

Multi-Level Programming Aimed at Gender Norms Transformation to Improve Capabilities of Young Adolescents in Ethiopia: A Cluster Randomized Control Trial

Pre-analysis Plan for the Third Follow-Up, and Evaluation of Longer-Duration Programming Targeting Adolescents

Joan Hamory, Sarah Baird, Saini Das, Nicola Jones, and Tassew Woldehanna

April 25, 2023

Abstract

Adolescence is considered a ‘critical period’ of development, and a window for interventions to improve economic and social well-being. However, it is also a time when unequal gender norms become entrenched, often reinforced by peers, families, communities and broader institutional structures. Without change in norms at each of these levels, improved outcomes for adolescents may not be sustained. Furthermore, the duration needed for such interventions to shift unequal gender norms is unclear. This study employs a multi-arm cluster randomized control trial to address two sets of research questions. First, we seek to understand the added value of specific components of multi-level programming aimed at gender norm transformation on medium-term young adolescent empowerment in Ethiopia, building on earlier work that explored the short-term impacts (Jones *et al.*, 2020 and Hamory *et al.*, 2023). Second, we study the added value of extended programming provided in a cross-cutting subset of program communities, in order to unpack the value of longer duration adolescent-centric programming aimed at gender norm change.

Keywords: Gender, Adolescence, Capabilities, Norms, Empowerment

JEL codes: O15, O12, I25, I15, I32

Study Pre-registration: This study has been registered on the AEA RCT Registry (#AEARCTR-0004024, <https://www.socialscisceregistry.org/trials/4024>).

Acknowledgements: We would like to thank our colleagues at Pathfinder, Courtney Boudreau and Tricia Petruney, for their helpful comments and inputs into the design of this evaluation; our quantitative data collection partners in Ethiopia, Laterite, for their hard work in gathering the data for this evaluation; and the Ethiopia qualitative research team, including Workneh Yadete, for their valuable insights as we put together this pre-analysis plan. We also thank Dr. Benjamin Avuwadah for his help in thinking through this pre-analysis plan. The evaluation component of this study is funded by UK Aid and the Bill and Melinda Gates Foundation, and the implementation component of this study is funded by the Bill and Melinda Gates Foundation.

Contents

1	Introduction.....	3
2	Research Design.....	3
2.1	Research Questions.....	3
2.1.1	Medium-term impacts of Multi-Level Programming for Very Young Adolescents.....	3
2.1.2	Impacts of Longer-Duration Multi-Level Programming for Adolescents	6
2.2	Evaluation Sample, Randomization, and Data Collection	8
2.2.1	Interventions for Very Young Adolescents (2019-2021).....	8
2.2.2	Interventions for Older Adolescents (late 2021-2022).....	9
2.2.3	Survey Data Collection	10
2.3	Outcomes	11
2.3.1	Girls' Education Outcomes	13
2.3.2	Girls' Bodily Integrity Outcomes.....	13
2.3.3	Girls' Sexual and Reproductive Health (SRH) Outcomes	13
2.3.4	Girls' Psychosocial Wellbeing Outcomes.....	14
2.3.5	Girls' Voice and Agency Outcomes.....	14
2.3.6	Girls' Economic Empowerment Outcomes.....	15
2.3.7	Girls' Cross-Cutting Outcomes.....	15
2.3.8	Boys' Outcomes	17
3	Empirical Analysis.....	17
3.1	Regression Specification.....	17
3.1.1	Regressions to Study Medium-term impacts of Multi-Level Programming for VYA.....	18
3.1.2	Regression to Study Impacts of Longer-Duration Multi-Level Programming for Adolescents	19
3.2	Attrition and Missing Data.....	20
3.3	Handling of Outliers	21
3.4	Multiple Outcome and Multiple Hypothesis Testing.....	21
3.5	Heterogeneous Effects	21
4	Administrative Information.....	21
4.1	Funding	21
4.2	Institutional Review Board	21
4.3	Declaration of interest.....	22
5	References.....	22
6	Appendix: Secondary Outcomes.....	23

1 Introduction

This pre-analysis plan follows on Jones *et al.* (2020), which detailed planned analysis to evaluate near-term (8- and 20-month) impacts of the multi-level *Act with Her (AWH)* program aimed at gender norm transformation and empowerment of very young adolescent (VYA) girls in rural Ethiopia. We review some background information related to that VYA impact evaluation here, but refer readers to Jones *et al.* (2020) for a detailed explanation of the study motivation, interventions, sampling design, and econometric methodology. Results of that near-term impacts analysis are presented in Hamory *et al.* (2023). The current pre-analysis plan outlines planned analysis of the medium-term (3.5 year) impacts of that same *AWH* VYA program, as well as evaluation of an additional cross-cutting intervention targeting older adolescents that was launched in a random subset of program sites after data collection on the near-term impacts for the VYA study was complete.

Data collection for the analysis proposed here began in mid-October 2022 and is expected to be completed in the second quarter of 2023. The authors of this pre-analysis plan certify that they have not yet analyzed any data from this third round of data collection (which is still ongoing) at the time of writing this pre-analysis plan.

2 Research Design

2.1 Research Questions

The aim of the planned analysis is two-fold. First, we seek to understand the medium-term (3.5 year) impacts of multi-level programming aimed at gender norm transformation and empowerment of very young adolescents (VYA) in rural Ethiopia. This portion of our planned analysis follows closely on that detailed in Jones *et al.* (2020), which described the near term (8- and 20-month) impacts of this same programming. We inform the medium-term planned analysis with findings from the near-term impact analysis, which are available in Hamory *et al.* (2023). Second, we seek to understand the added value of longer intervention duration, exploiting a cross-cutting design in which a subset of VYA program communities additionally received similar programming targeting older adolescents (OA; many of whom had been participants in the previous VYA program). In what follows, we provide further details on these research questions and the specific interventions studied.

2.1.1 Medium-term impacts of Multi-Level Programming for Very Young Adolescents

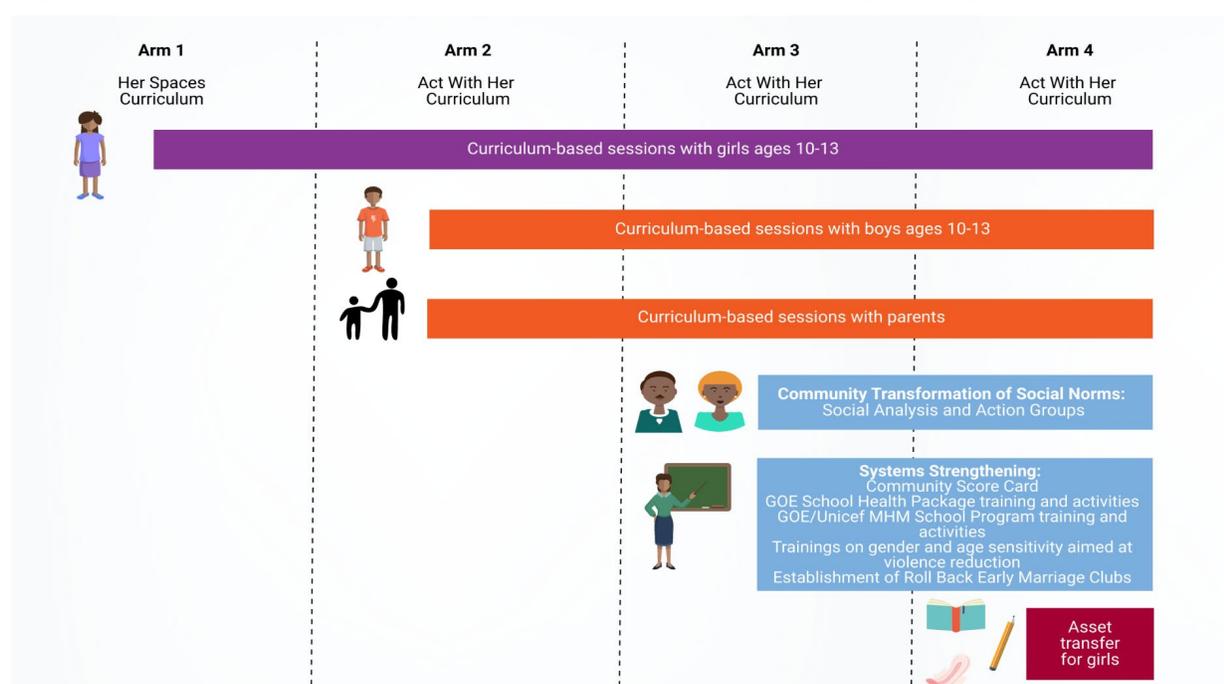
Our first set of research questions focus on the evaluation of the medium-term impacts of multi-level programming aimed at gender norm transformation and empowerment of VYA in rural Ethiopia, and unpacking of the added value of specific program components. Across two rural zones in Ethiopia (East

