that some interventions were likely to be more sustainable than others, with bursaries and income-generating activities noted as particularly lacking sustainability.

## Intermediate Outcomes

In addition to these three primary outcomes of interest, the endline report also examined intermediate outcomes and other auxiliary outcomes of interest. Intermediate outcomes include attendance, teaching quality, life skills, household support, and community behaviour and support, while auxiliary outcomes include school governance and management, economic empowerment, child protection, sexual and reproductive health, the impact of COVID-19, and attitudes towards STEM subjects.

Overall results for intermediate outcomes were mostly positive. Attendance rates were found to have increased from baseline and midline to endline, although slightly decreasing since the SAS round. Quality of teaching as measured through classroom observations was overall high and represents a success story for the programme; however, some gender biases in teachers' attitudes towards teaching and learning remain which warrant further improvement. Regarding life skills, we measured a slight increase from SAS to endline, a positive finding given the significant correlation between these skills and learning outcomes described above. For household support, we find high levels of caregiver support for girls' education; however, barriers such as lack of financial means to support girls in school remain a substantial limitation. Lastly, for community attitudes and support, we find relatively few formal community initiatives to support girls in education, but generally positive attitudes towards girls' education and willingness to support girls. We also find that caregivers expressed that early marriage and pregnancy was a major hindrance towards girls' education, with many suggesting that this practice should be done away with.

In comparison to these positive findings, results for auxiliary outcomes were somewhat more mixed, although with some success stories. Many caregivers had low perceptions of school governance and management, driven by low perceived levels of activity and effectiveness of BoMs and PTAs. Activities related to economic empowerment were limited, with relatively few caregivers expressing awareness of financial grants or receiving these grants. Over one-third of surveyed girls at endline were aware of instances of violence or harm against children in their communities, with the highest incidence rates in Mombasa and Nairobi, an alarming statistic. Knowledge of and attitudes towards SRH, however, were strong, although there is a need to continue strengthening attitudes towards use of contraceptives in order to ensure that girls avoid pregnancy and STDs. Lastly, girls expressed very positive attitudes towards STEM subjects, but analysis of enrolment suggests that they may particularly struggle with physics, which requires higher levels of mathematics ability.

## Concluding Remarks and Recommendations

Overall, these findings suggest that while progress has been made over the WWW implementation period, work is still needed to enhance learning, transition, and life outcomes for marginalised girls. Some of the impacts of the programme are likely to continue for years; for example, improved teaching quality will have downstream effects on learning and transition outcomes for girls (and boys) for as long as trained teachers remain in their careers. However, the impact of the COVID-19 pandemic on marginalized girls—in addition to boys, households, communities, and many other groups—cannot be understated. This impact will reverberate for years, and may necessitate further interventions in the future.

In light of these findings, it may be helpful for future education programming—whether government or NGO—to consider the following recommendations:

1. **Target ASALs for future learning-related interventions.** The endline evaluation found that ASAL counties lagged urban slums—particularly Nairobi—in terms of learning outcomes. Turkana in particular faced relatively weak learning outcomes. It is difficult to unpack the exact reason for these findings; they may be influenced by environmental factors, such as persistent droughts, economic challenges, and cultural practices which serve to marginalize girls and make it challenging for them to pursue their education. Regardless of the specific cause of these outcomes, however, it is clear that more work is needed in the ASAL counties to support girls' education.
2. **Support STEM teaching and learning.** Results for STEM assessments were uniformly low at both SAS and endline. However, girls expressed high levels of interest in and desire to pursue STEM subjects, often because these subjects were considered to be useful for future careers in areas such as medicine or engineering. Studying STEM subjects does, indeed, provide useful skills for the job market; girls participation—and, even more crucially, performance—in these subjects should thus be bolstered. In qualitative interviews, many girls mentioned the effectiveness of practical demonstrations for learning; strengthening teachers' abilities to provide practical science lessons may be one method to further engage girls in STEM learning.
3. **Continue supporting girls to acquire important life skills, such as confidence, self-esteem, and attitudes towards education.** These skills were found to be significantly correlated with learning outcomes for girls and, unlike demographic factors such as disability status or native language, can tangibly improve over time. Furthermore, strengthening these skills may not only improve students' learning, but also improve their overall life outcomes.
4. **Address the major impact of economic factors on transition.** In our analysis of transition, we found that the biggest challenge to transition was the ability of families to pay for the girl's education. This finding highlights the importance that scholarships, bursaries, and similar interventions have had to keep girls in school. However, it is also important to note that such activities are not highly sustainable. As such—while likely outside of the scope of any one organization to address—it will be important to acknowledge the need for economic stability within households and communities in the push for universal transition, and integrate education programmes with economic growth programmes as appropriate.
5. **Provide flexible “catch up” or remedial options for girls at the secondary school level, building off of the success of catch-up centres and remediation initiatives.** As noted in the section on transition, catch up centres have provided a useful option for girls to return to school with a flexible schedule. However, there are currently no similar flexible options at the secondary school level to which girls can transition once they have completed primary catch up courses. It may therefore be helpful to study possible options for instituting flexible study or catch up centres at the secondary school level to help marginalized students transition through education. This initiative may also help expand options for older girls who dropped out of school, or who have now fallen behind in their learning, due to COVID-19.
6. **Tailor teacher training interventions by county.** The evaluation found that during COVID-19, teacher attrition was higher in ASAL counties than in urban slums. As such, the number of trained teachers, and the amount of training exposure of teachers, varies substantially across counties. Furthermore, training on culturally-influenced factors such as gender sensitization may be more needed in ASAL counties. To deal with these issues, it is important that any future interventions do not treat teacher training as “one size fits all”, but rather provide training that is tailored to the teachers' experiences, background, and school location.
7. **Strengthen formal and informal mechanisms of community support for girls' education.** While around two-thirds of caregivers reported that girls in their community had received support from the community for education, only 40 percent stated that there had been community initiatives to strengthen girls' education in the past year. This suggests that while there may be

high levels of informal support for girls' education, it may not translate to community action plans or a similar, formalized structure. Given that economic barriers were noted as a main challenge for transition, it may be helpful to continue making efforts to formalize community support mechanisms to strengthen learning and transition outcomes.

8. **Expand and strengthen community structures providing support to girls affected by violence.** More than one-third of girls at endline were aware of incidents of violence or harm against children in their community. These incidents were rarely reported to occur at schools, a positive finding. However, there is still a need to support children affected by violence occurring within communities and households. Only around 45 percent of girls were aware of community structures to support children; as such, expanding and strengthening these structures may help provide more recourse for children affected by violence and provide a platform for further initiatives to decrease violence against children.
9. **Continue supporting/conducting government-led trainings on school management and governance.** Students' caregivers reported mixed views of school management and governance, particularly driven by low perceived levels of school council, BoM, and PTA communication and activity. However, in qualitative interviews, government officials noted that trainings were being held to strengthen school management. Given the need for further improvement of this area, these trainings and similar initiatives should continue to be supported.
10. **Improve perceptions of and access to contraceptives.** While girls generally showed high levels of SRH knowledge and confidence, willingness to use condoms and other forms of contraceptives were much lower. Furthermore, only around one-third of girls at endline stated that they would know where to get a contraceptive method. These findings suggest both stigmatization of contraceptives, as well as a lack of access. Expanding girls' ability to access contraceptives privately may help address both of these challenges, as girls will not feel shamed by their communities for pursuing safe sexual practices. Furthermore, expanding contraceptive use will help reduce early pregnancies and (for the case of condoms) STI transmission. The former outcome is particularly relevant to this evaluation, as pregnancy was described as a main reason for girls to drop out of school.





# WWW Endline Evaluation Annexes

## Table of Contents

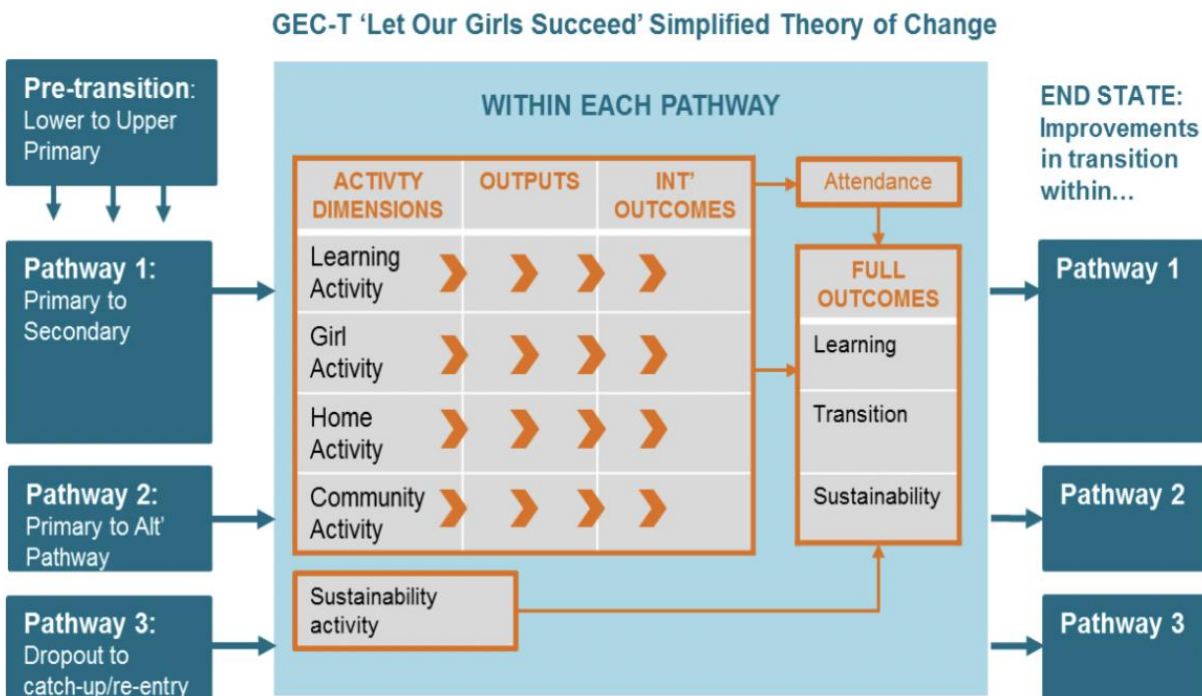
Annex 1: Project Design and Interventions.....	3
Annex 2: Endline Evaluation Approach and Methodology.....	4
Evaluation Questions.....	4
Evaluation Methodology.....	6
Data Collection Tools.....	9
Pre Data Collection.....	11
During Data Collection.....	12
Post Data Collection.....	21
Approach to Gender Equality and Social Inclusion.....	21
Challenges and Limitations.....	22
Attrition Analysis.....	25
Differences Between Midline and Endline.....	30
Differences Between Intervention and Comparison Girls.....	31
Annex 3: Learning Outcome Data Tables.....	32
Annex 4: Characteristics and Barriers.....	35
Annex 5: Logframe.....	37
Annex 6: Outcomes Spreadsheet.....	37
Annex 7: Beneficiaries Table.....	37
Annex 8: Inception Report.....	37
Annex 9: Data Collection Tools Used for Endline.....	37
Annex 10: Qualitative Transcripts.....	38
Focus Group Discussion with Girls, Kilifi, Int. 13.....	38
Key Informant Interview with Teacher, Kilifi, Int. 31.....	51
Annex 11: Quantitative Datasets, Codebooks, and Programs.....	55
Annex 12: Learning Test Pilot and Calibration.....	55
Tests Used.....	56
Methodology for Marking Tests.....	57
Enumeration Challenges.....	57
Post-Calibration of SeGRA and SeGMA.....	57
Annex 13: Sampling Framework.....	58
Annex 14: External Evaluator Declaration.....	58

Annex 15: Additional Introductory Tables .....	60
Activities.....	60
Project Beneficiaries .....	61
Assumptions .....	63
Annex 16: Supplementary Analysis of Learning Outcomes.....	67
Literacy Outcomes.....	67
Numeracy Outcomes.....	73
Predictive Analysis of Learning Outcomes.....	79
Annex 17: Supplementary Analysis of Other Primary Outcomes .....	82
Transition Outcomes .....	82
Sustainability Outcomes .....	86
Community Level.....	87
Annex 18: Supplementary Analysis of Intermediate Outcomes.....	88
Attendance .....	88
Teaching Quality.....	92
Life Skills.....	95
Household Support.....	99
Community Behaviour and Support .....	103
Annex 19: Supplementary Analysis of Other Outcomes of Interest.....	104
School Management and Governance.....	104
Child Protection .....	110
Sexual and Reproductive Health.....	111
COVID-19 Impact .....	116
Attitudes Towards STEM.....	118

## Annex 1: Project Design and Interventions

The project's abbreviated Theory of Change is pictured below.

Figure 1: Theory of Change



Specific project activities, intermediate outcomes, outcomes, and assumptions are described in the report introduction.

The evaluation suggests that the Theory of Change remains relevant. Learning-related activities, including trackers for school attendance, provision of bursaries and grants, training of coaching and teachers, support to provide learning materials, special needs learning training and materials, and capacity-building support for head teachers, are directly tied to Output 1 (enhanced teaching approaches for improved learning). In turn, this Output serves (somewhat tautologically) to strengthen Intermediate Outcome 2 (girls' performance in school) and Intermediate Outcome 3 (teacher quality). It is also linked to Intermediate Outcome 1 (attendance), both through interventions directly related to attendance and by improving the learning environment, thereby providing more incentives for girls to attend school. In turn, these intermediate outcomes are directly linked to Outcome 1 (learning) and Outcome 2 (transition).

Activities directed at girls, meanwhile, include strengthening girls' life skills and providing mentorship. As discussed in the report, an interesting finding was that life skills score was significantly correlated with improved learning outcomes for girls, providing strong evidence for the impact of these activities (and related outputs and intermediate outcomes) on Outcome 1.

Activities at the home and community level include training of community health volunteers, tracking learners, providing cash transfers, dispersion of solar lamps, and training of community members on social accountability and the importance of girls' education. These activities directly tie to Output 4 (household support for girls' education) and Output 5 (community awareness of benefits of girls' education), which in turn lead to Intermediate Outcome 5 (household support for transition) and Intermediate Outcome 6 (community support for transition). Improved household and community support then leads to improved transition rates (Outcome 2) and strengthens project sustainability, as community

members and caregivers see the benefits of girls' education, including the benefits of alternative learning pathways such as TVET. It may also lead to improved learning outcomes; the analysis showed some possible relationships between girls' household responsibilities and outcomes of interest, and it is plausible that if households greater support girls' education, they may be more likely to provide time and learning materials to their girls for studying, thereby leading to improved learning.

Lastly, assumptions detailed in the Introduction appear to have broadly held up during project implementation. We only note one assumption at the community level that was potentially under-considered by the project: that *theoretical* community support for girls' education would translate to *tangible* community support for girls' education. As shown in the report, while caregivers generally reported high levels of support for girls to stay in or return to school, many also reported that they did not have the financial resources to ensure that girls did so. As a result, interventions to increase the value communities and households place on girls' education may still face limitations in improving girls' transition rates.

## Annex 2: Endline Evaluation Approach and Methodology

This Annex expands upon the methodology described in the main body of the report.

### Evaluation Questions

The below questions guided the evaluation design.

#### Relevance/appropriateness:

- How well did the overall design of the WWW project address the intended objectives as outlined in the ToC? Specifically, how successful was the project in supporting 70,000 girls in marginalized communities to improve their learning outcomes, transition through key stages of education, and successfully move to the next phase of life?
- To what extent were the learning models used in the delivery of learning (Active Learning Model, Discovery, Product and Production, direct instructional model) implemented in the classroom? Have there been any changes in the learning and teaching process as a result of these models' implementation? How did these changes impact female students, specifically?
- What kind of community involvement is there in the schools? How have the schools collaborated with the community in the education system? How has community involvement changed over the course of the WWW project?
- How do the beneficiaries/stakeholders (namely, teachers and students) see themselves in their role as agents of change and their ability and willingness to take on this role? What are their constraints? Is the situation conducive enough for the beneficiaries to become agents of change? Is the situation sustainable (expanded in the sections, below)?
- How are school committees' roles blended to support project interventions? How well are they functional? What kind of improvements have they introduced?
- What and how have the innovative teaching approaches been implemented? How well have the activities contributed to the girls' performance in the targeted learning outcomes?

#### Effectiveness:

- To what extent did the WWW programme contribute to improved learning and transition outcomes?
- To what extent were the planned outputs and intermediate outcomes achieved? Did the project outputs significantly contribute to the achievement of the intermediate outcomes? To what extent

were services and items delivered in a timely manner, and to what degree was service provision adequately supported to achieve objectives on schedule?

- Is the capacity of the teachers who have been trained sufficient? How are the trained teachers utilizing the skills gained from the GEC-T?
- Did the project monitoring and evaluation system function satisfactorily? Were the indicators and targets used sufficient and adequate to evaluate the impact of the program?
- How well did the programme effectively link with the other GEC-T programs and other stakeholders in the county/sub-county? Were there consistent overlaps in programming and support?
- To what extent did the girl-targeted interventions (mentorship, Community Health Volunteer support, economic strengthening, etc) influence the girls to stay in school, learn, transition after school, and prepare them psychologically (i.e., life skills, self-confidence, self-esteem, and sexual and reproductive health)?
- How well have the community and household interventions influenced change of attitude and practices on the cultural barriers to girls' education? What are the other hindrances to girls' education that may have been affected by the interventions?

#### **Impact:**

- How have the learning and transition outcomes (Numeracy and Literacy) improved through the interventions?
- Has the WWW project design achieved the expected impact? Has it been an effective catalyst/influence in creating a friendly and safe space for the girls and their education? What factors remain that keep girls' from feeling safe in accessing education?
- How have the teachers' attitudes and skills developed to foster improvement of learning outcomes among school going children and especially girls? Are changes more obviously present in some subjects and less so in others?
- What are some of the short-term impacts from the interventions implemented by the GEC-T project across the 8 counties? How sustainable are they beyond the life of the project?
- What are some of the project successes in the girls' transitions? How viable were the three pathways of transition? What other transition pathways did the project achieve?

#### **Efficiency:**

- How did the project apply adaptive management practices towards ensuring the best use of resources?
- To what extent did the WWW programme contribute to internal efficiency of the overall education improvement in the eight counties through reduction of dropout and attendance enhancement?
- How cost-efficient were the implemented project interventions at the: school, household, community, system, and girl levels?
- Was the project implemented in the most efficient way compared to other alternatives?
- What aspects of the project didn't seem cost-efficient and why? How else could the project have implemented them to remain relevant and cost efficient?
- Did the evaluation reveal consistent overlaps between the WWW project and other concurrent projects?

#### **Sustainability:**

- In what ways did the WWW project work with and/or involve other key community stakeholders, county authorities, and TSC/MoE headquarters?
- How has the project embedded a sustainability mechanism during the implementation of its activities? What sustainability plans has the project put in place around school, community, households, and the girls themselves?

- How did the county governments participate in the project implementation? What kind of input can be expected from the county and national government beyond the life of the project?
- How is the WWW project perceived by the local population? What is the credibility of the programme? How would the local authorities sustain the project implementation strategies beyond the life of the project?
- How well did the project work with MoE/TSC/key stakeholders to achieve and sustain the education outcomes for the marginalized girls?
- To what extent did the WWW project align with the existing education policies and strategies? How did the project influence policy change, if any, based on the emerging needs from implementation?

## Evaluation Methodology

The endline evaluation utilised a mixed-methods evaluation with quantitative and qualitative tools. For the sake of brevity, fieldwork procedures and attrition rates from the sample are discussed in Annex 2; here, we discuss the basic design of the endline evaluation, data collection tools, and the achieved sample.

Our evaluation had three distinct cohorts:

- **Cohort 1:** This cohort included girls who participated in the midline evaluation. It included both intervention and comparison schools.
  - **Intervention:** This subsample was drawn from schools that were sampled in the midline and SAS evaluations, and in which EDT works.
  - **Comparison:** This subsample included girls who were sampled at midline in schools where no EDT interventions have taken place.
- **Cohort 2:** This cohort included girls who were sampled in the SAS evaluation only and is limited to girls in target counties and schools. We generally also constrained this sample to schools with at least six girls expected to remain in attendance for logistical and budgeting purposes.<sup>1</sup>
- **Cohort 3:** This cohort included girls outside of the sample schools in three subgroups. This sample was contacted via a Computer-Assisted Telephonic Interview (CATI) methodology, as many girls were no longer present in target schools and counties and were widely spread across the country, making in-person surveys infeasible.
  - **Tracer Study:** We attempted to recontact girls who were sampled in the SAS evaluation but who were no longer in target counties or schools. This did not include girls who had completed Form 4 and successfully graduated.
  - **Alternative Pathways:** This subsample included girls from TVET, apprenticeships, or in-school catch-up programmes.
  - **Kwale and Marsabit:** This subsample included girls reached at midline in Kwale and Marsabit counties, who could not be surveyed in person for logistical and budgetary reasons.

These three cohorts contribute to analysis in different ways. Cohort 1 data is critical for analysis related to programme impact, as it is the only cohort with comparison data. Difference in difference analysis around learning outcomes relies on midline data and Cohort 1 endline data. This design allows inferences to be drawn regarding changes in girls' assessment scores over time, by comparing the relative change among girls in intervention schools to the relative change among girls in comparison schools. However, Cohort 1 is limited in its small sample size, an issue discussed further below.

Cohort 2, as the largest sample, is required for subgroup analysis and disaggregation between groups. We also rely on Cohort 2 data to understand predictors (at the girl, household, school, and community levels) of key outcomes of interest. This is supplemented by endline data from Cohorts 1 and 3.

---

<sup>1</sup> This eliminated 25 schools and 180 girls from the sample.

Cohorts 2 and 3 are the primary data contributors to transition analysis. Analysis of Cohort 2 learners will provide insights into predictors of transition and re-enrollment. The Cohort 3 tracer subsample was intended to be used to track various pathways (and their predictors) of learners who have left their schools since the SAS evaluation; however, issues recontacting these girls, discussed further below, means that we do not conduct analysis for this subsample.

### ***Cohort 1 Analysis Methodology***

For analysis using Cohort 1, the evaluation utilises a quasi-experimental difference-in-differences design. As a preface to this approach, we note that sample size limitations mean that our chance of finding no programme effect when, in fact, there actually is a programme effect, is high.<sup>2</sup> In other words, if our analysis shows no programme effect on learning outcomes, this finding may only be due to small sample size. However, small sample size has no effect on our chance of finding “false positives”, or finding programme effect when, in fact, there actually was no programme effect.<sup>3</sup> In other words, if our analysis shows a statistically significant impact of the programme on learning outcomes for intervention girls as compared to comparison girls, small sample size does not detract from this finding.

With this limitation in mind, in order to utilise the difference-in-differences methodology, we incorporated an explicit comparison group of schools/communities that did not receive WWW interventions; school-age girls sampled from these comparison schools constitute the comparison group for assessment of learning outcomes. Data collection in comparison schools and intervention schools occurred at midline, and we tracked girls using information from this evaluation round. This design allows for inferences to be drawn regarding changes in girls’ assessment scores over time by comparing the relative change among girls in intervention schools to the relative change among girls in comparison schools.

Use of a difference-in-differences approach accounts for two important sources of bias when drawing inferences regarding programme impact: first, by comparing the same respondents across evaluation rounds, the design accounts for underlying cross-sectional differences between treatment and comparison groups. For instance, if programme schools already perform at a higher level than non-intervention schools at the time of the midline or in the absence of any intervention (i.e., under the counterfactual situation when no project were implemented in any school), this fact would bias straightforward cross-sectional comparisons of outcomes at the endline. By adjusting explicitly for gaps that existed prior to project implementation, the difference-in-differences approach ensures that pre-existing differences do not bias the conclusions drawn about project impact. Second, the design controls for systematic changes in outcomes over time that are not attributable to the programme itself. For instance, in the context of a widespread shock that reduced learning outcomes across most of the region (such as COVID-19), a simple longitudinal comparison of learning outcomes in programme schools over time would suggest that the programme reduced student performance. By incorporating an explicit comparison group, the design is able to control for systematic shocks that affect both intervention and non-intervention schools.

### ***Cohort 2 and 3 Analysis Methodology***

For cohorts 2 and 3, due to the methods employed in previous evaluation rounds, we had no ability to assess comparison groups against which progress in intervention schools can be measured. As such, we utilised these cohorts to provide descriptive analysis of the status of outcomes of interest, test the theory of change, and conduct a predictive analysis of the relationship between various student- and school-level characteristics and outcomes of interest (such as literacy, numeracy, and STEM assessment scores; attendance; life skills and attitudes; SRH knowledge and attitudes; and attitudes towards STEM subjects, among other outcomes).

---

<sup>2</sup> i.e., a Type II error.

<sup>3</sup> i.e., a Type I error.



For the predictive analysis, we utilise hierarchical linear modelling (HLM) to determine the extent to which student- and school-level characteristics determine variability in outcomes. HLM is a form of ordinary least squares regression that is used to analyse variance in outcome variables when the predictor variables are at varying hierarchical levels. In our case, because outcomes such as assessment scores of students vary according to their common teacher, classroom, and school, HLM is an appropriate approach, and can further be used to calculate the level of variation in students' outcomes due to their common factors (such as shared school).

Our initial approach utilises standard student-level predictors such as age, gender, and disability, which will be tested for statistical significance as predictors using stepwise regression. School fixed effects are also included (an HLM technique) to control for variation occurring at the school level which does not change (or changes at a constant rate) over time. This approach does not allow us to understand the specific school-level factors predicting literacy, but is the most robust model available to understand individual-level predictors.

Our second approach allows us to better understand school-level factors by excluding school fixed effects (as these would be colinear with school-level predictors of interest) and including variables such as use of engaging teaching practices, teacher attendance, and student-teacher ratios. This model is less rigorous for determining individual-level predictors, and we thus only report values of school-level variables in our analysis.

### ***Qualitative Data Analysis Methodology***

Our analysis of qualitative data from key informant interviews (KIIs) and focus group discussions (FGDs) focuses on validating results from the quantitative survey and providing information on programme outcomes that are difficult to measure through quantitative data, such as sustainability and efficiency. Qualitative interviews provide insight into the 'how' and 'why' of trends observed in the quantitative data and gleaned through secondary sources. Interviews allow for nuanced and open-ended responses to difficult and/or complex topics, thereby eliciting more information on attitudes, perceptions, and experiences that otherwise cannot be obtained by a structured survey.

For the qualitative data, we focus on thematic analysis to understand reasons underlying changes in outcomes of interest, how the WWW programme contributed to results, and what other factors may have influenced results. Overall, the goal of the qualitative analysis is to assess and validate the programme's impact, better understand whether impact varies across relevant subgroups, and understand the mechanisms of change.

Notes from qualitative interviews were collated into a master spreadsheet where all responses were coded by demographic information and research question. This process allowed us to analyse qualitative data systematically and to identify instances where qualitative data supports, or runs counter to, other findings. This approach also allowed us to understand whether findings represent general perceptions and experiences, or are specific to a limited number of respondents. Throughout our analysis, we note the extent to which a finding is likely to be representative (i.e., is expressed by a large percentage of respondents across demographic characteristics).

### ***Use of Secondary Sources***

Analysis of primary data was supplemented using secondary sources such as previous evaluation and monitoring reports, data from past evaluation rounds, programme documents (such as reports on activities undertaken and challenges faced), and literature (including academic and "grey" literature) on education in Kenya. These secondary sources were primarily used to answer research questions around value for money and, to a lesser extent, sustainability. Another key use (particularly of external academic and "grey" literature sources) was to further understand questions around COVID-19-specific challenges

faced by students and schools in Kenya, as well as issues surrounding education of marginalized groups such as girls with disabilities.

## Data Collection Tools

The design of this endline evaluation closely matches that of previous evaluation rounds in the tools used, with the exception of an additional classroom observation tool added to better understand use of positive and negative teaching practices. Full survey scripts and qualitative interview guides are annexed to this report. The tools included the following:

- Quantitative tools
  - Learning assessments: Secondary Grade Maths Assessment (SeGMA), Secondary Grade Reading Assessment (SeGRA), and science, technology, engineering, and mathematics (STEM) assessment
  - Girl survey
  - Caregiver survey
  - School (head teacher) survey
  - Classroom observation
- Qualitative tools
  - FGDs with girl students
  - KIIs with teachers
  - KIIs with government officials
  - KIIs with programme staff

Tools are described in detail in Annex 2.

### ***Learning Assessments***

Learning assessments included the SeGRA and SeGMA to evaluate literacy, numeracy, and English levels. A STEM assessment was additionally utilized to assess students' knowledge of chemistry, physics, and biology.

Consilient worked with the EDT team to modify learning assessments to adapt test forms from previous evaluations. Learning assessment equating—or the technical process conducted to establish comparable scores on a different version of a test—was integrated into tool design, data collection procedures, data treatment, and analysis. Modification of the assessment instrument was guided by internationally observed standards<sup>4</sup> for best practices on preparing equivalent test forms, which, for example, recommend making simple changes in names of story subjects, action, and adjectives. To prevent geographic biases from emerging, common words—those universal to Kenyan students regardless of area residence—were selected. Consilient ensured that reading subtasks considered word frequency, difficulty of sounds, cognates with other local languages, and potential biases in the presence (or absence of) certain letters and sounds in other local languages. The assessment also strived to ensure consistency in scoring and assessment administration instructions between rounds.

For the STEM assessment, to ensure comparability, the assessment utilised in the Special Assessment Study was adapted for this endline evaluation. The approach to adaptation mirrored the core principles and considerations described above. During the inception period, Consilient consulted with the EDT team to reflect on the relevance and applicability of the STEM assessment and used an approach to assessment development tailored to the specific objectives and elements of the curricula implemented in Kenyan schools. Similarly, results from the Kenyan Certificate of Secondary Education science exams were used to inform the development and adaptation of the STEM assessment.

---

<sup>4</sup> See: USAID EGRA Toolkit (2016). [https://pdf.usaid.gov/pdf\\_docs/PA00M4TN.pdf](https://pdf.usaid.gov/pdf_docs/PA00M4TN.pdf)

## ***Girl Survey***

Girl surveys were conducted in person for Cohort 2 and via CATI for Cohort 3. These tools generally remained consistent with previous evaluations to ensure comparability, and included modules on demographic and household information, perceptions of and attitudes towards STEM subjects, experiences in school and with teachers, disability (Washington group questions), life skills, teaching quality, COVID-19, sexual and reproductive health, and child protection. Adaptations and additions were only made to capture any important changes or experiences that may have occurred between evaluation rounds and that have the potential to interact with outcomes; for example, questions relating to NGO intervention (outside of WWW) in schools were added to better understand “contamination” of the sample and the extent to which students in comparison schools may have benefitted from other NGOs’ programming.

## ***Caregiver Survey***

During fieldwork, girls who responded to the girl survey were asked to provide contact information (phone numbers) for their caregivers. Consilient then used a CATI to survey these caregivers. The survey asked about demographic information of the household, the caregiver’s perception of the girls’ participation and performance in school, the girl’s household responsibilities, the caregiver’s support for the girl’s education, school management and PTA involvement, and community initiatives to support girls’ education. This included questions addressing Output 5 (community engagement), such as awareness and perceived benefits of community conversations, financial grants to community groups, strengthening of social accountability, and community champions.

## ***School Survey***

School survey tools remained consistent with past evaluation rounds. The survey included modules on the following:

**Table 1: School survey modules**

Module Name	Module Information	Methods and Sources
Teaching force	Number and gender of teachers	School records, observation
Support received by school	External support for school, school clubs	Interview
Attendance	Enrolment and head counts	School records, observation
Performance	School performance on KCPE/KCSE	School records

## ***Classroom Observation***

Consilient added a classroom observation tool not used in previous evaluation rounds. Improving teaching quality within schools was highlighted as a key pillar of the WWW programme. Given this, it was important to understand not only how teachers perceived training (through qualitative interviews), but also the extent to which teaching practices were being implemented in the classroom. To this end, we developed a classroom observation tool aimed at identifying the utilisation of training components such as pedagogical practices and curriculum elements. To balance tool effectiveness and simplicity of administration, the tool was comprised of a list of teaching practices tailored to the programme, such as student engagement in class, use of student-centred teaching techniques, and gendered behaviours of teachers. Researchers observed classes and measured the use or absence of a teaching practice through a binary variable of either “observed” or “not observed”.

Given the introduction of the tool at the endline, measuring changes in practices among trained teachers is not possible. However, even within a single evaluation round, teaching quality (as measured by the

classroom observation tool) can be tested as a potential predictor of learning outcomes and therefore contributes to the testing of the theory of change.

### **Qualitative Tools**

The qualitative component of the evaluation design serves multiple purposes. First, it complements quantitative data collected in learning assessments, school modules, and household surveys. In this sense, qualitative surveys provide context and depth to trends that emerge in quantitative data. Second, the qualitative component helps provide a more robust understanding of the processes of change and interactions between the learner, the home, the community, and the school. Utilizing both key informant interviews and focus group discussions, qualitative tools were designed based on respondent type and their associated expertise, point of view, and relationship to WWW programming.

FGDs were conducted with female students at target schools, and included groups of around six girls. Discussions were highly participatory. Our approach included a story-telling process in which girls were presented with short vignettes and asked to complete the story in their own words. Participants were presented with the premise of a story about a girl and asked to conclude the story, with probing about the constraints faced by the girl in the story and how things could have turned out differently. The goal of the vignette exercise was to elicit responses regarding learning outcomes and transition: what prevents girls from learning, how their learning can be accelerated, the likelihood of transition at key times in adolescence and for specific marginalized sub-groups, etc. Consilient's use of vignettes has been highly effective in eliciting information on sensitive or taboo subject areas, as it allows respondents to distance themselves from the discussion, reflecting on an imaginary girl and her experiences rather than being required to attribute feelings and experiences to themselves. This often results in more open and frank conversations about trends and common experiences that may be difficult to discuss, especially for children and/or young adults.

KIs conducted with teachers included questions on support provided to teachers and schools, the challenges facing girls' learning, perceptions towards STEM and girls' learning of STEM subjects, teaching practices, student attendance, challenges facing girl students and disabled students, and PTA engagement. KIs with project staff and government officials focused on issues related to the programme itself, including sustainability, value for money, and coordination with other partners, as well as broader issues related to girls' education.

### **Pre Data Collection**

The sampling framework was based on the list of schools and girls provided by EDT, including girls contacted at SAS and midline. As significant limitations were identified for data tracking girls from midline, the evaluation generally focused on schools targeted at SAS which also included midline girls. The full sampling framework is included in Annex 13.

For endline, a classroom observation tool (quantitative) was designed and used for the first time in order to provide improved information on teaching practices. All other tools were adapted from midline, including learning assessments; the girl, caregiver, and school surveys; and qualitative tools for girls and teachers. These tools are described in more detail in the main body of the report, and are included in full in Annex 9.

Enumerators were recruited from Consilient's list of trusted enumerators and from the pool of enumerators who participated in previous WWW project evaluations. Team leaders were recruited to manage teams, provide quality assurance, and conduct qualitative research. Enumerators participated in a five-day training, including a half-day pilot of survey tools. The training was conducted by Consilient's Research Manager with support from the Fieldwork Managers and Research Officers. Consilient developed training materials taking into account lessons learned from trainings for previous GEC projects

and the specific needs of the WWW endline evaluation. Over the course of five days our team covered the following:

- **Project introduction, ethics debriefing, and safety protocols:** The team covered project goals and objectives of data collection, and with the assistance of the EDT team, reviewed relevant protection policies and research ethics (confidentiality and consent).
- **Review of data collection instruments:** The training team provided enumerators and team leaders with training on the Ona/ODK platform and procedures for uploading data and updating surveys in the field. Enumerators and team leaders were all trained in the SeGRA/SeGMA and STEM tools. Enumerators were then trained in the girl survey tool, while team leaders were trained on the school survey, classroom observation, and qualitative tools. All participants were trained in how to approach sensitive questions and where we anticipated input errors. All participants also conducted mock interviews to improve their familiarity with tools.
- **Fieldwork planning and logistics:** Consilient’s Research Manager and Fieldwork Manager developed and reviewed fieldwork plans with teams, including travel logistics, data collection timing, respondent contact and recontact procedures, security concerns and mitigation plans, and expected communication plans and field updates required.

