Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

Improving Mental Health of Adolescent Girls in Low- and Middle-Income Countries

Causal Evidence from Life Skills Programming

Manisha Shah[r]
Sarah Baird[r]
Jennifer Seager[r]
Benjamin Avuwadah[r]
Joan Hamory
Shwetlena Sabarwal
Amita Vyas

ABSTRACT

This study provides causal evidence on the impact of life skills programming on the mental health of adolescent girls aged 10–19 in three distinct low- and middle-income countries: Tanzania, Bangladesh, and Ethiopia. Life skills interventions significantly improved a component of mental health in all three contexts, with reductions in depression in Tanzania and improvements in socio-emotional development in Bangladesh and Ethiopia. However, findings suggest substantial heterogeneity in impact. Programs that target both adolescent boys and girls appear more effective than those that target girls alone, and existing supportive environments are a necessary condition for programs to improve mental health.

Manisha Shah is Chancellor's Professor of Public Policy at the University of California, Berkeley and a Research Associate with the National Bureau of Economic Research (manishashah@berkeley.edu).
Sarah Baird is Professor of Global Health and Economics in the Department of Global Health at George

(continued on next page)

[Submitted December 2022; accepted August 2023]; doi:10.3368/jhr.1222-12707R2
JEL Classification: I15, I31, and O12
ISSN 0022-166X E-ISSN 1548-8004 © 2024 by the Board of Regents of the University of Wisconsin System
Color version of this article is available online at: https://jhr.uwpress.org.
Supplementary materials are available online at: https://jhr.uwpress.org.
This open access article is distributed under the terms of the CC-BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0) and is freely available online at: https://jhr.uwpress.org.
Special Issue: Causes and Consequences of Child Mental Health
Janet Currie, Guest Editor
Sponsored by the NOMIS Foundation and the Center for Health and Wellbeing at Princeton University.
Manisha Shah https://orcid.org/0000-0002-1514-3931
Sarah Baird https://orcid.org/0000-0001-5363-8559
Jennifer Seager https://orcid.org/0000-0001-8973-6599
Joan Hamory https://orcid.org/0000-0002-0071-8683

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

I. Introduction

Approximately one in seven adolescents experiences a mental health illness each year (WHO 2021), with many more suffering from adverse mental health. Poor mental health for young people is itself a concern (Patel 2013), but it also affects long-term human capital accumulation and can result in lower adult earnings and worse employment outcomes (Currie et al. 2010; Currie and Stable 2006; Hale, Bevilacqua, and Viner 2015; Richter et al. 2021; Smith and Smith 2010). In low- and middle-income countries (LMICs)—where 90 percent of adolescents live—conditions of adversity and lack of access to mental health services exacerbate both rates of poor mental health and consequences across the life course (Patel 2018; Reiss 2013; Zhou et al. 2020).

Importantly, mental health in young people goes beyond a diagnosable mental illness to a more comprehensive, multidimensional measure (Azzopardi et al. 2023). The World Health Organization (WHO) defines mental health as, "A state of mental well-being that enables people to cope with the stresses of life, to realize their abilities, to learn well and work well, and to contribute to their communities. Mental health is an integral component of health and well-being and is more than the absence of mental disorders" (WHO 2022, p. 8). Mental health thus includes not only mental illness but also constructs related to resilience, self-control, coping, life-purpose/goals, connectedness, self-efficacy, confidence, and grit (Manwell et al. 2015; Orth, Moosajee, and Van Wyk 2022; Pollard and Lee 2003; Renwick et al. 2022). These broader factors are strongly associated with mental illness among adolescents (Musumari et al. 2018; Tait, French, and Hulse 2003) and also influence human capital and labor outcomes during adolescence and adulthood (Cobb-Clark 2015; Heckman, Stixrud, and Urzua 2006; Krishnan and Krutikova 2012). This broader definition is particularly important for