Hararghe Zone in Oromia, and South Gondar Zone in Amhara), 155 study sites were randomly allocated to the control group or one of the following four treatment arms:¹

- (i) *Her Spaces* arm: curriculum-based group meetings for girls (aged 11-13) only;
- (ii) *AWH Essential* arm: curriculum-based group meetings for girls (aged 11-13), for boys (aged 11-13), and for their parents;
- (iii) *AWH Comprehensive* arm: the adolescent and parent group meetings, plus community-level systems strengthening;
- (iv) *AWH Comprehensive+* arm: the adolescent and parent group meetings, and community-level engagement, with the addition of in-kind transfers to girls.

Figure 1 provides a summary of these treatment arms.

Figure 1. Summary of Research Design for Very Young Adolescent Programming



Note: This figure was provided by Pathfinder International. Note that although curriculum based programming for adolescents was provided by Pathfinder and CARE for all adolescents living in a study site who were aged 10-13 and wanted to participate in the program, our analysis focuses only on the subset of adolescents who were randomly selected from a project-specific census style household listing, and who were aged 11-13 at the time the adolescent groups were launched.

¹ Treatment arm names have been updated since the writing of the Jones *et al.* (2020) pre-analysis plan, in order to provide more descriptive and meaningful titles. What we now refer to as *AWH Essential* was previous called *AWH-E Simple*. What we now refer to as *AWH Comprehensive* was previous called *AWH-E*. And lastly, what we now refer to as *AWH Comprehensive+* was previous called *AWH-E + Asset Transfers*. Importantly, the content of these treatment arms has not changed.

Participating girls living in sites allocated to the *Her Spaces* arm were engaged in 10 months of weekly near-peer mentor-led group sessions (40 sessions in total) following the *Her Spaces* curriculum, which covers a range of topics including nutrition, puberty and menstrual health, relationships, negotiation skills, harassment and safety in the community, community services, financial management, and creating an aspirational plan.² Participating girls and boys in all other treatment sites were engaged in 10 months of near-peer mentor-led group sessions following the *Act With Her (AWH)* curriculum, which was designed by Pathfinder International and others to build on the *Her Spaces* curriculum; it includes many of the same topics but goes into more depth on several issues – particularly sexual and reproductive health, sexual and gender-based violence and harmful traditional practices. *Act With Her* also places much more emphasis on changing attitudes and norms around gender equality. Girls in these groups met weekly for 40 sessions, while boys met approximately twice per month for a total of 18 sessions. Six caregiver sessions were held over the 10-month period in order to orient parents to topics covered in the adolescent curricula, and to help them create a supportive home environment for their adolescent child.

Community-level engagement in the *AWH Comprehensive* and *AWH Comprehensive+* treatment arms was operationalized by: (1) bringing together key decision-makers and stakeholders from the community for regular, structured meetings led by trained facilitators to establish locally led mechanisms for discussing social norms in ways that initiate shifts over time; and (2) enhancing local capacity for social accountability through increased participation, accountability and transparency between service users, providers and decision-makers. These meetings focused on: (1) enhancing menstrual hygiene management and violence reduction within schools; (2) improving local child protection and measures to counter harmful traditional practices; (3) fostering better linkages between school, health and social protection services; and (4) enhancing social systems and norms. Community group meetings were launched approximately 2 months after the first adolescent groups started, and continued for approximately 2 years (though implementation was disrupted for several months due to pandemic-related closures and restrictions).

The *AWH Comprehensive+* arm implemented the full *AWH Comprehensive* program, but with the addition of in-kind transfers to the participating girls. The transfers consisted of three equal-value (\$115) supply package options: one including school supplies, one including personal hygiene supplies, and one that is a combination of the first two. Each girl chose the package she wanted to receive within the first weeks of the group meetings (those who did not choose were assigned the combination package), and received three deliveries of that package over the course of the first 10 months of the project.

² The *Her Spaces* curriculum was developed through a collaboration between the Ethiopian Federal Ministry of Health and the international non-profit organization, Girl Effect. It was piloted with approximately 2,000 girls in four regions of Ethiopia (including our study regions) prior to the launch of the present study.

Using data collected for this impact evaluation, the proposed analysis seeks to answer six research questions:

- (1) What is the impact at 3.5 years post-program start of a multi-level program aimed at gender norm transformation on younger adolescent capability achievements and transitions? (compare all *AWH* program variations to control)
- (2) What is the impact at 3.5 years post-program start of a program aimed only at adolescent girls? (compare *Her Spaces* to control)
- (3) How do the 3.5 years impact of the multi-level, gender synchronized program compare with a program aimed only at adolescent girls? (compare all *AWH* program variations to *Her Spaces*)
- (4) Does the inclusion of boys and parents enhance the 3.5 year impacts of programming for younger adolescent girls? (compare *AWH Essential* to *Her Spaces*)
- (5) Does the inclusion of community norms transformation and systems strengthening activities enhance the 3.5 year impacts of gender-synchronized programming for younger adolescents and their parents? (compare *AWH Comprehensive* to *AWH Essential*)
- (6) Does the inclusion of economic support in the form of in-kind transfers enhance the 3.5 year impacts of multi-level programming for younger adolescents and their communities? (compare *AWH Comprehensive+* to *AWH Comprehensive*)

The precisely defined hypotheses we will study are defined in Section 3.1 below.

2.1.2 Impacts of Longer-Duration Multi-Level Programming for Adolescents

We additionally seek to understand the added value of longer duration multi-level gender-synchronized adolescent-centric programming. Here, we exploit a cross-cutting design in which, after the close of the second round of follow-up data collection for the VYA impact evaluation (described in Jones *et al.* (2020) and analyzed in Hamory *et al.* (2023)), a random half of *AWH Essential*, *AWH Comprehensive*, and *AWH Comprehensive+* communities received a program package similar to *AWH Comprehensive* that targeted older adolescents (OA; aged 14-18 at the time of program launch). Figure 2 provides a summary of this cross-cutting research design.