At the end of training, all enumerators and team leaders participated in a pilot data collection exercise in four schools in Nairobi.<sup>5</sup> Due to limitations on school and student availability, the pilot was scheduled for Friday afternoon. Because of this logistical limitation, we were unable to reconvene training with all enumerators and team leaders to discuss pilot results; specific issues identified by data collectors were rather shared by team leaders with the Consilient team, while issues identified in survey administration by the Consilient team were shared with team leaders for further discussion with enumerators.

Furthermore, due to the limitations on student availability and pilot timing, the Consilient team elected to pilot the girl survey and school survey, rather than learning assessments. This was done in order to test survey tools with higher potential for scripting error and due to the limited value of practicing administering learning assessments without being able to debrief afterward.

A total of 32 girl surveys (one per hired enumerator) and 3 school surveys<sup>6</sup> were completed during the pilot. No administration issues were identified by the Consilient team. Several survey scripting errors were identified and corrected within the girl survey and school survey, as described in the Inception Report. Overall, the need for revisions was minimal.

## During Data Collection

Fieldwork for the endline evaluation was conducted between October 5 and October 31, 2022. The fieldwork was conducted by Consilient’s research teams in order to conduct learning assessments with girls in project’s intervention and comparison schools, girl surveys with project beneficiaries, school (head teacher) surveys, and classroom observations. Qualitative interviews were also conducted with girls and teachers from project schools. Consilient employed 33 enumerators and 11 team leaders to undertake the fieldwork. Fieldwork was conducted by teams of four.

As described above, all enumerators and team leaders were trained in child protection and ethical standards prior to the fieldwork. Adherence to these standards was closely monitored by team leaders and by support staff from EDT who observed fieldwork to provide quality assurance. Only one issue was observed with adherence to these standards during fieldwork, with an enumerator who was sitting closely next to girls for surveys, rather than at an appropriate distance. This enumerator was reprimanded and the team leader was instructed to closely observe the enumerator to ensure that all policies were adhered to, after which the issue was resolved.

---

<sup>5</sup> These schools were selected as they are not included in the endline evaluation.

<sup>6</sup> Due to limited school and head teacher availability, team leaders were grouped together to complete pilot school surveys.

Rigorous recontact protocols were followed to track cohort girls from midline. Given the limited timeframe for fieldwork, teams spent only one day at most schools, limiting enumerators' abilities to trace girls who were not available within schools. However, enumerators still made substantial efforts to follow recontact procedures as follows:

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

Time constraints mainly limited enumerators' abilities to visit girls' households multiple times, particularly in schools where a large number of girls could not be found.

Because of limitations to the quality of tracking data from midline, this approach was of limited success in recontacting midline girls, but was far more successful for tracking girls from SAS. The achieved sample is described in detail below and in the main body of the report. Girls were not replaced, as there is limited value to replacement at endline.

During fieldwork, caregiver phone numbers were also collected from girls. Caregivers, as well as Cohort 2 and 3 girls, were then contacted following fieldwork using a CATI procedure which took place from October 19 to December 5. Call centre operators were trained for two days; the training covered the caregiver and girl surveys, CATI procedures, and how to successfully establish rapport and conduct a survey over the phone.

In order to ensure the largest possible sample size, call centre operators were instructed to call each number provided three times, or until the respondent answered. However, significant issues were identified with numbers collected from past evaluation rounds; as a result, call centre operators had difficulties reaching Cohort 2 and 3 girls, with many phone numbers no longer active or no longer answered by the desired respondent. To boost Cohort 2 and 3 sample size, during independent fieldwork conducted in Marsabit and Kwale by EDT, additional phone numbers for girls were collected.

### ***Quality Assurance***

During fieldwork, four main measures were taken to assure data quality. First, survey tools were designed to prevent as many errors as possible, with preventative measures scripted into survey design to minimize errors. The following non-exhaustive list of checks was scripted into quantitative tools:

- Geographic information entered at the start of each tool was choice-filtered to prevent logically inconsistent responses.
- Consistency checks were added to identifying information for the cohort respondent – such as ensuring a respondent's unique ID matches their name – making it almost impossible for an enumerator to mistakenly enter the wrong unique ID.
- Survey timers in the quantitative tools checked for administration that was unusually fast or slow, including module-specific durations.
- Unusual or contradictory responses in surveys generated flags warning the enumerator that they may have entered a response incorrectly and prompting them to double-check responses entered were correct (i.e., a respondent who reports being unemployed but also reports a high – or non-zero – personal income).

Second, Consilient's in-house team conducted daily checks of data quality on surveys submitted by enumerators the previous evening. Fieldwork progress was also monitored daily to ensure that teams

were on track with the fieldwork plan. Automated data quality checks included the following, among others:

- Checking for significantly short interviews (in terms of duration)
- Checking for enumerators who claimed they could not find girls to interview at higher than average rates, and verifying this with team leaders
- Checking GPS coordinates of interviews
- Checking responses to “other” questions to ensure this option was appropriately selected
- Checking responses to open-ended questions
- Checking integers and numeric answers to ensure no data entry errors were made
- Checking average scores by enumerator to the SeGRA/SeGMA to ensure assessments were marked fairly and accurately and there were no administration errors

Using these processes, Consilient identified one enumerator who gave unusually high marks for the SeGRA, and three who gave unusually low marks for the SeGMA (and in particular, for multiple point questions where enumerators had to mark girls based on their work, not just on their answers). Team leaders were asked to review these enumerators’ assessments to ensure they were correctly marked. For the SeGMA, for one enumerator, it was confirmed that marking was incorrect, and the assessments were flagged to be re-marked following fieldwork. Furthermore, one enumerator was identified with unusually short (duration) surveys; the team leader was asked to observe this enumerator to ensure he was correctly administering surveys, after which the issue was resolved.

Third, Consilient’s team leaders were instructed to occasionally observe enumerators, with a focus on teams and enumerators for whom potential data quality issues were identified. Team leaders were also occasionally asked to verify that surveys had been administered by recontacting respondents.

Lastly, EDT sent several teams to the field to observe the data collection process and flag any issues to Consilient for correction. The following issues were identified:

- One enumerator had girls self-administer the girl survey and sat too close to girls during these surveys, presenting a safeguarding issue
- Some teams were allowing girls to use calculators on the SeGMA
- Enumerators were not prompting girls with the time remaining for learning assessments, leading girls to improperly manage their time
- Some team leaders were observing teachers who had not been trained by the WWW project
- Marking errors were identified, particularly for the SeGMA, including enumerators marking questions incorrect when there were only minor differences in the method used to arrive at a correct answer (e.g., a student writing “ $5 \times 3 = 15$ ” rather than “ $3 \times 5 = 15$ ”)

These issues were corrected as follows:

- The enumerator who was having girls self-administer surveys and sitting too close was informed to immediately cease this behaviour or face dismissal. The team leader was instructed to closely observe this enumerator to ensure that the behaviour did not continue, which resolved the issue.
- All teams were reminded to not allow use of calculators during the SeGMA.
- All teams were reminded to prompt students about time remaining during learning assessments.
- Team leaders were instructed to only observe teachers who had been trained by the WWW project.
- Trained teachers with experience marking exams were hired to review assessment marking from all enumerators. These teachers were instructed to review three each of the SeGRA and SeGMA for all enumerators and verify that marking was completed correctly. For enumerators flagged as marking incorrectly, assessments will be re-marked.

We note that this last process of assessment review and re-marking is still ongoing. EDT will be provided with information on the results of this process in a subsequent update, and information will also be included in the final evaluation report.

## Achieved Sample

For the SeGRA and SeGMA, a total of 1,455 girls were successfully recontacted and consented to take the assessments. This represents a recontact rate of 54 percent.<sup>7</sup> By cohort, recontact rates were as follows:

- 525 girls were assessed within Cohort 1 (39 percent recontact rate)
  - 366 girls were assessed in intervention schools (47 percent recontact rate)
  - 159 girls were assessed within comparison schools (28 percent recontact rate)
- 930 girls were assessed within Cohort 2 (69 percent recontact rate)

It is worth noting the significantly lower recontact rates for Cohort 1 compared to Cohort 2. Cohort 1 included girls assessed two years prior at midline; during fieldwork, it was found that information on the locations and contacts for these girls was frequently out-of-date or incomplete, resulting in lower recontact rates. In contrast, for Cohort 2—girls contacted during the SAS assessment—recontact rates were around 69 percent, on par with previous GEC evaluations conducted by Consilient. These dynamics are discussed further below under Challenges and Limitations.

The below tables show the achieved sample for learning assessments by county, cohort, and class/form. We note that we do not calculate recontact rates by class/form as much of our data on expected class/form was found to be incorrect.

**Table 2: Achieved sample by county and class/form, SeGRA and SeGMA**

County	Cohort 1 Sample		Cohort 2 Sample	Total	Target	Recontact Rate
	Comparison	Intervention				
Kilifi	26	54	201	281	407	69.0%
Mombasa	20	24	46	90	205	43.9%
Nairobi	19	122	280	421	950	44.3%
Samburu	9	18	129	156	225	69.3%
Tana River	62	110	145	317	616	51.5%
Turkana	23	38	129	190	294	64.6%
<b>Total</b>	<b>159</b>	<b>366</b>	<b>930</b>	<b>1,455</b>	<b>2,697</b>	<b>53.9%</b>
Class/Form						
Class 7	0	1	10	11	-	-
Class 8	0	30	433	463	-	-
Form 1	0	19	78	97	-	-
Form 2	11	18	130	159	-	-
Form 3	11	34	143	188	-	-
Form 4	137	264	136	537	-	-
<b>Total</b>	<b>159</b>	<b>366</b>	<b>930</b>	<b>1,455</b>	<b>2,697</b>	<b>53.9%</b>

<sup>7</sup> An additional 32 girls were recontacted but did not consent to take the learning assessments.



The below table additionally shows the sample by age group. We note differences in the age distribution of intervention and comparison girls; these differences are discussed further below in a section comparing comparison and intervention groups.

**Table 3: Achieved sample by age, SeGRA and SeGMA, recontacted (Cohort 1) girls**

Age	Comparison (%)	Intervention (%)
n	159	366
12-13	0.0%	1.6%
14-15	1.3%	9.0%
16-17	12.7%	24.1%
18-19	62.7%	54.5%
20 or older	23.4%	10.7%

For the STEM assessment, 383 assessments were administered. The target sample for STEM assessments agreed upon in the inception phase was 551; as such, the achieved sample represents contact rate of 69.5 percent. However, in order to increase the likelihood of achieving the target sample, a total of 620 girls were targeted during fieldwork to take STEM assessments, giving us a contact rate of 61.8 percent. The below tables disaggregate by county and class/form for the total sample of 620 targeted during fieldwork; the tables do not disaggregate by cohort, as all STEM assessments were administered as intended with Cohort 2 girls.

**Table 4: Achieved sample by county and class/form, STEM assessment**

County	Achieved Sample	Target	Recontact Rate
Kilifi	105	166	63.3%
Mombasa	39	84	46.4%
Nairobi	65	104	62.5%
Samburu	55	63	87.3%
Tana River	62	107	57.9%
Turkana	57	96	59.4%
<b>Total</b>	<b>383</b>	<b>620</b>	<b>61.8%</b>
Class/Form			
Class 7	0	-	-
Class 8	0	-	-
Form 1	3	-	-
Form 2	126	-	-
Form 3	132	-	-
Form 4	113	-	-
Out of school	0	-	-
<b>Total</b>	<b>383</b>	<b>620</b>	<b>61.8%</b>

For the girl survey, 933 girls were successfully recontacted and consented to the survey, a recontact rate of 69 percent.<sup>8</sup> The below tables show recontact rates for the girl survey by county and class/form. As above, we do not provide recontact rates by class/form as data on expected class/form was frequently incorrect.

**Table 5: Achieved sample by county and class/form, girl survey**

County	Achieved Sample	Target	Recontact Rate
Kilifi	189	243	77.8%
Mombasa	33	79	41.8%
Nairobi	310	470	66.0%
Samburu	124	138	90.0%
Tana River	141	219	64.4%
Turkana	136	197	69.0%
<b>Total</b>	<b>933</b>	<b>1,346</b>	<b>69.3%</b>
Class/Form			
Class 7	8	-	-
Class 8	431	-	-
Form 1	83	-	-
Form 2	133	-	-
Form 3	148	-	-
Form 4	125	-	-
Out of school	5	-	-
<b>Total</b>	<b>933</b>	<b>1,346</b>	<b>69.3%</b>

For girls that were not found, enumerators were instructed to submit data on where the girl is now. While this data is incomplete, it provides some indication of why girls were unable to be recontacted. Enumerators were able to find families of 55 girls within communities surrounding target schools; among this group, 76 percent of girls now attended boarding school or had transferred to a new school. Enumerators were able to find information about an additional 52 girls; among this group, 46 percent were confirmed to be attending a new school in a different location or to have graduated from Form 4, while 33 percent were confirmed to have moved to a new location, but their schooling status could not be confirmed.

Additionally, 217 girls were contacted via CATI procedure. The CATI targeted girls in Marsabit and Kwale, ALP girls, and tracer girls, as described above. However, because phone numbers and data for girls was recorded in prior rounds of data collection, many girls had moved location or were no longer in ALPs at endline. Furthermore, no tracer girls were successfully contacted; this sample is thus omitted. In total, 17 girls were contacted who were participating in ALPs, and 109 girls were contacted in Marsabit/Kwale. The below table shows the distribution of girls contacted via CATI within counties.

<sup>8</sup> An additional 92 girls were recontacted but did not consent to take the girl survey, for a total recontact rate of 76 percent.

**Table 6: Achieved sample by county, girls contacted via CATI**

County	Achieved Sample	Target	Recontact Rate
Kilifi	49	-	-
Mombasa	1	-	-
Nairobi	26	-	-
Samburu	6	-	-
Tana River	23	-	-
Turkana	3	-	-
Marsabit	32	150	21.3%
Kwale	77	150	51.3%
<b>Total</b>	<b>217</b>	-	-

For all contacted girls, the below tables show breakdowns of disability status and disability severity. We note that Table 7 categorises girls as disabled if they have at least “some difficulty” with a task.

**Table 7: Disability status, endline**

		Intervention % (n) <sup>9</sup>	Variable name
<b>Girls with disability (% overall)</b>		33.9 (390)	-
<b>WG Child functioning questions</b>	<b>Domain of functioning</b>		
Difficulty seeing	Seeing	14.9 (171)	CS_D1S
Difficulty hearing	Hearing	9.4 (108)	CS_D2S
Difficulty walking or climbing steps	Walking	4.1 (52)	CS_D3S
Difficulty with self-care	Cognitive	1.0 (12)	CS_D5S
Difficulty with communication		2.7 (31)	CS_D6S
Difficulty remembering or concentrating		12.4 (142)	CS_D4S
Serious illness in last year <sup>10</sup>	Health	21.1 (243)	CS_D7S
Multiple impairments		8.5 (98)	-

<sup>9</sup> Girls who reported some difficulty, a lot of difficulty, or that they cannot do the task at all are included in the percentages reported.

<sup>10</sup> Not included in calculations of girls with any disability or multiple disabilities.

**Table 8: Disability severity, endline, intervention girls**

WG Child functioning questions	Domain of functioning	Some difficulty (%) (n)	A lot of difficulty (%) (n)	Cannot do at all (%) (n)
Difficulty seeing	Seeing	12.9 (148)	1.7 (20)	0.3 (3)
Difficulty hearing	Hearing	5.1 (59)	2.3 (26)	2.0 (23)
Difficulty walking or climbing steps	Walking	3.7 (43)	0.4 (4)	0.0 (0)
Difficulty with self-care	Cognitive	0.9 (10)	0.2 (2)	0.0 (0)
Difficulty with communication		2.4 (28)	0.3 (3)	0.0 (0)
Difficulty remembering or concentrating		11.8 (136)	0.5 (6)	0.0 (0)

Caregiver phone numbers were provided by girls during the girl survey. Using these numbers, 818 caregivers were successfully contacted via CATI procedure, or 60.8 percent of the target sample. The below table shows recontact rates for caregivers by county. We note that some caregivers of Marsabit/Kwale girls (from cohort 3) were contacted during the CATI using phone numbers obtained during the CATI with cohort 3 girls.

**Table 9: Achieved sample by county, caregiver survey**

County	Achieved Sample	Target	Contact Rate
Kilifi	163	243	67.1%
Kwale	32	0	-
Marsabit	9	0	-
Mombasa	28	79	35.4%
Nairobi	361	470	76.8%
Samburu	59	138	42.8%
Tana River	95	219	43.4%
Turkana	71	197	36.0%
<b>Total</b>	<b>818</b>	<b>1,346</b>	<b>60.8%</b>

This table shows, notably, that contact rates for caregivers in Nairobi and Kilifi were substantially higher than those for other counties. This may represent a limitation to the CATI procedure, whereby we are only able to survey caregivers who have access to a phone and a cellular network. This limitation is discussed more below.

All 158 target schools were visited by teams, but one school in Nairobi did not consent to a school survey and one school in Nairobi was closed. As a result, 156 school surveys were conducted. We note that during fieldwork, the EDT team provided information that several targeted schools were comparison, rather than intervention, schools. As such, the achieved sample was 129 intervention schools and 27 comparison schools. This corresponds to 108 primary schools and 48 secondary schools. The table below shows the achieved sample by county and intervention status.

**Table 10: Achieved sample by county, school survey**

County	Comparison	Intervention	Total	Target	Recontact Rate
Kilifi	3	25	28	28	100.0%
Mombasa	2	5	7	7	100.0%
Nairobi	13	48	61	63	96.8%
Samburu	2	13	15	15	100.0%
Tana River	5	23	28	28	100.0%
Turkana	2	15	17	17	100.0%
<b>Total</b>	<b>27</b>	<b>129</b>	<b>156</b>	<b>158</b>	<b>98.7%</b>

Classroom observations were conducted in 127 intervention schools;<sup>11</sup> two schools in Nairobi did not consent to the classroom observation. This corresponds to 106 primary schools and 48 secondary schools. The following table shows the achieved sample by county.

**Table 11: Achieved sample by county, classroom observation**

County	Total	Target	Recontact Rate
Kilifi	25	24	104.2%
Mombasa	5	5	100.0%
Nairobi	47	48	97.9%
Samburu	13	13	100.0%
Tana River	23	22	104.5%
Turkana	16	16	100.0%
<b>Total</b>	<b>129</b>	<b>128</b>	<b>100.8%</b>

For the qualitative data, in total, 19 focus group discussions (FGDs) with girl students and 21 key informant interviews (KIIs) with teachers were conducted, mostly aligned with the target sample. The following table shows the sample by county.

**Table 12: Achieved sample by county, FGDs and KIIs**

County	FGDs	KIIs	Total
Kilifi	3	4	7
Mombasa	2	2	3
Nairobi	7	7	14
Samburu	2	2	4
Tana River	3	3	6
Turkana	2	3	5
<b>Total</b>	<b>19</b>	<b>21</b>	<b>40</b>

<sup>11</sup> Two classroom observations were also conducted in comparison schools.

Sixteen additional KIIs were conducted with government officials and programme staff. In total, 11 KIIs were conducted with government officials (one in Kilifi and two in each of the remaining five counties), including County Directors and Sub-directors of Education, Curriculum Support Officers, and Quality Assurance and Standards Officers, and five KIIs were conducted with programme staff.

## Post Data Collection

All data was cleaned and checked for consistency following fieldwork. Several issues with duplicate respondents and incorrectly entered schools/locations were identified; these issues were corrected. All data is stored on Consilient's secure servers in Stata format. Cleaned data, a codebook, and cleaning Stata (.do) files are also provided in Annex 11.

Qualitative interviews were recorded by researchers, from which summary notes were written. Full transcriptions and translations of all interviews were not completed, as this was deemed to be of limited value relative to notes. However, two full transcriptions/translations are included in Annex 10.

## Approach to Gender Equality and Social Inclusion

Given the WWW programme's focus on girls' education and marginalised groups, the endline evaluation necessitated a GESI-sensitive research strategy. To achieve this, Consilient first reviewed all data collection tools with GESI considerations in mind. While tools were generally kept highly similar to those used in past evaluation rounds, examples of GESI-sensitive tool design occurred for the classroom observation tool, which included explicit questions around gender sensitive or biased teaching practices, as well as teachers' approaches towards engaging students with (visible) disabilities.

Qualitative tools were also designed with GESI considerations in mind. The tool for FGDs with girls, for example, included a number of vignettes which sought to understand attitudes towards marginalised and disabled girls. These vignettes presented scenarios—for example, of a disabled girl who started school late—to which girls were asked to respond. By presenting hypothetical scenarios, rather than asking girls for their personal experiences with marginalised groups, the vignettes allow for richer, GESI-sensitive data to be collected, as girls are more free to express their genuine opinions or perceptions without fearing stigma.

Second, Consilient strived to achieve a gender-balanced enumeration team, with at least one woman in every data collection team. This was achieved for all but one team, where language barriers meant that four male enumerators were selected.

Third, training of enumerators and team leaders also focused on gender and social inclusion considerations, such as methods to ensure that girl children felt comfortable speaking with male enumerators and ways to approach sensitive questions about SRH, among other topics.

Fourth, a Kenyan Sign Language-speaking enumerator was assigned to visit special schools for deaf students. In special schools for blind students, assessments were translated into and printed in braille in order to ensure all students could participate. We note, however, that students at special schools were erroneously not given equivalent time to complete assessments as at midline, but were rather given the same amount of time to complete assessments as students without disabilities. This limitation is discussed under *Learning Outcomes* in the main body of the report.

Lastly, data analysis used a GESI approach. As the evaluation measures outcomes of interest to girls, analysis of girls' learning, transition, and life outcomes has an inherent girl-centred approach. While data was generally not collected for boys, in some instances, we also strive to better understand differences between outcomes for boys and girls; key examples of this include perceptions of whether girls or boys are better at different subjects, and gendered behaviours in classrooms. Furthermore, for all individual-level outcomes, disability is included as a predictor of interest, as are other dimensions of marginalization such as whether girls live in a female-headed household or speak a minority language.

## Challenges and Limitations

A number of key limitations apply to our methodology and achieved sample.

**Difference-in-differences assumptions and limitations:** Inferences drawn via difference-in-differences rely on two key assumptions. The first assumption is that under the counterfactual condition—i.e., in the absence of intervention—the change in outcomes over time in treatment and comparison schools will be similar. This assumption is often referred to as the “parallel trends assumption” because it assumes that, in the absence of treatment, trends in outcomes in the treatment group would evolve in a fashion parallel to those in the comparison group. This first assumption is, in general, untestable under a difference-in-differences framework. In the absence of randomisation, the parallel trends assumption may or may not hold, though the selection process leaves no reason to expect divergent trends under the counterfactual.

The second assumption is that comparison group respondents are not exposed to the treatment or to similar interventions affecting learning and transition. Informally, this assumption is often stated as the “no spillover” assumption. The second assumption is testable *ex post*, by directly measuring whether comparison group respondents received any of the project’s interventions.

If both assumptions are satisfied, inferences drawn using difference-in-differences are expected to be unbiased. However, we note that, in the context of this endline evaluation, there is expected to be some contamination among comparison schools from other programme (i.e., not WWW) interventions.

Our school survey tool included explicit questions probing whether comparison (and intervention) schools have received support from other NGOs in order for us to better understand the extent of spillover effects. We find that among all endline schools, 94 out of 156 (60.3 percent) reported receiving investments from NGOs. This includes 5 out of 27 comparison schools, or 18.5 percent, who reported benefitting from interventions targeting teacher training, provision of learning materials and school supplies, and strengthening reading and mathematics—similar to those interventions provided by the WWW programme. This poses a limitation to the validity of the difference in differences approach, although only a minority of comparison schools were affected.

We also note that 38 intervention schools (29.5 percent of all endline intervention schools) reported receiving assistance from NGOs other than EDT (or Concern Worldwide, an implementing partner). This included interventions from Asante Africa, Kesho Kenya, UNICEF, the Safaricom Foundation, and the School Based Teacher Support System, among others. Given that interventions related to learning, teaching, and school quality were not solely implemented by the WWW programme in target schools, this may cause us to overestimate the impact of the WWW programme, some of which may be attributable to other programmes.

**Tracking across evaluation rounds:** No tracking data on girls sampled at baseline was available at endline. This means that our analysis is not able to measure programme impact on girls from the full lifecycle of the programme. Similarly, because our past data extends at most to the midline evaluation (two years prior to the endline) and often only to the SAS evaluation (one year before endline), our analysis will only show change over a limited period of time. For some expected programme impacts, such as learning outcomes, we may still expect a measurable change. However, for some expected impacts that tend to be gradual—such as impacts on confidence and leadership—we may measure only a limited programme impact. This would not necessarily be due to the efficacy (or lack thereof) of the programme, but rather due to the limited timeframe for analysis.

**Sample size of cohort 1:** The sample size of cohort 1 is small due to limited tracking data on girls sampled in the midline evaluation. Because attrition rates of girls from our sample were high—which may have occurred because tracking information on girls was incorrect, if girls had dropped out of school (and especially if they moved to new locations not targeted in this evaluation), or if girls had transferred to

new schools not included in the sample—the statistical power of the evaluation is low, reducing our ability to draw statistically significant conclusions using a difference-in-differences methodology for analysis.

Our approach included substantial efforts to maximize sample size within constraints on timeline, budget, and tracking information from past rounds. However, our overall sample size remained low. Low recontact rates, particularly for Cohort 1 (midline) girls, were primarily due to incorrect information on the current location of the girl. Many girls were no longer present in the school recorded at midline; anecdotally, team leaders and enumerators often reported from asking the head teacher and community members that these girls had since graduated or moved to new schools.

The low recontact rates for Cohort 1 present a limitation for the difference-in-differences analysis of learning outcomes, as the sample size for this cohort is under-powered. This means that the chance of a “Type II” error in statistical analysis—accepting the null hypothesis of “no treatment effect” when there is, in fact, a treatment effect—is higher. However, it is important to note that the small sample size does not increase the risk of a “Type I” error—finding a treatment effect when there is no actual treatment effect. In other words, if our analysis shows no effect of the programme on learning outcomes, we cannot confidently say that this is accurate due to low sample size. However, if our analysis shows a significant effect of the programme on learning outcomes, the small sample size does not detract from this finding.

**Downstream effects of girl survey sample size:** Recontact rates for the girl survey had downstream effects on the sample size of the caregiver survey. Among the 69 percent of girls who were recontacted for the girl survey, only around 70 percent provided a phone number for their caregiver. This means that the total sample of caregiver phone numbers collected during fieldwork for the CATI is only around 650 numbers. Given that response rates to the CATI were inevitably not 100 percent, these dynamics mean that the caregiver sample was substantially smaller than the target sample.

**Recontact biases:** There may be some biases in our sample from girls that could not be recontacted. For example, girls who dropped out of school were more difficult to recontact, and may also have relatively lower learning, transition, or life outcomes leading them to drop out. As such, since these girls are excluded from our sample (because they could not be recontacted), we may overestimate our measures of programme impact. In contrast, girls who have since graduated or moved to new schools (such as boarding schools) were also difficult to recontact, but may have relatively higher learning, transition, or life outcomes which led them to successfully graduate or continue in school; this would lead us to underestimate programme impact. Since we do not have information on girls who were not recontacted, we cannot definitively say the direction or extent of this bias.

The difference-in-differences approach to analysis of cohort 1 is designed to mitigate some of this impact, as we would expect similar types of girls to drop out of both comparison and intervention schools. However, this bias is cumulative from previous rounds, and is thus an important limitation.

Annex 2 includes analysis on attrition characteristics and predictors of recontact. While we generally find that differential attrition by demographic characteristics is relatively small, there do appear to be some differences in girls who were recontacted compared to those who were not, including relatively higher recontact rates in Samburu, lower rates in Mombasa and Tana River, and higher recontact rates for disabled girls. We also find some indicative differences in household responsibilities between recontacted and attrition girls, although small sample size limits our ability to draw conclusions about these differences. Lastly, we find that recontacted students had relatively higher SeGRA scores and (for intervention students) SeGMA scores. All of these findings suggest potential bias in our difference in differences analysis.

The CATI procedure may also introduce some biases, as we were unable to contact caregivers or girls who do not have phone access (who are also more likely to belong to marginalized groups and have high levels of poverty). Indeed, the above analysis of the achieved sample finds much higher contact rates in



Nairobi and Kilifi than in other counties. This may limit the applicability of our conclusions to other counties with lower contact rates.

**Marking of learning assessments:** Due to time and logistical constraints, enumerators marked the SeGRA and SeGMA assessments themselves during fieldwork. This could potentially affect our measures of programme impact if enumerators did not accurately mark assessments, particularly if assessments were consistently marked lower (or higher) than appropriate.

To address this potential issue, first, enumerators and team leaders were extensively trained on marking procedures during training. For the SeGMA, where answers are objective (i.e., either correct or incorrect), this involved a review of all questions and the marking scheme. For the SeGRA, some answers were subjective (i.e., have multiple possible correct answers, or are marked on a scale, rather than simply marked as correct/incorrect). As such, for the SeGRA, questions and the marking scheme were reviewed, but enumerators and team leaders also underwent extensive practice marking example answers to subjective questions. All examples were discussed at length to ensure that enumerators understood not only the correct mark, but also the rationale for the mark, in order to enable them to mark SeGRA assessments accurately in the field.

Second, quality control conducted during fieldwork identified enumerators whose marks were consistently below or above the average for all students (see Annex 2 for more detail). For these enumerators, we further investigated potential reasons for the deviations, such as the age and form of the students whose assessments were marked. If the deviations were not explainable by such factors, we requested that team leaders review the assessments in order to ensure that marking was accurate.

Third, after fieldwork concluded, all enumerators returned copies of their assessments to our offices. Experienced maths and English teachers were hired to review the assessments at random, ensuring that at least three SeGRA and three SeGMA assessments were reviewed for each enumerator. If a substantial deviation from the appropriate marking was found, the enumerator's assessments would be re-marked. Using these procedures, one enumerator was identified whose assessments were then re-marked.

Given the technicality of the STEM assessment, copies of this assessment were kept by team leaders, and then marked later by individuals with education in science. As such, the above limitations do not apply to the STEM assessment.

**Format of learning assessments:** The format of learning assessments was adjusted slightly from midline/SAS. At midline/SAS, girls were given learning assessments which included both questions and spaces for answers on the same sheet. In contrast, at endline, girls were provided with separate question and answer sheets. This may have slightly increased the difficulty of learning assessments as girls may have made mistakes matching question and answer numbers, copying numbers for maths problems, or coming back to questions previously skipped (although this issue was not noted by enumerators during fieldwork). However, we note that the same learning assessments were administered to both comparison and intervention girls, meaning that any change in difficulty is equivalent across groups and will not affect analysis. Furthermore, several additional steps were taken to reduce the increase in difficulty, including by providing every girl with a question sheet (so that she did not have to copy from one "master sheet" used by multiple girls) and instructing enumerators to mark answers correctly if they had been consistently copied to the wrong question numbers of the answer sheet.<sup>12</sup> Given these steps and the fact that girls are in upper primary or secondary schools, we expect a minimal, at most, impact of these changes in format.

---

<sup>12</sup> For example, if a girl accidentally answered questions 3 through 12 in the answer boxes for questions 4 through 13, these were marked as if they were written in the correct answer boxes.

**Fieldwork challenges:** Outside of these limitations to the analysis, the main challenges faced during fieldwork were logistical. In Turkana, Samburu, and Tana River, long distances between many schools made transportation challenging and costly. This issue was resolved in part through the provision of transportation by EDT teams.

**Soliciting information through CATI:** In general, it is more difficult to establish rapport over the phone than in person, which may reduce the quality and depth of the information solicited through CATI approaches. Our CATI training addressed these issues to mitigate their impact on data. We note that enumerators did not report substantial difficulties establishing rapport during the CATI.

## Attrition Analysis

The below table shows the endline learning sample recontact and attrition rates.

**Table 13: Endline learning sample and attrition**

Cohort group	Endline sample (treatment)	Recontacted (treatment)	Attrition (treatment)	Endline sample (comparison)	Recontacted (comparison)	Attrition (comparison)
Cohort 1	159	27.8%	72.2%	366	47.0%	53.0%
Cohort 2	930	69.0%	31.0%	-	-	-

We now analyse whether attrition—i.e., girls who fell out of the endline sample due to inability to recontact—occurred in systematic ways. Attrition from the panel sample poses three challenges to drawing inferences regarding the impact of the WWW programme. First, attrition reduces the sample size for analysis, reducing the power of statistical tests—their ability to distinguish real programme impact from null or zero effect. As discussed above, this is a real challenge for our analysis, but affects only our likelihood of claiming zero effect when there was an effect, not vice versa.

Second, attrition is unlikely to be random. As shown above, there are significant differences across counties in attrition and individual-level factors also likely cause differential attrition. These biases, whereby girls (or caregivers) with certain characteristics are more likely to face attrition, can bias our estimates of programme impact. This is a particularly acute challenge when attrition is correlated with outcomes of interest. For example, if girls who are falling behind in school (thus performing worse on learning assessments) are more likely to drop out and thus less likely to be located and interviewed, this may lead us to overstate the impact of the programme on learning outcomes. However, using a difference-in-differences methodology, this is only problematic if there are differences in this pattern between intervention and comparison groups. For example, if intervention schools are better at facilitating recontact, even for girls who have dropped out, this may drive down estimates of impact on learning outcomes.

Third, even if attrition occurs at similar rates and in similar patterns across intervention and comparison schools, it alters the underlying sample for analysis. For example, as older girls drop out of the sample due to graduation from school, the sample becomes younger. If marginalised girls are more likely to drop out and thus face attrition, the sample may become less representative of these communities.

While our difference-in-differences analysis focuses on girls surveyed at ML, we did not target ML girls for the girl survey, and thus have very little overlap in our ML and EL girls' samples: Only around 125 girls were administered a girls' survey at both ML and EL. As such, we briefly analyse attrition from the SAS sample. To do so, we compare the characteristics of girls and households within the full SAS sample to the characteristics of those from the SAS sample who were also recontacted at endline. This comparison allows us to see the impact that attrition has had on sample composition outside of any programme

impacts. Although use of the SAS sample presents a limitation for attrition analysis, it still provides an indication of potential risks of utilising our panel sample. Furthermore, we supplement this analysis by comparing the characteristics of the ML and EL samples to further identify any potential areas of concern.

### **Attrition and Recontact: Girl Survey Data**

The below table shows differences in characteristics between the full SAS sample and the SAS sample with attrition (i.e., the sample of girls recontacted at EL).<sup>13</sup>

**Table 14: Attrition characteristics, girl survey**

Characteristic	Full SAS Sample	SAS Sample with Attrition
Age (mean)	14.9	14.9
Minority language speaker <sup>14</sup>	51.0%	53.4%
Reported disability <sup>15</sup>	4.4%	5.2%
Female head of household	22.0%	22.2%
Orphan	11.2%	10.8%
Feels safe travelling to school	91.5%	92.3%
Life skills (mean score) <sup>16</sup>	13.9	14.3

We find few differences between the characteristics of the full SAS sample and the sample with attrition. The mean age of girls in the two groups is nearly identical, and most other characteristics vary by less than one percentage point. Overall, while we again note the limitations of this analysis, it suggests that differential attrition by demographic characteristics—at least those measured here—was relatively small.

To better understand the actual impact of attrition, we analyse differences in recontact rates as a function of geography, girls' demographic characteristics, household, and school characteristics. This analysis employs linear regression models to predict a binary model of successful or unsuccessful recontact at EL.

The below figure shows predictors of recontact by county and age. Girls in Samburu were significantly more likely to be recontacted than girls in other counties, girls in Mombasa and Tana River were significantly less likely to be recontacted than girls in other counties, and there were no significant differences in recontact rates by age.