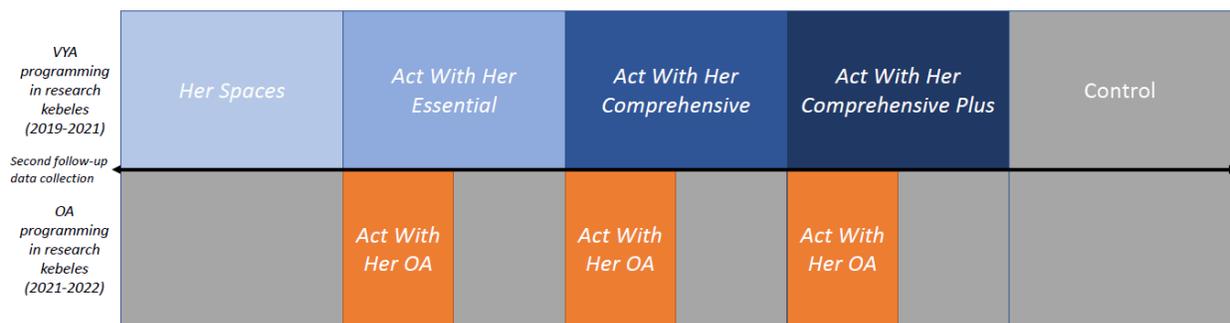
Communities assigned to this additional OA intervention received an additional 25 weeks of near-peer mentor-led weekly group meetings for older adolescent girls (aged 14-18), the same for older adolescent boys (aged 14-18), and community level systems strengthening work for 12 months. The systems strengthening work was similar to that from the VYA *AWH Comprehensive* intervention, with one notable exception. In Ethiopia, sexual and reproductive health (SRH) services are primarily provided through health centers (which serve multiple kebeles in a woreda) rather than through health posts (which generally serve a single kebele). Because SRH was seen as a key part of the systems-strengthening work

for interventions focused on older adolescents, that work happened at a level higher than the kebele, and could potentially impact surrounding kebeles that were part of the study control group.

The curriculum used in the older adolescent group meetings was based on the AWH VYA curriculum, but updated with learning from the VYA work as well as some new topics relevant to older adolescents. The shorter length of programming for this age group was deemed appropriate due to the higher demands on their time, and providing boys and girls with the same number of meetings was a response to fairness issues raised in some communities during the earlier VYA program.

Importantly, it should be noted that many of the adolescents who participated in the earlier VYA programming additionally participated in this intervention, so that some (but not all) of these older adolescents actually received both VYA and OA programming. However, there are also adolescents who did not participate in the VYA programming, but did participate in the OA programming. Thus, the “duration” impacts we are able to measure experimentally are community-level duration of programming, rather than adolescent-level duration of programming. We will conduct exploratory secondary analysis in order to measure the latter using quasi-experimental methods.

Figure 2. Summary of Research Design for Older Adolescent Programming



Using this cross-cutting randomized intervention, we are able to additionally causally answer the following research questions:

- (7) Does three years of a multi-level, gender synchronized program targeting first VYA and then OA lead to sustained impacts on adolescent capability achievements and transitions? (compare *OA* to control, where the ‘control’ are OA age-eligible adolescents living in the control communities)
- (8) Does three years of a multi-level, gender synchronized program targeting first VYA and then OA lead to larger impacts on adolescent capability achievements and transitions than two years of similar programming targeting VYA only? (compare AWH communities not assigned to *OA* to AWH communities assigned to *OA*)

The precisely defined hypotheses we will study are defined in Section 3.1 below.

2.2 Evaluation Sample, Randomization, and Data Collection

2.2.1 Interventions for Very Young Adolescents (2019-2021)

Jones *et al.* (2020) provides a detailed summary of study site selection, randomization, adolescent sampling, and statistical power for the study of the initial set of interventions for VYA. We provide a short overview here, but refer readers to that pre-analysis plan for more details.

In late 2017, 155 kebeles (communities) across 10 purposely selected woredas in South Gondar, Amhara and East Hararghe, Oromia were block randomized (by woreda, and kebele marginalization status) into control and four ‘hypothetical’ treatment arms. Age-eligible adolescents were identified through a household census (listing) that was performed in each of these kebeles, and a total of 15 girls and 11 boys aged 10-12 were randomly sampled in each site from this census list for participation in the research. Baseline data collection was conducted during November 2017 through February 2018, in order to collect information from adolescents, their caregivers, and their households. After accounting for (randomly chosen) replacements, a total of 3,962 individuals (2,281 girls and 1,681 boys) completed a baseline interview (99% of the selected sample). This sample, which we call ‘Sample 1’, is used to study the medium-term impacts of the VYA programming.

Each of the four hypothetical treatment arms was later assigned to a treatment group – *Her Spaces*, *AWH Essential*, *AWH Comprehensive*, or *AWH Comprehensive+* – and interventions were launched in treatment sites in March 2019. At this time, the research sample of adolescents were aged 10-13, and thus fully eligible for programming. The non-governmental organization (NGO) Pathfinder managed implementation of all program variations in South Gondar, and CARE Ethiopia managed implementation of all program variations in East Hararghe. Programme recruitment was done separately from the research study recruitment, but also used a household listing methodology aiming to locate all adolescent girls and boys (where appropriate for the intervention arm) aged 10–13. Staff implementing the programme aimed to include as many eligible adolescents from each community as possible, and so most communities had two girls’ groups and two boys’ groups (where applicable), with up to 35 members each. Approximately 84% of girls in the research sample living in treatment communities in South Gondar were enrolled in programming, and 81% of boys. Recruitment rates among the research sample in East Hararghe were somewhat lower, with 64% of girls enrolling, and 68% of boys. Weekly adolescent group meetings began in March 2019 and continued for 10 months, with parent group meetings starting shortly thereafter. Community-level and higher-level systems strengthening began by June 2019. Although this component was meant to continue for 24 months, in practice there was disruption for a number of months due to pandemic-related closures. The adolescent and parent group meetings were completed by January 2020, so were not affected by those closures.

Two rounds of follow-up data were collected to evaluate the near-term impacts of this VYA programming. The first round was conducted from November 2019 to March 2020, just as adolescent group meetings in the treatment sites were wrapping up. This data collection round provides evidence on the 10-month impacts of layered programming for VYA.³ All sampled adolescents and their caregivers were sought for re-interview, and attempts were made to track adolescents no longer living at their baseline residence. Follow-up survey data was collected for 89% of the adolescent sample; refusal rates were low (2.5%), and most of those who were not interviewed were simply unable to be found (7%, most likely due to migration). A second round of follow-up data collection was conducted in two waves, from March to May 2021 and from October 2021 to May 2022.⁴ The research sample was randomly divided into these two waves in expectation of work stoppages around a national election that was scheduled for mid-2021. Furthermore, a two-stage tracking methodology was implemented in this data collection round in order to limit bias due to survey attrition. This second follow-up provides evidence on the 24-to-36-month impacts of layered programming. As before, all sampled adolescents were sought for re-interview, and attempts were made to track adolescents no longer living at their baseline residence. Follow-up survey data was collected for 88% of the adolescent sample; refusal rates were low (3.3%), and most of those who were not interviewed were simply unable to be found (8.5%, again, most likely due to migration). Analysis of data from these first two follow-up rounds is presented in Hamory *et al.* (2023). The present pre-analysis plan lays out plans for analysis of a third round of follow-up survey data collection, described in detail in Section 2.2.3.

2.2.2 Interventions for Older Adolescents (late 2021-2022)

In anticipation of additional interventions targeting older adolescents that were being considered in a subset of the original treatment sites described in Section 2.2.1, the research team recruited a cohort of approximately 500 older adolescents to the study in late 2019 and early 2020. For this additional study cohort, adolescents aged 15-17 living in all study kebeles were recruited purposefully, rather than randomly, due to budget constraints. Pre-intervention information was collected on these adolescents and their households. The subset of these individuals who were aged 18 or younger at the start of AWH OA programming in each zone, which we call ‘Sample 2’, is used in combination with ‘Sample 1’ (described above) to study the impacts of longer duration programming.

³ The pre-analysis plan for this near-term impacts study (Jones *et al.*, 2020) specifies that near-term impacts will be measured at 8 months post-programme launch; data collection was actually conducted over a 2-month period, so 8–10 months post-programme launch. We refer to the 10-month end point of the follow-up data collection here for brevity.

⁴ This second follow-up data collection to study near-term program impacts was conducted later than originally envisioned in Jones *et al.* (2020), due to delays related to the COVID-19 pandemic as well as armed conflict in Ethiopia.

Once it became clear that our NGO partners were going to be able to implement multi-level programming targeting older adolescents (aged 14-18) in some communities, the research team randomly selected 50% of *AWH Essential*, *AWH Comprehensive*, and *AWH Comprehensive+* communities (additionally blocked on woreda, or district) to receive this programming. We refer to the program that was launched as *AWH-OA*, and the details of this program are described in Section 2.1.2 above. A total of 45 rural kebeles were assigned to the *AWH-OA* treatment (in addition to the AWH VYA treatment variation already received), and the remaining 42 rural kebeles are kept with their original VYA-only treatment status.⁵

AWH-OA launched in selected sites in East Hararghe in November 2021, and in sites in South Gondar in January 2022. At that time, adolescents in the original randomly sampled research cohort (of 3,962 adolescents) were primarily between the ages of 13-16, and many were thus eligible to participate in this OA programming. Some, but not all of them, would have already participated in the VYA programming. In addition, the purposefully recruited additional older adolescents were primarily aged 17-21. These adolescents would not have been eligible for the VYA programming in their kebeles, but many of them were eligible for the new OA programming.

2.2.3 Survey Data Collection

The primary source of data used for this analysis will be survey data collected through face-to-face individual interviews with adolescents. In particular, we will employ four rounds of data collection in our analysis (a baseline and three post-programming follow-up rounds). Jones *et al.* (2020) describes data collection for the baseline and first two follow-up rounds, and Hamory *et al.* (2023) presents the results of analysis of these two follow-ups (which measured the near-term impacts of the VYA programming, and were completed prior to the launch of any *AWH-OA* programming). Here, we focus on planned analysis for the third follow-up round, which was launched as the *AWH-OA* programming was coming to a close.

The third follow-up round of data collection was launched in October 2022, and is currently ongoing. Our data collection partner in Ethiopia, Laterite, is seeking to conduct face to face interviews with all adolescents who were part of the original random sample of young adolescents (recruited in late 2017 and early 2018), as well as the newer sample of older adolescents who were purposefully recruited in late 2019 and early 2020. Adolescents who have moved from their baseline location will be tracked, and data collection will be undertaken using the two-stage tracking methodology described in Baird *et al.* (2016) in

⁵ Note that one of the 45 “treatment” kebeles will not be used in the quantitative impact evaluation analysis. This kebele was assigned to OA treatment status *purposefully* rather than *randomly*, in order to provide a marginalized kebele where qualitative data was already being collected for a more nuanced understanding of the OA programming sites. Thus, 44 out of the 45 treatment kebeles will be include in the impact evaluation analysis, as well as all of the non-treatment kebeles.

order to reduce concerns from survey attrition (this methodology was also used during the second round of follow-up data collection).

Laterite uses experienced survey enumerators who speak the local languages of the study areas, and were well-trained in interviewing adolescents. Survey enumerators are closely supervised to ensure the quality of their work. Daily telephonic communications between field supervisors and the in-country data collection partner monitor the progress and quality of the data collection. Interviews are conducted on tablet computers using the SurveyCTO software, with functionality that checks for internal validity and reasonableness of responses. Interviewers ensure completion and accuracy of the survey before terminating each interview and transmitting the data back to the data collection firm. At the end of each day, the field supervisor have de-briefing meetings with interviewers and discuss any problems or shortcomings, and they take necessary actions in consultation with the data collection firm and the research team.

Once data collection is complete, the in-country data collection firm and the research team will review the data collected during field visits for completeness and consistency. Back checks will be conducted for a targeted subset of the sample and will include re-interviews asking a subset of questions to respondents (for an interview lasting 5-10 minutes per respondent), generally conducted by supervisors. Basic data cleaning and management will be performed by the research team; this will primarily include checks on uncoded answers to questions with coded response sets.

Data collection is expected to be completed in the second quarter of 2023. The authors of this pre-analysis plan certify that they have not yet analyzed any data from this third round of data collection (which is still ongoing) at the time of writing this pre-analysis plan.

For secondary analysis, we will also receive adolescent group attendance information from Pathfinder and Care.

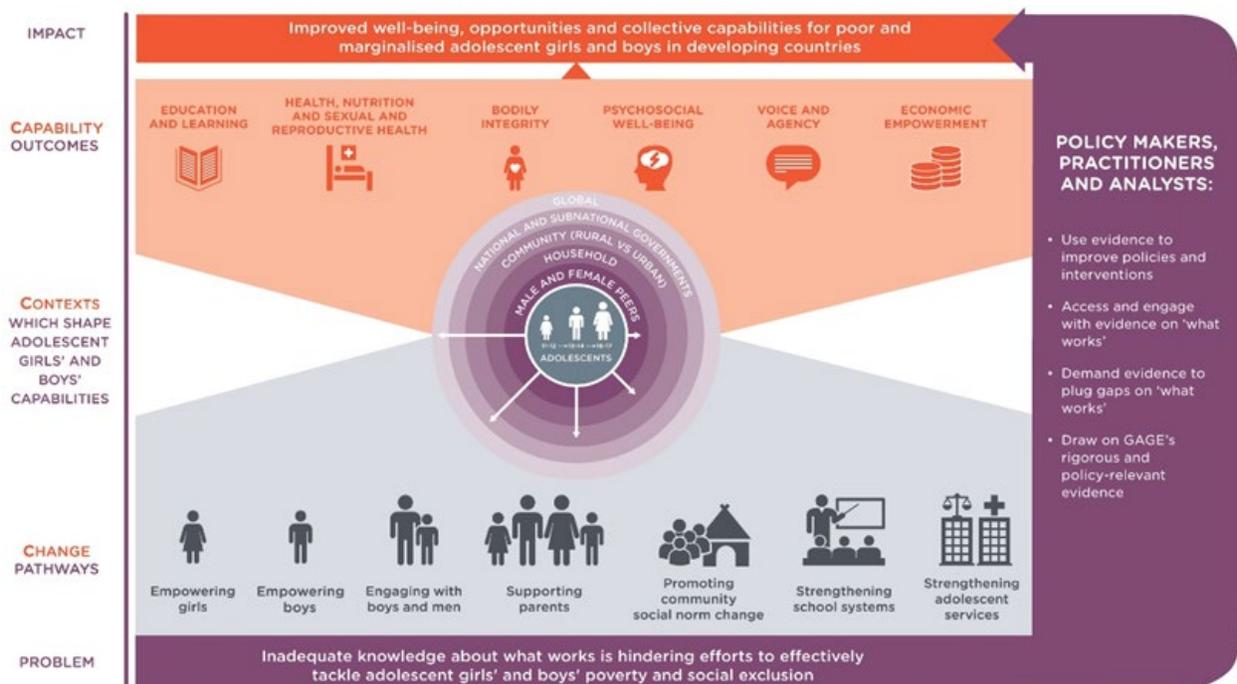
2.3 Outcomes

In addition to answering the research questions we have already discussed, this study will add to the evidence base on adolescent-centric programming by moving beyond simply measuring gains in education, health, and/or income-generating activities, and instead attempting a more comprehensive measure of adolescent empowerment embedded in the *Gender and Adolescence: Global Evidence* (GAGE) conceptual framework (GAGE Consortium, 2017) (Figure 3). Gender inequities and power imbalances in families, communities, and institutions drive poor outcomes—especially for girls—across all aspects of physical health, education, bodily integrity, psychosocial well-being, voice and agency, and economic empowerment, the six primary domains within the GAGE framework that will form the basis of our empowerment measure.