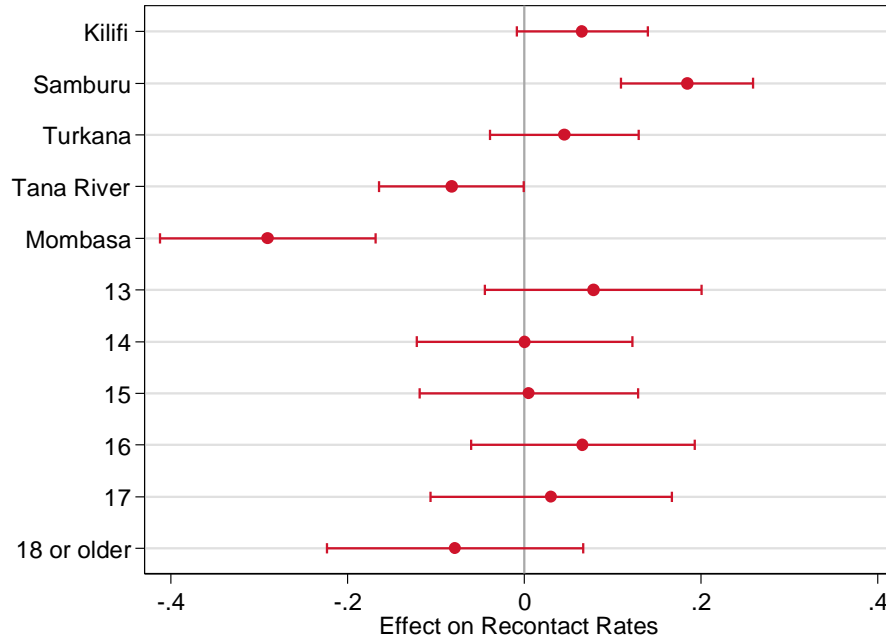
<sup>13</sup> T-tests of significance are defined only for disjoint sets and paired values. As such, we do not calculate significance of differences as these samples are neither disjoint nor paired values

<sup>14</sup> Defined as a girl who does not speak Swahili, or does not speak Turkana in Turkana or Samburu in Samburu.

<sup>15</sup> From Washington group questions; includes difficulty seeing, hearing, walking, with memory, with self-care, and communicating.

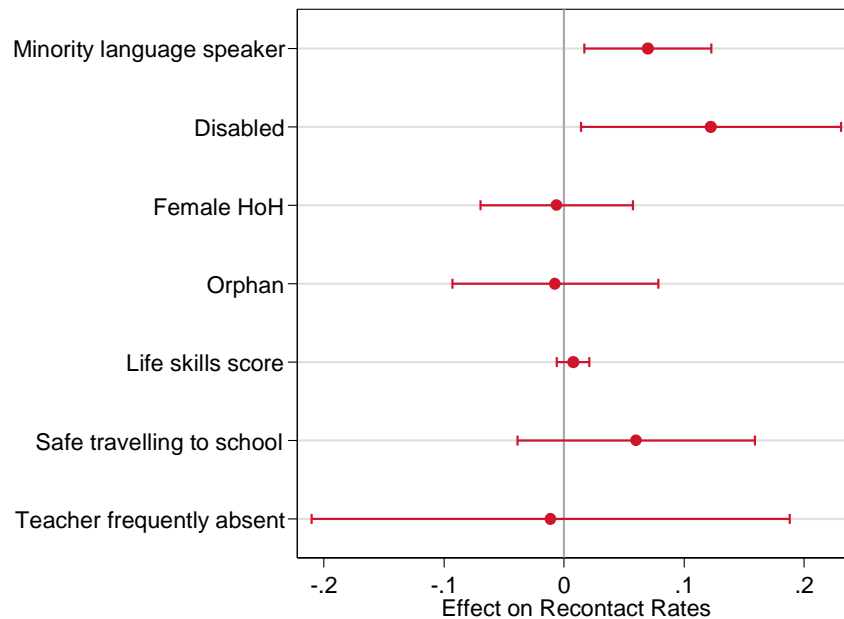
<sup>16</sup> Calculated out of a maximum of 17 points.

**Figure 2: Predictors of recontact rates: county and age**



The below figure further analyses results by other demographic, household, and school characteristics. We find that minority language speakers and disabled girls were significantly more likely to be recontacted than majority language speakers and non-disabled girls respectively, controlling for other characteristics including county and age. These findings are notable as it could be expected that minority language speakers and disabled girls would be more likely to drop out of school due to their marginalised status, and thus less likely to be recontacted. The data suggest that this was not the case.

**Figure 3: Predictors of recontact rates: other factors from girl survey**



Null findings are also notable; for example, it is somewhat surprising that there is no correlation between life skills and recontact rates. Life skills measure aspects including self-confidence, decision-making, problem-solving, and leadership skills; we would expect these skills to be correlated with a girl's likelihood of staying in school, and thus recontact, but find no such effect. It may be that life skills are correlated with higher likelihood of staying in school, but that these girls are also more likely to transfer to boarding schools or schools in other areas, and are thus less likely to be found and interviewed.

### ***Attrition and Recontact: Caregiver Survey Data***

We now utilise the smaller caregiver survey sample to further understand attrition. The below table shows, as in Table 14, characteristics of the full SAS sample and the SAS sample with attrition (i.e., caregivers recontacted at EL). We note that the panel of caregivers contacted at both SAS and EL is small – only 118 respondents. As such, this comparison is only indicative of potential differences, but is not definitive, and we do not utilise the data for testing predictors of recontact or the significance of differences.

**Table 15: Attrition characteristics, caregiver survey**

<b>Characteristic</b>	<b>Full SAS Sample</b>	<b>SAS Sample with Attrition</b>
Caregiver had no education	31.3%	32.2%
Caregiver partially or fully completed primary school	46.9%	44.1%
Caregiver partially or fully completed schooling above primary level	21.8%	23.7%
Caregiver speaks language of instruction well	25.5%	28.8%
Difficult to afford for girl to go to school	54.1%	50.9%
Girl has ever been pregnant	1.4%	3.5%
Girl has ever been married	1.0%	1.7%
Girl spends time caring for family members	53.4%	62.7%
Girl spends a few hours or more per day doing housework	32.3%	24.5%
Girl does agricultural work	21.8%	15.3%
Girl helps with a family business or other non-agricultural work	17.0%	23.7%
Girl spends a few hours per day or more doing the above tasks	35.5%	24.5%

Looking first at comparisons across the full SAS sample and the SAS sample with attrition, we find that most attributes are similar across these samples. Caregivers have relatively similar education levels, speak the language of their girl child's instruction at roughly equivalent levels, and find it similarly difficult to afford for their girl to go to school. The frequency with which caregivers reported that their girl has ever

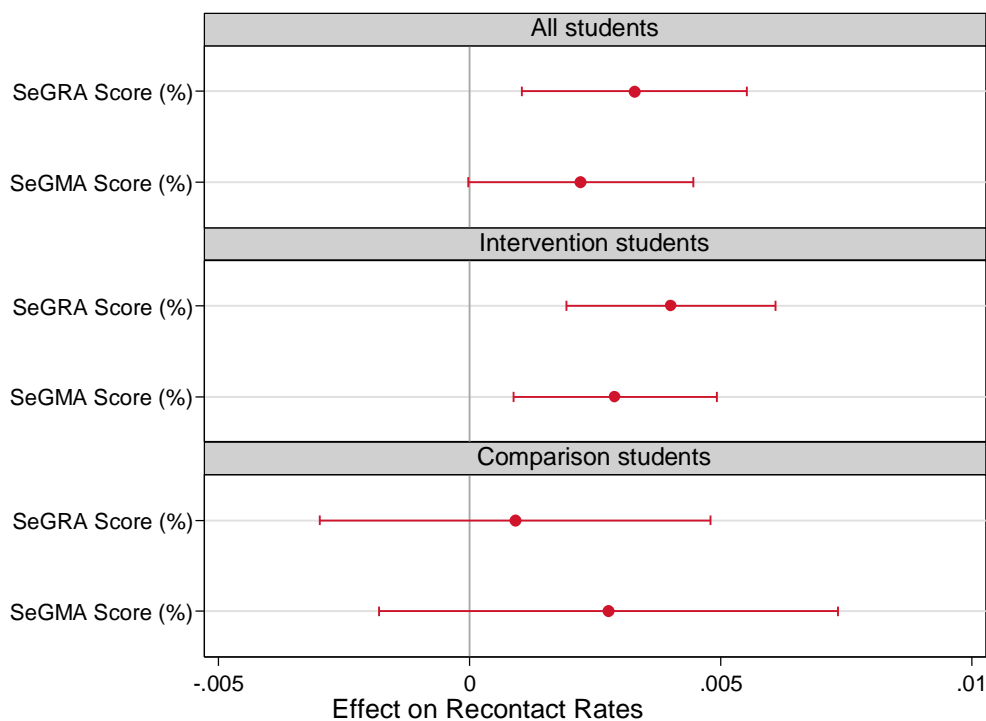
been pregnant or married was somewhat, but not substantially, higher among the SAS sample with attrition. The most notable differences come in girls' household and non-school responsibilities. We find relatively higher rates of girls participating in caregiving and helping with a family business among girls who were not recontacted. In contrast, we find relatively lower housework burdens and levels of participation in agricultural work among girls who were not recontacted. We cannot draw definitive conclusions about the differential characteristics of recontacted girls from this analysis, but it does suggest that there are some differences between girls (and their caregivers) who were recontacted compared to those who were not.

Looking at the midline sample, we find substantial differences in caregiver education levels, the affordability of schooling, and girls' household responsibilities. Girls' reported involvement in chores is particularly high at midline, with over 90 percent of respondents stating that the girl spends a few hours a day or more doing household chores. As in our analysis above, since we did not target caregivers surveyed at midline for recontact at endline, we expect differences in the samples. However, these findings suggest that there may be some systematic differences across ML and EL which effect our analysis.

**Attrition and Recontact: Learning Assessment Data**

We now analyse whether attrition rates correlated with learning assessment scores. For this, we utilise the ML-EL sample, as this is the sample primarily used for the learning outcome measurement. The below figure shows that among all students, girls with higher SeGRA scores were significantly more likely to be recontacted. In contrast, there was no significant relationship at the  $p < 0.05$  level between SeGMA scores and recontact, although there was a positive association. These findings suggest—in line with expectations—that stronger learners were relatively more likely to be recontacted at endline. This could occur if, for example, stronger learners are relatively less likely to drop out of school, and thus easier for the fieldwork teams to find and survey.

**Figure 4: Predictors of recontact rates: learning assessment scores**



Looking at just intervention students, we find a slightly stronger relationship between SeGRA and SeGMA scores and recontact rates, with both associations now significant. In contrast, among comparison students, we find no significant relationship between SeGRA/SeGMA scores and recontact rates. In other words, we are more likely to recontact strong students among the intervention group, but not more likely to recontact strong students among the comparison group. This may potentially bias our estimates of programme impact, as it makes us more likely to overestimate the WWW programme’s positive impact on learning outcomes.

Overall, the most important takeaway from this analysis is that differential attrition is likely to have affected the panel sample. The panel sample tracking girls from SAS to EL tends to slightly underrepresent Mombasa and Tana River and slightly overrepresent Samburu. Disabled girls and minority language speakers are also slightly overrepresented. Furthermore, girls with stronger English literacy skills were somewhat more likely to be recontacted at EL. However, along the other dimensions analysed above, attrition was relatively evenly spread across subgroups, without significant impact on sample composition.

## Differences Between Midline and Endline

We now briefly examine differences between the midline and endline samples. Differences are to be expected, as we did not attempt to recontact ML girls with the girl survey or caregiver survey. However, a basic analysis of differences is useful to better understand how the types of girls found and surveyed at endline may affect our analysis. The below table presents various characteristics for the midline and endline samples.

**Table 16: Girls' characteristics, midline and endline**

Characteristic	ML	EL
	Girl Survey Data <sup>17</sup>	
Age (mean)	14.3	16.6
Minority language speaker <sup>18</sup>	35.5%	51.0%
Reported disability <sup>19</sup>	3.6%	4.5%
Female head of household	37.9% <sup>20</sup>	22.0%
Feels safe travelling to school	90.9%	90.8%
	Caregiver Survey Data <sup>21</sup>	
Caregiver had no education	22.2%	17.2%
Caregiver speaks language of instruction well	23.6%	41.3%
Difficult to afford for girl to go to school	74.4%	83.6%
Girl has ever been married	0.5%	1.2%

<sup>17</sup> Values are based on the “actual midline” data provided to the evaluation team, which included a sample of 2,901 girls, from which most questions had responses from around 2,420 girls.

<sup>18</sup> Defined as a girl who does not speak Swahili, or does not speak Turkana in Turkana or Samburu in Samburu.

<sup>19</sup> From Washington group questions; includes difficulty seeing, hearing, walking, with memory, with self-care, and communicating.

<sup>20</sup> Based on the midline caregiver data for 2,891 caregivers, of which values were only reported for 1,770 caregivers (i.e., 38.8 percent were missing values).

<sup>21</sup> Values are based on the midline caregiver sample of 2,891 caregivers, from which most questions had responses from around 1,770 caregivers.

Girl spends time caring for family members	73.7%	75.3%
Girl does agricultural work	27.8%	38.8%
Girl helps with a family business or other non-agricultural work	23.2%	31.5%
Girl spends a few hours per day or more doing the above tasks	91.8%	37.3% <sup>22</sup>

As expected, we find that girls are, on average, around two years older at endline than at midline. Although this is expected, it is still important to note that because girls are older, we expect relatively stronger performance on learning assessments and better life skills, confidence, and sexual and reproductive health knowledge due to maturation effects. This is one of the main reasons, in fact, why we rely on difference-in-differences analysis to better understand learning outcomes; without using a control group, we would expect to find better learning (and other) outcomes, but this would simply be due to aging.

Outside of this, our endline sample includes relatively more girls who speak minority languages and slightly more girls with disabilities. If we expect girls who speak minority languages or have disabilities to face greater challenges in school and life due to their marginalisation, then this may bias estimates of programme impact found by analysing the full samples of ML and EL students downwards.

However, we also find that girls were relatively more less likely to have a female head of household at endline and that caregivers were relatively less likely to have no education and more likely to speak the language of instruction well at endline. All of these effects may, in contrast, bias naïve estimates of programme impact upwards.

Lastly, we find that caregivers were somewhat more likely to report that it is difficult to afford to send their girl to school. This is potentially correlated with the finding that girls were more likely to report involvement in agricultural work or another family business at endline than at midline. If involvement in these activities detracts from the time the girl has to spend studying and learning, this may have an impact on learning outcomes. However, notably, caregivers were much less likely to report that girls spent a few hours a day or more on chores at endline than at midline, although given a high prevalence of “don’t know” answers at endline, these results should not be overinterpreted.

## Differences Between Intervention and Comparison Girls

This section examines systematic differences between girls and schools in intervention and comparison groups that may impact our difference-in-differences analysis. We note that because girl surveys were not conducted with Cohort 1 girls, we have very limited demographic data with which to conduct this analysis. As such, we focus on differences in age and school characteristics presented in the below table.

---

<sup>22</sup> However, 25.4 percent of caregivers responded “do not know” to this question.



**Table 17: Differences between intervention and comparison girls, endline**

Characteristic	Intervention	Comparison
Age (mean) <sup>23</sup>	17.8	18.7
Primary school	70.5%	63.0%
Total number of teachers (mean)	16.7	17.4
Student-teacher ratio (mean)	12.2	13.2
Number of days per week that teachers miss school (mean)	1.5	1.6
Number of school clubs (mean)	3.3	2.4

The most notable differences between our intervention and comparison samples comes from the average age of girls: The average age of comparison girls is nearly one year older than that of intervention girls. Similarly, substantially fewer primary schools were included in the comparison sample than in the intervention sample. Because we generally expect older girls to perform better on learning assessments and other outcomes due to maturation effects, we may find that outcomes are relatively worse for intervention students than comparison students simply due to demographic variance.

We also find some differences in school characteristics across the two samples, although these characteristics may bias findings in the opposite direction as the differences in age. Comparison schools have relatively higher student-teacher ratios (i.e., larger class sizes) despite having higher average total numbers of teachers. Comparison schools also have fewer school clubs, on average, which may reduce students' abilities to participate in extracurricular activities that strengthen life skills. It is worth noting, however, that this latter difference may be directly attributable to WWW programme interventions, rather than inherent differences in school characteristics.

### Annex 3: Learning Outcome Data Tables

Our basic regression analysis of learning outcomes for girls tracked from midline took the following form:

$$Y_i = \alpha + \beta T_i + \gamma R_i + \delta(T_i \cdot R_i) \quad (1)$$

Where:

$Y_i$  is the outcome of interest

$\alpha$  is a constant

$T_i$  is a variable controlling for whether the individual is in the intervention or comparison group

$\beta$  is the intervention group specific effect

$R_i$  is a variable controlling for the round of the observation (midline or endline)

$\gamma$  is a time trend common to intervention and comparison groups

$\delta$  is the true effect of treatment

Additional models used to calculate the robustness of results included (1) a control variable for age and (2) control variables for age and school fixed effects. Controlling for age provides an important robustness check given the differences in age across comparison and intervention groups discussed above. Adding school fixed effects additionally controls for any school-specific differences that may have varied across

<sup>23</sup> Calculated from data collected in learning assessments, hence the variance from results presented in the table above.

groups and biased results, such as teacher quality or access to learning materials, that may affect student learning. We note that inclusion of these fixed effects also serves as a control for differences across counties and across ASALs and urban slums. These models are specified as follows, where model 2 controls for age ( $A_i$ ) and model 3 controls for age and school fixed effects:

$$Y_i = \alpha + \beta T_i + \gamma R_i + \delta(T_i \cdot R_i) + \vartheta A_i \quad (2)$$

$$Y_i = \alpha + \beta T_i + \gamma R_i + \delta(T_i \cdot R_i) + \vartheta A_i + \mu_i \quad (3)$$

In model 3,  $\mu_i$  represents our school fixed effects, and can be interpreted as school-specific intercepts, defined as follows, where  $\rho_0$  is a constant,  $\rho_1$  is the coefficient for school characteristics, and  $Z_i$  is a variable for school characteristics:

$$\mu_i = \rho_0 + \rho_1 Z_i \quad (4)$$

The following tables represent the results of this regression analysis for the SeGRA, SeGMA, and subtasks, as also reported under Literacy Outcomes in the body of the report. We note that we do not include baseline results, as girls were not recontacted from baseline to endline.

**Table 18: SeGRA difference-in-differences results**

Model	Intervention		Comparison		Diff-in-Diffs	P-value
	ML	EL	ML	EL		
No controls	54.8	50.5	50.1	41.4	4.3	0.26
Controlling for age	-	-	-	-	4.2	0.28
Controlling for age and school	-	-	-	-	4.9	0.22

**Table 19: SeGRA subtask 1 difference-in-differences results**

Model	Intervention		Comparison		Diff-in-Diffs	P-value
	ML	EL	ML	EL		
No controls	74.5	76.2	68.9	68.3	2.3	0.58
Controlling for age	-	-	-	-	2.0	0.64
Controlling for age and school	-	-	-	-	3.4	0.43

**Table 20: SeGRA subtask 2 difference-in-differences results**

Model	Intervention		Comparison		Diff-in-Diffs	P-value
	ML	EL	ML	EL		
No controls	47.2	47.5	42.1	37.5	4.9	0.30
Controlling for age	-	-	-	-	4.7	0.32
Controlling for age and school	-	-	-	-	4.7	0.35

**Table 21: SeGRA subtask 3 difference-in-differences results**

Model	Intervention		Comparison		Diff-in-Diffs	P-value
	ML	EL	ML	EL		
No controls	42.7	41.7	39.3	33.1	5.2	0.30
Controlling for age	-	-	-	-	5.3	0.29
Controlling for age and school	-	-	-	-	6.0	0.24
No zero scores	45.2	49.0	42.9	42.5	4.2	0.28

**Table 22: SeGMA difference-in-differences results**

Model	Intervention		Comparison		Diff-in-Diffs	P-value
	ML	EL	ML	EL		
No controls	34.2	36.5	28.2	29.3	1.3	0.65
Controlling for age	-	-	-	-	1.3	0.66
Controlling for age and school	-	-	-	-	1.0	0.75

**Table 23: SeGMA subtask 1 difference-in-differences results**

Model	Intervention		Comparison		Diff-in-Diffs	P-value
	ML	EL	ML	EL		
No controls	52.4	55.6	44.6	49.1	-1.3	0.76
Controlling for age	-	-	-	-	-1.3	0.77
Controlling for age and school	-	-	-	-	-1.7	0.71

**Table 24: SeGMA subtask 2 difference-in-differences results**

Model	Intervention		Comparison		Diff-in-Diffs	P-value
	ML	EL	ML	EL		
No controls	35.6	38.5	27.9	29.4	1.4	0.63
Controlling for age	-	-	-	-	1.4	0.64
Controlling for age and school	-	-	-	-	1.2	0.69

**Table 25: SeGMA subtask 3 difference-in-differences results**

Model	Intervention		Comparison		Diff-in-Diffs	P-value
	ML	EL	ML	EL		
No controls	14.4	14.3	12.0	8.1	3.9	0.08
Controlling for age	-	-	-	-	3.9	0.09
Controlling for age and school	-	-	-	-	3.6	0.13

**Table 26: STEM assessment results**

	Intervention		
	SAS	EL	Diff.
Overall score	11.7	11.6	-0.1
Subtask 1 score (chemistry)	15.2	12.8	-2.4
Subtask 2 score (physics)	10.1	7.3	-2.8
Subtask 3 score (biology)	9.9	14.7	4.8

## Annex 4: Characteristics and Barriers

The below tables show girls' characteristics and potential barriers to learning and transition. We note that at endline, there is no data for comparison girls as only girls tracked from SAS were administered girl surveys (from which the below data is drawn), and the SAS evaluation did not include a comparison sample.

**Table 27: Girls' characteristics**

	Intervention			Comparison			Source
	Baseline	Midline	Endline	Baseline	Midline	Endline	
Single orphans (no mother) (%)	3.4	3.4	3.7	3.8	3.8	-	Girl survey, CS_11b
Single orphans (no father) (%)	12.5	13.0	18.0	11.7	13.5	-	Girl survey, CS_11b
Double orphans (%)	4.5	1.5	1.7	4.0	1.7	-	Girl survey, CS_11b
Living in female headed household (%)	34.7	38.0	18.0	35.5	41.2	-	Girl survey, CS_11b
Married (%)	0.8	0.6	1.2	0.7	0.9	-	Caregiver survey, PCG_22g
Mothers (%)	1.1	1.7	3.8	0.2	0.9	-	Caregiver survey, PCG_23g

Difficult to afford girl's school (%)	65.0	66.3	83.6	63.0	66.1	-	Caregiver survey, PCG_7enr
Language of instruction different from mother tongue (%)	84.2	88.7	58.2	85.4	87.5	-	Girl survey, Q_21 = 1 or Q_21 = 2
HoH has no education (%)	29.1	32.5	18.2	34.4	28.2	-	Caregiver survey, PCG_6 if HH_7 == 1
Primary caregiver has no education (%)	-	-	17.2	-	-	-	Caregiver survey, PCG_6

**Table 28: Potential barriers to learning and transition**

	Intervention			Comparison			Source
	Baseline	Midline	Endline	Baseline	Midline	Endline	
<b>Home – community</b>							
<b>Safety:</b>							
Doesn't feel safe travelling to/from school (%)	10.8	9.5	10.9	10.7	6.3	-	Girl survey, Q_40
<b>Sufficient time to study:</b>							
High chore burden (more than quarter of day) (%)	29.9	24.2	17.5	40.1	34.8	-	Girl survey, CS_16s
<b>School level</b>							
<b>Attendance:</b>							
Attends school half or less than half time (%)	1.9	3.4	10.7	1.8	2.3	-	Caregiver survey, PCG_6enr
Doesn't feel safe at school (%)	0.0	1.1	0.7	0.0	0.9	-	Girl survey, Q_41
<b>Teaching quality:</b>							
Disagrees teachers make them feel welcome (%)	2.2	1.8	0.9	1.2	1.8	-	Girl survey, T_Q_42_1
Agrees teachers treat boys and girls differently in the classroom (%)	24.1	19.2	6.4	25.6	21.2	-	Girl survey, T_Q_42_2
Agrees teachers often absent from class (%)	24.1	19.8	11.1	20.7	23.7	-	Girl survey, T_Q_42_5

The below table shows learning assessment scores by some characteristics of interest reported on in the main body of the report. Data is included for SAS and endline intervention students (rather than baseline/midline) because, as stated above, girl surveys from which characteristic data is drawn were administered with girls tracked from SAS.

**Table 29: Learning outcomes by key subgroups**

Characteristic	SeGRA Score		SeGMA Score		STEM Score	
	SAS	EL	SAS	EL	SAS	EL
Minority language speaker	28.6	38.0	23.1	28.4	12.3	10.2
Disability (any)	32.3	24.4	21.9	17.4	10.8	5.1
Female HoH	36.5	37.9	26.2	28.4	11.8	10.1
Orphan (single or double)	35.3	38.3	25.7	26.9	11.4	9.4
High housework burden (more than two hours per day)	-	35.8	-	27.7	-	9.9

## Annex 5: Logframe

Annex 5 is provided as a separate document.

## Annex 6: Outcomes Spreadsheet

The Outcomes Spreadsheet is provided as a separate document.

## Annex 7: Beneficiaries Table

The beneficiaries table is to be completed by the project.

## Annex 8: Inception Report

The Inception Report is provided as a separate document.

## Annex 9: Data Collection Tools Used for Endline

- Quantitative data collection tools:
  - Learning assessments
    - Secondary Grade Reading Assessment (SeGRA)
    - Secondary Grade Mathematics Assessment (SeGMA)
    - Science, Technology, Engineering, and Mathematics (STEM) assessment
  - Girl survey
  - Caregiver survey
  - School (head teacher) survey
  - Classroom observation
- Qualitative data collection tools:
  - Focus group discussion tool for girls
  - Key informant interview tool for teachers
  - Key informant interview tool for government officials

- Key informant interview tool for project staff

**For all tools, respondents only provided consent that data be used for the purposes of the WWW endline evaluation.**

Full data collection tools are provided as a separate document.

## Annex 10: Qualitative Transcripts

This annex include two qualitative transcripts.

**Transcript 1** is from a FGD conducted with girls in Kilifi in the afternoon of October 8, 2022.

**Transcript 2** is from a KII conducted with a teacher in Kilifi in the afternoon of October 20, 2022.

### Focus Group Discussion with Girls, Kilifi, Int. 13

*Note: "M" is moderator; "R" are respondents.*

M: Do you all agree to participate in this research project?

R: Yes (*all the participants agreed to participate and be recorded*)

M: Do you think you are learning in schools?

R: *All responded: Yes*

M: What are your favourite subjects?

R: Maths

R: English

R: English and biology

R: Maths, physics

R: Chemistry, geography

M: You?

R: Physics, geography

R: History, English

R: History, maths

M: Okay, which subjects do you find most difficult?

R: Chemistry

M: Eeh, which one?

R: Chemistry

M: You?

R: Chemistry

R: Biology

M: Why do you find chemistry difficult?

R: Due to peer pressure

M: From who?

R: Other students

M: Any other reason... you?

R: I don't understand some concepts

M: Okay, do you ask the teacher to explain more?

R: Yes

M: And then when they explain you don't get it?

R: Yes

M: So it's hard for you kabisa?

R: Sometimes you can read but get nothing

M: You read and you don't understand?

R: Yes

M: Okay, who else mentioned a different subject?

R: Maths and Biology

M: Why?

R: Some of the words in biology are difficult to write

M: You mean some biological terms are difficult to write?

R: Yes

M: Any other reason?

R: Yes, and even pronunciation of the some of the words are difficult

M: Any other thing part from biological words being difficult to pronounce and difficult to read? Any other person? Someone mentioned geography, yes you, why?

R: Biology, because some concepts are hard to understand

M: Why?

R: Pronunciation of the words

M: What else?

R: Only that

M: Okay, to what extent are you interested in science subjects? Engineering or maths subjects? Are you interested too much or less?

R: Am too much interested in engineering subjects

M: You, to what extent?

R: 100% interested

M: Yes, to what extent?

R: 100% interested in sciences

M: And you?

R: Technology

M: To what extent?

R: 100%

M: And you?

R: Sciences

M: To what extent?

R: 100%

M: And you?

R: Maths

M: To what extent?

R: 100%

M: Have you become interested in them in the last year? Two years? And why? You said science?

R: Yes

M: When did you become more interested in science in the last one year or two?

R: In the last 2 years

M: Why?

R: Am interested in practicals

M: What else? You have you become interested in them in the last one year or two?

R: In the last one year

M: Why? You said technology.

R: Yes, the exams are not that difficult

M: Who else mentioned mathematics? Did you become interested in mathematics in the last one year or 2?

R: In the last 2 years

M: Why?

R: Because when I revise them and the teacher comes to teach I understand, and also the marks awarded for the questions are good.

M: You get it well?



R: Yes

M: Okay, when you struggle with a subject at school, does the teacher spend time with you to explain?

R: Yes

R: Yes

M: You?

R: Yes

R: Yes

R: Yes

R: Yes

M: If yes, can you tell me some examples? You said yes – start with you, can you tell me some examples?

R: In mathematics

M: You said mathematics is your best.

R: Yes but if there is something I don't understand I ask questions, then when I revise again and I don't understand I ask the teacher individually.

M: What about you, which subject?

R: Chemistry

M: Mhmm.

R: The teacher normally finds that extra time where he can come and meet us all. Sometimes I go and see him individually.

M: Okay and he explains more?

R: Yes

M: What about you?

R: Biology

M: Mhmm.

R: Am doing it in terms because I consult a lot with my teacher.

M: Okay what about you?

R: Mathematics

M: Mhmm.

R: For example the teacher has taught but if somebody has not understood, you say exactly where you have not understood and then he explains more.

M: Okay, and does the teacher ever punish students who are struggling to learn?

R: No (*all responded*)

M: Okay, when do you have time for homework? Yes this is a boarding school but you have assignments, okay?

R: Yes. After having family classes at 10am.

M: What are family classes?

R: Form 3 and Form 4, usually, we have combined classes, from each and every class we have some students going to the other class so those who are left do their homework.

M: What about you, where are Form Twos?

R: Even during our free time.

M: You?

R: I extend class from 9:30 to 10

M: And for each subject how much homework do you typically have a day? Let's say like mathematics, per day how much homework?

R: Maybe 2 assignments

R: 3 assignments

M: You?

R: Three

M: What about in a week?

R: Because there are some subjects we don't have assignments, but like Swahili and English when the teachers always give us assignments.

M: Let's concentrate on the ones the teachers gave you assignments, like how many in a week?  
R: Twelve  
M: Yes, you?  
R: Around 15  
M: Yes, in a week?  
R: Twelve  
M: Does anyone help you with homework at home? When you are given homework over the holidays does anyone help you with the assignments at home?  
R: Yes  
R: Yes  
M: Okay, and do you feel that you are able to focus when doing homework at home?  
R: Yes  
M: You, are you able to focus?  
R: Yes  
R: Yes  
R: Yes  
M: Okay, are there times when you miss school?  
R: No  
R: No  
R: No  
R: Sometimes  
M: Okay, does someone in your family ever prevent you from missing school? Anybody?  
R: Yes  
M: Who?  
R: My father  
R: My mum  
M: Can you tell me any stories about missing school?  
R: Maybe like last year I was missing school when we were sent home for school fees  
M: Anyone with another story? Okay we can proceed, when you don't understand something is the teacher explaining what do you do?  
R: I ask him questions  
M: Mhmm, you, what do you do?  
R: I tell him to repeat  
M: What about you?  
R: I ask the teacher to repeat

M: When you don't understand something the teacher is explaining, what do you do? Do girls normally ask questions when they don't understand?  
R: Yes (*all responded*)  
M: Okay, what happens when a girl asks a question?  
R: The teacher replies to the questions  
M: Mhmm, what about you?  
R: The teacher explains  
R: The teacher repeats again to make sure you understand  
M: What about you?  
R: The teacher elaborates more  
M: Okay, thank you. How do teachers punish students in this school?  
R: Ask the student to wash the dining hall  
M: Mhmm, any other?  
R: Watering some plants in the compound  
R: Sweeping the compound  
M: Any other form of punishment?

R: Some teachers do cane us

M: Which students does the teacher punish most often?

R: The ones who have failed to do assignments

M: Mhmm, any other group of students that is punished?

R: Yes, those who disobey the school rules

M: Yes?

R: Indiscipline students

M: Any other reason for punishment? How do they respond to the punishment?

R: They respond positively

R: They do (*all agreed*)

M: Okay, how do students respond to physical punishment? Like caning?

R: They accept

M: They just accept?

R: Yes because if you don't accept you are given a letter to go home and come with your parent and then when you come back you have to do the punishment

M: Any other way in which students respond to punishment? Other than just accepting? Okay do you think they act out more or do you think they can skip school or become embarrassed?

R: They feel embarrassed

M: Do they tend to skip school because of the physical punishment?

R: No

M: Do they act out more?

R: *No response*

M: How would you describe the attendance of the teachers over the last one year?

R: They were always present in the school

R: They were present

R: They were ever present in school

M: Do you think teachers' attendance has gotten better or worse in the last one year?

R: It has gotten better

M: Mhmm, what about you?

R: It has gotten better

M: Okay, in what ways do you think your learning is impacted when teachers do not come to class?

R: Yes

M: In what ways?

R: Maybe he or she has some topic to cover that day then he or she does not come, then it's like he is taking us back and it will be hard for us to cover the syllabus.

M: Any other way you feel you are impacted when teachers do not come to class, you?

R: *No response*

M: Okay do you face any danger going to school or returning home?

R: No

M: And you?

R: No

M: Do you have friends from your village who go to different schools?

R: Yes I have

M: Do you know anything about their school?

R: Yes, they are always struggling to cover the syllabus

M: Any other thing you know about them?

R: One of my friends was telling me that their school is the best but to me I know our school is better than theirs

M: You, do you know anything about your friends' school?

R: Yes, they usually lead in KCSE

M: What about you?

R: They are struggling to improve their performance

M: Do they like their school?

R: Yes, they like it

M: Mhmm, do you think they are learning different things from you?

R: No

M: So you are learning the same things?

R: Yes

M: Now am going to read a story for you, listen carefully.

Adea is 14 years old. She lives with her mother and her younger sister and brother. Adea loves to play with her siblings. She is in Grade 5. She started school a bit late because her mother did not have money to pay the school fees.

Adea loves to learn Swahili, but she is struggling with English. She feels a bit discouraged when she doesn't understand what Mr. Mwangi, the teacher, is explaining on the board. When she is at home with her sister and brother, she loves to talk, but she feels very shy with Mr. Mwangi and her colleagues. When the school closed because of COVID, Adea got very worried about studying English alone for the exams.

M: What do you think happened to Adea during the lock down? What happened after the schools reopened?

R: Her results in English kept declining since she was not going to school and English was a difficult subject to her

M: Okay

R: Her performance in English went down because during Covid she didn't have enough time to study English

M: Okay, you tell us something

R: Her performance decreased because she didn't have anyone to consult with

M: Okay, you wanted to say something? What happened after the schools reopened?

R: It was difficult for her to catch up with the other students

M: Mhmm, you want to say something?

R: She was forced to struggle more

M: Okay, you have not said something

R: She struggled much more

M: Do you think Adea learned English?

R: No

M: Why?

R: it was a difficult subject to her and she could not handle it alone

M: Mhmm, you?

R: She feels bad for not understanding English

M: Let's go to the next question. Do you think Mr. Mwangi helped Adea when the schools were closed?

N: No

M: Okay, and after the schools reopened do you think Mr. Mwangi helped her?

R: No

M: Why?

R: Because Adea herself was to have courage to talk to Mr. Mwangi but she was shy

M: Okay, she was shy?

R: Yes

M: Any other reason? Do you think WWW could help Adea while the school was closed?

R: Yes

M: Mhmm, why?

R: Because WWW they are always producing some papers of English and maths for students to do it well in this holiday so if Adea was attending this she could have been helped

M: Okay, imagine you are Adea, or that she lives in your village, how would Adea's story end?

R: By looking for person who knows English, maybe the concept of that teachers when he is teaching in front but another person could help her by making her to understand better

M: Let's assume that she gets someone and she is taught better, now how will it end, what will be the outcome now?

R: She will be able to understand English

M: Another outcome?

R: She would look for her friend and discuss about English

M: Another outcome? You have talked of positive outcome, any negative one?

R: Her performance would continue to go down

M: Why?

R: Because she was shy and didn't have anyone to help her

M: Any other possible outcome? We can proceed, is there anything that would change Adea's story if she lived in your village?

R: Yes

M: What?

R: The struggling would stop and she would put more efforts in English and stop being shy

M: Anything else that would change her story if she lived in your village

R: *No response*

M: Do you know any girls like Adea?

R: Yes

M: And what happened to them when schools closed?

R: They always say they don't understand that subject and they have a negative attitude towards that subject, they say that even if they study they may not pass

M: What about when school reopen again?

R: The attitude remains the same, they have a negative attitude towards the subject

M: Someone else who knows a girl like Adea? it doesn't have to be English

R: Yes

M: Tell us about her

R: They keep on relaxing back at home

M: Do they do revision?

R: No

M: So when the schools reopen?

R: It's difficult for them to catch up

M: Anyone who has a friend like Adea?

R: I don't have

M: Can you tell us about what happened to you when the school closed? Someone tell us something, let's start here

R: I was just reading the subjects I understood and those that I did not understand I leave them.

M: You leave them?

R: Yes, because I was lacking someone to consult

M: Okay, you?

R: During the Covid time I used to meet with my former primary school friends and we met in our former school and discuss together

M: Did you have a teacher or someone to guide you?

R: Some of them were ahead of us in class so they used to help us

M: So they were coaching you?

R: Yes

M: What about you?

R: I was reading my notes alone

M: You didn't have someone to consult?

R: No

M: What about you?

R: I was reading my notes and ask my uncle questions where I didn't understand and if he doesn't know I leave it

M: What about you?  
R: We had a group discussion with my former colleagues from primary  
M: Mhmm, did you have someone to coach you?  
R: No, we were discussing on our own  
M: What about you?  
R: We used to meet in our form primary school, but one of the administration staff told us not to gather so we studied on our own and it was difficult, I could not concentrate when I was studying, my mother calls to go and help  
M: We are interested about how things were in this village, tell us.  
R: Many girls in my village became pregnant  
M: Okay, then?  
R: They are just at home  
M: Learning stopped until now?  
R: Yes  
M: Anyone who has a story from their village?  
R: Some girls got pregnant but after birth they went back to school  
M: Okay, tell us a story from your village  
R: Others were discouraged to go back to school because Covid took a long time and they had grown big so they were ashamed of themselves  
M: Any other stories? What did you do during the Covid 19 time?  
R: I kept on reading till schools reopened  
M: What about you?  
R: I read past question papers and assisted my parents  
R: I helped my mother who was running a hotel  
M: What about you?  
R: I was reading but my mother could send me to my grandmother to wash clothes for her and some small activities  
M: Any other person?  
R: I was reading and helping my parents  
M: Okay, did your teachers help students to learn?  
R: Yes, they helped  
M: In what ways?  
R: When I came back to school I heard some girls say they formed a WhatsApp group and they were discussing with their teachers, the teachers send some past papers through the group to the students  
M: Someone else, apart from the WhatsApp group? Did your families encourage you to keep on learning?  
R: Yes  
R: Yes  
M: What about yours?  
R: Yes  
R: Yes  
R: Yes  
M: And in what ways did they encourage you to keep learning?  
R: They used to wake me up at 4am to begin my study  
M: What about you?  
R: They always encouraged me to go for group discussions with my friends  
M: What about you?  
R: They created time for me to do my personal studies  
M: And you?  
R: They kept reminding me of my studies  
M: What about you?  
R: They give me enough time to go and discuss with my friends  
M: Mhmm, what about you?

R: They kept reminding me about studies

M: Okay that's nice, during the Covid 19 times if you continue to study, where did you continue to study from?

R: At home

M: And you?

R: At home

M: What about you?

R: In my former primary school

M: Who were you with?

R: Alone

R: Alone

M: You?

R: I had a group

M: And you?

R: Alone

M: You?

R: I had a group

M: So only 2 people had a group, the rest were studying on their own?

R: Yes

M: Okay, when about when the schools reopened again please tell us a little about what happened to you after the Covid time?

R: It was hard for me catch up

M: Mhmm, did you ask the teachers for assistance?

R: Yes

M: And they assisted?

R: Yes

M: Mhmm tell us something

R: It was difficult for me to wake up during morning preps because I was used to sleeping too much at home

M: What about you, tell us something

R: Here at home I was always used to eating and now when I came back to school there is a scheduled time to eat. It was difficult for me to cope, my head used to ache because I am hungry

M: Mhmm, what about you?

R: Some concepts were hard to catch up because others students were learning through the WhatsApp group and I had to consult from my friends

M: Did they help you?

R: Yes, after some time

M: Now am going to read another story for you, please listen carefully.  
Akeyo and Mariam are neighbours. They are both in Grade 6. Akeyo and Mariam come from poor families. Their parents have never attended school and cannot read or write. Akeyo's mother has been ill and Akeyo has been taking care of the chores at home as her siblings are still very young. She has had to miss some school days to help her mother when she got sick. She told Mariam, "I'm struggling a bit to learn Swahili and mathematics."

What do you think will happen to Akeyo?

R: She will get help from her friends

M: Mhmm, somebody else, what do you think will happen to Akeyo?

R: Her performance in English and Mathematics will go down

M: Yes?

R: She might drop out of school

M: What about you?

R: Her performance will go down

M: Why do you think that the answers you have given could be the most likely outcome? Let's start here, why do you think Mariam will help her?

R: Since they have been discussing the subject

M: What is the reason for your answer?

R: Because she doesn't have someone to help her, her parents do not have any skill to help her

M: You said she might drop out of school, why?

R: Her parents were poor

M: So?

R: Could be they didn't have enough money to pay her school fees. She used to miss school so there is likelihood to drop

M: You said her performance will go down, why?

R: Mariam she could not understand some concepts in Swahili and maths so she could not help Akeyo

M: Do you think Mariam could do anything to help Akeyo?

R: Yes, by explaining to her the concepts that she could understand but the concepts that Akeyo could not understand and were difficult to Mariam, she could have consulted the teacher then come to explain to Akeyo

M: Yes, what about you, what do you think Miriam could have done to help Akeyo?

R: By advising her to find someone who knows those subjects so they should help her

M: Okay, you want to say something?

R: Yes, by teaching her the concepts that she understands better

M: Anyone who wants to add something? We can proceed. Do you think Akeyo's teacher Ms. Otieno could do something to help Akeyo?

R: Yes

M: Mhmm, what could she do?

R: Giving her exercises

M: What else? Anything else?

R: Telling her to remove the negative attitude towards those subjects

M: What about you, you had said yes

R: Yes by visiting Akeyo's home she may be to see the difficulties that Akeyo is going through, maybe she could help

M: Yes, any other thing?

R: Explaining where Akeyo does not understand

M: Okay we can proceed. Do you think WWW would do anything to anything to help Akeyo?

R: Yes

M: What could they do?

R: By visiting her and giving her some papers so that she could understand the concept that she was not understanding

M: What else could WWW do to help Akeyo, tell us

R: As her parents were poor maybe WWW could her pay school fees and continue

M: What else could WWW project do?

R: They could support her family financially for Akeyo's education

M: Who else could help Akeyo?

R: Some family members who have gone to school

M: Who else could help Akeyo?

R: The education officers

M: Who else? We can proceed?

R: Yes

M: Imagine you were Akeyo or that Akeyo lives in your village. How would Akeyo and Mariam's story end?

R: Their performance would improve

M: Why?

R: They could have gotten assistance from the teacher



M: You want to say something?

R: Maybe the performance of Akeyo would still go down

M: Why?

R: Because Miriam could decide not to help Akeyo her performance would drop down

M: Anyone else would say how their story would end? What is different in this village that would make the story end differently?

R: Maybe the village members would find people to support that family in Akeyo's education

M: Okay, do you any girls like Akeyo and Mariam in this village?

R: Yes I know

M: So what happened to their case?

R: They could drop out of school, and search for support to assist their family

M: Mhmm, you know of some?

R: Yes

M: Tell us about their story

R: Their parents are struggling and the girls are selling fruits to the market to help raise some money to help their parents

M: Let me read another story for you

Zuleikha is a smart girl. She loves to learn new things. She is always ready to help her mother and likes to take care of her siblings. Zuleikha needs a crutch to walk. The crutch is a bit uncomfortable and she struggles if she has to walk fast. She started school a bit late and feels a bit shy at school. She is now 14 years old and is in a catch-up education class.

What will happen to Zuleikha?

R: She might drop out of school

M: Why?

R: She is struggling to walk and she feels tired

M: Mhmm, you want to say something?

R: Yes, she will be discouraged by the other students because she could be the only big girl and the others are small

M: Any other thing that will happen to Zuleikha?

R: She would be discouraged and be isolated by the other girls

M: Why?

R: Maybe they would say she is too old to be in that class

M: Who else?

R: I would say she may drop out of school

M: Why?

R: Because she is the only girl who is bigger in the class than the other girls. She might not have a friend there because others are small

M: You want to say something?

R: Maybe those other girls are not disabled and so she may be discouraged

M: And how do you think Zuleikha's classmates would treat her?

R: They might treat her negatively

M: How?

R: She wants to stand up and the crutches are far and she might need assistance from the other students, then they can refuse it and will force her to struggle to get her crutches to get up

M: Someone else to say something?

R: The other small classmates might come and take the crutches and then they are like bullying her

M: Any other thing that might happen?

R: They might be stubborn to her

M: What else, okay we may proceed. Do you think Zuleikha's class teacher would treat her any differently from other students?

R: Yes

M: Why?

R: Because she is disabled the teacher will not ask her to do tasks like sweeping

M: Any other reason?

R: Yes

M: Why?

R: Because some activities should be carried out by walking from one place to another and because Zuleikha cannot walk fast she will not be able to that

M: What else?

R: Teachers know know that disability is not inability, she could do better in exams than every other student, so they will treat her fairly like the other students

M: We can proceed; do you think Zuleikha will face any problem in the school? We start from here

R: Yes

R: Yes

R: Yes

R: Yes

R: No, I don't think so

M: Why?

R: Teachers in the school will always take care of her

M: We can proceed, what could be some of these problems?

R: They might mistreat her by bullying her taking her crutches

M: Any other?

R: She might be discouraged

M: How?

R: Maybe she can fail to do something, then they tell her the way you are old and you can do something

M: You want to say something?

R: Yes, as she can't walk maybe they will discriminate against her

M: Do you think Zuleikha faces any problems in the community?

R: No

R: Not sure

R: No

M: Anyone else?

R: In the community is where she is coming from, now those people understand her situation, so she will not face any challenges

M: And please imagine Zuleikha lives in this village. How would her story end and why?

R: She will end up performing poorly in her studies

M: Why?

R: She will perform better

M: Why better?

R: Since knowing where she is coming from, they will be more caring, they will help her

M: Okay what about you?

R: She will perform better

M: Why?

R: From her story the people will encourage her to do better

M: Who else?

R: She will be encouraged by members of the community

M: Who else?

R: She might improve in her study

M: Why?

R: She may get financial support from the education officer and WWW

M: What about you?

R: I don't have an idea

M: Can you share stories of your own now? You can tell us stories about your own, school, friends, family, do you have stories you can tell?

R: Yes

M: Mhmm, continue

R: There is in our village, there is a person who cannot speak, people are not used to how he speaks so people write down for him and he is able to understand, when he was taken to school he blends well with people

M: That was interesting, another person, tell us about your drawing, okay if you don't have any we can proceed. We have been working in this school for several years. What do you think the project has done well?

R: It has boosted us girls here in school through the payment of the school fees

M: Mhmm, anything else?

R: Through the exam that they give us it enables us to do well

M: What else?

R: They have published books which contain English stories that improves our creativity

M: What other area has the project done well?

R: Helped some of the girls in their studies through offering financial support

M: You have not talked for so long, tell us something, any other area they have done well?

R: During the holidays they guided the girls on how they can improve their performance and achieve their goals

M: What else okay which activities should have been prioritised?

R: Payment of school fees

M: Why?

R: Majority of the students in the school who are sponsored by WWW are still being send home for school fees sometimes, if they could pay all the school fees.

M: What else?

R: Create time for meeting with all the girls in the county and have a session with them

M: What else?

R: Allow us to give them the challenge that they face by visiting us here in school and we share with them, come and listen to us

M: Any other areas? Okay, what activities should have been done differently, do you think there are activities that should have been done differently?

R: *No response*

M: Are there certain girls you think still need more help than others?

R: Yes

M: They are there? and who are they?

R: Those that need financial support (*most of the girls had this answer to give*)

M: What are your plans for your future education?

R: After completing my secondary education I want to go to the university and study law because I want to become a lawyer

M: Who else?

R: After finishing my form 4 I want to take and become a teacher

M: Okay, what about you?

R: After completing my form 4 education I want to become a doctor, I want to study medicine

M: What about you?

R: After finishing my studies here in school I want to join university and do electrical engineering

M: What about you?

R: Join university and study nursing

M: What about you?

R: I want to join university and do civil engineering

M: So you all guys want to join university. Congratulations and work hard! Okay, now when do you plan to finish school, you?

R: I want to finish my secondary school after completing form 4

M: You?

R: After form 4  
M: So all you want to finish form 4?  
R: Yes  
M: Thank you, does anyone of you want to join TVET programs?  
R: No  
M: Why not?  
R: I want to join university and not TVET, 95% of the learners said they want to join university  
M: Anyone with a reason why?  
R: Those TVET colleges don't offer teaching  
M: What about you?  
R: I will join the TVET, as am waiting for the results to be out, I will enrol and study as I wait to join the university  
M: What would like to study in TVET?  
R: Like tailoring, I want to be able to make my own clothes  
M: What about you?  
R: I will join the TVET and do a beauty course, I will start my business  
M: It has taken us 1 hour 39 minutes, I am so happy I have learnt a lot from you and it has been an interactive and interesting session, thanks once again and have a good lunch.  
R: Welcome

## Key Informant Interview with Teacher, Kilifi, Int. 31

*Note: "F" is facilitator, "R" is the respondent*

F: To begin please introduce yourself by telling me your name, age. How long you have lived here in this location and for how long have you been a teacher? And where did you transfer from? What is your role in [removed for anonymity] school?

R: My name is [removed for anonymity]. Am [removed for anonymity] years old. I teach mathematics and business studies here. I was the HOD Mathematics but currently I was given a responsibility of being the Deputy Principal so I gave the responsibility to another teacher 3 weeks ago

I have been in this school [removed for anonymity] for the last 8 years. I didn't transfer from any school although I was teaching somewhere else. My role in this school is to teach the learners maths and business studies.

F: What are some of the challenges you have identified in the community? How do they affect your work?

R: Mhmm, in the community here there is not much challenge but the challenge that I identified was on land issues. [Removed for anonymity] university were claiming that this land where [removed for anonymity] school is built is their land and that we took the land without their consent, so you find that there are some disagreement. So when I am in school, I am not sure if they will send their people to come and chase us away. In the community around us there is not much interaction so I can't say there is any interference or not.

In the school here another challenge is the learners sometimes sneaking out at night to go and buy some food stuffs, as a community we should control that.

There are other challenges, also is on the students, the learning resources. For the students when they come here they come with very little marks, so you find that most of them cannot grasp what you are teaching them. Sometimes you go back to the content as low as grade 4 so that you teach them what was taught in grade 4 – and they are in form 1 – so that they can now understand what you are teaching them, you go back and this calls for extra work like individual coaching.

Another challenge is that most students here come from poor families and so they report to school after a few weeks of opening. You find week one only a third of the students have reported. Week 3 is when  $\frac{3}{4}$  of the class will have reported, and this is a big challenge. You start teaching at week 2 and the students are absent. Even if the students are sponsored you find the sponsors do not pay for their transport to school. The challenge for the parents is how their children could reach the school. The third week is when some are reporting, already you have taught a concept and only a third of the class are present, so at the end of the term most of them will have missed and therefore their performance is not good at all so as a teacher I feel so bad. I put a lot of effort and at the end of it am demoralised.

About teaching resources, initially we had a lot of challenges, but we have received assistance from the sponsors and the government and I can say that the WWW has given us a lot of resources especially in mathematics and Sciences. We have several teaching materials which have made learning easier and interesting to the students. They have given us revision materials, the whiteboard, the smart board, a projector. Teaching has become easy and the learners now enjoy learning, even their interest in school has improved, they are reporting early.

F: Thinking about the last 3 years, have you changed anything about the way you teach?

R: Yes, I think I have changed a lot, one is that there are teaching models that we have acquired from trainings that have enabled us to use different teaching approaches, like now we can use the ICT to integrate the teachings, we can use group work model and they interact with it. Initially we didn't have such things, it would be lecture, question and answer, but now because we have the resources we can give the learners the resources, they interact with it, then they can do a presentation.

F: Were any subjects less or more affected by these changes than others? What subjects were they?

R: Yes, I think generally there has been some improvement in the school in most of the subjects, I will tell you the resources have translated to the performance of the learners.

F: Have you received mentoring support from anyone to help you improve as a teacher?

R: Yes, we have received mentorship programs from WWW. We have seen people come and talk to us in areas of teaching and also in personal life. The sponsors also have their programmes, they come and mentor the teachers.

Now that this is a special school there are special challenges other than teaching, you find that the learners can pose a challenge to you personally, you look at a student, you see the way they are struggling, you want to assist but you find you totally can't do anything to help, this directly affects you. As such mentorship helps us.

F: Has the head teacher or the district education officer mentored you?

R: Yes, we have received mentorships from the district education officer. Okay, you find that the ministry together with the TSC they always organise workshops, they bring mentors to talk to us, sometimes they visit the school and talk to us as a group and if you have a personal issue you can follow up with them, there is that opportunity.

F: Did you provide support to the students while schools were closed as a result of the Covid -19 lockdown?

R: We could not access our students so we didn't support them, only the form 4 who came back early after Covid, but the ones who were at home, we could not assist them.

F: What do you think are the main challenges that girls are facing in learning basic literacy and numeracy?

R: Some of the challenges I think is attitude, they perceive some subjects especially the numeracy and the sciences are male dominant. At the same time some of the challenges are from the parents, you know this is a mixed school with boys, also even if they don't perform well the parents won't encourage them. You should give them a challenge so that they can improve, even if they don't meet the expectation, but at least they improve.

Again the role models, this will also be a challenge, the ladies who have succeeded in such areas especially those who are deaf. You find they may have succeeded in other areas but in numeracy there are very few.

F: Have you grown more confidence in teaching stem subjects? Do you feel there has been change on how confident you feel in the last 3 years?

R: Personally I feel encouraged to teach maths because the students are following on the concepts closely and they can grasp when am teaching. The ICT integration has really helped us as I said earlier, the classes are participatory not a lecture anymore.

F: Think about a student who asks you a difficult question in a STEM subject. Would you handle a difficult question differently now than 3 years ago?

R: There are many ways you can use to make sure you solve a problem. Initially it was hard but through the mentorships and the programmes we have interacted with, the trainings and the workshops, you find you can try different ways to ensure the students get the information at once.

F: How would you describe the participation of girls in STEM subjects in this school?

R: Yes, in the last 3 years, the girls now have started having interest in the STEM subjects. One is the resources they have received. There is a program called girls in the STEM, they provide laptops, it has revived the girls to like STEM subjects. I would say the girls are average, not bad and not good.

F: Who are likely to take up the STEM subjects, boys or girls? What factors do you think contribute to different take up rates between boys and girls?

R: Generally I can say we don't give the students the opportunity to choose the STEM subjects because the number of teachers we have are not enough. What we can talk about now is the attitude, we have compulsory subjects of which maths is one of them and 2 science subjects, so whether you like them or not you have to take them. So what is remaining is how do you feel about these subjects.

This year we have a class of physics, we have 3 boys and 2 girls so we can say their perception towards sciences has improved.

F: Have you changed any practices during the past year when you are teaching students who are struggling to learn basic mathematics?

R: Yes, I have tried to change the practice where I put the students together, those who are fast learners and those who are struggling, they sit in pairs or in groups so that they can assist them. I sit them together and I take them through the steps. Then I give them assignments and I allow them to do in pairs so that they can discuss together.

F: If you notice a student is struggling to understand a lesson in English, do you try to explain it in their mother-tongue? Do you end up explaining most or all of a lesson in the student's mother-tongue?

R: No, I don't use mother-tongue in my class lessons because I don't even understand the mother-tongue that is spoken here. Secondly, the learners we admit here are from different ethnicities so the only language we stick to is Kiswahili. If I find a student is struggling to understand, Kenyan Sign Language is what the students understand better and that is what I teach them.

F: How would you describe the attendance rate of students over the last year? What factors have affected this change in attendance?

R: There are those learners who are not sponsored, so their parents pay their school fees, however some learners here are from very poor families. The poverty level has made some to drop out. Some do not have the necessary resources required for them to be in school so they end up dropping from school. Others lack even the transport to report back to school. This is a boarding school and a special school for the deaf, in that case, and our catchment area is from the whole country, some come from very far and the economic situation in the country is not very good so this has affected their attendance very much.

F: Are girl's attendance rates affected when they are menstruating? Have you noticed a change in this effect for the last 3 years?

R: I think it affects them at some level because a few of them don't come but we have a program where we give them sanitary towels so they cannot say they don't have them, but all them have access to sanitary towels. But how they feel is what might make them not to come to class. We have the dormitory mother and the boarding mistress who takes care of them, they assess the situations of the girls. So only a few are not able to come but most of them do come.

F: Are there particular kind of girls who are struggling to learn in this school/class?

R: Yes, there are, especially those girls who come with very low marks, they struggle to learn but they put in more effort. The entry marks really affect them.

F: Are there particular types of girls who get punished more often than others in this school?

R: No, we don't have particular girls who are punished here in school, but we had one who had some funny behaviour but it was corrected. But generally just a few who might do one or a few mistakes but once they are corrected you don't find them doing those mistakes again.

F: Which girls face the greatest barriers/challenges in enrolling or staying in schools?

R: Those from poor backgrounds are the ones who face the greatest challenges in enrolling in school.

F: What are the greatest challenges that schools in your community face in providing quality education and a friendly, safe learning environment, particularly for girls?

R: Poverty is the challenges that hinders the girls from accessing quality education. Also cultural beliefs of the community. This may pose as a challenge because of lack of sensitization that the girl child needs to be educated as well as the boy child. They believe the girls place is the home. Illiteracy is also another challenge. Most of the parents in the community never went to school or dropped out of school so they don't see its value and importance.

F: If a girl has difficulties walking or moving her arm/ hand, does she face any challenges to attend school in this community? Why? Do teachers welcome children with disabilities?

R: No, this school we take learners with hearing impairment mostly. We have no facilities to take care of other needs. We have learners with physical disabilities as well, others with autism. For those with extreme difficulties in walking we are not able to admit them because they need special care and we don't have the manpower. The ones we don't admit are the blind because we don't have structure in place to take care of them and their learning is different.

F: Are there teachers equipped to help students with these difficulties?

R: Yes, like here all the teachers have gone through a training on special children, so we are equipped in teaching them and handling other issues. Every teacher here is conversant with Kenyan Sign Language.

F: Is there a PTA in this school?

R: Yes we have PTA here in the school.

F: Does the PTA support girls' education? What are they doing to improve girls' education? Do you think their efforts are effective?

R: Yes, they partly support girls' education because there are times they provide the sanitary towels, there is a time the learners had an issue with entertainment, they wanted at least to have entertainment here at school and the PTA followed this up and it was put in place. Whenever the learners have issues we forward them to the PTA, then it's them that decide. Also having talks with the learners, encouraging them to work hard. Through this they have improved in class.

F: In the past 3 years do you think the roles of the girls at home have changed?

R: Yes, there are a lot of things the girls have learnt here in school. Other than the academics we teach them morals and behaviour and responsibilities. So you find that there are students that when they came the parents used to see them as people who cannot do anything, but after they came to school we explain that they are able to go and fetch water, do cleaning, wash clothes, go to the farm, can be sent to the shop, initially they could be sent to the shop by the parents. When we have the PTA meetings we explain to them these learners are like any other children, the only difference is the way of communication. The girls can enrol in different courses as long as there is a college that can accommodate them.

F: As a teacher do you feel the community supports your effort to better girls' education? Can you provide some examples or stories?

R: Yes, I think they are doing a lot because I have seen several programs by the government and the NGOs to support the girls, ensuring that the girls come to school, giving them sanitary towels, teaching them about adolescence.

## **Annex 11: Quantitative Datasets, Codebooks, and Programs**

Merged, cleaned, and labelled quantitative datasets, including the learning assessments, girl surveys, caregiver surveys, school surveys, and classroom observations are provided as separate files. Codebook and Stata cleaning program are also provided as separate files.

## **Annex 12: Learning Test Pilot and Calibration**

Learning assessments (SeGRA, SeGMA, and STEM) were designed based on the assessments used in the SAS study. Questions were generally designed to be directly comparable to SAS questions. For example, for the SeGRA, girls were given written passages of comparable length as midline for subtasks 1 and 2 and asked comparable questions about both passages. For the SeGMA, questions which asked about, for example, two-digit multiplication at midline were retained at endline with the numbers adjusted.

Initial versions of the learning assessments were sent to EDT, and both the SeGRA and SeGMA were flagged as slightly more difficult at endline than at SAS, while no comparability issues were flagged with the STEM assessment. For the SeGRA, issues were flagged with the length of the passages for subtask 1 and 2 and with the difficulty of several questions. With the SeGMA, several questions were flagged as requiring more complex calculations (such as three-digit as compared to two-digit multiplication) at endline than at SAS. These issues were addressed by adjusting the length of passages, simplifying



questions flagged as overly difficult, and in one case, replacing a question in the SeGRA with a less difficult question. After these adjustments, the assessments were judged to be comparable.

While a pilot of the learning assessments was initially scheduled for training, scheduling issues with schools targeted for the pilot meant that the learning assessments could not be piloted, with the focus pivoting to the girl and school survey. However, assessments were extensively reviewed and practiced by the external evaluator and data collectors during training, and no comparability or other issues were identified.

## Tests Used

As described in the report, the following tests were used, including descriptions of subtasks. Tools are included in Annex 9.

**Table 30: Learning Assessment Description**

Assessment/Subtask	Number of questions	Total points	Timing
SeGRA	19	48	45 minutes
SeGRA Subtask 1: Passage reading and simple reading comprehension	7	9	15 minutes (suggested)
SeGRA Subtask 2: Passage reading and complex reading comprehension	11	19	15 minutes (suggested)
SeGRA Subtask 3: Essay writing	1	20	15 minutes (suggested)
SeGMA	16	45	45 minutes
SeGMA Subtask 1: Two- and three-digit multiplication and division, simple geometry	7	16	15 minutes (suggested)
SeGMA Subtask 2: Simple algebra	5	14	15 minutes (suggested)
SeGMA Subtask 3: Word problems, complex algebra	4	15	15 minutes (suggested)
STEM	37	89	60 minutes
STEM Subtask 1: Chemistry	12	30	20 minutes (suggested)
STEM Subtask 2: Physics	13	30	20 minutes (suggested)
STEM Subtask 3: Biology	12	29	20 minutes (suggested)

## Methodology for Marking Tests

As the SeGRA, SeGMA, and STEM were written assessments, enumerators were given marking guides and marked the SeGRA and SeGMA immediately after administration. Some challenges were flagged with marking of the SeGRA—particularly subtask 3—as described in the Introduction. After review by trained teachers hired by both the external evaluator and the project implementer, however, no major issues with SeGRA marking were identified.

For the STEM, given that this assessment was technical, trained STEM teachers were hired to mark these assessments following the conclusion of fieldwork.

Subtask scores for all tasks were calculated as a simple percentage out of the total possible points, as shown in the above table.

## Enumeration Challenges

Outside of the challenge with enumerators marking SeGRA tests described above, several other challenges were identified during QC. First, it was found that some fieldwork teams were allowing girls to use calculators on the SeGMA. Teams were advised that calculators were not permitted, after which this practice no longer occurred. Second, within special needs schools, students were given an extra 20 minutes to complete learning assessments at SAS, but this practice was not continued at endline. As such, learning assessment results within special needs schools are not directly comparable from SAS to endline.

## Post-Calibration of SeGRA and SeGMA

Following fieldwork, the SeGRA and SeGMA were calibrated with students in Nairobi in a school not included in the endline evaluation. The SeGRA was calibrated with ten students, while the SeGMA was calibrated with eight students. These students were asked to take both the endline and midline SeGRA or SeGMA, in order to compare their scores on the assessments and calibrate relative difficulty. Half of the students were asked to take the midline assessment first, while half took the endline assessment first. In other words:

- Ten students took the SeGRA, of which:
  - Five students took the midline SeGRA followed by the endline SeGRA
  - Five students took the endline SeGRA followed by the midline SeGRA
- Eight students took the SeGMA, of which:
  - Four students took the midline SeGMA followed by the endline SeGMA
  - Four students took the endline SeGMA followed by the midline SeGMA

Exams were marked by the external evaluator.

As a result of this exercise, the endline SeGRA was calculated to be slightly less difficult overall than the midline SeGRA. Scores for subtasks 1 and 2 were slightly lower, on average, for endline assessments than midline assessments, while scores for subtask 3 were slightly higher on average.<sup>24</sup> In contrast, the endline SeGMA was calculated to be more difficult at endline than at midline; average scores for both subtask 1 and 2 were lower, on average, for endline than for midline. We note that four students received zero points for endline subtask 3 (and three students for midline subtask 3); as such, there was insufficient data to calibrate this subtask. Subtask 3 is, however, included in the overall SeGMA calibrations.

---

<sup>24</sup> One student scored 0 on midline subtask 3; this was included in the overall test calibration.

The below table presents the calibration factors calculated for each of the two assessments. The factor is presented as the endline assessment average as a percent of the midline assessment average; in other words, calibrated endline scores are calculated as raw endline scores divided by the below calibration values.

**Table 31: Assessment calibration factors**

Assessment	Overall	Subtask 1	Subtask 2	Subtask 3
SeGRA	1.048	0.999	0.960	1.014 <sup>25</sup>
SeGMA	0.888	0.971	0.838	-

Overall, this exercise revealed that the SeGRA was relatively well calibrated, particularly subtask 1, for which the calibration factor was almost 1.0. However, SeGMA difficulty was more variable across rounds. Within the *Learning Outcomes* section of the main report, we thus present calibrated scores unless otherwise noted.

## Annex 13: Sampling Framework

The sampling framework is provided as a separate document.

## Annex 14: External Evaluator Declaration

**Name of Project:** Endline Evaluation of Wasichana Wetu Wafaulu Programme

**Name of External Evaluator:** Consilient Research

**Contact Information for External Evaluator:** payce.madden@consilientresearch.org;  
erin@consilientresearch.org

**Names of all members of the evaluation team:** Payce Madden, Ryan Geitner, Iñaki Alvarez Camps, Nicolas Wicaksano, Samuel Nyabwari, Sadiq Abdiqadir

Consilient Research certify that the independent evaluation has been conducted in line with the Terms of Reference and other requirements received.

The following conditions apply to the data collection and analysis presented in the endline report:

- Quantitative (learning assessment, girl survey, caregiver survey, school survey, and classroom observation) and qualitative (interviews with girls, teachers, government officials, and project staff) data was collected independently by the EE and quantitative (learning assessments with girls in Marsabit and Kwale) data was provided by the project for analysis.
- Was data analysis conducted independently by the EE and does it provide a fair and consistent representation of progress? Yes *RM*

<sup>25</sup> Excludes values from one student who scored 0 on this subtask at midline.

- Data quality assurance and verification mechanisms agreed in the terms of reference with the project have been soundly followed (Initials: *PM*)
- The recipient has not fundamentally altered or misrepresented the nature of the analysis originally provided by Consilient Research (Initials: *PM*)
- All child protection protocols and guidance have been followed (Initials: *PM*)
- Data has been anonymised, treated confidentially and stored safely, in line with the GEC data protection and ethics protocols (Initials: *PM*)

Payce Madden



Consilient Research

10 March, 2023

## Annex 15: Additional Introductory Tables

### Activities

The WWW project was designed to provide five main interventions or activities, each of which is mapped to a target output. Activities are briefly summarised by their respective output in the tables below.

**Table 32: Project activities**

Intervention types	What is the intervention?	Relevant output	Relevant intermediate output	Means of contribution to outcomes
Digital tracker, school attendance, provision of bursaries, grants	To retain learners in productive learning pathways; see progression and transition to higher learning cycles	Output 1	Girls' attendance in productive learning pathways improves	Digital monitoring of attendance will provide real-time data for decision making to prevent/reduce dropping out.  Community Health Volunteer (CHV) visits will support early intervention.
Training of coaches and teachers, materials support, classroom observation, community of practice, use of ICT in learning, special needs learning training and materials, infrastructure support, capacity building of head teachers.	To create conducive learning environments for improved learning outcomes.  Apply lessons learnt and best practice.  Drive learning outcomes.	Output 1	Schools and ALPs become enabling environments for girls' learning and continuing in education at all levels.	Improved learning outcomes will enhance retention, progression, and transition.  Sustainability is built-in as teachers train teachers, resulting in continuous professional development.
Training of CHVs, household data collection, tracking of learners, cash transfers, and dispersion of solar lamps	Encourage the household to actively support girls' education by addressing socio-economic barriers, attitudes, and knowledge.	Output 4	Households actively support the transition of girls into productive education pathways.	Change of attitude, allocation of chores and resource allocation will enhance girls' prospects of remaining in productive learning pathways.

In-school and community-based mentorship, girls' kits, start-up kits, life skills, bursaries/scholarships	Girl empowerment to succeed in life through enhanced self-esteem, aspiration, and awareness.	Outputs 2, 3, and 4	Girls improve their aspirations to pursue productive education pathways	Increase understanding of education benefits and rights and reduce household barriers (economic/time) for study or re-engagement.
Refurbishment of catch-up centres, enrolment of subtenants in ALPs (catch-up and TVET), bursaries	Girls' re-enrolment and transition to ALPs.	Output 2	Girls improve their aspirations to pursue productive education pathways.	Drop-out girls will re-enrol in accelerated learning centres and transition to mainstream or alternate pathways.
Improving sustainability mechanisms for the project interventions.	Training of community members on social accountability and training national/sub-national MoE officials on gender analysis and project buy-in	Outputs 5 and 6	Communities develop more positive attitudes to enable girls' learning and transition	Once trained, communities will conduct social accountability forums. MoE staff will start taking up the supervision of project activities at different levels.

## Project Beneficiaries

The below table shows student beneficiaries at baseline and midline, disaggregated by grade and age.

**Table 33: Student beneficiaries, baseline and midline**

	Ages	Class 4	Class 5	Class 6	Class 7	Class 8	Form 1	Total	
<b>Baseline (BL)</b>	8-10		3,858	1,508	361			<b>5,727</b>	
	11-13		4,318	5,700	5,582	3,844		<b>19,444</b>	
	14-16		640	1,350	2,612	4,553	5,938	<b>5,093</b>	
	17-19		34	57	169	571	1,054	<b>1,885</b>	
	20-22		1	1	3	19	32	<b>56</b>	
	23-25				2		3	11	<b>16</b>
	26-31				1			2	<b>3</b>
	<b>BL Total</b>		<b>8,851</b>	<b>8,619</b>	<b>8,727</b>	<b>8,990</b>	<b>7,037</b>	<b>42,224</b>	
<b>Midline (ML)</b>	8-10		3,670	1,511				<b>5,181</b>	
	11-13		6,832	15,547	7,033	4,246	30	<b>33,688</b>	
	14-16		1,186	3,834	4,488	6,203	6,238	<b>21,949</b>	
	17-19		76	189	392	984	1,441	<b>3,082</b>	

	20-22		5	6	8	39	49	<b>107</b>
	23-25			3		3	11	<b>17</b>
	26-31			1		2	3	<b>6</b>
	<b>ML Total</b>		<b>11,769</b>	<b>21,091</b>	<b>11,921</b>	<b>11,477</b>	<b>7,772</b>	<b>64,030</b>

Additionally, the below table shows data on target and reached populations—including girls, boys, teachers, head teachers, households, government officials, and other relevant populations—by year. We note that information on the reached population as of December 2022 was not yet available for almost all indicators.