Primary and secondary outcomes at approximately 3.5 years post-program start (58 months post-baseline) will be measured for adolescent girls across the six GAGE capability domains and a set of cross-cutting themes (including gender, support, and knowledge). In addition, a number of outcomes across these same domains and themes will be measured at 3.5 years post-program start for adolescent boys. In what follows, we present primary outcomes by GAGE capability area and cross-cutting theme as measured in the quantitative survey. Complementary qualitative data is also being collected, with the aim of triangulating and nuancing responses (see Jones et al., 2017). It is important to note that the measures included were informed by qualitative research undertaken during 2017-2022 in order to strengthen context-relevant questions. Additionally, we frame our primary outcome measures around impacts we could reasonably expect to see from the gender-sensitive life skills training that we study.

Primary outcomes are indicated below in *bold italics* (and numbered). We will consider components of the primary outcome indices as secondary outcomes and may do exploratory analysis on additional secondary outcomes not presented here. A planned set of secondary outcomes is presented in Appendix A.

Figure 3. GAGE Conceptual Framework



To construct indices, we employ the following procedure: a) for each index component, create a normalized measure by subtracting the control group mean and then dividing by the control group standard

deviation; b) construct the index by calculating the raw mean across all normalized component variables, and c) create the normalized index value by subtracting the control group mean of the index and then dividing by the control group standard deviation of the index. We have signed outcomes such that higher values of indices indicate ‘better’ outcomes.

2.3.1 Girls’ Education Outcomes

1. ***Index of educational attainment:*** Index consists of two components:
 - a. Indicator for ever enrolled in secondary school (*CR Module, Education, Q4b and Q5*)
 - b. Highest grade attended (zero for kindergarten or less) (*CR Module, Education, Q4b and Q5*)

2.3.2 Girls’ Bodily Integrity Outcomes

2. ***Index of exposure to violence:*** Index consists of three components:
 - a. Peer violence scale: ranges from 0-6; constructed as the sum of indicators for has experienced –at least once in past 12 months– either in person or digitally, violence from peers, including: (i) hurtful words, (ii) exclusion, (iii) theft/damage of personal property, (iv) physical violence, (v) being forced to do something, or (vi) threats to self or someone close to self; inverse so higher values indicate better outcomes (*CR Module, Violence, Q1i-vi*)
 - b. Indicator for did not experience violence in the household, or witness violence against female caregiver, in past 12 months (*CR Module, Violence, Q3-7*)
 - c. Indicator for did not experience sexual violence in past 12 months (*CR Module, Violence, Q11a and Q12a*)
3. ***Indicator for never married*** (*CR Module, Marriage and Relationships, Q1 and 1b and previous surveys*)

2.3.3 Girls’ Sexual and Reproductive Health (SRH) Outcomes

4. ***Index of menstrual hygiene:*** Measured among those who have reached menarche. Index consists of two components:
 - a. Indicator for self-reports normal activities not affected during menstruation (*CR Module, Sexual and Reproductive Health, Q4*)
 - b. Index of improved menstrual hygiene practices: ranges from 0-2; constructed as the sum of indicators for (i) has improved menstrual hygiene management (uses re-usable pad, sanitary pad, or other modern sanitary product) and (ii) practices appropriate disposal at home (latrine, rubbish heap, burn, or wash and reuse if a reusable pad) (*CR Module, Sexual and Reproductive Health, Q3 and Q3a*)

5. **Index of SRH:** Index consists of two components:
 - a. Indicator for never had sex (*CR Module, Sexual and Reproductive Health, Q21*)
 - b. Indicator for not ever pregnant (zero if never had sex) (*CR Module, Sexual and Reproductive Health, Q23*)

2.3.4 Girls' Psychosocial Wellbeing Outcomes

6. **Patient Health Questionnaire-9 (PHQ-9) score** (measure of psychosocial distress; inverse) (*CR Module, Psychosocial and Mental Health, Q1-9*)
7. **Child and Youth Resilience Scale score** (*CR Module, Psychosocial and Mental Health, Q12-23*)
8. **Locus of Control** (*CR Module, Physical Health and Nutrition, Q2*)

2.3.5 Girls' Voice and Agency Outcomes

9. **Index of voice and agency:** Index consists of three components:
 - a. Index of say in household (0-8, indicators for has a little bit or a great deal of say in how much time spent helping around the house, how much education will get, when to marry, who to marry, who to be friends with, what to do in free time, whether to leave the house, whether to be involved in income generating activities) (*CR Module, Mobility, Voice, and Agency, Q6a-6h*)
 - b. Index of comfort expressing oneself: index consists of two components:
 - i. Indicator for comfortable expressing opinion with agemates (*CR Module, Mobility, Voice, and Agency, Q3*)
 - ii. Indicator for comfortable expressing opinion with those who are older (*CR Module, Mobility, Voice, and Agency, Q4*)
 - c. Index of voice: composed of six measures, including:
 - i. "My parents or guardians ask for my opinion on things" (0-2, more is better) (*CR Module, Mobility, Voice, and Agency, Q3a*)
 - ii. "My parents or guardians listen when I share my opinion" (0-2, more is better) (*CR Module, Mobility, Voice, and Agency, Q3b*)
 - iii. "My friends ask my advice when they have a problem" (0-2, more is better) (*CR Module, Mobility, Voice, and Agency, Q3c*)
 - iv. "If I see something wrong in school or the neighborhood, I feel I can tell someone and they will listen" (0-2, more is better) (*CR Module, Mobility, Voice, and Agency, Q3d*)
 - v. "I can speak up when I see someone else being hurt" (0-2, more is better) (*CR Module, Mobility, Voice, and Agency, Q3e*)

- vi. “I can ask adults (others) for help when I need it.” (0-2, more is better) (*CR Module, Mobility, Voice, and Agency, Q3f*)

2.3.6 Girls’ Economic Empowerment Outcomes

- 10. ***Index of economic empowerment:*** Index consists of three components:
 - a. Indicator for had money she controls in past 12 months (*CR Module, Financial Inclusion and Economic Empowerment, Q1*)
 - b. Indicator for paid work in the past 12 months (*CR Module, Paid Work, Q1*)
 - c. Indicator for has a goal she would like to achieve in the next year (*CR Module, Financial Inclusion and Economic Empowerment, Q2*)
- 11. ***Index of economic aspirations:*** Index consists of two components:
 - a. Indicator for aspires to be employed in skilled or professional work when adult (*CR Module, Financial Inclusion and Economic Empowerment, Q3*)
 - b. Indicator for aspires to have employment or own a business when adult (*CR Module, Financial Inclusion and Economic Empowerment, Q3*)

2.3.7 Girls’ Cross-Cutting Outcomes

- 12. ***Index of gender equitable attitudes:*** Index consists of two components:
 - a. Global Early Adolescent Study’s Index of Gender Stereotypical Traits⁶: Index consists of five components, including indicators for:
 - i. Disagrees with “Girls should avoid raising their voice to be lady like” (*CR Module, Education, Q18*)
 - ii. (Partially) agrees with “Boys should be able to show their feelings without fear of being teased” (*CR Module, Education, Q19*)
 - iii. Disagrees with “Boys who behave like girls are considered weak” (*CR Module, Education, Q22*)
 - iv. Disagrees with “Girls need their parents’ protection more than boys” (*CR Module, Education, Q21*)
 - v. Disagrees with “Boys should be raised tough so they can overcome any difficulty in life” (*CR Module, Education, Q20*)
 - b. Global Early Adolescent Study’s Index of Gender Stereotypical Roles: Index consists of seven components, including indicators for:

⁶ For more information on the GEAS, see geastudy.org.

- i. (Partially) agrees with “Women should have the same chance to work outside of the home as men” (*CR Module, Education, Q26*)
 - ii. (Partially) agrees with “Girls and boys should share household tasks equally” (*CR Module, Education, Q23*)
 - iii. Disagrees with “A woman’s most important role is to take care of her home and cook for her family” (*CR Module, Education, Q24*)
 - iv. Disagrees with “A man should have the final word on decisions in his home” (*CR Module, Education, Q25*)
 - v. Disagrees with “A woman should obey her husband in all things” (*CR Module, Sexual and Reproductive Health, Q16*)
 - vi. Disagrees with “It is okay to tease a girl who acts like a boy” (*CR Module, Sexual and Reproductive Health, Q12*)
 - vii. Disagrees with “It is okay to tease a boy who acts like a girl” (*CR Module, Sexual and Reproductive Health, Q13*)
13. **Index of supportive network:** Index consists of three components, including: indicators for has female friend can trust and talk to, has male friend can trust and talk to, and has an adult can trust and talk to (*CR Module, Social Inclusion, Q1-2*)
14. **Index of service knowledge:** Index consists of four components, including: indicators for knows a place where services could be received for alcohol and drug addiction (*CR Module, Health and Nutrition, Q6-6a*), mental health (*CR Module, Psychosocial and Mental Health, Q11-11a*), pregnancy prevention (*CR Module, Sexual and Reproductive Health, Q10-10a*), violence (*CR Module, Violence, Q10-10a*)
15. **Index of service accessibility:** Index consists of four components, including: indicators for believes an adolescent in his/her area could access services for alcohol and drug addiction (*CR Module, Health and Nutrition, Q6b*), mental health (*CR Module, Psychosocial and Mental Health, Q11b*), pregnancy prevention (*CR Module, Sexual and Reproductive Health, Q10b*), violence (*CR Module, Violence, Q10b*)
16. **Knowledge index:** Index consists of ten components, including indicators for correct answers to the following:
- a. Approximately how often does menstruation usually occur? By menstruation, I mean the bleeding girls get when they reach a certain age. (Correct = “once per month” or similar response) (*CR Module, Sexual and Reproductive Health, Q6*)
 - b. True or False: The beginning of menstruation means that girls have the ability to become pregnant. (Correct = true) (*CR Module, Sexual and Reproductive Health, Q7*)

- c. True or False: Getting pregnant early in puberty can be bad for the health of the girl and the baby. (Correct = true) (*CR Module, Sexual and Reproductive Health, Q8*)
- d. At what age is it legal for women to get married in Ethiopia? (Correct = 18) (*CR Module, Marriage and Relationships, Q20-20a*)
- e. At what age is it legal for men to get married in Ethiopia? (Correct = 18) (*CR Module, Marriage and Relationships, Q21-21a*)
- f. Does the practice of FGM/C have risks? (Zero for those who have never heard of FGM/C; Correct = “yes”) (*CR Module, Sexual and Reproductive Health, Q30 and Q30a*)
- g. Where can people seek help if they are experiencing violence? (Correct = coded response, 0 if DK a place) (*CR Module, Violence, Q10 and Q10a*)
- h. If you wanted to save money, where could you keep that money, other than a safe space at home? (Correct = coded response) (*CR Module, Financial Inclusion and Economic Empowerment, Q4*)
- i. Which of these things is *not* part of negotiation skills? (Correct = “don’t let the other person have a chance to say their point of view”) (*CR Module, Mobility, Voice and Agency, Q7*)
- j. True or False: Boys are biologically smarter than girls. (Correct = “false”) (*CR Module, Education, Q27*)

2.3.8 Boys’ Outcomes

1. ***Index of gender equitable attitudes:*** (as described above)
2. ***Index of violence exposure and perpetration:*** composed of indicator for not perpetrating peer violence, and peer violence index (as described above)
3. ***Patient Health Questionnaire-9 Score*** (as described above)
4. ***Indicator for has a supportive adult*** (as described above)
5. ***Index of service knowledge*** (as described above)
6. ***Index of service accessibility*** (as described above)
7. ***Knowledge index*** (as described above)

3 Empirical Analysis

3.1 Regression Specification

The main analysis will focus on the intention to treat (ITT) program impacts in Amhara and Oromia across the set of primary outcomes listed in Section 2.3, in order to answer the research questions outlined in

Section 2.1. Our main specification will explore impacts on Amhara separately from Oromia, though we will additionally provide results for both regions combined. We will utilize linear probability models in the case of binary outcomes.

3.1.1 Regressions to Study Medium-term impacts of Multi-Level Programming for VYA

To answer our first six research questions, we will use the following simple reduced form linear models, studying only the “Sample 1” individuals (randomly selected VYA recruit prior to the launch of the VYA programming):

$$y_{ic} = \alpha_1 + \beta_1 HS_c + \beta_2 AWH_c + \beta_3 OA_c + X'_{ic} \beta_4 + \varepsilon_{1,ic} \quad (1)$$

$$y_{ic} = \alpha_2 + \gamma_1 HS_c + \gamma_2 AWH-Essential_c + \gamma_3 AWH-Comprehensive_c + \gamma_4 AWH-ComprehensivePlus_c + \gamma_5 OA_c + X'_{ic} \gamma_6 + \varepsilon_{2,ic} \quad (2)$$

where y_{ic} is the outcome of interest for individual i in community c , HS_c is a binary indicator for being in a *Her Spaces* community, AWH_c is a binary indicator for being in a community having any version of the *Act With Her* model, and $AWH-Essential_c$, $AWH-Comprehensive_c$, and $AWH-ComprehensivePlus_c$ are binary indicators of community assignment to a specific *Act With Her* program variation described in Section 2.1.1. OA_c is a binary indicator for being in a community assigned to the *AWH-OA* treatment, which we include as a control in regressions (1) and (2), but are not directly interested in here.⁷ The standard errors ε_{ic} are clustered at the kebele level, which account for both the design effect of the kebele-level treatment and the heteroskedasticity inherent in the linear probability model. In the basic model, X_{ic} includes block indicators used in the randomization (where blocks combine woreda and marginalization status), adolescent age (in years), an indicator for whether there were multiple eligible adolescents in the household, and month of survey indicators. We will also provide results for an expanded control set that includes additional baseline household characteristics of interest, including: household size, an indicator for household head literate, an indicator for female headed household, household asset index, and an indicator for household receives PSNP benefits.⁸ Following McKenzie (2012), we will control for baseline measures of the outcome of interest (denoted by $y_{ic,0}$) when available. Gender and community specific sampling weights will be used to make the results representative of the target population in the study area.