**Figure 5. Target and reached beneficiaries by output indicator**

Output Indicator	Population	2018		2019		2020		2021		Apr. 2022		Dec. 2022	
		T	R	T	R	T	R	T	R	T	R	T	R
Indicator 1.1	Teachers	500	0	1,000	2,345	1,500	2,302	1,500	2,269	2,000	1,196	2,300	
Indicator 1.2	School Heads & Deputies	0	0	100	342	250	436	250	467	400	437	490	
Indicator 1.3	Teachers	0	0	122	0	81	0	108	0	135	94	189	
Indicator 2.1	Girls	0	0	1,000	357	1,500	701	1,000	874	1,000	2,089	2,000	
Indicator 2.2	Girls	0	0	2,000	3,124	2,680	3,524	3,320	2,792	3,600	3,048	3,600	
Indicator 3.1	Girls & Boys	17,000	0	18,000	17,702	15,000	15,542	12,000	20,245	8,000	9,915	3,600	
Indicator 3.2	Facilitators & Girls	-	-	-	-	-	-	-	-	75%		100%	
Indicator 3.3	Girls & Boys	5,400	5,760	7,200	8,946	10,800	12,420	13,500	13,680	1,102	1,148	932	886
Indicator 4.1	Households	6,000	0	10,000	0	12,000	7,395	12,000		12,000		12,000	
Indicator 4.2	Households	100	0	900	772	1,500	4,936	1,800	5,296	2,000	3,935 <sup>26</sup>	2,500	
Indicator 4.3	Girls	0%	-	68%	-	98%	-	20%	43%	30%	23%	-	
Indicator 5.1	Groups & Individuals	10	0	20	151	40	202	300	271	450	466	-	
Indicator 5.2	Groups	50	0	300	0	400	110	500	266	500	397	500	

<sup>26</sup> Bursary beneficiaries, both TVET and secondary.

Indicator 6.1	MoE Officials	0	0	60	189	80	189	120	189	270	189	270	329 <sup>27</sup>
Indicator 6.2	MoE/TSC Officials	0	0	200	188	250	188	270	188	270	360	400	437

## Assumptions

The below table details assumptions for each output. Also included are means by which the EDT team verified that these assumptions held true throughout the lifecycle of the WWW programme. Verification of assumptions is vital to ensure that programme activities remain relevant and useful to achieving WWW goals.

**Figure 6. Project output assumptions and means of verification**

Output	Output Indicator	Assumptions	Means of Verification
OUTPUT 1 Teachers and school leaders in primary and secondary schools demonstrating gender-sensitive and enhanced teaching approaches (ICT and pedagogy) for improved learning.	<b>Output Indicator 1.1</b>		
	Number of primary and secondary teachers utilising improved teaching approaches.	<ul style="list-style-type: none"> <li>- Teachers are willing to integrate ICT into their teaching approaches.</li> <li>- Safety of ICT equipment and infrastructure is a primary objective of the schools.</li> </ul>	<ul style="list-style-type: none"> <li>- Quarterly reports collected via teacher evaluations</li> <li>- Quarterly formative assessment scores</li> <li>- Bi-annual KAP surveys</li> <li>- Bi-annual qualitative case studies</li> </ul>
	<b>Output Indicator 1.2</b>		
	Number of head teachers implementing action plans from the leadership mentor programme.	<ul style="list-style-type: none"> <li>- Teachers are open to learning and implementing new approaches to their leadership.</li> </ul>	<ul style="list-style-type: none"> <li>- Pre/post-test scores from Gender Champion facilitator training</li> <li>- Quarterly observation notes</li> <li>- Action point logs from the mentorship forums</li> <li>- Bi-annual KAP surveys</li> <li>- Bi-annual case studies</li> </ul>
	<b>Output Indicator 1.3</b>		
	Percent of primary and secondary school teachers utilising improved teaching approaches in STEM subjects.	<ul style="list-style-type: none"> <li>- Community engagement and mobilisation efforts are sufficient to mitigate resistance, stimulate demand for education, and enable girls' participation in all components of the programme.</li> <li>- Project content is engaging and meaningful to the girls.</li> <li>- Teachers use</li> </ul>	<ul style="list-style-type: none"> <li>- Pre/post-test scores from Gender Champion facilitator training</li> <li>- Quarterly observation notes</li> <li>- Bi-annual KAP surveys</li> <li>- Bi-annual case studies</li> </ul>

<sup>27</sup> Sixty-five officials (25 CSOs and 40 QASOs) were added to the 189 recorded at Q21, resulting in an actual class of 254 trained and monitoring MoE officials. An additional 9 CSOs were involved in school monitoring in Kilifi during Q22, resulting in an actual class of 263 officials. Finally, 66 additional MoE officials were trained and involved in monitoring at Q23, resulting in an actual class of 329.



		innovative teaching approaches that create demand for and uptake of the Literacy and STEM subjects.	
<b>OUTPUT 2</b>	<b>Output Indicator 2.1</b>		
Alternative learning pathways established or expanded for girls outside or at risk of leaving school.	Number of girls enrolled and continuing or have completed education in an alternate learning pathway (catch-up program, TVET, apprenticeship, etc.). (cumulative).		
	<b>Output Indicator 2.2</b>		
	Proportion of girls with improved perceptions on the viability of the alternative learning pathways.	<ul style="list-style-type: none"> <li>- The project will establish apprenticeships in 50% of the private sector contacts pursued.</li> </ul>	<ul style="list-style-type: none"> <li>- Routine monitoring</li> <li>- Project and partner reports and updates data</li> <li>- Special assessments</li> <li>- Routine and focused internal girl assessments and observations</li> </ul>
<b>OUTPUT 3</b>	<b>Output Indicator 3.1</b>		
Improved self-confidence and aspirations among the girls in mentorship and scholarship programmes.	Number of girls completing the mentorship programme.	<ul style="list-style-type: none"> <li>- The training and support package sufficiently equips Gender Champion facilitators to deliver high-quality sessions.</li> <li>- Adolescents can be recruited and retained to participate in the Gender Champion programme.</li> <li>- Caregivers are supportive of the programme content and allow adolescents to attend.</li> </ul>	<ul style="list-style-type: none"> <li>- Routine monitoring</li> <li>- Project/partner reports and updates</li> <li>- Special assessments</li> <li>- Focused studies</li> </ul>
	<b>Output Indicator 3.2</b>		
	Number of project girls and boys regularly attending girls' clubs or disability clubs.	<ul style="list-style-type: none"> <li>- School management will allow gender and reproductive health education/discussion to be the main pillars of club activities.</li> <li>- Since girls will be transitioning to other education pathways, the number of project girls attending the clubs will decrease.</li> <li>- Girls will be courageous enough to open up during the evaluations that are</li> </ul>	

		done by external personnel with whom the girls are unfamiliar.	
	<b>Output Indicator 3.3</b>		
	Percent of girls with improved understanding regarding their reproductive health risks and needs.	<ul style="list-style-type: none"> <li>- The training and support package sufficiently equips Gender Champion facilitators to deliver high-quality sessions.</li> <li>- Facilitators have buy-in into the content.</li> <li>- Will contribute to reducing drop-out rates to 30% of the estimated 26,000 girls already dropping out.</li> <li>- Households are willing to make investment decisions that create more/additional income or savings, which is used to financially support the girls' education.</li> </ul>	<ul style="list-style-type: none"> <li>- Routine monitoring</li> <li>- Project/partner reports and updates</li> <li>- Special assessments</li> <li>- Focused studies</li> </ul>
<b>OUTPUT 4</b>	<b>Output Indicator 4.1</b>		
Continued household support for girls' education, including in alternative learning pathways	Number of households with improved attitudes toward supporting and investing in girls' education.	<ul style="list-style-type: none"> <li>- Will contribute to reducing drop-out rates to 30% of the estimated 26,000 girls already dropping out.</li> <li>- Households are willing to make investment decisions that create more/additional income or savings, which is used to financially support the girls' education.</li> </ul>	<ul style="list-style-type: none"> <li>- Routine monitoring</li> <li>- Project/partner reports and updates</li> <li>- Special assessments.</li> <li>- Focused studies</li> </ul>
	<b>Output Indicator 4.2</b>		
	Number of households reporting that financial and other material support from the WWW project has helped them keep their daughters in school ( <i>disaggregated by support package</i> )	<ul style="list-style-type: none"> <li>- Will contribute to reducing drop-out rates to 30% of the estimated 26,000 girls already dropping out.</li> </ul>	<ul style="list-style-type: none"> <li>- Routine monitoring</li> <li>- Project/partner reports and updates</li> <li>- Special assessments</li> <li>- Focused studies</li> </ul>
	<b>Output Indicator 4.3</b>		
	Percentage of girls who have successfully been referred for services against the impacts of traditional initiation and		

	other retrogressive cultural practices.		
<b>OUTPUT 5</b>	<b>Output Indicator 5.1</b>		
Compared to baseline, school catchment communities are more aware of the importance, benefits, and opportunities available to support girls' education	Number of community groups conducting accountability and tracking the utilisation of education funds available to schools		
	<b>Output Indicator 5.2</b>		
	Number of groups from the catchment communities that have received funding and established functional income generating activities (IGAs) that support girls' education	<ul style="list-style-type: none"> <li>- The group's investments will yield profitable income.</li> <li>- The groups have the desire and will to utilise their funds/proceeds and that of the project towards improving the perception and support for education, especially girls' education.</li> </ul>	<ul style="list-style-type: none"> <li>- Routine monitoring</li> <li>- Project/partner reports and updates</li> <li>- Special assessments</li> <li>- Focused studies</li> </ul>
<b>OUTPUT 6</b>	<b>Output Indicator 6.1</b>		
The WWW project aligns to the WWW models, in turn informing emerging MoE gender and teaching approaches	Number of MoE officials trained on and conducting learning and gender responsiveness.	<ul style="list-style-type: none"> <li>- Quality assurance, curriculum support officers, and county education officials will be available for training, as well as for planning and adaptation meetings.</li> </ul>	<ul style="list-style-type: none"> <li>- Special studies</li> <li>- Training reports</li> <li>- Routine monitoring</li> </ul>
	<b>Output Indicator 6.2</b>		
	Number of MoE/TSC officials demonstrating appropriateness of the project system-led interventions as a means of improving learning and school governance structures	<ul style="list-style-type: none"> <li>- County governments are interested in investing and expanding TVET training opportunities.</li> <li>- Quality assurance officers will be willing to incorporate the project interventions as part of their monitoring.</li> <li>- Some of the interventions will be pursued as part of the Teacher Performance Appraisal and Development mechanisms for the MoE and TSC.</li> </ul>	<ul style="list-style-type: none"> <li>- Special studies.</li> <li>- Training reports</li> <li>- Routine monitoring</li> </ul>

## Annex 16: Supplementary Analysis of Learning Outcomes

### Literacy Outcomes

The below table presents a summary of overall scores for all endline and midline students. We note that the below scores are not calibrated. Among all endline students (including those surveyed at both midline and endline, as well as those who were surveyed only at endline) in intervention schools, the average overall literacy score was 43.8 percent, while the average literacy score for intervention students at midline was 53.7 percent. Overall literacy scores declined for comparison students as well as those in intervention schools. This pattern of decreasing scores from midline to endline holds across almost all subtasks for both groups; in other words, on average, students performed worse at endline on almost every subtask compared to midline.

**Table 20: Summary of SeGRA scores, all learners**

	Intervention			Comparison		
	ML	EL	Diff.	ML	EL	Diff.
Number of respondents	917	1,265	-	555	190	-
<b>Overall score</b>	ML	EL	Diff.	ML	EL	Diff.
Mean score	53.7	43.8	-9.9	47.2	41.0	-6.2
Standard deviation	19.9	22.0	-	20.1	20.3	-
<b>Subtask 1</b>	ML	EL	Diff.	ML	EL	Diff.
Mean score	73.2	69.7	-3.5	67.6	68.2	0.6
Standard deviation	23.2	23.7	-	25.2	21.1	-
<b>Subtask 2</b>	ML	EL	Diff.	ML	EL	Diff.
Mean score	45.7	40.1	-5.6	37.9	36.8	-1.1
Standard deviation	27.7	22.9	-	25.7	21.2	-
<b>Subtask 3</b>	ML	EL	Diff.	ML	EL	Diff.
Mean score	42.1	35.6	-6.5	35.9	32.7	-3.2
Standard deviation	20.4	29.0	-	20.9	28.9	-

The standard deviation for both overall scores and subtasks was moderate, ranging from around 20 to almost 30 points. Standard deviation was relatively higher for subtasks 2 and 3 compared to subtask 1 and the overall score. This indicates that scores tended to vary moderately, with some students receiving scores far from the mean. Particularly for subtasks 2 and 3, relatively high standard deviations suggest that learning outcomes are widely distributed among learners; in other words, our sample includes students encompassing the full range learning outcomes, including both a large number with relatively low scores and a large number with fairly high scores.

The above scores are not calibrated; calibrated scores are presented in the following table. We note that subtask 3 was not individually calibrated, although it was included in the overall score calibration.

**Table 34: Summary of calibrated SeGRA scores, all learners**

	Intervention			Comparison		
	ML	EL	Diff.	ML	EL	Diff.
Number of respondents	917	1,265	-	555	190	-
Overall score	53.7	41.8	-11.9	47.2	39.1	-8.1
Subtask 1 score	73.2	69.8	-3.4	67.6	68.3	0.7
Subtask 2 score	45.7	41.8	-3.9	37.9	38.3	0.4
Subtask 3 score	42.1	35.6	-6.5	35.9	32.7	-3.2

The difference-in-differences results for overall scores are discussed in the main body of the report. To test the robustness of findings, we first add control variables to the regression of overall (subtask 1-3) calibrated SeGRA results to control for differences in student age; the results are shown in Table 22 below.<sup>28</sup> In general, older students perform better on learning assessments due to their additional years of exposure to school and general learning; as such, any differences in these demographic variables across comparison and intervention groups may bias our results. After adding this control variable, we find that there remains a positive but insignificant effect for the intervention group compared to the comparison group; in other words, when controlling for differences in age, at endline the intervention group performed around 4 percentage points better than would be expected compared to the comparison group. We also find a moderate positive relationship between age and literacy scores, as was expected.

We then add school fixed effects to the regression to control for any school-specific differences that may have varied across groups and biased results (see Table 22). These fixed effects control for observed or unobserved differences across schools, such as teacher quality or access to learning materials, that may affect student learning. We note that inclusion of these fixed effects also serves as a control for differences across counties and across ASALs and urban slums. Inclusion of these variables has little effect on findings; our estimate of programme impact remains positive, with a coefficient of around 4, and insignificant.

**Table 22: SeGRA difference-in-differences results**

Model	Intervention		Comparison		Diff-in-Diffs	P-value
	ML	EL	ML	EL		
No controls	54.8	48.2	50.1	39.5	4.3	0.26
Controlling for age	-	-	-	-	3.8	0.31
Controlling for age and school	-	-	-	-	4.5	0.25

The below table additionally adds these checks for robustness into the subtask 1 and 2 results.

<sup>28</sup> No other control variables can be added as not all girls who were administered a learning assessment were also administered a girl survey. We note that midline girls' ages were extrapolated from endline data, as midline data on age was found to be unreliable during fieldwork.

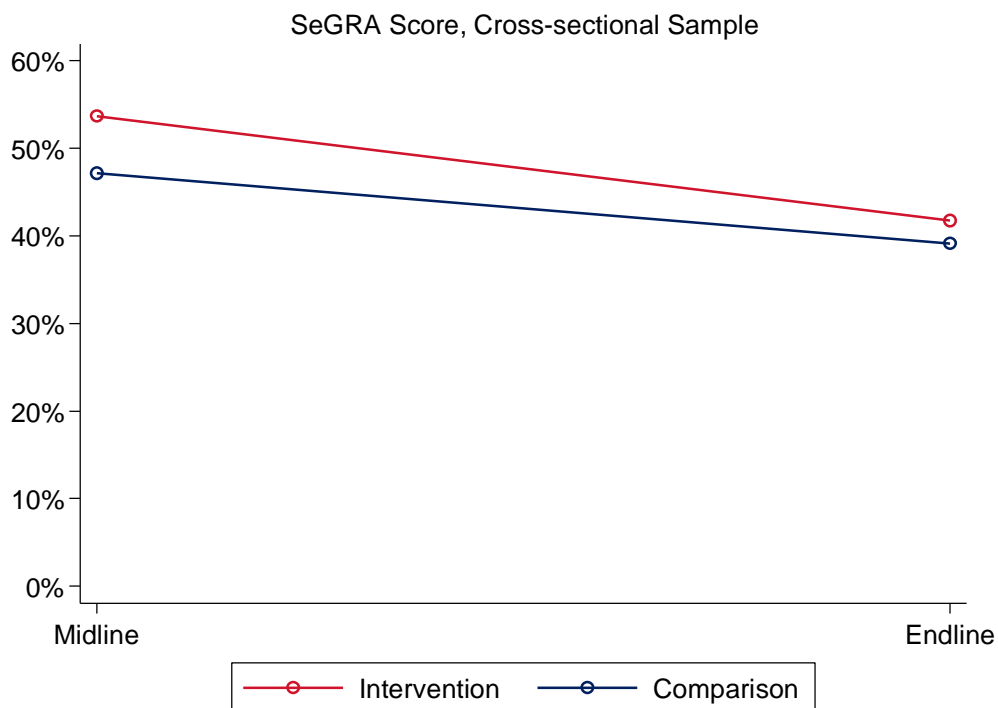
**Table 35: SeGRA subtasks 1 and 2 difference-in-differences results**

Model	Intervention		Comparison		Diff-in-Diffs	P-value
	ML	EL	ML	EL		
No controls	61.5	63.6	56.6	53.2	5.5	0.16
Controlling for age	-	-	-	-	5.2	0.19
Controlling for age and school	-	-	-	-	5.5	0.20

Overall, these findings suggest that while WWW programming may have had some effect on mitigating the negative impacts of COVID-19 on learning outcomes, these findings are not conclusive.

For further analysis, we now look at the cross-sectional sample.<sup>29</sup> The below figure shows change in SeGRA scores for this sample of students. With no control variables added, this figure shows that scores for the intervention group declined by around 3.8 percentage points more than scores for the comparison group, although the difference was not significant. However, as noted elsewhere in the report, since the cross-sectional sample includes different girls assessed in different rounds, the results found may be driven more by observable or unobservable differences in assessed students, rather than by actual programme impact. As such, we now add several control variables to examine the robustness of results.

**Table 36: Change in SeGRA score, cross-sectional sample**



<sup>29</sup> This sample includes 555 comparison students and 917 intervention students from midline, and 190 comparison students and 1,265 intervention students from endline.

First, we add control variables for age and school fixed effects. As this information was collected for all girls assessed at both midline and endline, this does not affect sample size. Controlling for just age, we find a small negative effect of 1.2 percentage points for intervention students relative to comparison students; this result is not significant. Controlling for both age and school, however, we find a positive effect for intervention students, whose scores improved by 3.2 percentage points more than expected given the results of comparison students. This result is again insignificant.

Next, we build off of this last model by adding weights based on the number of beneficiaries in each county targeted in the evaluation.<sup>30</sup> Use of sampling weights has a minimal effect on the overall results; we find that in this model, intervention students scored an average of 2.9 percentage points higher than expected given the results of comparison students, all else held constant. This result remains insignificant.

Overall, this analysis shows that for the cross-sectional sample, results are generally similar to or less positive than the difference-in-difference results. As mentioned above, however, we judge that these results are highly unlikely to represent actual programme impact, but rather to represent differences between students assessed in different rounds. As such, in our subtask-specific analysis, we primarily focus on the recontacted cohort and use the difference-in-differences analysis.

Examining learning bands for overall calibrated scores (including subtask 3), the below table shows that for intervention students, among both the panel sample and the full sample, there was a decrease in non-learners, established learners, and proficient learners at endline compared to midline. For the comparison sample, while there were small increases in the number of proficient learners, there was also an increase in emergent learners similar to that of the intervention group. In general, these findings reinforce the findings above about a regression in learning due to COVID-19.

**Table 23: SeGRA learning bands**

Panel Students	Intervention		Comparison	
	ML	EL	ML	EL
Non-learner (0%)	1.1%	0.3%	1.9%	0.0%
Emergent learner (1% - 40%)	23.2%	38.0%	25.8%	54.1%
Established learner (41% - 80%)	68.6%	59.3%	72.3%	44.0%
Proficient learner (81% - 100%)	7.1%	2.5%	0.0%	1.9%
All Students	ML	EL	ML	EL
Non-learner (0%)	1.4%	0.4%	1.1%	0.0%
Emergent learner (1% - 40%)	23.2%	50.0%	35.0%	54.7%
Established learner (41% - 80%)	68.9%	48.3%	61.6%	43.2%
Proficient learner (81% - 100%)	6.4%	1.3%	2.3%	2.1%

<sup>30</sup> For midline, Nairobi and Tana River were overrepresented relative to the number of beneficiaries in these counties, while Kilifi, Samburu, Turkana, and Mombasa were underrepresented. For endline, Samburu, Turkana, and Tana River were overrepresented relative to the number of beneficiaries, while Nairobi, Kilifi, and Mombasa were underrepresented. Adding sampling weights serves to “boost” the results of counties underrepresented in the evaluations.

Lastly, looking at the cross-sectional sample for just subtask 1 and 2 scores, the below table shows an increase in proficient learners among both intervention and comparison samples, but a decrease in established learners. This could be potentially suggestive of a gap in learning, whereby stronger students are moving ahead and weaker students falling behind. However, the similarity of results across the intervention and comparison groups suggests that these patterns may not have been strongly affected by WWW programming.

**Table 37: SeGRA subtasks 1 and 2 learning bands, cross-sectional sample**

All Students	Intervention		Comparison	
	ML	EL	ML	EL
Non-learner (0%)	0.4%	0.6%	0.7%	0.0%
Emergent learner (1% - 40%)	20.1%	25.7%	28.5%	29.5%
Established learner (41% - 80%)	61.3%	53.5%	61.8%	57.4%
Proficient learner (81% - 100%)	18.2%	20.2%	9.0%	13.2%

### **Subtask 1 Outcomes**

The below table shows difference-in-differences regression results and robustness checks for the panel sample. We find positive but insignificant results for all three models. Comparing scores, we see that among comparison panel students, subtask 1 scores slightly decreased from midline to endline; in contrast, among intervention panel students, subtask 1 scores increased by 1.7 percentage points.

**Table 24: SeGRA subtask 1 difference-in-differences results**

Model	Intervention		Comparison		Diff-in-Diffs	P-value
	ML	EL	ML	EL		
No controls	74.5	76.2	68.9	68.3	2.3	0.58
Controlling for age	-	-	-	-	2.0	0.64
Controlling for age and school	-	-	-	-	3.4	0.43

Lastly, the below table shows learning band results for the cross-sectional cohort (i.e., all students assessed at midline and endline). In contrast to the panel students, among all students, there was a decline in non-learners but also a decline in proficient learners among the intervention group, while there was a small increase in proficient learners among the comparison group. As discussed above, this finding is likely indicative of a change in the average type of students assessed at endline compared to midline as well as the impact of COVID-19 on learning.



**Table 38: SeGRA subtask 1 learning bands, cross-sectional sample**

All Students	Intervention		Comparison	
	ML	EL	ML	EL
Non-learner (0%)	2.3%	1.0%	2.7%	1.1%
Emergent learner (1% - 40%)	9.4%	11.2%	14.2%	9.0%
Established learner (41% - 80%)	51.6%	52.4%	56.2%	61.6%
Proficient learner (81% - 100%)	36.8%	35.4%	26.9%	28.4%

### Subtask 2 Outcomes

The below table shows results for the difference-in-differences analysis of the panel dataset. Our difference-in-differences coefficient is higher than those found for overall results or subtask 1; at endline, the intervention group performed 5.3 percentage points better than would be expected compared to the comparison group. Furthermore, as with subtask 1, we find that average subtask 2 scores decreased among comparison students at endline, while increasing slightly among intervention students. However, these results are not significant. Models which include controls for age and school fixed effects show similar results, with insignificant and positive coefficients of 5.1.

**Table 26: SeGRA subtask 2 difference-in-differences results**

Model	Intervention		Comparison		Diff-in-Diffs	P-value
	ML	EL	ML	EL		
No controls	47.2	49.5	42.1	39.1	5.3	0.26
Controlling for age	-	-	-	-	5.1	0.29
Controlling for age and school	-	-	-	-	5.1	0.32

Lastly, the below table shows learning bands for the cross-sectional cohort. Findings are broadly similar to those of the panel sample. We additionally note that only 3.2 percent of all intervention students scored 0 percent of this subtask at endline compared to 7.1 percent at midline, although decreases in non-learners among the comparison group were even more substantial.

**Table 39: SeGRA subtask 2 learning bands, cross-sectional sample**

All Students	Intervention		Comparison	
	ML	EL	ML	EL
Non-learner (0%)	7.1	3.2	10.5	0.0
Emergent learner (1% - 40%)	44.3	47.2	51.2	59.0
Established learner (41% - 80%)	38.6	46.6	34.4	37.9
Proficient learner (81% - 100%)	10.0	3.0	4.0	3.2

### Subtask 3 Outcomes

Before proceeding with supplementary analysis, we note, as in the main body of the report, that the high percent of zero scorers on this task is unlikely to reflect the reality of students' writing abilities.

The table below shows the difference-in-differences results for the panel dataset. We find a positive coefficient ranging from 5.2 for the model with no controls to 6.0 for the model controlling for student age and school, showing a potential positive impact of the programme on learning outcomes for intervention students relative to comparison students. However, as with all results thus far, these findings are not significant.

**Table 28: SeGRA subtask 3 difference-in-differences results**

Model	Intervention		Comparison		Diff-in-Diffs	P-value
	ML	EL	ML	EL		
No controls	42.7	41.7	39.3	33.1	5.2	0.30
Controlling for age	-	-	-	-	5.3	0.29
Controlling for age and school	-	-	-	-	6.0	0.24
No zero scores	45.2	49.0	42.9	42.5	4.2	0.28

Lastly, the below table shows learning bands for the cross-sectional cohort of all students assessed at either midline or endline. Results are broadly similar to those found for the panel sample, showing an increase in the percent of non-learners as well as proficient learners, and a decrease in emergent and established learners for both intervention and comparison groups.

**Table 40: SeGRA subtask 3 learning bands, cross-sectional sample**

All Students	Intervention		Comparison	
	ML	EL	ML	EL
Non-learner (0%)	1.4	21.3	1.1	20.5
Emergent learner (1% - 40%)	45.7	37.5	56.8	44.2
Established learner (41% - 80%)	51.4	35.7	41.6	30.0
Proficient learner (81% - 100%)	1.5	5.5	0.5	5.3

### Numeracy Outcomes

The below table presents a summary of overall scores for all endline and midline students. Among all endline students in intervention schools, the average overall mathematics score was 32.4 percent, an improvement of 1.3 percentage points over the average score of 31.1 percent for all intervention students at midline. Overall mathematics scores improved slightly for both intervention and comparison students, as did scores on subtask 1 for both groups and subtask 2 for comparison students. Scores for subtask 3, however, declined slightly from midline to endline for both intervention and comparison students. This may suggest that at endline, students had improved at basic mathematics skills but struggled with more advanced skills; this potential finding is explored in more detail below.

**Table 30: Summary of SeGMA scores, all learners**

	Intervention			Comparison		
	ML	EL	Diff.	ML	EL	Diff.
Number of respondents	903	1,265	-	548	190	-
<b>Overall score</b>	ML	EL	Diff.	ML	EL	Diff.
Mean score	31.1	32.4	1.3	24.5	28.1	3.6
Standard deviation	22.5	20.4	-	19.4	19.2	-
<b>Subtask 1</b>	ML	EL	Diff.	ML	EL	Diff.
Mean score	48.2	51.8	3.6	40.0	47.0	7.0
Standard deviation	29.1	26.0	-	26.5	25.3	-
<b>Subtask 2</b>	ML	EL	Diff.	ML	EL	Diff.
Mean score	32.1	32.1	0.0	24.4	27.9	3.5
Standard deviation	30.0	23.4	-	25.9	24.3	-
<b>Subtask 3</b>	ML	EL	Diff.	ML	EL	Diff.
Mean score	13.0	12.0	-1.0	9.0	8.0	-1.0
Standard deviation	17.4	20.7	-	13.7	16.1	-

The standard deviation for overall scores and subtasks was high; in particular, for subtask 3, standard deviations exceeded mean scores. This indicates that scores varied widely, with many students receiving scores far from the mean. In other words, our sample encompasses students with a wide range of mathematics abilities, including those who received very low scores and those who received very high scores.

The above results are not calibrated; calibrated results are presented in the below table. As with the SeGRA, subtask 3 is not individually calibrated due to low sample size for calibration. However, subtask 3 is incorporated into the overall score calibration.

**Table 41: Summary of calibrated SeGMA scores, all learners**

	Intervention			Comparison		
	ML	EL	Diff.	ML	EL	Diff.
Number of respondents	903	1,265	-	548	190	-
Overall score	31.1	36.5	5.4	24.5	31.6	7.1
Subtask 1 score	48.2	53.3	5.1	40.0	48.4	8.4
Subtask 2 score	32.1	38.3	6.2	24.4	33.3	8.9
Subtask 3 score	13.0	12.0	-1.0	9.0	8.0	-1.0

Difference-in-differences analysis is presented in the main body of the report. To test the robustness of these findings, we now add control variables to the regression model. We first control for differences in student age; after adding this control variable, we find little change in the overall regression results, with a

small, positive, and insignificant effect size of 2.2 percentage points. We also find a moderately positive relationship between age and SeGMA scores: For every one-year increase in age, we find a 2.0 percentage point increase in overall SeGMA score, controlling for intervention status and round.

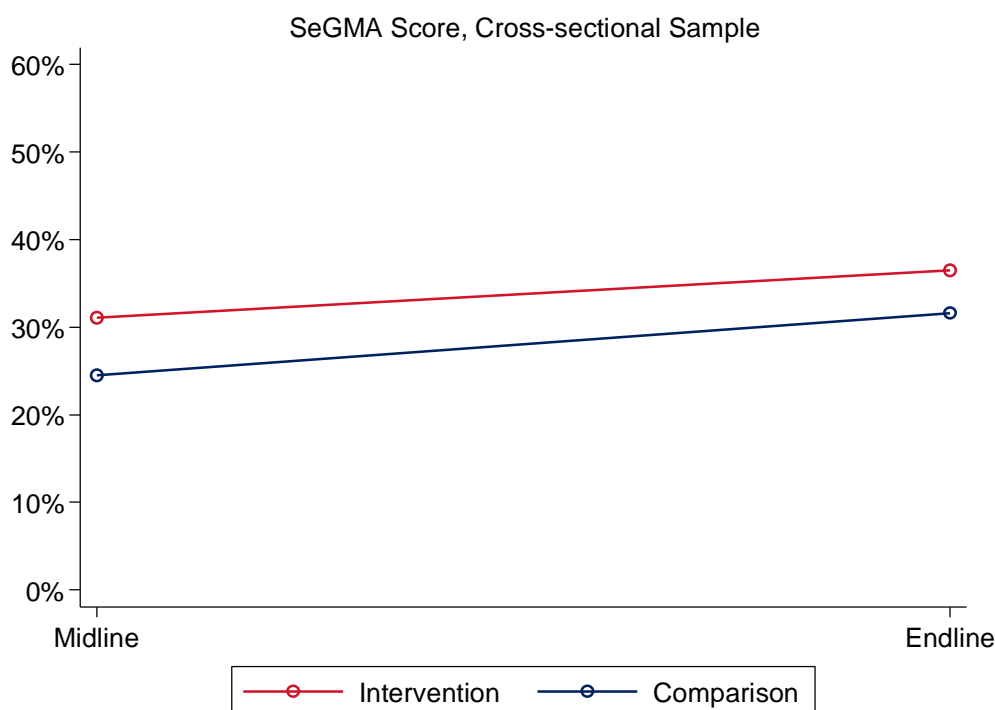
Adding school fixed effects to this regression to control for school- or county-level differences that may have biased results, we find a slightly lower effect size of 1.9. This value remains insignificant, as shown in the below table.

**Table 32: SeGMA difference-in-differences results**

Model	Intervention		Comparison		Diff-in-Diffs	P-value
	ML	EL	ML	EL		
No controls	34.2	41.1	28.2	33.0	2.2	0.46
Controlling for age	-	-	-	-	2.2	0.47
Controlling for age and school	-	-	-	-	1.9	0.55

To further analyse SeGMA results, we now use the cross-sectional sample of all students assessed at either midline or endline. The below figure shows the change in scores among comparison and intervention students, with no control variables added. We find that overall scores for both intervention and comparison students increased. Average scores increased slightly more, however, for comparison students than intervention students; at endline, intervention students' average scores fell 1.8 percentage points below the scores that would be expected from results in comparison schools. This difference, however, was insignificant.

**Table 42: Change in SeGMA scores, cross-sectional sample**



As with the cross-sectional analysis for the SeGRA, we now add control variables for age and school, as well as weights to account for sampling differences between project beneficiaries and the midline and endline evaluations. Controlling for just age, we find a very small and insignificant positive treatment effect. Controlling for school fixed effects, the size of this treatment effect increases slightly, to 1.4 percentage points, but remains insignificant. Inclusion of weights again reduces the size of the treatment effect; in this model, intervention students scored 0.4 percentage points higher than expected given the results of the comparison group, all else held constant. These results are again insignificant.

As with the SeGRA, because the cross-sectional sample faces significant flaws for analysis and because results do not widely vary from the difference-in-differences analysis, we proceed with the panel sample for the remainder of the analysis.

Lastly, the below table shows learning bands for the cross-sectional sample. Similar to the panel sample, we find that there was an increase in the percent of students classified as proficient learners and a decrease in the percent of emergent learners among both intervention and comparison students. These increases/decreases did not vary substantially depending on whether a student benefitted from WWW interventions or not, suggesting that the results may rather represent a general rebound in mathematics learning post-COVID.

**Table 43: SeGMA learning bands, cross-sectional sample**

All Students	Intervention		Comparison	
	ML	EL	ML	EL
Non-learner (0%)	2.3%	1.8%	2.3%	1.6%
Emergent learner (1% - 40%)	63.9%	57.0%	76.6%	66.3%
Established learner (41% - 80%)	29.6%	35.5%	18.7%	27.9%
Proficient learner (81% - 100%)	4.3%	5.7%	2.3%	4.2%

### Subtask 1 Outcomes

The below table presents the results of regression models with and without control variables. Among just the panel sample, the average subtask 1 score for intervention students improved by 4.8 percentage points from midline to endline, while the average score for comparison students improved by 6.0 percentage points. This means that comparison students did indeed improve their subtask 1 performance relative to intervention students; in other words, intervention students underperformed by around 1.1 percentage points relative to what would be expected given the results of comparison students. These results, however, are not significant.

**Table 34: SeGMA subtask 1 difference-in-differences results**

Model	Intervention		Comparison		Diff-in-Diffs	P-value
	ML	EL	ML	EL		
No controls	52.4	57.2	44.6	50.6	-1.1	0.79
Controlling for age	-	-	-	-	-1.1	0.80
Controlling for age and school	-	-	-	-	-1.6	0.74

The addition of control variables for age and school has little effect on the difference-in-differences coefficient, which remains small and negative. The coefficients in these regression models are also not significant. Overall, due to the very high p-values and small coefficients, no definitive conclusions about programme impact on subtask 1 performance can be drawn from these results.

Looking now at learning bands for the cross-sectional sample, we find generally similar patterns to the panel sample reported in the main body of the report, although there is an increase in the number of proficient learners among comparison students at endline. As among the panel sample, however, we generally find a consolidation of learners to the emergent or established levels, showing a potential improvement of the weakest learners but also a potential decline in the strongest learners.

**Table 44: SeGMA subtask 1 learning bands, cross-sectional sample**

All Students	Intervention		Comparison	
	ML	EL	ML	EL
Non-learner (0%)	2.9%	2.5%	3.1%	1.6%
Emergent learner (1% - 40%)	40.7%	32.8%	50.3%	41.6%
Established learner (41% - 80%)	35.4%	45.8%	34.4%	43.2%
Proficient learner (81% - 100%)	20.9%	18.9%	12.3%	13.7%

### **Subtask 2 Outcomes**

The below table shows the difference-in-differences analysis for the panel sample. In contrast to subtask 1, we find a small positive improvement in the intervention group compared to the comparison group, with intervention students' calibrated scores improving by 10.4 percentage points while comparison students' calibrated scores improved by only 7.1 percentage points. These results, however, are not significant. Results remain similar and insignificant when controls for age and school are included.