For these first six research questions specified in Section 2.1.1, we are primarily interested in six hypothesis tests. First, we seek to understand whether programming for girls only (*Her Spaces*), and

⁷ We will also run the regressions as specified in equations (1) and (2), but including only the random subset of sites that did not receive an OA interventions. These regressions will have lower power, so we do not them as our primary analysis.

⁸ We will also explore the possibility of integrating machine learning for choice of control variables.

whether the gender-synchronized, multilevel *Act With Her* programming for girls, boys, and parents, impact adolescent outcomes.

(1) $\beta_2 \neq 0$: comparing *AWH* (all variations) the control group

(2) $\beta_1 \neq 0$: comparing *Her Spaces* to the control group

Then we seek to understand any added value of the former programming relative to the latter:

(3) $\beta_1 \neq \beta_2$: comparing *Her Spaces* to *AWH* (all variations)

We are also interested in understanding the added value of specific components of the gender-transformative programming:

(4) $\gamma_1 \neq \gamma_2$: comparing *Her Spaces* to *AWH-Essential*

(5) $\gamma_2 \neq \gamma_3$: comparing *AWH-Essential* to *AWH-Comprehensive*

(6) $\gamma_3 \neq \gamma_4$: comparing *AWH-Comprehensive* to *AWH-Comprehensive+*

3.1.2 Regression to Study Impacts of Longer-Duration Multi-Level Programming for Adolescents

To answer the final two research questions related to longer duration programming (questions (7) and (8) in Section 2.1.2), we will use the following simple reduced form linear model, studying the “Sample 1” individuals (randomly selected VYA recruited prior to the launch of the VYA programming) and the “Sample 2” individuals (purposefully selected OA recruited prior to the launch of the OA programming, and aged ≤ 18 at start of OA programming) living in control and *AWH* treatment sites only:⁹

$$y_{ic} = \alpha_3 + \theta_1 AWH_c + \theta_2 OA_c + X'_{ic} \theta_3 + \varepsilon_{3,ic} \quad (3)$$

where y_{ic} is the outcome of interest for individual i in community c , AWH_c is a binary indicator for being in a community having any version of the *Act With Her* VYA model, and OA_c is a binary indicator for being in a community assigned to the *AWH-OA* treatment (so also received the VYA model). The standard errors ε_{ic} are clustered at the kebele level, which account for both the design effect of the kebele-level treatments and the heteroskedasticity inherent in the linear probability model. In the basic model, X_{ic} includes block indicators used in the randomization (where blocks combine woreda and marginalization status), adolescent age (in years), an indicator of whether the adolescent was part of the original random VYA sample, an indicator for whether there were multiple eligible adolescents in the household, and month of survey indicators. We will also provide results for an expanded control set that includes additional

⁹ Importantly, it should be noted that many of the adolescents who participated in the earlier VYA programming additionally participated in this intervention, so that some (but not all) of these older adolescents actually received both VYA and OA programming. However, there are also adolescents who did not participate in the VYA programming, but did participate in the OA programming. Thus, the “duration” impacts we are able to measure experimentally are community-level duration of programming, rather than adolescent-level duration of programming. We will seek to measure the latter using quasi-experimental methods.

baseline household characteristics of interest, including: household size, an indicator for household head literate, an indicator for female headed household, household asset index, and an indicator for household receives PSNP benefits.¹⁰ Following McKenzie (2012), we will control for baseline measures of the outcome of interest (denoted by $y_{ic,0}$) when available. Gender and community specific sampling weights will be used to make the results representative of the target population in the study area.

Thus, for the final two research questions specified in Section 2.1.2, we are interested in two hypothesis tests. First, we seek to understand whether communities that received longer-duration adolescent-centric programming have better adolescent outcomes than those that did not receive any adolescent-centric programming.

(i) $\theta_2 \neq 0$: comparing *OA* to the control group

Then we seek to understand any added value of the longer-duration programming relative to the communities that got VYA programming only:

(ii) $\theta_1 \neq \theta_2$: comparing *OA* to VYA-only sites

As a secondary analysis, we will conduct treatment on the treated analysis for the primary outcomes.

3.2 Attrition and Missing Data

We are confident that we have put into place protocols that will lower attrition, nevertheless the risk of differential attrition across treatment arms is a possibility. To this end, we will conduct attrition analysis for the third follow-up data collection, looking at overall differential attrition by treatment arm, as well as interactions of treatment with a pre-specified set of baseline characteristics, including age, indicator for female adolescent, indicator for multiple eligible adolescents per household, household size, indicator for household head literate, indicator for household is female headed, household asset index, and an indicator for whether the household received benefits from the large scale social safety net program in Ethiopia known as the Productive Safety Net Programme (PSDP). If differential attrition by treatment arm is problematic (either in levels or by baseline characteristic), we will follow Karlan and Valdivia (2011), using three different methods to address the potential for differential attrition: Lee (2009) bounds, Kling and Liebman (2004) bounds and Inverse Probability Weighting (IPW). While the Lee bounds rely on an assumption of monotonicity, these two additional methods do not.

For individual outcomes with missing data (not from attrition), missingness less than 20% will simply be ignored. If 20% or more is missing across all arms, we will no longer utilize that variable. For missing data on covariates, we will follow Lin, Green and Coppock (2016). Observations with missing covariate values will be included in the regressions that estimate average treatment effects, as long as the

¹⁰ We will also explore the possibility of integrating machine learning for choice of control variables.

outcome measure and treatment assignment are non-missing. We will use the following approach: (1) if no more than 10% of the covariate's values are missing, recode the missing values to the overall mean (do not use arm-specific means); (2) if more than 10% of the covariate's values are missing, include a missingness dummy as an additional covariate and recode the missing values to the overall mean.

3.3 Handling of Outliers

We do not expect to have outliers among our predefined set of outcomes, all of which are composed of indicators and/or measures with well-defined ranges.

3.4 Multiple Outcome and Multiple Hypothesis Testing

To be conservative, given the number of hypothesis tests and primary outcomes we pre-specify here, we will construct FDR q-values, as described in Anderson (2008). This multiple-hypothesis testing will be conducted across the primary outcomes (16 for girls and 7 for boys) for each hypothesis and sample separately. These q-values will be presented alongside the standard per-comparison p-values.

3.5 Heterogeneous Effects

This evaluation is being conducted across two different regions in Ethiopia (Amhara and Oromia) – our main analysis explores impacts in each of these regions separately. However, we also include areas with very different marginalization statuses, and will explore this heterogeneity in secondary analysis. Additional secondary analysis may explore other dimensions of interest.

4 Administrative Information

4.1 Funding

The impact evaluation proposed here is supported by UK Aid and the Bill & Melinda Gates Foundation. The programming implementation described here is supported by the Bill & Melinda Gates Foundation, Seattle, Washington.

4.2 Institutional Review Board

The study design was approved by the George Washington University Committee on Human Research, Institutional Review Board (071721), the ODI Research Ethics Committee (02438), and the Addis Ababa University College of Health Sciences Institutional Review Board (113/17/Ext). A Memorandum of Understanding (MoU) was signed with the Ethiopia Federal Ministry of Women, Children and Youth. The study design was also approved by different regional-level bodies, including the Ethical Committee of the

Amhara Region Bureau of Health (Amhara Region), the Ethical Committee of the Oromia Region Bureau of Health (Oromia Region), and Semera University Ethical Board (Afar Region).

4.3 Declaration of interest

The other authors have no conflicts of interest to declare.