**Table 36: SeGMA subtask 2 difference-in-differences results**

Model	Intervention		Comparison		Diff-in-Diffs	P-value
	ML	EL	ML	EL		
No controls	35.6	46.0	27.9	35.0	3.2	0.33
Controlling for age	-	-	-	-	3.1	0.34
Controlling for age and school	-	-	-	-	3.0	0.38

Lastly, the table below shows learning bands for the cross-sectional sample. These findings are broadly aligned with those for the panel sample, except for a decline in the percent of proficient learners among intervention students.

**Table 45: SeGMA subtask 2 learning bands, cross-sectional sample**

All Students	Intervention		Comparison	
	ML	EL	ML	EL
Non-learner (0%)	24.0%	13.4%	29.2%	17.9%
Emergent learner (1% - 40%)	41.7%	40.7%	45.2%	44.7%
Established learner (41% - 80%)	24.0%	37.6%	20.2%	30.0%
Proficient learner (81% - 100%)	10.4%	8.2%	5.4%	7.4%

**Subtask 3 Outcomes**

The below table shows difference-in-differences results for the panel sample of students. We note that scores for this subtask are not calibrated; four out of eight students during the calibration exercise scored 0 on this subtask, resulting in an insufficient sample size for calibration.

The table shows a much more substantial decline in scores for comparison students than intervention students. While this difference is not significant at a 5 percent level (our benchmark for significance), it is significant at the 10 percent level (i.e.,  $p < 0.1$ ) for our regression model without controls and the model that controls for age, as shown in the table below.

**Table 38: SeGMA subtask 3 difference-in-differences results**

Model	Intervention		Comparison		Diff-in-Diffs	P-value
	ML	EL	ML	EL		
No controls	14.4	14.3	12.0	8.1	3.9	0.08
Controlling for age	-	-	-	-	3.9	0.09
Controlling for age and school	-	-	-	-	3.6	0.13

Looking at learning bands, for the cross-sectional sample, we find increased percentages of non-learners at endline for both intervention and comparison groups. However, as with the panel sample, we also see increasing percentages of established and proficient learners, suggesting a widening of learning outcomes whereby weaker students are falling behind relative to stronger students.

**Table 46: SeGMA subtask 3 learning bands, cross-sectional sample**

All Students	Intervention		Comparison	
	ML	EL	ML	EL
Non-learner (0%)	44.5%	57.2%	53.0%	65.3%
Emergent learner (1% - 40%)	45.5%	30.3%	42.3%	27.4%
Established learner (41% - 80%)	7.7%	10.0%	2.9%	6.3%
Proficient learner (81% - 100%)	2.3%	2.6%	1.8%	1.1%

## Predictive Analysis of Learning Outcomes

Given the above findings for subgroup learning outcomes, we now seek to understand factors which are predictive of learning outcomes. This analysis can help untangle the effects that intersecting forms of marginalisation may have on learning. For example, in the above analysis, we find that learning outcomes are relatively lower for students in ASALs compared to those in urban slums. However, it is possible that this could occur because students in ASALs spend, on average, more time on household chores than students in urban slums, or are relatively younger. The predictive analysis seeks to better understand such interacting effects.

To do so, we utilise hierarchical linear modelling (HLM) to determine the extent to which student- and school-level characteristics determine variability in learning outcomes. HLM is a form of ordinary least squares regression used to analyse variance in outcomes variables when the predictor variables are at varying hierarchical levels. In our case, these hierarchical levels refer to the common teacher, classroom, and school that some students share, all of which impact their learning outcomes.

Our model includes five student-level predictors analysed above: age, whether the student speaks a minority language, disability status, whether the head of household is female, and time spent on housework. We do not include orphan status as a predictor as it is somewhat collinear with female HoH. Additionally, we include two variables measuring characteristics more inherent to the student: their life skills score and their decision-making score. These scores were calculated from relevant questions in the girl survey; life skills generally measures the girl's confidence, desire to do well in school and life, and ability to plan for the future, while decision-making measures the girl's agency in making decisions about her own life. In general, we would expect girls with stronger life skills and decision-making abilities to also have better learning outcomes.

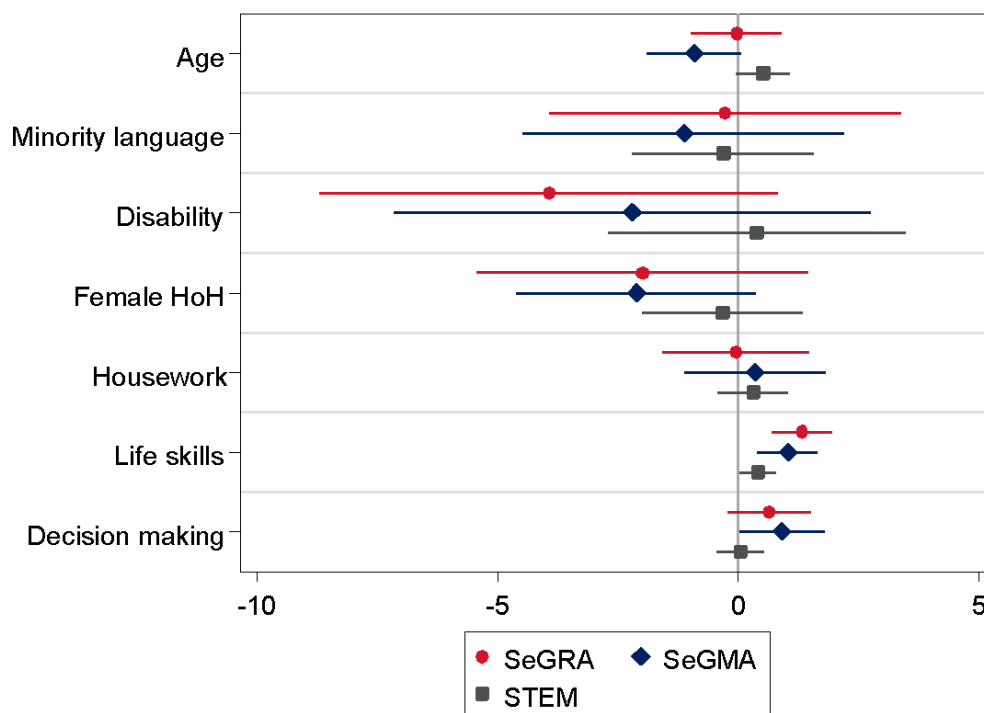
In our first model, which seeks to identify individual-level predictors of learning outcomes, we also include school fixed effects. This HLM technique allows us to control for any variation which occurs at the school level and which does not change, or changes at a constant rate, over time. This model, however, does not allow us to understand the school-level predictors of learning outcomes, but it is the most robust model available to understand individual-level predictors. In a subsequent model, we exclude school fixed effects and include school-specific variables to better understand these predictors.

For our model, we utilise data from endline. Figure 5 below shows regression results for the SeGRA, SeGMA, and STEM assessments; specific values are included in a below table. Interestingly, we find very few significant relationships between learning outcomes and the included individual-level predictors. For the SeGRA, we find a significant relationship only between life skills and literacy scores; as students' life skills scores increase by one point, their SeGRA scores also tend to increase by an average of 1.3 percentage points, all other factors held constant. For the SeGMA, similarly, we find a significant and positive relationship between life skills and mathematics scores; we also find a positive relationship between decision-making abilities and mathematics scores. For the STEM assessment, we similarly only find a significant and positive relationship between life skills and assessment scores.

These findings suggest that demographic characteristics may actually have relatively less effect on learning outcomes than the previous section would suggest. It is important to emphasise, however, that findings may be interrelated. For example, a student who speaks a minority language or who is disabled may feel less confident participating in class, and thus have both a lower life skills score and weaker learning outcomes. Indeed, in further regression analysis, we find significant and negative relationships between life skills and minority language and disability status, suggesting that this may be the case. Furthermore, when life skills and decision-making are dropped from the above regression model, we find a significant and negative relationship between SeGRA scores and disability status: Disabled students score, on average, 4.5 percentage points lower on the SeGRA than their peers without disabilities, all other factors held constant.



**Figure 5: Predictors of learning outcomes, individual level**



While we cannot fully untangle causality of learning outcomes, we note that the inconclusive relationship between factors expected to result in lower learning scores—disability, minority language, and female HoH in particular—is potentially a positive finding for two reasons. First, this finding emphasises that students with disadvantaged backgrounds do not inevitably face poorer learning outcomes. If given appropriate support at the individual, household, and school level—and perhaps in particular, as the findings above suggest, support to strengthen life skills—then these students are fully capable of succeeding in school.

Second, this finding suggests that WWW programme support may have helped improve learning for the most marginalised students. While this is by no means conclusive, the findings above for endline students could be explained if WWW programming has been particularly successful in targeting disabled, minority, and otherwise disadvantaged students, thus increasing their learning scores relative to their more-advantaged peers. Indeed, running the same regression model<sup>31</sup> on SAS data, we find a significant and negative relationship between minority language and SeGRA scores and a near-significant ( $p = 0.051$ ) and negative relationship between minority language and SeGMA scores. In line with the results reported in Table 45, this may suggest that the WWW programme has helped these students catch up to their peers over the past year.

In order to now better understand school-level factors which may predict learning outcomes, we run a model including the above individual-level predictors and four school-level variables: teacher attendance, student-teacher ratio, the number of school clubs, and whether the school was a WWW intervention school.<sup>32</sup> The number of school clubs is included as a proxy for the resources available to students within

<sup>31</sup> Excluding housework, for which there is no comparable SAS data.

<sup>32</sup> Not included for the STEM assessment model, as these assessments were only administered with intervention students.

schools; school clubs may also help strengthen learning outcomes directly, if the clubs address topics of relevance to learning assessments, or indirectly by strengthening students' life skills, such as time management. We note that we do not include any variables representing teaching quality, as our measures of teacher quality occurred at the classroom observation level and thus may not apply to all students in a school. Lastly, we note that we do not include school fixed effects in this model, as these would be colinear with the school-level predictors of interest. As a result, this model is less rigorous for determining individual-level predictors; we thus only report the values of school-level predictors below.

**Figure 6: Predictors of learning outcomes, school level**

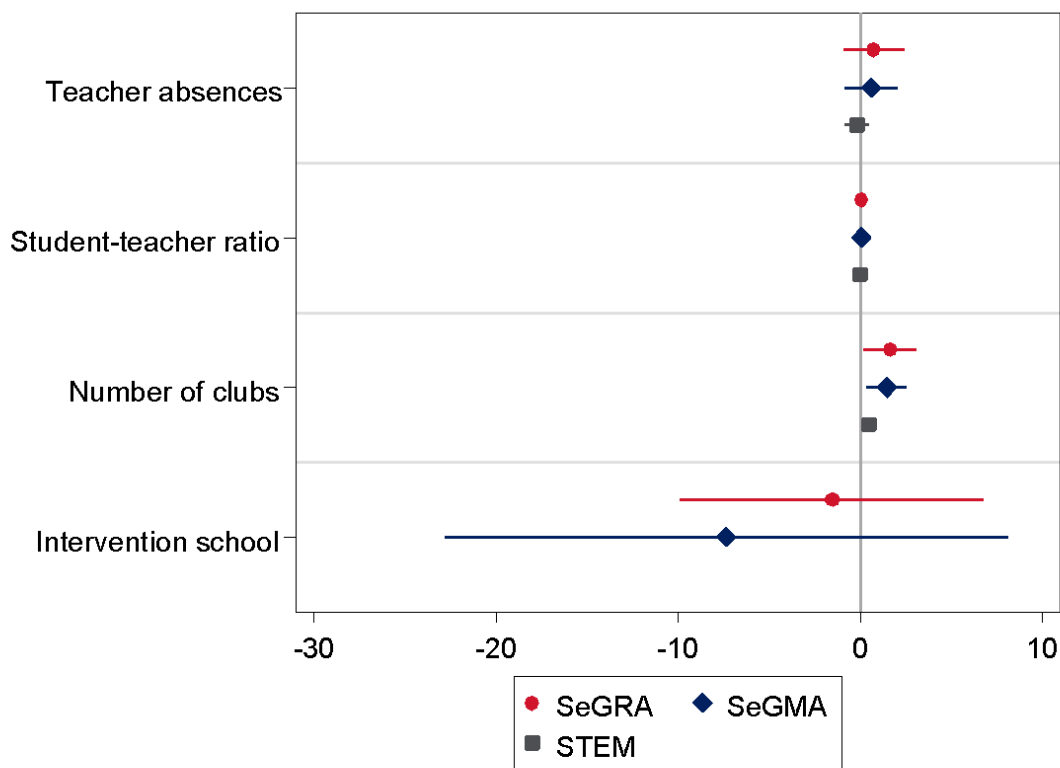


Figure 6 shows that using this model, we find no significant relationship between teacher absences or student-teacher ratio and learning outcomes. However, we find a significant and positive relationship between the number of school clubs and results on all three learning assessments. We note that we also do not find a significant relationship between the number of school clubs and teacher absenteeism or student-teacher ratio; in other words, the relationship shown in the figure below is not simply a result of collinearity. Rather, these findings may suggest that the level of resources available to students within a school have a strong effect on learning outcomes.

To summarise, the below table shows all regression results for predictors of learning outcomes. We find that life skills are a significant predictor of all assessment scores, with a higher life skill score associated with improved learning outcomes. We also find that the number of school clubs is associated with significantly higher scores on all assessments; this may be because schools with more clubs have more resources for students, or because students benefit substantially from participation in clubs.

**Table 52: Summary of predictors of learning outcomes**

Individual Predictor	SeGRA		SeGMA		STEM	
	Coef.	P-value	Coef.	P-value	Coef.	P-value
Age	-0.03	0.95	-0.9	0.07	0.5	0.07
Minority language	-0.3	0.89	-1.1	0.51	-0.3	0.75
Disability	-3.9	0.11	-2.2	0.38	0.4	0.80
Female HoH	-2.0	0.26	-2.1	0.10	-0.3	0.70
Housework	-0.04	0.96	0.4	0.63	0.3	0.39
Life skills	1.3	<0.001	1.0	0.002	0.4	0.04
Decision-making	0.7	0.14	0.9	0.04	0.06	0.82
<b>School Predictor</b>						
Teacher absences	0.8	0.37	0.6	0.38	-0.1	0.67
Student-teacher ratio	0.1	0.58	0.08	0.58	0.001	0.99
Number of clubs	1.7	0.03	1.5	0.01	0.5	0.02
Intervention	-1.5	0.72	-7.3	0.35	-	-

## Annex 17: Supplementary Analysis of Other Primary Outcomes

### Transition Outcomes

Table 47 documents potential pathways girls can take, and what is classified as a successful or unsuccessful transition.

**Table 47: Transition pathways**

	Grades	Successful Transition	Unsuccessful transition
Upper primary	Enrolled in Grade 7 or 8	<ul style="list-style-type: none"> <li>- In-school progression</li> <li>- Moves into secondary school</li> </ul>	<ul style="list-style-type: none"> <li>- Drops out of school</li> <li>- Moves into work, but is below legal age</li> </ul>
Secondary school	Enrolled in Form 1-4	<ul style="list-style-type: none"> <li>- In-school progression</li> <li>- Enrols into technical &amp; vocational education &amp; training (TVET)</li> <li>- Gainful employment if of legal age</li> </ul>	<ul style="list-style-type: none"> <li>- Drops out of school</li> <li>- Moves into work but is below legal age</li> <li>- Moves into employment, but is paid below minimum wage</li> </ul>
Out of school (age 10 to 18)	Dropped out	<ul style="list-style-type: none"> <li>- Re-enrol in appropriate grade level in basic education</li> <li>- Enrols into TVET</li> </ul>	<ul style="list-style-type: none"> <li>- Remains out of school</li> </ul>

The sample analysed for transition outcomes consists of the Cohort 2 girls (including the Marsabit and Kwale girls), as well as Cohort 3 girls from the alternative pathways sample which will be used as supplementary data. The Cohort 2 sample includes 933 girls, which goes up to 1,076 girls when including the Cohort 3 girls.

In this annex, we first discuss in more detail limitations to the transition rate of the SAS sample, which was very high at 99.7 percent. These results may be misleading, because the SAS sample is the largest sample in this analysis and skews the results to the right for a number of reasons. Firstly, the SAS took place in December 2021 and therefore, only one year has passed since this study. The SAS sample was made up solely of in-school girls at the time and it is likely that in that one-year time span not many girls would have dropped out of school. Secondly, the endline SAS sample is mostly made up of girls that we were able to track at the schools, which means that there is some bias in our results. Although we attempted to follow up with girls who we could not find at the school, we were only able to find 933 girls of the 1,469 in the original sample. The other girls may have moved to other schools which were not part of the sample, or may have dropped out of school. Therefore, the aggregate transition rate may not give the true transition rate due to the limitations of the data.

Of the 536 girls who we were not able to survey, we were able to find information about 192 of the girls by speaking to the community, family members, or the school. Out of these 192 girls, we found that 73 percent of them had moved to different schools or graduated from high school. We could only confirm that 8 percent of girls had dropped out; girls who dropped out did so for a variety of reasons including to begin working, getting married, due to differences with their teachers, or due to sickness. The remaining girls who we were able to find information about had moved away from the county with their families or to live with other family members, but we could not confirm whether they had transitioned into a new school or not. It is possible that many of the girls who we were not able to interview have transitioned to other non-intervention schools; however, it also suggests that the SAS sample's transition rate is somewhat lower than what the table above suggests, as we were not able to reach many of the girls who had dropped out.

However, it is important to reiterate that we were not able to interview these girls and the data is somewhat anecdotal as it does not come directly from them. Therefore, it is not possible to make definitive conclusions about their whereabouts or transition outcomes.

### ***Subgroup Transition Rates***

Subgroup transition rates for girls in the SAS-ML sample are reported in the main body of the report. When this data is supplemented with data from girls who we were able to track but not survey, the results remain similar as transition rates for most characteristics are below the average. Interestingly, the counties of Turkana and Nairobi had the lowest transition rates at 89 percent and 91 percent respectively. The data does not provide a definitive reason for this; in FGDs with girls from Turkana and Nairobi, all girls expressed desire and intention to transition to secondary school and later attend university. However, teachers in Turkana and Nairobi noted several common challenges to girls' attendance and transition. In Nairobi, all seven interviewed teachers stated that family issues, such as coming from a single parent household, and economic issues, which could lead families and students to struggle to pay school fees, made it more difficult for some students to continue in school. In Turkana, all three interviewed teachers similarly noted that lack of household and community support reduced transition rates. For example, one teacher stated that:

*One of the challenges we face is a cultural problem within this community. Whenever they see a teenage girl in school they discourage her and tell her she should be at*

*home or be married off. So that cultural practice has affected adolescent girls in this school.*<sup>33</sup>

The other noticeable characteristic was that the transition rate for male headed households was much lower than that of female headed households (92 percent and 98 percent respectively). Although this goes against conventional theory, this could be due to the fact that a large majority of the households in our sample were headed by men. Furthermore, neither of these differences were statistically significant.

### ***Predictive Analysis of Transition Outcomes***

In this section, we will test the relationship between various student- and school level-characteristics and the transition rate of girls.<sup>34</sup> The goal of this section is to assess whether the programmes outcomes are correlated with transition rates. For example, we may expect that higher life skills score by girls will be correlated with higher transition rates, all else equal. We first analyse the individual-variables such as the girls' age and then analyse variables measured at the school level such as teacher attendance.

At midline only the education level of the caregiver had a statistically significant relationship with the transition rates of girls. In our model in this endline evaluation we have identified five student level predictors including age, life skills score, decision making score, disability status, and minority language speaker.<sup>35</sup> The results of this regression are displayed in Table 48. According to this analysis none of the predictor variables have a significant relationship with transition rates. However, this null effect is very likely to be due to the very high transition rates.

To better understand the relationship between school level predictors and transition outcomes we run a second model including the same individual level variables and 5 school level variables: number of trained teachers at the school, number of lessons taught by teachers in a week, teacher attendance, presence of children's clubs at the school, and support by remedial teachers after the COVID pandemic. When running Model 2 the relationship between predictors remains weak and statistically insignificant.

Finally, a third model is used using data from the caregiver survey. This model has a smaller sample as we only use those respondents for which we have both girl survey data, as well as caregiver data. This model uses the same variables as used in Model 2 but also includes a number of variables from the caregiver survey including the education level of caregivers, female head of households, and if the caregiver is unemployed. Although the relationships between caregiver characteristics and transition rates are what we would theoretically expect, i.e., caregivers with higher education predict a higher transition rate and female head of households predict a lower transition rate, the results show weak and statistically insignificant relationships between the predictors and transition rates.

---

<sup>33</sup> KII with teacher, Turkana, Int. 42

<sup>34</sup> As in other sections, we utilize HLM for the predictive analysis.

<sup>35</sup> The model also includes a school fixed effects control variable.

**Table 48: Predictive analysis with individual-level predictors**

Variables	Coefficient	p
<b>Model 1: Individual level characteristics</b>		
n	924	
Age	<0.001	0.720
Disability status	0.013	0.118
Life Skills Score	0.004	0.245
Decision-making score	<-0.001	0.931
Minority Language speaker	<0.001	0.560
<b>Model 2: School characteristics</b>		
Number of trained teachers at the school	<-0.001	0.536
Average number of lessons taught by a teacher a week	<0.001	0.092
Average teacher attendance	0.013	0.101
School has children's clubs School received support from remedial teachers	0.011	0.482
School has children's clubs	0.025	0.182
<b>Model 3: Individual characteristics (Caregiver Data)</b>		
n	522	
Caregiver Education	0.005	0.391
Female head of household	-0.01	0.353
Unemployed caregiver	-0.002	0.451

A second predictive analysis was also used to test the relationship between repetition rate and the same predictor variables. We find that the relationship between individual-level variables remains weak and statistically insignificant. However, school level predictors have a strong and significant relationship with the repetition rates of girls. The strongest relationship found in the analysis was that between the teacher attendance and repetition rate.<sup>36</sup> The positive relationship suggests that the more often teachers miss school the more likely girls are to repeat their grade. This is to be expected as teachers' attendance is key for girls' learning and substitute teachers are not always available at these schools. Therefore, if the teacher is not able to attend, girls may go without guidance in class that day.

Furthermore, the number of trained teachers<sup>37</sup> and number of lessons taught by teachers a week<sup>38</sup> also had statistically significant relationships with the dependent variables, although the magnitude of the relationships was somewhat weaker. These results suggest that more trained teachers at the school leads to a lower repetition rate. This is to be expected as these teachers should be able to use better and more child-friendly teaching methods than those who have not been trained. Secondly, there is a positive

<sup>36</sup> Coefficient = 0.26, p < 0.001.

<sup>37</sup> Coefficient = -0.014, p < 0.001.

<sup>38</sup> Coefficient = 0.022, p < 0.001.

relationship between the number of lessons taught a week and higher repetition rate. Again, this is unsurprising as these teachers may be overworked and therefore are less able to implement individualized teaching practices into their classes.

### **Alternative Learning Pathways**

Table 49 shows the results of the survey on statements related to their ALP courses as well as where they had learned about the programme and their future hopes. The third column shows the percentage of respondent who either agreed or highly agreed with the statements. Overall, the quality of the TVET programmes seem to be high from the perspective of the girls. Over 90 percent of respondent agreed that instructors were able to explain lessons and answered learners' answers. Although we are not able to assess the quality of TVET programmes in previous rounds, this analysis suggests that the current TVET programmes are of high quality from the girls' perspectives. However, findings also suggest that there is room for improvement for these programmes. The data suggest that it was still common for instructors to be absent from class and for teachers to discriminate against disabled students. Nonetheless, the majority of respondents believed that the programme was preparing them for work and would allow them to obtain a job once they graduated. This is key as it shows girls understand the doors that can be opened after completing TVET courses and the benefits this can have. Furthermore, there is evidence that the programme is preparing girls well for the labour market

**Table 49: Alternative learning pathways quality**

Statement	Response (%)
My teachers/instructors explain the lessons well in class	92%
My teachers/instructors respond to learner's questions well in class	96%
My instructors are often absent from class	37%
My teachers/instructors do not discriminate against students with functional difficulty	67%
I believe my current programme is preparing me well for work	94%
Having completed the programme will help me secure a job	98%
I would have joined the programme if I hadn't learned about it from the WWW project	39%

## **Sustainability Outcomes**

Our methodology for calculating scores differs somewhat from midline and baseline. In these evaluation rounds, informants were asked to self-report scores for various components with reasons for their rating. We assess this method as highly vulnerable to significant and unpredictable bias. Respondents may, for example, be incentivised to rate components overly high in order to make their organisation, school, or community appear more favourably to evaluators or the WWW programme. In contrast, respondents may also be incentivised to rate components overly low, in an attempt to solicit further support from the programme. As such, self-ratings are a highly unreliable means of assessing progress on sustainability.

We instead utilise, where possible, less subjective methods to measure indicators that do not rely on respondents self-rating progress. This includes the use of verifiable data where it exists; for example, rather than requesting that respondents rate levels of participation in community action plans, we directly asked caregivers about their awareness of these plans and, for those who were aware, the efficacy of

plans. Where possible, we also recalculate midline values using the same methodology as endline; however, due to lack of availability of girl and caregiver survey data from baseline, baseline values are not recalculated. As such, baseline values are not directly comparable to endline values, and are included just as an indication of the initial status of indicators. The specific methods used, and any implications for the comparability of data across rounds, is described in relevant subsections in the main body of the report.

## Community Level

### **Community Action Plans**

Caregivers most frequently reported attending one, two, or three CCs, although nearly half of caregivers expressed that they did not know how many CCs they had participated in. Respondents reported that CC groups were relatively well-attended, with a median of 20 participants.

Respondents described a variety of topics discussed during the CCs; the most frequent topics included general support for girls’ education, mentorship of girls in school, empowerment of girls, support for girls to remain in school or return to school (for those who have dropped out), and support for girls to buy menstrual products.

### **Support for Girls’ Education**

Examining average community support scores by county in the below table, we find that at endline, levels of support were lowest in Turkana (with an average score of 2.0 out of 4) and highest in Samburu (with an average score of 2.8 out of 4). This difference was significant. In general, we do not find significant differences in community support for education by county groupings (i.e., ASALs and urban slums), with two ASAL counties having the highest and lowest average scores. As a result, it is difficult to draw clear conclusions about these findings.

**Table 50: Community support for girls' education, by county**

	Kilifi	Kwale	Marsabit	Mombasa	Nairobi	Samburu	Tana River	Turkana
n	162	32	9	28	359	59	94	71
Score	2.3	2.6	2.6	2.2	2.5	2.8	2.6	2.0

The qualitative data does not fully explain the low results found for Turkana, but sheds light on some barriers to support for girls’ education in this county. All three interviewed teachers in Turkana described limitations to parental involvement in education; one teacher, for example, stated that, “When you have students who don’t perform well, and the administration calls for a meeting to discuss student issues, only a few parents will turn up”. This teacher also stated that some parents are not willing to educate their girls, and instead ask them to get married so they can receive a dowry.<sup>39</sup> In contrast, in Samburu—a comparable ASAL county with a much higher score—both interviewed teachers stated that parents were supportive and active, although economic challenges limited the extent of support provided.

<sup>39</sup> KII with teacher, Turkana, Int. 5



## Annex 18: Supplementary Analysis of Intermediate Outcomes

### Attendance

#### ***Methodological Considerations***

The first part of our analysis on attendance, which relied on head count measures, must be caveated by the fact that the attendance rates calculated through this methodology have several limitations. Notwithstanding the limitations of observation in a single point in time, a major limitation is that cross-referencing headcounts against formal enrolment records was only done for Classes 7 and 8 students, whereas for Form 1 and above, only formal records were checked as part of this and the SAS evaluations, meaning attendance rates were not calculated for the latter. This can have significant implications on findings. As grade level tends to correlate with age, the observed trends above are not representative of all age groups and ability levels. Indeed, in our sample, the mean age of girls in Class 7 was 13.7 years, whilst the Form 1 mean age was 19.7.

On one hand, should (on average) older Form 1 and above learners be included in the analysis above, the findings might reveal even more dramatic declines in attendance (and indeed, retention in schools), as older adolescent learners may be able to be involved in a wider range of income-generating activities than their younger peers, which might make a higher proportion of learners sacrifice school-time in favour of work. The reverse – increasing attendance rates – might also be true. More committed or capable learners might be the ones who stay on to higher and more difficult grade-levels, as their peers may have unenrolled from school entirely during earlier grade levels or be. In this case, we might expect attendance rates at higher grade levels to be higher than at lower grade levels (and potentially improving) due to the selection of more committed or capable learners into such grade-levels.

Gendered differences in attendance rate changes may also be affected by this methodological limitation. Recall that boys registered steeper declines in their attendance rates in our analysis using this measure of attendance. However, this gendered finding might not be applicable to older cohorts. Multiple qualitative respondents have cited pregnancies (and often marriages by extension) as some of the reasons why girls might not be attending school regularly, as these girls experience a precipitous jump in their household responsibilities and/or substantial social stigma from peers and instructors.<sup>40</sup> While early pregnancies and marriages are still serious concerns in the geographic areas of focus in this evaluation, it is reasonable to expect that younger learners in Classes 7 and 8 (on average) are still less likely than their older peers to be sexually active or to be married. Incorporating attendance data for Form 1 and above may reveal lower attendance rates and steeper declines between rounds for girl learners compared to boy learners; alternatively, such girls might register steeper increases if the challenges posed by early pregnancies and marriages were tackled more effectively in the context of the WWW programme.

#### ***Predictive Analysis of Attendance Outcomes***

To test for whether certain characteristics can serve as predictors of attendance outcomes, we further ran a number of statistical tests with the attendance outcome as the dependent variable. In this instance, we primarily rely on attendance as indicated by caregivers, due to the possible limitations with the head counts. Predictive analysis using the headcount data are included in Annex 18. In this sub-section, we specifically use the likelihood that caregivers assess their dependents as having attended most days of school since re-opening, as well as the likelihood that caregivers report improved attendance over the past year as our dependent variables.

---

<sup>40</sup> FGD with Girls, Mombasa, Int.19; FGD with Girls, Mombasa, Int. 20.

**Figure 7: Effect of Individual-level predictors on assessed attendance**

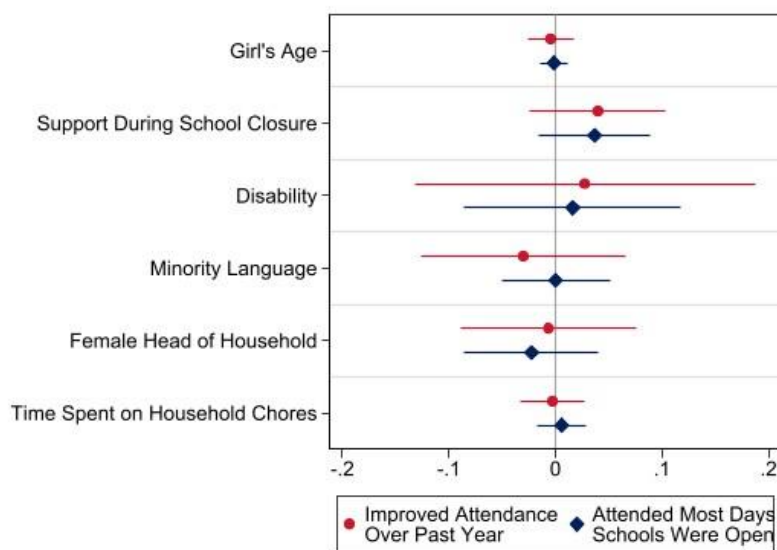


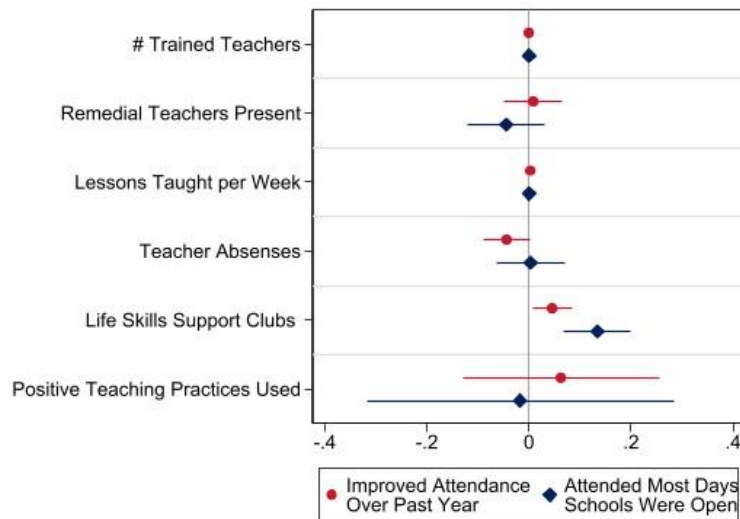
Figure 7 highlights the relationship between several individual-level predictors on both sets of assessed attendance outcomes, as judged by caregivers, after controlling for school-level predictors and school fixed effects. The red and blue points show the correlation coefficient for the relationship, while the lines show the confidence interval. For example, the above figure shows a positive relationship (controlling for other factors) between disability and attendance rates with a correlation coefficient of around 0.02, but this effect is not significant, as the confidence interval crosses 0.

Generally, most predictors affect outcomes in sensible ways, though most are not significant and have small effect sizes. Additional support provided during COVID-19 increases the likelihood of better attendance, though our previous analysis when using the school survey measure and the fact that the coefficient in this model is not significant suggests that this finding in the current model is not conclusive. A girl's primary language being a minority language also has a negative effect on one measure of attendance, as the language discrepancy might mean that they are less likely to be able to keep up with lessons in schools and are thus more at risk of missing classes, though we note again the small effect size and lack of statistical significance precludes us from making definitive claims.

Interestingly, having a female-headed household is associated negatively with attendance outcomes, though this is not statistically significant. Our qualitative data do not indicate whether female heads of households are necessarily more or less willing to support girls in their educational pursuits. However, there is an existing body of research suggesting that female-headed households tend to allocate more of the household's resources on children's needs compared to male-headed households, but that this sometimes means allocating more resources to boys rather than girls.<sup>41</sup> This is especially true of female household heads who are single, as they see their male children as having more potential to be breadwinners and support the household in the future. In turn, this might mean fewer resources devoted to girls, including for education fees or general consumption that is necessary for girls to attend school and learn.

<sup>41</sup> Theophilline Bose-Duker, Michael Henry, and Eric Strobl. (2021). "Children's Resource Shares: Male Versus Female Headed Households," *Journal of Family and Economic Issues* 42, 573 – 585.

**Figure 8: Effect of school-level predictors on assessed attendance**



Using the same two models as the above analysis of individual-level predictors, but with the elimination of the school fixed effects, we find that the presence of clubs that facilitate the development of girls' life skills has a significant and positive impact on attendance outcomes. Girls may find these clubs to be useful in imparting valuable skills which girls can utilise in their daily lives outside of school. Similarly, extracurricular clubs may be an avenue for girls to explore topics about which they are passionate, but which may not be offered in the formal curricula of schools. In turn, the opportunities to explore other topics or activities in which girls are interested may provide them with additional motivation to attend school, as the benefits these clubs offer can only be experienced in school premises. Though not statistically significant, we also observe a positive relationship between attendance and the use of positive teaching practices, such as skills on which the WWW programme trained teachers, with the effect of using positive teaching practices on attendance improvements being the second biggest coefficient in this model. The correlation is logical: as teaching quality increases, so too might student willingness and ability to remain focused on their studies.

One final note worth mentioning is that the disbursement of bursaries may help increased attendance. The total number of caregivers in our sample who claimed to have received financial grants from WWW is small, which necessitates excluding this variable in the multivariate model. While a regression on whether a household received grants, along with a school fixed effects control, did not result in a statistically significant association with both measures of attendance, some qualitative data point to a positive relationship, with one teacher claiming that "attendance has increased because of the bursaries that were given in Galole,"<sup>42</sup> while many others claimed that financial barriers posed a formidable challenge to regular attendance, even enrolment.<sup>43</sup> This finding thus suggests that financial grants and bursaries could have a positive effect on attendance if scaled.

### ***Predictive Analysis Using Headcount Data***

In this subsection, we use four different models for our predictive analysis: school survey attendance rates for 1) boys, 2) girls, and for 3) all genders as the dependent variable. A fourth model, measuring girls' attendance but with additional individual-level factors incorporated, is also included in the analysis.

<sup>42</sup> KII with Teacher, Tana River, Int. 19

<sup>43</sup> FGD with Girls, Tana River, Int.17

The first three models are intended to determine whether school factors might have statistically significant predictive power for student attendance, and whether this differs for girls and boys. Specifically, the predictors utilised in the models are: 1) number of trained teachers in the school, 2) use of positive teaching techniques, 3) whether the school received remedial teaching support, 4) average number of lessons teacher delivers a week, 5) frequency of teacher absenteeism, and 6) whether the school has extracurricular clubs that support girls' life-skills development. An additional county fixed effects control was added to account for differences between counties. However, because individual-level predictors were only collected through surveys with girls and their caretakers, these models which include boys do not account for individual-level factors when running the regression. The table below summarizes the main results of the tests.

**Table 51: School-level predictors of attendance rates of girls and boys**

	Model 1: Girls		Model 2: Boys		Model 3: All Genders	
	Coef.	Coef.	Coef.	Coef.	Coef.	P-Value
# of trained teachers	0.00	0.15	0.01	0.06	0.00	0.05
Remedial teaching support given	-0.03	0.37	-0.02	0.70	-0.03	0.45
# of lessons taught per week	0.00	0.67	0.00	0.80	0.00	0.61
Teacher absenteeism	-0.02	0.63	0.02	0.57	0.00	1.00
Presence of life-skills development club	0.06	0.22	0.01	0.86	0.02	0.64
Use of positive teaching practices	0.19	0.31	0.20	0.20	0.19	0.19
County fixed effects	Omitted from presentation					

Of these predictors, only the number of trained teachers in the school registered a statistically significant result at at least the 90 percent confidence level. This is theoretically logical, as a higher number of trained teachers means increased capacity to attend to learners' needs, thus reducing the likelihood of their missing school. However, two important notes are worth emphasising. First, the coefficient remains small, meaning that the practical effect on attendance for each additional trained teacher in the school is small. However, given the previous analysis suggesting that attendance rates already hover around the 80 – 95 percent range, marginal increases may remain a worthwhile investment. Second, and crucially, this figure is only significant for boys and when aggregating genders, but not when looking at attendance rates of girls alone. This might suggest that while the linkage between more teachers and more careful attention paid to students may be beneficial for boys, that increased number of teachers may have minimal effects on girls, possibly due to teachers not paying as much attention to girls' needs and unique challenges in educational settings.

Our fourth model for girls' attendance (as measured through school surveys), we include additional individual-level predictors into our analysis: 1) whether support was provided during COVID-19, 2) disability status, 3) whether the girl's primary language is a minority language in the school, 4) whether the household head is female, and 5) how much time a girl is asked to do household chores. In this adjusted model, the p-value of the number of trained teachers in school further increases, while the coefficient decreases, meaning that it is even less predictive of attendance rates. Instead, the only significant predictor was whether the girl was provided with support by the school during COVID-19 closure; this predictor in fact registered a negative coefficient, which suggests that girls receiving support were more likely to have lower attendance rates. This may be due to the fact those most in need of support were girls who already struggled academically to begin with, thus making them among the least likely to be in classes regularly once schools re-opened.

## Teaching Quality

### ***Girl Survey Subscores***

Beginning with the gender subscore, in order to only compare responses from girls who attended mixed-gender schools, we dropped observations that responded to any of the three gendered questions with “Not applicable”. In doing so, no girls from SAS were dropped from the first gendered question (G1) as the answer choice “not applicable” was not provided at SAS. However, 226 girls were dropped from the EL sample. As girls from the SAS sample did respond to similarly gendered questions (G2 and G3) with “not applicable,” we dropped the observations of 108 girls from the SAS sample who responded with “not applicable” to G3.

In response to the second question (G2), 106 girls from SAS responded with “not applicable”, as well as 274 from EL, totaling 380 observations. In response to the third question (G3), 108 girls from SAS were responded with “not applicable,” as well as 239 from EL, totaling 347 observations. Though we would expect the number of girls responding with “not applicable” to be consistent across all three questions at SAS and EL, respectively, there was some variation. As a result, any observation that selected “not applicable” in response to G2 or G3 was dropped from the gender subscore questions, a total of 349 observations.

Once the observations were cleaned, we selected the answer choices that correlated with either a teacher’s lack of a gender bias or a preferential treatment of girls. As the WWW project worked to increase *girls’* education and outcomes, our scores positively evaluate observations reporting preferential treatment vis-à-vis girl students.

Looking at girls’ characteristics, girls who reported having a disability at endline responded to the question “my teachers are often absent from class/lessons” with “disagree” or “strongly disagree” 66.7 percent more often at EL than at SAS, while girls who did not report having a disability responded to the same 85.4 percent more often at EL than SAS. The difference between the two rates may be the result of girls who have a disability interacting with specialised teachers who are in turn more likely to stay at a school and not transition to other schools. Across all other characteristics, the margin of difference between SAS and EL was minimal, further emphasising the effect that COVID-19 – an effect experienced by all – may have played on teachers’ attendance.

Examining the rate at which teachers explain in languages other than English, we disaggregate results by minority language. In doing so, we find that the rate at which girls who speak a minority language at home responded to the question with “often” or “sometimes” actually decreased by a marginal 0.04 percentage points between SAS and EL. In contrast, the rate at which girls who do not speak a minority language at home responded with “often” or “sometimes” increased by 1.6 percentage points. Though this tells us that the responses of girls who identified themselves as speaking a minority language at home do not correlate with a higher rate of responding to the question “If you don’t understand something, do your teachers use a different language to help you understand?” with “often” or “sometimes”, our theory is not disproven. Rather, we note that languages like Turkana that, relative to national numbers, are a minority dialect, make up a majority in certain counties, largely ASAL counties. As such, we can still hypothesise that girls who do not speak English or Kiswahili at home, girls largely enrolled in schools outside of the urban slum counties, were more likely to respond to the question with “often” or “sometimes”.

### ***The Subject Subscore Methodology***

Calculating the subscore’s evaluation, a score of 100 percent was awarded if the girl said that there were no topics that the teacher needed to refocus on for the girl to better understand. In answering that there *were* such topics, the girl was given four subsequent questions about English, Kiswahili, mathematics, and science, and asked to identify the topics within each subject that the teacher needed to refocus on. For each subject, the girl was given a set number of topics from which to choose from, as well as the

option to add an “other” topic. As a result, there was the possibility for the girl to name four English topics, four Kiswahili topics, nine mathematics topics, and 41 science topics. As such, the girl was given a point for each topic within one of the four subjects that she identified as needing refocusing. Each question score was then calculated by dividing her points by the number of possible subject topics, subtracting the quotient from one, and multiplying the difference by 100. In doing so, we found the average subject subscore to be 81.5 percent at SAS and 78.9 percent at EL, resulting in a negative difference of 2.6 percentage points.

### ***The Teacher Attitudes Subscore Methodology***

Teachers were asked thirteen questions, four of which were framed negatively, and nine positively. Within the teacher attitudes sub-subscore, ten questions are evaluated, with the remaining three concerning corporal punishment evaluated within the corporal punishment sub-subscore. As the ten questions were evaluated along the same four-point agreement scale, the questions that were framed negatively were reversed so that the final scoring logic aligned with that of the nine positively framed questions. In other words, the responses that “agree[d]” or “agree[d] strongly” with positive statements earned a point, while those that “disagree[d]” or “disagree[d] strongly” with negative statements earned a point. To calculate the teacher attitude score for each observation, the total number of points was divided by ten – the number of questions – and the quotient was multiplied by 100. As a result, the scores were situated on a 100-point scale.

Of the thirteen questions, the four negatively framed questions are “Coaching is a burden to my schedule”, “Boys and girls are better at different subjects”, “When a student is struggling, they just need to work harder”, and “I use the same strategies to teach the students who are performing well and the students who are struggling”. This means that unlike the other questions where an affirmative response earned a point, a point was earned for responses of “disagree” or “disagree strongly.”

Examining results for the notably low-scoring question “boys and girls are better at different subjects”, disaggregating by county, only 14.6 percent of teachers from Samburu disagreed with the statement, while teacher disagreement from all other counties ranged from 23.7 percent in Nairobi to 40.6 percent in Turkana. As a result of the low level of disagreement from teachers in Samburu County, we may assume that the society may be more gendered, as displayed by cultural practices such as beading, where young men, morans, select a girl with whom they can have sex with before either marries.<sup>44</sup> In addition, as noted above, during the COVID-19 pandemic, teacher attrition proved higher in ASAL counties than those of urban slums. As such, the number of trained teachers working in ASAL schools at Endline is lower than that of those working in urban slum schools. With this differential in mind, future interventions may take care to diversify intervention focus by county, accounting for the difference in the level of training completed by teachers after this intervention round. For instance, teacher training that includes gender sensitization may be emphasised in ASAL counties, specifically Samburu.

---

<sup>44</sup> “The Unspoken Vice in Samburu Community: Laikipia, Marsabit & Samburu Counties,” Samburu Women Trust, April 2016, accessed 20 December 2022, [https://www.ngeckkenya.org/Downloads/SWT%20Girl-Child%20beading%20Research%20in%20\(Laikipia,%20Samburu%20and%20Marsabit\)%20Counties.pdf](https://www.ngeckkenya.org/Downloads/SWT%20Girl-Child%20beading%20Research%20in%20(Laikipia,%20Samburu%20and%20Marsabit)%20Counties.pdf).

## Corporal Punishment

**Table 52: Teachers' self-reported attitudes vis-a-vis corporal punishment**

Self-reported Attitudes	Response	Nairobi	Kilifi	Samburu	Turkana	Tana River	Mombasa
Corporal punishment is sometimes necessary in class	Agree/Agree strongly	19.6	22.7	19.7	23.9	25.5	0.0
	Disagree/Disagree somewhat	80.4	77.3	80.3	76.1	74.5	100.0
Corporal punishment is never acceptable	Agree/Agree strongly	72.9	72.1	75.9	52.3	93.3	100.0
	Disagree/Disagree somewhat	27.1	28.9	24.1	47.7	16.7	0.0
Corporal punishment slows down learning	Agree/Agree strongly	91.7	62.4	65.0	55.3	76.0	66.7
	Disagree/Disagree somewhat	18.3	37.6	35.0	44.7	24.0	33.3

## Predictive Analysis of Teaching Quality

We analyse predictors of the Teaching Quality – Girl Survey Subscore using individual level characteristics as well as school-level characteristics. Results are shown in the table below. At the individual level, no characteristics proved significant. At the school level, the characteristic describing the average number of lessons taught by a teacher in a week showed a positive significant correlation with the Teaching Quality – Girl Survey Subscore at the 90 percent confidence interval.

**Table 53: Predictive analysis: Teaching quality - girl survey subscore**

Variables	Coef.	p
<b>Individual (Girl) Level Characteristics</b>		
Age	-1.6	0.12
Disability	-6.2	0.14
Life Skills Score	0.07	0.92
Decision-making score	0.7	0.18
Minority Language speaker	1.6	0.51
<b>School Level Characteristics</b>		
Trained teachers	0.2	0.16
School received support of remedial teachers from the WWW project	-0.8	0.59
Average number of lessons taught by a teacher in a week	0.3	0.08

Average number of school days per week that teachers miss	2.2	0.46
---	-----	------

The positive correlation between girls' beliefs regarding teacher quality and the average number of lessons taught by teachers per week at her school is not totally surprising. Generally speaking, the correlation seems to indicate that the more classes taught by a teacher per week, the higher the girls' perception of her teachers' quality. While we might expect that teachers who teach many classes each week may provide a lower level of quality, unable to devote required time to lesson planning, it seems that teachers who teach a full but not overloaded course load provide, in the girls' view, a high quality of teaching. This may be because they teach multiple subjects, perhaps indicating greater experience or qualification. It may also mean that girls have multiple classes with the same teacher, resulting in a level of comfort or familiarity that results in a higher perception of the teacher's quality.

## Life Skills

**Table 54: Adapted Life Skills Index questions**

	Adapted LSI Question
<b>Learning to Learn</b>	I want to do well in school
	I get nervous when I have to speak in front of a group of people my age
	I get nervous when I have to read in front of an adult
	I get nervous when I have to do maths in front of others
	I feel confident answering questions in class
	I can stay focused on a goal, despite things getting in the way
	When I succeed at school it is because I worked hard
<b>Learning for Life</b>	I would like to continue studying / attending school after this year
	I would like to continue learning a vocation or trade
	I am aware of the different transition pathways in the education system
	I am willing to transition through the education pathways until I reach my goal
	I am aware of the alternative pathways of education (TVETs, Apprenticeships, etc.)
	I recognise when choices I make today about my studies can affect my life in the future
	I can put a plan in place and stick with it
	I ask the teacher if I don't understand something
	I study well in a group with other peers during the non-school hours
	When I have/had the opportunity, I can/could organise my peers or friends to do an activity
<b>Agency</b>	Who decides: Whether or not you will go back to school
	Who decides: Whether or not you will continue in school past this year
	Who decides: When / at what age you will get married



	Who decides: If you will work after you finish your studies
	Who decides: What type of work you will do after you finish your studies
	Who decides: What type of work you do while at home
	Who decides: What topics you study while at home

### Calculating the LSI Scores

The LSI is calculated by asking respondents a series of questions about self-esteem (“learning to learn”/L2L), life skills (“learning for life”/L4L), and agency. After each of the questions, the girl is given a Likert scale in terms of agreement and asked to select whether she “strongly agrees”, “agrees”, “neither agrees nor disagrees”, “disagrees”, “strongly disagrees”, or “doesn’t know.” Scores are calculated so that a response of “agree” or “strongly agree” earns one point.

In parallel, a divisor is developed to match the number of questions asked. For instance, the divisor of the total LSI score is 24, reflecting the section’s 24, while that of the “learning for life” subscore is 10 and the “learning to learn” and “agency” subscores are seven. The divisors decrease by one point each time a respondent answers a relevant question with “don’t know” or their response is missing from the data. As a result, girls are not penalised for not knowing how to respond to a question, and in theory the scores are not unduly affected by missing data. Once a girl’s score is summed it is divided by her relevant divisor before being multiplied by one hundred. The result is a standardised score on a 100-point scale.

**Table 55: L2L subscore by county**

<i>L2L Subscore</i>			
<b>Characteristic</b>	<b>SAS</b>	<b>EL</b>	<b>Difference</b>
<b>Counties</b>			
Kilifi	85.5	83.5	1.0
Mombasa	79.3	90.5	11.2
Nairobi	84.4	88.8	4.5
Samburu	86.0	92.9	6.9
Tana River	81.3	88.3	7.1
Turkana	85.6	91.8	6.2

**Table 56: Difference between L2L question scores between SAS and EL, Mombasa County and overall**

Question	Mombasa Diff.	Overall Diff.
When I succeed at school it is because I worked hard.	16.7	8.5
I want to do well in school.	0.0	0.2
I get nervous when I have to read in front of an adult.	5.0	0.5
I get nervous when I have to speak in front of a group of people my age.	8.3	-2.8
I get nervous when I have to do maths in front of others.	10.0	-2.1
I feel confident answering questions in class.	-1.7	-0.4
I can stay focused on a goal despite things getting in the way.	1.7	1.2

### **Agency Subscores**

Disaggregating the agency subscore by county, Kilifi, Mombasa, and Nairobi saw very little changes to their aggregate agency subscore between the two rounds.<sup>45</sup> In contrast, Samburu, Tana River, and Turkana saw notable negative changes, with those of Samburu and Turkana declining by 9.9 and 12.1 points, respectively. Like the trend described above, it's possible that these declines are also the result of parents' waning support for their daughters to remain in school as the girls mature.

**Table 57: Agency subscore by county**

<i>Agency Subscore</i>			
Characteristic	SAS	EL	Difference
Kilifi	85.5	85.6	0.1
Mombasa	87.6	87.5	-0.1
Nairobi	83.9	83.9	0.0
Samburu	93.4	83.5	-9.9
Tana River	91.5	88.5	-3.0
Turkana	88.6	76.6	-12.1

For instance, a teacher from Samburu described that men's fear of "girls who are learned" may cause those who continue their education to face discrimination from within their community.<sup>46</sup> In addition, through the practice of "beading" a girl may be selected for marriage while still quite young, despite Kenya's Marriage Act that mandates 18 as the legal age of marriage. Due to the prevalence of marrying while young, while girls may wish to remain in school, pressure from their parents to leave may increase

<sup>45</sup> Girls from Marsabit and Kwale were not asked the 24 LSI questions at SAS and as such their responses given at EL are not considered in this section.

<sup>46</sup> KII with Teacher, Samburu, Int. 25.

as they increase in age, resulting in a situation where older girls, even just a year older, face opposition from their parents when making decisions regarding staying in school.

According to a teacher from Turkana, pressure to leave school and marry falls on girls in Turkana, too. Described in the quote below, this pressure is not purely cultural, but influenced by the dowry that comes with marrying a daughter.

*There is the issue of some parents not willing to educate their girls. They ask the girls to get married instead of going to school. They see the girls in terms of dowry. The community does not value education. Mostly the illiterate are a threat to girl education.<sup>47</sup>*

In addition, both Samburu and Turkana are sprawling counties in terms of area. As a result, girls often travel far distances to reach their primary or secondary school. Considering this sprawl and the more limited number of post-secondary options, the idea of a girl continuing her education after secondary school may seem logistically impossible to some parents. As a result, the logic may follow that transitioning to secondary school is an end goal in and of itself and once this transition is achieved, the girl has completed her education and should be ready for marriage.

### **Predictive Analysis of Life Skills**

We analyse predictors of the Life Skills Index and subscores. We focus on individual-level characteristics predicting the Life Skills Score, Learning to Learn Subscore, Learning for Life Subscore, and Agency Subscore, as the school-level variables included in evaluation data (such as teacher attendance and student-teacher ratio) were deemed to be of limited theoretical relevance to these outcomes. However, as in previous predictive analyses, we include school fixed effects.

Results are shown in the table below. We find a significant relationship between whether the girl is disabled and two of the outcomes of interest: the L2L subscore and the L4L subscore. For both subscores, disability status is significantly correlated with negative subscores, and the effect size is relatively large. We note, however, that disability status is actually correlated (although not significantly at the  $p < 0.05$  level) with agency; in other words, all else held constant, girls with disabilities are actually more likely to report, on average, that they make decisions for themselves.

**Table 58: Predictive analysis of SRH confidence and perceptions**

Characteristic	Life Skills		L2L Subscore		L4L Subscore		Agency Subscore	
	Coef.	P-value	Coef.	P-value	Coef.	P-value	Coef.	P-value
Age	0.4	0.24	0.7	0.16	0.4	0.24	0.08	0.92
Minority language	-0.7	0.46	-1.0	0.60	-0.2	0.81	-1.2	0.56
Disability	-0.4	0.82	-5.6	0.01	-4.4	0.01	10.7	0.08
Female HoH	0.2	0.81	0.3	0.85	-0.07	0.94	0.6	0.82
Housework	-0.4	0.52	0.2	0.79	1.1	0.03	-3.1	0.06

Outside of disability, the only other significant predictor found is housework, which has a significant positive relationship with the L4L subscore. Given that the L4L subscore included several questions about

<sup>47</sup> KII with Teacher, Turkana, Int. 4.

non-university tertiary pathways, it may be that girls with more household responsibilities are also more inclined to want to pursue TVET or other ALPs. It may also be that girls with more household responsibilities also have, by necessity, greater time management or planning skills, as these skills are more crucial for their lives.

## Household Support

The Household Support Score is composed of eighteen questions. To better understand household support, the overall Household Support Score is broken down into four subscores: the Leave School Subscore (LSS), Future School Subscore (FSS), Girl's Will Subscore (GWS), and Value Girl's Education Subscore (VGES). The four subscores are composed of eleven, three, two, and two questions, respectively, and were balanced when feeding into the Household Support Score. This balance means that a subscore like FSS which is composed of three questions made up only three eighteenthths of the overall Household Support Score; while LSS, with eleven questions, made up eleven eighteenthths of the overall Household Support Score.

### *Calculating the Leave School Subscore*

The Leave School Subscore is characterised by eleven possible points, giving it the heaviest weight within the Household Support Score. Specifically, the caregiver was asked “under which of the following conditions do you think it is acceptable for a child to not attend school?” and then given ten scenarios, plus the opportunity to “think of any other condition where you think it is acceptable for a girl to not attend school.”

**Table 59: “Under which of the following conditions do you think it is acceptable for a child to not attend school?”**

	Scenario
1	“The child may be physically harmed or teased at or on the way to/from school”
2	“The child may physically harm or tease children at school”
3	“The child needs to work”
4	“The child needs to help at home”
5	“The child is married/is getting married”
6	“The child is too old”
7	“The child has physical or learning needs that the school cannot meet”
8	“The child is unable to learn”
9	“Education is too costly”
10	“The child is a mother”

The caregiver earned a point when they responded to a scenario with “Not acceptable” and lost a point when they responded with “Acceptable”. The Leave School Subscore was then calculated by summing each caregivers’ “Acceptable” answers and dividing the sum by eleven. If for some reason a caregiver’s data was missing, the divisor, which defaults at eleven, decreased so that in theory the scores were not unduly affected by missing data.

### ***Calculating the Future School Subscore***

A caregiver earned a point if they responded to one of the three questions (1) “would you like [girl] to achieve”, 2) “would you be willing to support your [girl] to achieve”, and 3) “are you able to support your [girl] to achieve?”) with “College/TVET” or “University” and did not earn a point if they responded with “Primary School” or “Secondary School.” Considering that the subscore is composed of three questions, as a default, in order to standardise the score on a 100-point scale, respondents’ total number of points was divided by three. That said, a caregiver’s divisor decreased by one point each time they answered a relevant question with “Do not know” or their response was missing from the data. As a result, caregivers were not penalised for not knowing how to respond to a question, and in theory the scores are not unduly affected by missing data. We note that particularly for the last question, many caregivers answered “do not know”.

### ***Calculating the Girl’s Will Subscore***

The two questions comprising the Girl’s Will Subscore are “Do you discuss with [girl] what she wants to become in the future?” and “Do you listen to the views of [girl] when you make decisions about her education or are those decisions made by adult members of the family only?” For the question regarding decisions about the girl’s education, the caregiver earns a point for responding with “Girl’s opinion is considered”. For the question asking about discussing with the girl what she “wants to become in the future”, the caregiver must respond with “Yes” in order to receive the point.

With the two main questions of import in mind, the caregiver has the potential to earn two points. Regardless of whether the caregiver earns no points, one, or two points, as a default their accrued point value is divided by two. That said, a caregiver’s divisor decreases by one point each time they answered one of the two relevant questions with “Do not know” or their response was missing from the data. As a result, caregivers were not penalised for not knowing how to respond to a question, and in theory the scores are not unduly affected by missing data.

### ***Calculating the Value of Girl’s Education Subscore***

The Value of Girl’s Education Subscore (VGES) is composed of two questions, “Even when funds are limited it is worth investing in [girl’s] education” and “A girl is just as likely to use her education as a boy”, both of which are scored on a Likert scale of agreement, with responses of “Agree” or “Strongly agree” earning a point and responses of “Neither agree nor disagree”, “Disagree”, and “Strongly disagree” earning no points. Considering that the subscore is composed of two questions, as a default, in order to standardise the score on a 100-point scale, respondents’ total number of points was divided by two. That said, a caregiver’s divisor decreased by one point each time their response was missing from the data. As a result, in theory the scores are not unduly affected by missing data.

### ***Predictive Analysis of Household Support***

In this section, similar to sections prior, we test the relationship between caregiver- and student-level characteristics against the Household Support Score, Leave School Subscore, Future School Subscore, Girl’s Will Subscore, and Value of Girl’s Education Subscore. The goal of this section is to assess whether certain caregiver or girl characteristics are correlated with the household support score or subscores. For example, were all other characteristics equal, we may expect that a higher level of completed education by the caregiver may correlate with a higher Future School Subscore. We first analyse the individual caregiver variables such as the caregiver’s relationship to the girl and the highest grade of school completed by the caregiver. We then analyse variables measured at the girl-level such as her age and grade in school. We utilise all SAS and endline data to increase sample size and predictive power.

To evaluate the Household Support Score we select eight individual caregiver-level characteristics, namely: the county in which the caregiver lives; the caregiver’s relationship to the girl; the gender of the head of the household in which the caregiver and girl are a part; the highest grade in school completed by

the caregiver; whether the caregiver is unemployed; whether the girl and caregiver come from a pastoralist household; the division of domestic chores in the home as described by the caregiver; and whether the caregiver is a traditional, village, or government leader.

To better understand the relationship between the individual girl-level predictors and the household support outcomes, we run a second model including five girl-level variables: the girl's age; her grade in school; whether she speaks a minority language at home; whether the girl has been enrolled in school since January 2021; the time he spends on chores, housework, helping the family business, etc.; and whether the girl is a mother.

Considering the relationship across the Household Support Score, three caregiver-level characteristics stand out as significant, all of which have a positive relationship with the score. In terms of employment, we see that pastoralists and traditional, village, or government leaders correlate significantly and positively with the Household Support Score. Considering pastoralists, caregivers who are pastoralists may want and imagine a different life for their girls. That said, while the characteristic correlates positively with the Leave School and Future School Subscores, with pastoralist caregivers projecting imagining and supporting the girl in attaining higher levels of education, there is a negative (though not significant) relationship between pastoralist caregivers and the Girl's Will Subscore, suggesting that though caregivers who are pastoralists imagine and support a higher level of education for their girl, they do so without meaningful discussion and input from the girl herself.

**Table 60: Predictive analysis with caregiver-level predictors**

Variables	Household Support Score		LSS		FSS		GWS		VGES	
	Coef.	p	Coef.	p	Coef.	p	Coef.	p	Coef.	p
County	0.5	0.21	0.9	0.13	-0.01	0.99	-0.3	0.51	-0.01	0.96
Relationship to the girl	-0.03	0.48	0.003	0.97	0.03	0.55	-0.3	0.09	-0.09	0.31
Gender of head of household	-0.3	0.79	-1.0	0.56	-0.8	0.69	4.1	0.02	1.2	0.30
Highest completed school grade	0.8	0.04	0.04	0.72	3.5	<0.001	0.1	0.74	0.2	0.13
Unemployed	1.8	0.18	2.9	0.17	-1.9	0.55	0.4	0.87	-0.6	0.70
Pastoralist	7.3	<0.001	9.3	<0.001	13.1	<0.001	-7.6	0.08	1.0	0.72
Domestic chores in the home	-0.9	0.79	-6.7	0.28	14.4	<0.001	3.3	0.39	2.7	0.14
Traditional, village, or government leader	9.9	<0.001	12.2	<0.001	7.1	0.03	11.2	<0.001	4.5	<0.001

Caregivers who serve as traditional, village, or government leaders, on the other hand, correlate significantly and positively with all subscores. These significant relationships follow as community leaders at any level may be more financially stable, have higher levels of education, and also have a broader view

of the community and beyond, better positioning them to foresee the value of an education for a girl and her future.

Looking at individual girl-level characteristics, we see that none have a significant relationship with the Household Support Score. However, the Future School Subscore has a significant positive relationship with the girl's age, whether she speaks a minority language at home, whether the girl has been enrolled since January 2021, and whether the girl is a mother. The Girl's Will Subscore, too, shows a significant negative relationship with the girl's age, grade in school, whether she had been enrolled in school since January 2021, and the amount of time she spends on chores, housework, helping the family business, etc.

Considering age, caregivers whose girls are older reported higher Future School Subscores, but lower Girl's Will Subscores. This follows as caregivers of girls who are older are more actively considering the girl's future and future in education, indicating that caregivers may become more supportive of girls' attainment of higher education as she matures. It may also indicate that caregivers of girls who are older than is usual for their grade in school may be more supportive of the girl's educational attainment, perhaps having re-enrolled her in school after a period of hardship or disruption, a period that offered the caregiver more perspective on the value of an education to the girl and her future. That said, girls' age and grade in school reported a significant negative relationship with the Girl's Will Subscore, suggesting somewhat counter-intuitively that as the girl matures in age and school, her will is considered less and less frequently by her caregiver. Though this could mean that caregivers override girls' will when the girl does not want to pursue higher education, the trend points to a lack of girls' personal autonomy vis-a-vis the household.

**Table 61: Predictive analysis with individual girl-level predictors**

Variables	Household Support Score		Leave School Subscore		Future School Subscore		Girl's Will Subscore		Value of Girl's Education Subscore	
	Coef.	p	Coef.	p	Coef.	p	Coef.	p	Coef.	p
Age	-0.05	0.56	-0.08	0.56	0.2	<0.001	-0.3	<0.001	-0.02	0.81
Grade in school	-0.2	0.11	-0.2	0.23	-0.05	0.70	-0.1	0.03	0.01	0.73
Minority language speaker	-2.3	0.13	-2.1	0.35	-5.0	0.02	-2.4	0.15	0.7	0.56
Has the girl been enrolled since January 2021	-9.6	0.17	-11.0	0.27	-14.8	<0.001	-11.5	<0.001	8.0	0.31
Time spent on chores, housework, helping the family business, etc.	-0.03	0.20	-0.03	0.35	-0.2	0.55	-0.05	0.04	0.02	0.36
Is the girl a mother	0.4	0.69	2.8	0.09	-8.7	<0.001	1.5	0.14	0.4	0.65

We also see a significant negative relationship between whether the girl speaks a minority language at home and the Future School Subscore. As girls who speak a minority language at home may also struggle more than their peers who are more comfortable with the language of instruction, their caregivers may be less likely to support their future education, perhaps citing the girls' struggling as a lack of reasonable acumen for continued investment in her education.

In addition, whether the girl had been enrolled in school since January 2021 also recorded significant negative relationships with the Future School and Girl's Will Subscores. While only thirty-three girls were reportedly not enrolled in school since January 2021, it appears those thirty-three recorded higher Future School and Girl's Will Subscores. This may be due to the fact that these girls were out of school for extenuating circumstances, which led their caregivers to imagine higher levels of education attainment for the girl and consider the girl's will at a higher frequency than caregivers whose girls were consistently enrolled in school.

Finally, we note that whether the girl is a mother has a positive significant relationship with the Leave School Subscore at the 90 percent confidence interval and a negative significant relationship with the Future School Subscore. Beginning with the Leave School Subscore, regardless of their age, girls who are mothers may be seen as less vulnerable or fragile than girls who are not. Such a perception may be the result of some protective ideas about girls having to do with "honour" and ideas of virginity. As the threat of being physically harmed or teased on the way to school was the second-most-cited scenario as an acceptable reason for leaving school, the threat of harassment or assault, often defined in terms of sexual assault and/or rape in key informant interviews with teachers, may be weighed as less consequential. What's more, though being a mother was the most cited scenario as an acceptable reason for a girl to leave school, the positive correlation between a girl's status as a mother and the Leave School Subscore shows us that though caregivers may imagine a girl who is a mother will need to leave school, in practice, caregivers of girls who are mothers support their girl in staying school, citing few - if any - reasons to leave. That said, the motherhood variable correlates negatively with the Future School Subscore, suggesting that though caregivers may envision fewer reasons for a girl who is a mother to leave school, they also limit the girl's education timeline; more frequently reporting that the caregiver would not like, is not willing, and/or is not able to support the girl through college/TVET and/or university.

## Community Behaviour and Support

### *Predictive Analysis of Community Behaviour and Support*

We conduct a predictive analysis of community support utilising methods similar to those described in previous sections. As our demographic variables exist for the caregiver her/himself, we utilise a caregiver-level measure of support for girls' education as our outcome of interest, specifically whether caregivers are willing to support girls and have supported girls to continue their education. Our predictor variables include whether the caregiver is unemployed, whether they are a pastoralist, their level of education, whether they are a traditional or community leader, and the gender of the head of household. We also include fixed effects for county, which we do not report in the table below for the sake of brevity.



**Table 62: Predictive analysis of caregiver support for girls' education**

Variables	Willing to support		Has supported	
	Coef.	P-value	Coef.	P-value
Female HoH	-0.01	0.53	0.03	0.68
Caregiver education	-0.003	0.49	-0.004	0.13
Unemployed	-0.03	0.49	-0.04	0.47
Pastoralist	0.01	0.48	-0.3	0.09
Community leader	-0.3	0.11	-0.3	0.20

The table above shows no significant relationships between the analysed variables and the two outcomes of interest, although we note that there were significant relationships between the county fixed effects and the outcomes, emphasizing the variation noted above by county. Interestingly, we do not find a significant correlation between measures of economic ability, such as unemployment and whether the caregiver is pastoralist, and whether the caregiver has supported a girl to continue her education. However, the direction of the effect is negative as expected; in other words, unemployment and pastoralism are associated (though not significantly) with lower provision of support to girls to continue education.

Overall, given that no variables are found to be significant, it is difficult to draw conclusions from this analysis. Hypothesized relationships, such as female head of households or well-educated caregivers being more likely to support girls' education, are not found in the above analysis. However, we note that this does not mean that these relationships do not exist; rather, there may be confounding factors or we may simply not have a large enough sample to find these relationships.

## Annex 19: Supplementary Analysis of Other Outcomes of Interest

### School Management and Governance

The first additional outcome of interest is school governance and management. In this section, we investigate how caregivers and head teachers evaluated school governance during the endline evaluation. In doing so, we analyse observations from caregivers whose girls attend 158 different schools that were the site of WWW project interventions across eight counties.

In order to evaluate school management and governance, we used questions from the caregiver survey concerning school management and school governance to create respective school management and school governance subscores which are calculated on a 100-point scale. The school management subscore is comprised of four questions while the school governance subscore is comprised of five questions. The combined school management and governance score is calculated by multiplying the school management subscore by 4/9 and the school governance subscore by 5/9 and summing the two products. As a result, the school management and governance score is also standardised on a 100-point scale.

Considering the way in which each question score was calculated, the process depends on the question's corresponding answer choices. The first question regarding school management measures the percent of the caregiver respondents who said that their girl's school was managed "well" or "extremely well". The second management question asks how the management of the school has changed over the last 12 months and measures the ratio of caregiver respondents who said that school management had "improved". The third management question score measures what percentage of caregiver respondents

said that the performance of the girl’s head teacher or school principal was “good” or “excellent”. The final management question measures the percentage of caregiver respondents who stated that the performance of the girl’s head teacher or principal had “improved” over the last twelve months.

Considering the governance questions, the first governance question score measures the percent of caregivers who responded that the school council, Board of Management (BoM), or Parent-Teacher Association (PTA) communicates with them either “weekly” or “monthly”. The second governance question measures the percentage of caregivers who listed at least two activities undertaken by the school’s school council, BoM, or PTA in the last twelve months. The third governance question measures the percentage of caregivers who responded “yes” to the question “Were any of these initiatives useful for improving the quality of schooling [girl] received?” The fourth governance question measures the percentage of caregivers who rate the performance of their school’s BoM as “good” or “excellent”. The fifth and final governance question score measures the percentage of caregivers who said that the performance of the BoM had “improved” over the past twelve months.

### **School Management and Governance Score and Subscores**

In order to evaluate school management and governance, we used questions from the caregiver survey concerning school management and school governance to create respective school management and school governance subscores which are calculated on a 100-point scale. The school management subscore is comprised of four questions while the school governance subscore is comprised of five questions. At first glance, we see that at 88.7 percent, the Management Subscore drove the overall score up, countered by the Governance Subscore at 45.5 percent.

**Table 63: School management and governance score, subscores, and questions**

Score Type	Score at EL
<b>School Management and Governance Score</b>	62.9
Management Subscore	88.7
Governance Subscore	45.5

Breaking the subscores down to the question level, we see that the question that caregivers responded to most confidently across the board was that concerning how well the girl’s school is managed, with 66.1 percent of the caregiver respondents responding with “well managed” or “extremely well managed”. When we look at the MQ1 score sorted by how the school management has changed over the past twelve months, we see that the mean MQ1 score at “improved” is 98.0 percent. This means that of the 586 schools that caregivers said “improved”, most of them were described as “well managed” or “extremely well managed” by the caregiver. Such a mean MQ1 score corresponding with reports of improved school management over the past twelve months could be the positive indication of WWW project initiatives.