5 References

- Anderson, M.L. (2008). "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects", *Journal of the American Statistical Association*, 103(484): 1481-1495.
- Baird, S., J. Hamory Hicks, M. Kremer and E. Miguel (2016). "Worms at work: Long-run impacts of a child health investment", *Quarterly Journal of Economics*, 131(4): 1637-1680.
- GAGE Consortium (2017). *Gender and adolescence: why understanding adolescent capabilities, change strategies and contexts matters*. London: Gender and Adolescence: Global Evidence.
- Hamory, J., S. Baird, S. Das, N. Jones, T. Woldehanna, and W. Yadete (2023). "Do layered adolescent-centric interventions improve girls' capabilities? Evidence from a mixed-methods cluster randomized controlled trial in Ethiopia." Working Paper.
- Jones, N., B. Tefera, G. Emirie, W. Yadete, K. Gezahegne, K. Tilahun, K. Birhanu (2017). "Exploring Ethiopian adolescents' gendered experiences and perspectives." London: GAGE/ODI, Retrieved from <https://www.gage.odi.org/publications/exploring-ethiopian-adolescents-gendered-experiences-and-perspectives>.
- Jones, N., S. Baird, J. Hamory, and T. Woldehanna (2020). "Multi-Level Programming Aimed at Gender Norms Transformation to Improve Capabilities of Young Adolescents in Ethiopia: A Cluster Randomized Control Trial" Pre-analysis Plan, published on the AEA RCT Registry, 9 October 2020.
- Karlan, Dean, and Martin Valdivia (2011). "Teaching Entrepreneurship: Impact of Business Training on Microfinance Clients and Institutions." *Review of Economics and Statistics*, 93(2): 510–527.
- Kling, Jeffrey R., and Liebman, Jeffrey B. (2004). "Experimental Analysis of Neighborhood Effects on Youth." Princeton IRS Working Paper 483, March.
- Lee, D. (2009). "Training, Wages, and Sample Selection: Estimating Sharp Bounds on Treatment Effects." *Review of Economic Studies*, 76: 1071–1102.
- Lin, W., D.P. Green, and A. Coppock (2016). "Standard operating procedures for Don Green's lab at Columbia." Available at: <https://github.com/acoppock/Green-Lab-SOP>.
- McKenzie, D. (2012). "Beyond Baseline and Follow-up: The Case for More T in Experiments." *Journal of Development Economics*, 99: 210-221.

6 Appendix: Secondary Outcomes

As in the main text, note that to construct indices, we employ the following procedure: a) for each index component, create a normalized measure by subtracting the control group mean and then dividing by the control group standard deviation; b) construct the index by calculating the raw mean across all normalized component variables, and c) create the normalized index value by subtracting the control group mean of the index and then dividing by the control group standard deviation of the index. We have signed outcomes such that higher values of indices indicate ‘better’ outcomes.

A.1 Girls’ Education Secondary Outcomes

- ✓ Components of primary outcome index
- ✓ Index of school attendance: Index consists of two components (measured only among those aged 18 or younger at the time of survey):
 - a. Indicator for enrolled in school at time in survey (or if school not in session at survey, at time of most recent session) (*CR Module, Education, Q4*)
 - b. Share of school days attended in past two weeks (zero for those not enrolled, missing if no school) (*CR Module, Education, Q4 and Q14-14a*)

A.2 Girls’ Bodily Integrity Secondary Outcomes

- ✓ Components of primary outcome index
- ✓ Indicator for did not experience peer violence (as defined above) in last 12 months (*CR Module, Violence, Q1*)
- ✓ Indicator for did not witness or experience corporal punishment in school in past 12 months (among those enrolled) (*CR Module, Education, Q8-10a*)
- ✓ Age of first marriage (among married) (*CR Module, Marriage and Relationships, Q4a or previous rounds’ data*)
- ✓ Age at first pregnancy (among ever pregnant) (*CR Module, Sexual and Reproductive Health, Q23a or previous rounds’ data*)

A.3 Girls’ Physical Health, Nutrition, and SRH Secondary Outcomes

- ✓ Components of primary outcome indices
- ✓ Indicator for self-reported health very good or good (*CR Module, Health and Nutrition, Q1*)
- ✓ Proportion of meals yesterday with meat/chicken/fish/egg (zero if no meals yesterday) (*CR Module, Health and Nutrition, Q4a*)
- ✓ Proportion of meals yesterday with shiro/lentils/beans etc. (zero if no meals yesterday) (*CR Module, Health and Nutrition, Q4b*)

- ✓ Indicator for has not ever been hungry because not enough food in past four weeks (*CR Module, Health and Nutrition, Q3*)
- ✓ Indicator for use of modern conception (among those who have ever had sex) (*CR Module, Sexual and Reproductive Health, Q27 and Q27b*)
- ✓ Desired fertility (measured so that fewer children has higher values) (*CR Module, Sexual and Reproductive Health, Q29*)
- ✓ Ideal age of first child: Measured among those who have not yet had a child; code “don’t know” values to median age, code “never” values to max age. (*CR Module, Sexual and Reproductive Health, Q28*)

A.4 Girls’ Psychosocial Wellbeing Secondary Outcomes

- ✓ Indicator for minimal depression (i.e., little sign of depression; PHQ-9 score ≤ 4) (*CR Module, Psychosocial and Mental Health, Q1-9*)

A.5 Girls’ Voice and Agency Secondary Outcomes

- ✓ Components of primary outcome indices (both subindices and their components)
- ✓ Index of has discussed issues with mother (0-7, includes indicators for has discussed education, future work, romantic relationships, menstruation, puberty, age of marriage, bullying/harassment at school; missing if married or no mother in household) (*CR Module, Social Inclusion, Q4a-g*)
- ✓ Index of has discussed issues with father (0-6, includes indicators for has discussed education, future work, romantic relationships, puberty, age of marriage, bullying/harassment at school; missing if married or no father in household) (*CR Module, Social Inclusion, Q4a-c and Q4e-g*)
- ✓ Index of safety in community: index consisting of two components:
 - a. Indicator for feels safe walking in community during day (*CR Module, Violence, Q9b*)
 - b. Indicator for feels safe walking in community during night (*CR Module, Violence, Q9c*)
- ✓ Index of collective action: index consisting of two components:
 - a. Indicator for talked with others about serious problem in community (*CR Module, Social Inclusion, Q6a*)
 - b. Indicator for took action with others about serious problem in community (*CR Module, Social Inclusion, Q6b*)

A.6 Girls’ Economic Empowerment Secondary Outcomes

- ✓ Components of primary outcome indices

A.7 Girls' Cross-Cutting Secondary Outcomes¹¹

- ✓ Components of all primary outcomes indices
- ✓ (Partially) agrees with “I think it is possible to change how people react to my being a girl” (*CR Module, Sexual and Reproductive Health, Q14*)
- ✓ Index of gendered attitudes toward education: Index of three components, including:
 - a. Disagrees with “If a family can afford for one child to go to secondary school, it should be the boy only” (*CR Module, Education, Q16*)
 - b. Disagrees with “Girls should be sent to school only if they are not needed to help at home” (*CR Module, Education, Q17*)
 - c. (Partially) agrees with “A girl’s marriage can wait until she has completed senior secondary school” (*CR Module, Sexual and Reproductive Health, Q15*)

A.8 Boys' Secondary Outcomes

- ✓ Components of all primary outcome indices

¹¹ We note that the GAGE survey asked a number of additional attitude questions on school violence, politics, and female empowerment. We intend to explore these questions in supplementary analysis.