Sorting MQ1 by MQ3, or head teacher/principal performance ratings, we see that when a caregiver rated the school’s principal or head teacher as good or excellent, the mean MQ1 score was 94.8 and 93.7 percent, respectively. What’s more, when the head teacher or principal was rated as poor or very poor, the mean MQ1 was 44.4 and 0.0 percent, respectively. The positive correlation between highly rated school leaders and well to extremely well managed schools, and the inverse correlation between poorly rated school leaders and more poorly managed school scores seems to indicate that the calibre of a principal or head teacher may factor critically into how well that school’s management is perceived.

**Table 64: School management and governance questions**

Subscore	Question	Score at EL	Baseline <sup>48</sup>
Management (MQ1)	How well is the school [ <i>girl</i> ] attends managed?	66.1	88.9 <sup>49</sup>
Management (MQ2)	Compared to 12 months ago, how has management of the school changed?	52.7	67.4 <sup>50</sup>
Management (MQ3)	How would you rate the performance of the school head teacher or principal of the school that [ <i>girl</i> ] attends?	53.7	42.2 <sup>51</sup>
Management (MQ4)	In the last 12 months, how do you think the performance of the head teacher or principal of the school that [ <i>girl</i> ] attends has changed?	55.3	-
Governance (GQ1)	Does the school/school council/BoM/PTA communicate with you regularly about its plans and activities?	31.7	20.5 <sup>52</sup>
Governance (GQ2)	What kinds of actions or initiatives did this school council/BoM/PTA or similar group take in the last 12 months?	22.2	-. <sup>53</sup>
Governance (GQ3)	Were any of these initiatives useful for improving the quality of schooling [ <i>girl</i> ] received?	48.9	-
Governance (GQ4)	How would you rate the performance of the BoM of the school that [ <i>girl</i> ] attends?	40.4	-
Governance (GQ5)	In the last 12 months, how do you think the performance of the BoM of the school that [ <i>girl</i> ] attends has changed?	44.2	-

Moving to governance questions, we do the same kind of cross-question evaluation to better understand the relationship between different factors contributing to the general idea of school governance. For instance, considering the frequency with which BoMs routinely communicate with caregivers, 6 percent of caregivers reported weekly communication, 45 percent of caregivers reported monthly communication, 33 percent of caregivers reported annual communication, and 7 percent of caregivers reported no communication (with the remainder stating that this didn't know). While weekly communication was associated with the highest mean BoM rating (72.7 percent) and no communication is associated with the lowest BoM score (37.3 percent) the difference between BoM ratings when communication was monthly

<sup>48</sup> The baseline scores come from the total intervention group scores, reported in the baseline report. As the panel of caregivers interviewed differed between endline and baseline, we're unable to attribute change from responses. However, where available, baseline scores are included to give some perspective of change over time.

<sup>49</sup> "GEC - T Baseline Report," Education Development Trust, May 2018, page 95.

<sup>50</sup> Ibid., 96.

<sup>51</sup> Ibid.

<sup>52</sup> Ibid., 98.

<sup>53</sup> We are unable to compare endline to baseline as the endline evaluation awards a point when the caregiver listed two or more initiatives, whereas the baseline report only notes the frequency that caregivers responded with one of the six given initiatives.

versus annually was nearly null. Specifically, caregivers who reported monthly communication reported a mean BoM score of 68.8 percent, while caregivers who reported annual communication reported a mean BoM score of 66.5 percent. As such, while weekly communication between the BoM and caregivers may be indicative of an active and productive BoM and no communication may be indicative of an inactive and ineffective BoM, other intermediate frequencies of communication may not be an appropriate marker of BoM efficacy.

Disaggregating by county, we see that with a score of 70.1 percent, Samburu had the highest school management and government score of all the counties. This high score was largely driven by the management subscore, which at 97.3 percent is a near perfect score. Samburu’s governance subscore, on the other hand, was the second lowest, at 30.7 percent. Considering the nearly perfect nature of the management subscore and relatively low governance subscore, the mean Samburu County data point serves as another indicator of the fact that a school may be very well managed, as reported by caregivers, but lack an active or involved BoM or PTA.

**Table 65: School management and governance score and subscores by county**

County	School Management and Governance Score	Management Subscore	Governance Subscore
Kilifi	65.6	89.0	31.9
Kwale	61.8	83.6	47.2
Marsabit	69.5	80.6	62.1
Mombasa	59.0	90.2	38.2
Nairobi	59.3	86.0	33.9
Samburu	70.1	97.3	30.7
Tana River	67.1	92.7	39.9
Turkana	64.9	91.4	25.4

Considering BoMs and PTAs in Samburu, specifically, when asked about the PTA at his school, one head teacher from a Samburu school said “yes we have a PTA in our school but they are not very supportive. They visit the school and talk and listen to [girl’s] grievances when they can.”<sup>54</sup> In contrast, a teacher from Mombasa described the PTA at her school, stating “[the PTA] supports the girls. Last year they organised and we got sanitary pads for the girls.”<sup>55</sup> In Tana River, a teacher shared that the PTA at his school mostly “ensures that the girls are in school because they are coming from the communities. [The PTA] talks to the parents to ensure that the girls report to school on time. The effort has been effective.”<sup>56</sup> Another head teacher from Tana River shared “yes, we have a PTA and it’s working to support girls’ education. That’s their responsibility. By talking to the girls frequently and spearheading having more dormitories and classes, I think their efforts are very effective.”<sup>57</sup>

Though just four data points, the difference between the Samburu teacher’s experiences and that of the three other teachers with PTAs at their school is clear. The difference between the teacher’s experiences is also reflected in their respective governance subscores. Hypothesising why Samburu’s governance subscore may be so low, in line with a Samburu teacher’s KII, it’s possible that the insecurity in Samburu

<sup>54</sup> KII with Teacher, Samburu, Int. 12.

<sup>55</sup> KII with Teacher, Mombasa, Int. 8.

<sup>56</sup> KII with Teacher, Tana River, Int. 6.

<sup>57</sup> KII with Teacher, Tana River, Int. 24.

that affects girls' feelings of safety and ability to get to school also affects community members who may wish to otherwise contribute to or participate in the schools' PTAs. For instance, one teacher from Samburu said that one of the greatest challenges that girls face in enrolling and staying in school comes from the fact that many of the girls come from pastoralist families.<sup>58</sup> As a result, the girls have to travel far distances to get to school and there is no way to communicate with them before they come to school or with their parents because they aren't reached by internet or cellular networks. Given the difficulties associated with girls' commute to school, it makes sense that fewer caregivers would be able or willing to participate in PTAs, versus in communities like Nairobi or even Tana River, where the community is more concentrated within the vicinity of the schools, making PTA and BoM activities and initiatives easier to develop and pursue.

Lastly, despite challenges to school management and governance, we note that in KIIs, government officials mentioned holding trainings to strengthen school management, including the capacity of head teachers, financial management, and trainings targeting the BoM. These trainings may help improve the sustainability of programme activities and further improve school management and governance beyond the close of the WWW programme.

### ***Predictive Analysis of School Management and Governance***

In this section, as in sections prior, we test the relationship between several student- and school level-characteristics against the school management and governance score, management subscore, and governance subscore. The goal of this section is to assess whether certain school or girl characteristics are correlated with management and governance scores or subscores. For example, were all other characteristics equal, we may expect that a higher number of trained teachers will correlate with a higher management subscore. We first analyse the individual-variables such as the caregivers' girls' age and life skills score. We then analyse variables measured at the school level such as teacher attendance.

To evaluate the school governance and management score we select six individual level characteristics, namely: whether the girl received support from the teacher while school was closed during the COVID-19 pandemic, whether the girl identifies as having a disability, whether the girl speaks a minority language at home, whether the head of the girl's household is a female, and the county in which the girl attends school. The model also includes a school fixed effects control variable to control for variation which occurs at the school level, and which does not change, or changes at a constant rate, over time. We use the cohort 2 sample to conduct this analysis and use data from the girl's survey. The results of this regression are displayed in the table below. According to this analysis, whether the girl speaks a minority language at home has a statistically significant relationship to the school management and governance score generated by her caregiver's responses.

To better understand the relationship between school level predictors and school governance outcomes, we run a second model including the same individual level variables and five school level variables; the number of trained teachers at the school, the number of lessons taught by teachers in a week, teacher attendance, the presence of children's clubs at the school, and support by remedial teachers after the COVID pandemic. When running Model 2 the relationship between predictors and the school

---

<sup>58</sup> KII with Teacher, Samburu, Int. 39.

agement and governance subscore remains weak and statistically insignificant.

**Table 66: Predictive analysis with individual- and school-level predictors**

Variables	School M&G		Management		Governance	
	Coef.	P-value	Coef.	P-value	Coef.	P-value
<b>Model 1: Individual-level Characteristics</b>						
Support during COVID-19	-1.2	0.54	-2.3	0.22	0.6	0.80
Disability	3.9	0.33	4.8	0.17	-3.0	0.53
Minority Language speaker	-8.8	<0.001	-8.6	0.001	-5.6	0.09
Female HoH	0.2	0.94	-0.7	0.78	-2.7	0.38
County	25.9	<0.001	29.2	<0.001	3.3	0.24
<b>Model 2: School-level Characteristics</b>						
Number of trained teachers at the school	0.08	0.32	0.01	0.87	0.3	0.02
Average number of lessons taught by a teacher a week	0.03	0.82	-0.1	0.43	-0.3	0.06
Average teacher attendance	-1.8	0.29	-2.6	0.21	-0.5	0.81
School received support from remedial teachers	1.2	0.58	1.7	0.39	-0.5	0.87
School has clubs	1.9	0.62	4.1	0.06	-11.7	0.001

Considering relationships across the school management and governance score, the school management subscore, and the school governance subscore, one individual level characteristics and two school-level characteristics stand out with statistically significant relationships between the characteristics and scores. As mentioned above, at the individual-level, whether a girl is a minority language speaker has a statistically significant relationship with the school management and governance score, as well as the management subscore. Given that the management and governance subscores are created from questions supposing good communication between a caregiver and a school, it makes sense that if a family speaks a minority language at home, the parents especially may have a harder time communicating with or knowingly receiving communications from the school, resulting in negative perceptions of management and governance.

Moving to school-level characteristics, the number of trained teachers at the school has a significant and positive effect on the governance subscore. The relationship between the governance subscore and number of trained teachers at a school may be because some PTAs and BoMs lobby for improved teacher training as well as new, better trained teachers. The relationship may also indicate that girls are more successful in school when there are more trained teachers, and caregivers and community members tend to invest more in the school and the students when they see them already succeeding.

The second and final school-level characteristic that has a strong, statistically significant with one of the three scores of interest measures whether the school has clubs, a characteristic that has a statistically significant relationship with the governance subscore. Similar to hypotheses that BoMs and PTAs may lobby for more trained teachers, resulting in a relationship between the governance subscore and number of trained teachers indicator, the relationship between the presence of children’s clubs at a school and the governance subscore may be the result of lobbying by the BoMs or PTAs for children’s clubs. That said, it

could indicate a reverse causality, where better-managed schools are more likely to have more clubs and hire more trained teachers.

### **Points of Interest**

The positive correlation between highly rated school leaders and well to extremely well managed schools, and the inverse correlation between poorly rated school leaders and more poorly managed school scores seems to indicate that the calibre of a principal or head teacher may factor critically into how well that school's management is perceived.

Considering the nearly perfect nature of the management subscore and relatively low governance subscore, the mean Samburu County data point serves as another indicator of the fact that a school may be very well managed, as reported by caregivers, but lack an active or involved BoM or PTA.

## **Child Protection**

### **Predictive Analysis of Child Protection Outcomes**

In this predictive analysis, we test for predictors of three measures related to child protection: 1) whether girls are aware of incidents of violence or harm done onto children in their communities, 2) whether they generally feel that children are safe in their communities, and 3) whether they are aware of community structures that support children on protection issues. We include a series of our regressors used throughout analysis in this evaluation, such as age, disability status, life skills and decision-making scores, and whether respondents belonged to households with a female head.

**Table 67: Predictors of exposure to violent or harmful incident information, feelings of safety, and familiarity with community structures**

	Knows of Incidents		Believes Neighbourhood is Safe for Children		Believes There Are Community Structures to Support Children	
	Coef.	P-value	Coef.	P-value	Coef.	P-value
<b>Individual Level Predictors</b>						
Age	-0.02	0.48	0.01	0.38	-0.01	0.64
Disability	-0.03	0.83	0.15	0.25	0.01	0.97
Female HoH	0.00	0.99	-0.09	0.08	0.08	0.34
Life Skills	-0.04	0.03	0.00	0.85	-0.01	0.33
Decision Making	0.01	0.63	-0.02	0.38	0.00	0.89
Supports Agricultural Tasks	-0.01	0.00	0.01	0.00	-0.02	0.74
Support Business/Work for Family	-0.01	0.00	0.00	0.54	0.00	0.00
<b>School-Level Predictors</b>						
Teacher Uses Corporal Punishment	0.55	0.00	-0.29	0.00	0.34	0.04
Girl's Attendance Rate	-2.70	0.00	2.20	0.00	-0.67	0.00

Turning first to the school-level predictors, the coefficients move in the direction we expect, and all four are significant at the 99 percent confidence level. As expected, belonging to a school where a teacher had been observed using corporal punishment against students predicts a higher likelihood that a girl will be aware of violent or harmful incidents against children in their communities, especially if these acts of

bodily harm are witnessed directly by female learners.<sup>59</sup> More crucially, and interestingly, being enrolled in a school known to use corporal punishment also decreases the likelihood that girls believe the community writ large is safe for children. This suggests that although girls generally believe that violent incidents in schools are rare, as we described in an earlier sub-section, violent incidents in school leave a strong impression on girls and affects their perceptions of their communities.

We also find a positive relationship between the attendance rate of girls in school and feelings of community safety, as well as a negative relationship with knowledge of violent or harmful incidents. It is possible that girls who are in school more frequently are less exposed to community news about violent incidents against children, as a greater share of their time is used for studying and doing schoolwork, rather than engaging with community happenings. Lesser exposure to news of incidents against children may in turn prompt girls to feel that their neighbourhoods are safe for children.

Girls spending time outdoors and supporting household agricultural activities registered small but statistically significant negative effects on exposure to information, while the predictor on supporting the household's agricultural activities is also a significant predictor increased feelings of safety. This may be a case of reverse causality, in that girls who have less knowledge of violent incidents and who already feel that the community is relatively safe for children are the ones who are likelier to be more active outside their households and in their communities.

Finally, we turn to the protection measure on whether girls believe that community support structures exist to support children who have faced violent/harmful incidents. Only school-level predictors appear to be more influential in shaping girls' beliefs on this matter. Those belonging to schools where corporal punishment is used are more likely to believe that support would be available for children, and the effect of this predictor – 0.34 for a binary 0 to 1 outcome – is considerable. This correlation may be because those exposed to violent incidents, including the use of corporal punishment, may be more attentive to possible sources of support which they can use in the future, if they have not already sought such support.

However, controlling for other factors, attending more days of school may reduce the likelihood of girls and learners knowing of protection support mechanisms in their communities because the relative safety of schools compared to the workplace or outdoor agricultural areas might allow learners to devote more of their energies to their studies, rather than worrying about finding support resources. From a programmatic perspective, this might suggest that increased awareness raising of available community mechanisms to support children and youth may be valuable for learners. Even if they are relatively safe in schools compared to other environments, safety concerns still remain for young female learners, including possible threats to their safety while travelling to and from school.

## Sexual and Reproductive Health

### *SRH Awareness and Knowledge*

Girls were asked about the effects of engaging in early sexual activities and their knowledge about STDs. The below table shows responses to the question “What are the effects of engaging in early sexual activities?” Girls were allowed to state more than one effect, and indeed, all the options listed below are potential consequences; responses thus do not sum to 100 percent.

The table first shows that only a small percentage of girls—0 percent at SAS and under 4 percent at endline—stated that they did not know the consequences of early sexual activity. This shows that girls are

---

<sup>59</sup> Prior to analysis, one might consider the possibility that corporal punishment is seen as “normal” disciplinary action by students, and may not register as an incident of violence or harm done towards children, in the eyes of observing students. The findings of our predictive model suggests that these incidents are seen appropriately as negative acts of harm.



generally aware of the risks of early sex. Second, the table shows that girls most frequently stated that pregnancy is a potential consequence of early sexual activity, with over 80 percent of girls at both SAS and endline stating that this was a consequence. Interestingly, however, only a small percentage of girls—13 percent at SAS and 19 percent at endline—stated that single parenthood was also a consequence. In other words, while girls perceive a real risk of getting pregnant from early sex, they also believe that they are unlikely to raise a child conceived during early sex as a single mother. This may, in part, be due to social pressure which encourages boys (or men) to marry girls that they impregnate; however, only 35 percent of girls at SAS and 42 percent at endline stated that early marriage was a potential consequence of early sexual activity. As such, the risks of single parenthood appear to be underestimated by girls.

**Table 68: Consequences of early sexual activity**

Consequence	SAS	EL	Difference
Single parenthood	13.4%	19.0%	5.6
Dropping out of school	59.6%	57.7%	-1.9
Early marriage	34.9%	41.6%	6.7
STDs	66.4%	74.8%	8.4
Pregnancy	84.5%	82.4%	-2.1
Other	0.0%	6.5%	6.5
Don't know	0.0%	3.6%	3.6

We also note, as in the main body of the report, that while fewer girls reported that dropping out could be a consequence of early sex than pregnancy, in the qualitative, data, many girls expressed very negative views about whether a pregnant girl could return to school. One girl from Tana River stated the following:

*[She] would drop out of school because of being pregnant, her classmates would stop talking to her, and some teachers would treat her differently but others would understand. [She] would face challenges at school because she could have a lot of stress and could not concentrate on her studies and would also have morning sickness. The community would also point a finger at her due to her pregnancy and [she] might even leave school or commit suicide, she could be discriminated against and given as an example. The story would end with [her] dropping out from school, [she] could also be married off.<sup>60</sup>*

Lastly, in addition to these dynamics, we find that a somewhat higher percentage of girls stated that STDs were a potential consequence of early sexual activity at endline than at SAS. However, when subsequently asked to rate their knowledge of STDs and how they are transmitted, 93 percent of girls at SAS and 92 percent of girls at endline stated that they had “good” or “some” knowledge. This suggests that there has been little change in knowledge of STDs among project beneficiaries over the past year.

Disaggregating by county below, we find that in Turkana, girls had higher levels of knowledge than other counties, on average, about the consequences of early sexual activity and methods to prevent STDs. However, notably, only 18 percent of girls in Kwale stated that condoms could prevent STDs and only 58 percent stated that STDs could be a consequence of early sexual activity. While the reason for this

<sup>60</sup> FGD with girls, Tana River, Int. TRG3

finding in Kwale is unknown, it suggests a need to focus on strengthening SRH education in Kwale particularly.

**Table 69: SRH knowledge by county**

	Kilifi	Kwale	Marsabit	Mombasa	Nairobi	Samburu	Tana River	Turkana
<b>Consequences of early sexual activity</b>								
Single parenthood	9.7%	11.7%	43.8%	47.1%	16.7%	4.6%	28.7%	34.5%
Dropping out of school	49.2%	36.4%	68.8%	73.5%	60.4%	50.0%	50.0%	87.1%
Early marriage	25.6%	27.3%	68.8%	70.6%	31.6%	52.3%	51.8%	65.5%
STDs	82.8%	58.4%	62.5%	70.6%	80.1%	65.4%	57.3%	90.7%
Pregnancy	78.2%	63.6%	68.8%	76.5%	87.5%	80.0%	82.3%	94.2%
Other	8.8%	5.2%	3.1%	8.8%	11.3%	3.9%	0.0%	2.2%
Don't know	5.5%	6.5%	6.3%	2.9%	1.8%	6.9%	3.1%	0.0%
<b>Methods to prevent STDs</b>								
Abstinence	80.7%	75.3%	81.3%	94.1%	92.3%	81.5%	82.9%	97.1%
Use of condoms	46.2%	18.2%	40.6%	58.8%	29.2%	50.0%	39.0%	61.9%
Faithfulness	15.6%	6.5%	37.5%	38.2%	22.9%	10.0%	23.8%	55.4%
Other	7.1%	18.2%	9.4%	14.7%	7.7%	9.2%	4.9%	2.2%

### **SRH Perceptions**

The below table shows responses to the SRH perceptions questions. We first note that at endline, the highest positive response rates occurred in response to questions about whether personal cleanliness was important for reproductive health, whether pregnant girls should be allowed to return to school, whether schools should teach SRH and have supportive SRH policies, whether unprotected sex would lead to pregnancy, and whether girls should not have sex before marriage. For all of these questions, over 90 percent of girls responded positively. There was generally little change in positive response rates to these questions from SAS to endline, with the exception of the question about whether SRH education should take place in classrooms, to which 7.5 percentage points more girls agreed at endline than at SAS.

**Table 70: SRH perceptions**

	SAS	EL	Difference
"A girl should be allowed to come to the same school after delivery of a baby to complete her education" (% agree)	90.4%	93.0%	2.6
"Sexual and reproductive health education should be taught in the classroom" (% agree)	85.0%	92.5%	7.5
"Schools should have supportive adolescent and youth sexual and reproductive health policies" (% agree)	92.6%	94.3%	1.7
"I feel embarrassed talking about sex with my parents" (% disagree)	44.6%	45.4%	0.8
"Any girl who falls pregnant while still in school should be expelled" (% disagree)	73.5%	76.4%	2.9
"Girls should not have sex before marriage" (% agree)	92.5%	90.4%	-2.1
"When girls have unprotected sex, they will get pregnant" (% agree)	94.1%	93.3%	-0.8
"Personal cleanliness is important for girls' reproductive health" (% agree)	99.3%	99.2%	-0.1
"I can speak freely with friends/family about sexual and reproductive health" (% agree)	66.4%	75.7%	9.3
"It's suitable to teach unmarried youth sexual and reproductive health" (% agree)	91.6%	91.5%	-0.1
"I would like to make friends with girls who have had pregnancies while in school" (% agree)	42.2%	58.9%	16.7
"Young girls who get pregnant should be allowed to return to school and transition" (% agree)	91.9%	94.3%	2.4

In contrast, we note relatively low positive response rates to questions about whether girls were comfortable talking about sex with their parents and friends, whether girls would befriend a schoolmate who had gotten pregnant, and whether girls who had gotten pregnant should be expelled. This suggests that stigma around SRH and pregnancy remains, as also evidenced by the qualitative data referred to in the above subsection *SRH Awareness and Knowledge*. Stigma against pregnant adolescent girls was brought up repeatedly in FGDs; the following response, provided after being read a vignette about a pregnant schoolgirl, provides another example of this:

*[For her classmates,] some will be making fun of her and insulting her and telling her dirty things. Some will be insulting her, telling her that now you are a mother, leave us children alone. Some will understand and others will make her a topic for discussion. [At the school she will face] discrimination because children will be saying, this one has gotten pregnant, don't play with her or you will also get pregnant. Some will say,*

*this one is badly behaved, don't play with her or you will also possess bad behaviour.*<sup>61</sup>

Although positive response rates to these questions were low, several showed substantial increases since SAS. In particular, the percent of girls agreeing that they would befriend a pregnant girl increased by 16.7 percentage points from SAS to endline and the percent of girls saying that they could speak freely with friends and family about SRH increased by 9.3 percentage points at endline. This suggests a notable improvement in some aspects of SRH perceptions and attitudes over the past year.

### **Predictors of SRH Confidence and Perceptions**

Lastly, we analyse predictors of SRH confidence and perceptions. We focus on individual-level characteristics predicting confidence and perceptions, as the school-level variables included in evaluation data (such as teacher attendance and student-teacher ratio) were deemed to be of limited theoretical relevance to these outcomes. However, as in previous predictive analyses, we include school fixed effects.

For SRH perceptions, our outcome of interest is the SRH Perceptions Index score described above. For SRH confidence, we similarly calculate an index score based on positive responses to the six questions listed in the subsection *SRH Confidence and Self-advocacy*. For this index, we note that nearly 70 percent of girls scored 100 percent (i.e., a positive response to all six questions), and over 96 percent of students had positive answers to at least five out of six questions. This lack of variance may challenge our ability to find predictive relationships.

Results are shown in the table below. For SRH confidence, we find a significant and negative relationship between disability and this index, and a significant and positive relationship between life skills and SRH confidence. For SRH perceptions, we find significant and positive relationships between disability and perceptions, life skills and perceptions, and decision-making and perceptions. Having a female HoH is also borderline significant, with a p-value slightly over 0.05.

**Table 71: Predictive analysis of SRH confidence and perceptions**

	SRH Confidence		SRH Perceptions	
	Coef.	P-value	Coef.	P-value
n	918	-	924	-
Age	-0.2	0.48	-0.3	0.48
Minority language	0.2	0.82	0.9	0.43
Disability	-4.5	0.02	2.7	0.02
Female HoH	0.9	0.35	-1.9	0.05
Housework	-0.1	0.74	0.7	0.21
Life skills	0.7	0.001	0.8	0.002
Decision-making	0.3	0.13	0.5	0.05

The correlation between life skills and SRH confidence/perceptions is not surprising. In general, we would expect girls with stronger general self-confidence and self-esteem to be better able to advocate for healthy sexual practices and to be more comfortable learning about and discussing sex. Likewise, the

<sup>61</sup> FGD with girls, Nairobi, Int. NG2

relationship between decision-making ability and SRH perceptions is not surprising, as a girls' ability to make decisions regarding her own life may also translate into greater comfort discussing sex and an increased desire to learn about healthy sexual practices.

The relationship between disability and SRH confidence and perceptions, however, is discouraging. While having a disability is associated with higher SRH perceptions, all else held constant, it is also associated with lower SRH confidence. This suggests that although girls with disabilities may support and have benefitted from sexual education and conversations about sex, they still remain less confident about their ability to refuse unwanted sexual activity. We note that this relationship exists despite controlling for life skills and decision-making ability, suggesting that this lack of SRH confidence may be predominantly due to their disability status (or other, unobserved factors which correlate with disability).

Further examining this dynamic, Table 72 shows average responses to specific SRH confidence indicators by disability status. We find a significant differences in response rates for only one question, regarding the girl's right to complain about a schoolmate who touches her inappropriately. For this question, on average, girls with disabilities were 9.6 percentage points less likely to say they could complain than girls without disabilities.

**Table 72: SRH confidence metrics, by disability status**

	No disability	Disability	Difference	P-value
Right to complain about teacher	97.7	97.5	-0.2	0.28
Right to complain about adult	98.5	100.0	1.5	0.44
Right to complain about schoolmate	97.0	87.3	-9.6	<0.001
"I am confident that I could refuse to engage in sexual intercourse if I did not want it" (% agree)	97.1	96.2	-0.9	0.13
"I am confident that I could resist peer pressure to participate in risky behaviours such as pre-marital sex" (% agree)	95.2	94.9	-0.3	0.58
"My friends would laugh at me for refusing to have sex" (% disagree)	68.6	62.0	-6.6	0.16

Overall, these results emphasise the marginalisation faced by girls with disabilities and its resultant effects across all aspects of these girls' lives. While most girls—87 percent—still stated that they would be able to complain if touched inappropriately by a schoolmate, it is concerning that 13 percent of these girls do not feel that they have this right. This emphasises the need to continue addressing the unique issues faced by girls with disabilities.

## COVID-19 Impact

### ***Predictive Analysis of COVID-19 Impacts***

In this sub-section, we investigate whether certain school- and individual- level factors can predict the impact of COVID-19 on girl learners. Specifically, our outcomes of interested are 1) the likelihood that a girl will claim that she and her family are not at all affected by the pandemic, and 2) the likelihood that a girl will drop out before re-enrolling as a result of the pandemic. In addition to our standard individual-level predictors, we add whether a girl had received support from her schoolteacher as an individual-level

predictor, as we expect that the amount of additional support provided may impact learners' resilience against the pandemic's effects. Furthermore, we add three school-level predictors: the school's student-teacher ratio, the use of positive teaching practices, and whether teachers had previously been trained on the use of information and communications technology (ICT). We expect that the three-school level predictors all affect teachers' and school's capacity to adapt to a hybrid or fully remote model of learning and may thus help increase girls' resilience against COVID-19 shocks.

**Table 73: Effects of predictors on COVID-19 impact outcomes**

	Likelihood to Claim Not Affected at All		Likelihood to Dropout and Re-Enrol	
	Coef	P-Value	Coef	P-Value
<b>Individual Level Predictors</b>				
Age	-0.03	0.08	-0.01	0.25
Disability	-0.01	0.92	0.04	0.02
Female HoH	0.04	0.35	-0.01	0.47
Life Skills Score	-0.01	0.22	0.01	0.33
Decision Making Score	-0.02	0.16	0.00	0.78
Hours Worked on HH Chores	-0.06	0.01	-0.02	0.13
Received Support from Teacher During Closure	-0.03	0.55	0.02	0.32
<b>School-Level Predictors</b>				
Student-Teacher Ratio	0.00	0.68	0.00	0.20
Use of Positive Teaching Practices	0.18	0.43	-0.07	0.53
Teacher Received ICT Training	-0.07	0.23	-0.04	0.07

No school-level indicator registered significant effects on the likelihood of girls and their families being not affected at all by the pandemic. This is not entirely surprising, given that the impacts of COVID-19 extend beyond girls' education; economic effects on the household were by far the most frequently cited negative impact which girls and their families faced as a result of the pandemic. Improvements in the school's capacity to adapt to a remote or hybrid learning environment, while beneficial for girls' learning, is unlikely to be able to wholly shield learners from the wide array of economic shocks caused by the pandemic. In fact, economic shocks may also be a plausible explanation for why age and hours worked on household chores are both negatively and significantly correlated with the likelihood to claim not being affected by the pandemic at all. As livelihoods opportunities become more restricted due to pandemic, older learners may be more affected than their younger peers as they are expected to take on bigger roles in helping the household earn income. A similar logic may be applicable to the number of hours worked on household chores. As lockdowns increased, learners would increasingly find themselves at home and might thus be asked to do more to support the household's daily activities, making the number of hours worked on household chores an indicator of being impacted by lockdowns and the pandemic.

Turning to the COVID-19 impact measure on whether a learner is at risk of dropping out during school closures, we find two predictors to have significant effects. The first is the learners' disability status, with girls who have disabilities being slightly more likely than their peers to have dropped out at some point during closures. The difficulties which girls generally face with remote or hybrid learning permeates across different demographics in our sample, but it is not difficult to imagine that learners with disabilities may have an even steeper learning curve when trying to adapt: the switch away from in-person learning may mean that students with disabilities lose access to special needs programs or other accommodative

initiatives. This is especially true when community and household attitudes towards a girl with disability's learning is less supportive than the attitudes of teachers and other school officials: the return to household and community environments in which learners with disabilities are stigmatised or neglected compounds the already daunting challenges of home-based learning.<sup>62</sup>

The positive insight is that training teachers on ICT use can have a mitigating effect. Our model found that girls from schools where teachers were trained on ICT use had a lower likelihood of dropping out and re-enrolling. Given the emphasis in which many teachers and schools placed on using technology to disseminate lessons and engage with students, such as through the use of WhatsApp groups, an ability to use such platforms effectively would be a key factor in being able to help girls continue to learn and stay on top of class materials. While doing so is not enough to protect girls from the full brunt of the pandemic, it can have a critical impact on those most vulnerable to falling behind in their learning.

## Attitudes Towards STEM

For this analysis, we utilise data from the girls' survey conducted at SAS and endline. The table below shows positive response rates to four questions about girls' pursuit of STEM subjects. We find that girls express very positive attitudes about whether STEM subjects are meant for both girls and boys and whether they like science classes. However, attitudes about girls' performance in STEM subjects are much weaker. At endline, only around 30 percent of surveyed girls agreed that girls understand and perform better in STEM subjects than boys. Furthermore, at endline, 52 percent of girls disagreed that girls perform better in these subjects than boys, while 51 percent disagreed that girls understand these subjects better than boys.<sup>63</sup> Overall, this suggests that while girls may be interested in and enjoy pursuing STEM subjects, many girls may still lack confidence about their ability to succeed in these subjects.

**Table 74: Attitudes towards STEM subjects, SAS and EL**

	SAS	EL
"I believe science and mathematics subjects are not meant for girls but for boys" (% disagree)	87.0%	89.1%
"Girls perform better in mathematics and science subjects than boys" (% agree)	35.3%	30.1%
"Girls understand mathematics and science subjects better than boys" (% agree)	34.1%	28.1%
"Do you like science classes?" (% "yes, all" or "yes, some")	93.7%	93.0%

We also note that when asked about the specific STEM subjects they like, girls were much more likely to state that they liked biology (59.6 percent) or chemistry (44.5 percent) than physics (11.3 percent). Physics generally requires much stronger mathematics abilities than biology and, to a lesser extent, chemistry. This suggests that girls may still lack confidence in their mathematics skills.

Girls were then asked about their reasons for liking and disliking science classes. The below table shows reported reasons for liking science classes at SAS and endline; girls were allowed to state multiple reasons, so numbers do not sum to 100 percent. Interestingly, in both rounds, the most frequently selected option was that the girl finds science classes easy; at endline, over 50 percent of respondents stated this reason. While it is not surprising that girls enjoy subjects at which they feel they can excel,

<sup>62</sup> KII with Teachers, Turkana, Int. 4

<sup>63</sup> The remainder of responses were either "neither agree nor disagree" or "don't know".

given the very low STEM assessment results reported in *Learning Outcomes*, it is unexpected that many girls would state that these classes are easy.

The second most frequently stated reason at both SAS and endline was that it is easy to see how science classes will be useful for the future; at endline, around one-third of girls similarly stated that they would like to pursue a future career in the sciences. This is a positive sign, as it suggests that girls are planning for their futures and are willing to take challenging subjects in order to pursue their goals.

**Table 75: Reasons for liking STEM classes**

Girls were asked: “Thinking about your own experiences in STEM subjects, why do you like science classes?”

	SAS	EL
“I found science classes easy”	48.5%	50.4%
“It is easy to see how science would be useful for the future”	41.6%	36.6%
“I felt that I belonged in science classes”	-	12.8%
“I liked the labs and hands-on learning experiences”	14.2%	9.4%
“I had a lot of support at home or after school to help me do well in these subjects”	5.9%	7.7%
“I would like to pursue a career in the sciences”	-	32.0%

Girls were then asked why they themselves and their fellow students dislike science subjects; results are reported in the table below.<sup>64</sup> Girls were more likely to self-report that they did not dislike anything about science classes, particularly at endline; this may reflect a desire to show confidence and interest in challenging subjects to enumerators asking questions. However, girls also frequently stated that both for themselves and other students, science classes were disliked because they were hard. This runs somewhat counter to the findings above, in which around half of respondents reported that they found science classes easy; this, too, may reflect girls’ desire to appear confident during the survey.

Outside of difficulty, we note that relatively few girls reported that they or other students felt like they did not belong in science classes or felt like they did not have support to succeed in science classes. This is again a positive finding, as it suggests that biases towards girls in STEM may not currently have a large effect on the girls targeted by the programme.

**Table 76: Reasons for disliking STEM classes**

	The girl herself		Other students	
	SAS	EL	SAS	EL
“Science classes are hard”	62.8%	22.7%	60.7%	63.6%
“It is not easy to see how science would be useful for the future”	1.4%	1.6%	11.3%	5.9%
“I/Students feel that they do not belong in science classes”	1.4%	1.5%	10.9%	10.3%

<sup>64</sup> Girls were allowed to select more than one option; as such, numbers do not sum to 100 percent.



"I/Students do not like the labs and hands-on learning experiences"	6.1%	0.6%	4.3%	2.4%
"Lack of support at home or after school to help me/students do well in these classes"	3.7%	2.8%	7.9%	4.1%
"Do not dislike anything"	41.4%	70.9%	25.8%	24.1%

Lastly, girls were asked about ways to improve teaching of STEM subjects and girls' performance in these subjects. At both SAS and endline, girls most frequently responded that the use of physical demonstrations and practicals would help improve teaching; at endline, over 50 percent of girls stated this option. Girls similarly were most likely to state that the use of physical demonstrations and practicals would help improve girls' performance in STEM. Provision of more books was also frequently mentioned as a way to improve teaching; this suggests that there is still a substantial need for better provision of STEM learning materials. Additionally, around one-third of girls in both rounds stated that innovative teaching methods could also help improve girls' performance; at endline, around one-third of respondents also stated that individualised teaching of particular topics could also help improve performance. This suggests not only a need for learning resources, but also strengthened STEM teaching practices.