Washington University. Jennifer Seager is an Associate Professor of Global Health and Economics in the Department of Global Health at George Washington University. Benjamin Avuwadah is a Post-Doctoral Scientist in the Department of Global Health at George Washington University. Joan Hamory is an Associate Professor of Economics at the University of Oklahoma. Shwetlena Sabarwal is a Senior Economist at the Education Global Practice of The World Bank. Amita Vyas is an Associate Professor at George Washington University. r denotes Manisha Shah, Sarah Baird, Jennifer Seager, Benjamin Avuwadah, and Joan Hamory are joint first authors. The order in which these authors' names appear has been randomized. *The authors thank participants at the Causes and Consequences of Child Mental Health conference hosted by the Center for Health and Wellbeing at Princeton University and the Ottawa Applied Microeconomics Lab, for their valuable comments and feedback; in particular, they thank Corrine Low for her insightful discussion. They thank David McKenzie, Berk Özler, and Luca Parisotto for useful conversations. Tania Ismail expertly assisted in the creation of the conceptual framework figure, and Saini Das assisted with the literature review. The authors gratefully acknowledge funding for the three underlying impact evaluations from The Hewlett Foundation, the Africa and South Asia Gender Innovation Labs at the World Bank, the Global Financing Facility at the World Bank, and UK Aid through the Gender and Adolescence: Global Evidence program. This research would not have been possible without the dedicated in-country partners, BRAC Maendeleo in Tanzania, Innovations for Poverty Action and Room to Read in Bangladesh, and Laterite and the Ethiopian Development Research Institute in Ethiopia, as well as implementation partners in Ethiopia, Pathfinder International and CARE. Most importantly, the authors are thankful to the young people and their families, who generously shared their time and experiences for the purpose of this research. The replication files for this analysis are posted at Jennifer Seager's personal website (https://sites.google.com/view/jennifer-seager) under the publications tab, and at https://drive.google.com/file/d/1waZh5PYjCfbdmaUf1LekUzPOlKJEcoH_/view. If you have any questions, please email Jennifer Seager (jseager@gwu.edu).*

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

adolescents in LMICs, given limited resources to diagnose mental illness in these contexts and supply-side constraints on care (WHO 2022; Gureje and Oladeji 2022). Thus, interventions that target adolescent well-being more broadly may be a critical tool in tackling adverse mental health in LMICs.

Drivers of mental health among young people are also gendered, with adolescent girls having worse mental health than adolescent boys in many contexts (Campbell, Bann, and Patalay 2021). For young women, particularly those in LMICs, adverse mental health outcomes are often linked to increasingly restrictive gender norms (Baird et al. 2019; Kapungu and Petroni 2017; Koenig et al. 2021). This suggests that interventions that tackle restrictive gender norms may have the potential to improve mental health.

This work provides evidence on the short-term causal impacts of life skills programming (described in detail in Section II.B) on the mental health of adolescent girls aged 10–19 in three diverse LMIC settings (Tanzania, Bangladesh, and Ethiopia), utilizing randomized controlled trials (RCTs). Life skills programs provide adolescents, typically females, with safe spaces and education on practical topics, such as navigating relationships with boyfriends and parents, coping with unequal gender norms, reproductive and menstrual health, and building confidence. They also often include direct links to employment and skill building for older participants. There is heterogeneity in how programs bundle various features, but all have the ultimate objective of empowering girls to live better lives. Online Appendix Table 1 provides an overview of the recent impact evaluation literature on life skills programming for adolescent girls in LMICs.[1]

We employ a conceptual framework (Figure 1) to understand how life skills programming could impact the mental health of adolescent girls and to help synthesize findings across our varied interventions and contexts. We further explore heterogeneity in treatment effects both traditionally along dimensions identified in the literature (household wealth, education, community gender norms, risk exposure, and parental investment) and using machine-learning techniques to identify treatment heterogeneity across a rich set of observable adolescent and household characteristics.

In all three countries, we find that at least one variant of life skills interventions significantly improves a dimension of mental health, with reductions in depression in Tanzania and improvements in socio-emotional development in Bangladesh and Ethiopia. There is also considerable evidence of important heterogeneity by baseline community, household, and adolescent characteristics. Both traditional heterogeneity analysis using interactions and machine-learning techniques highlight the important role of baseline wealth, which suggests that benefits accrue to the relatively better

1. These programs are offered both in the community, providing safe spaces for both in- and out-of-school adolescents (Bandiera et al. 2019, 2020a, 2020b; Buehren et al. 2017; Buchman et al. 2022), as well in school settings, incorporating programming into the curricula (Dhar, Jain, and Jayachandran 2022; Ashraf et al. 2020; Edmonds, Feigenberg, and Leight 2023). While findings from life skills programs are mixed (Buehren et al., 2017), they have been shown to increase engagement in income generating activities (Bandiera et al. 2019, 2020a; Buchman et al. 2022), decrease unintended teen pregnancy and early entry into marriage or cohabitation (Bandiera et al. 2019, 2020a), protect adolescents from the worst effects of negative systemic shocks (Bandiera et al. 2020b), reduce school dropout and improve educational outcomes (Edmonds, Feigenberg, and Leight 2023; Buchman et al. 2022; Ashraf et al. 2020), reduce experience of intimate partner violence (Shah et al., 2023), improve gender attitudes (Dhar, Jain, and Jayachandran 2022; Edmonds, Feigenberg, and Leight 2023), and increase agency (Edmonds, Feigenberg, and Leight 2023).

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

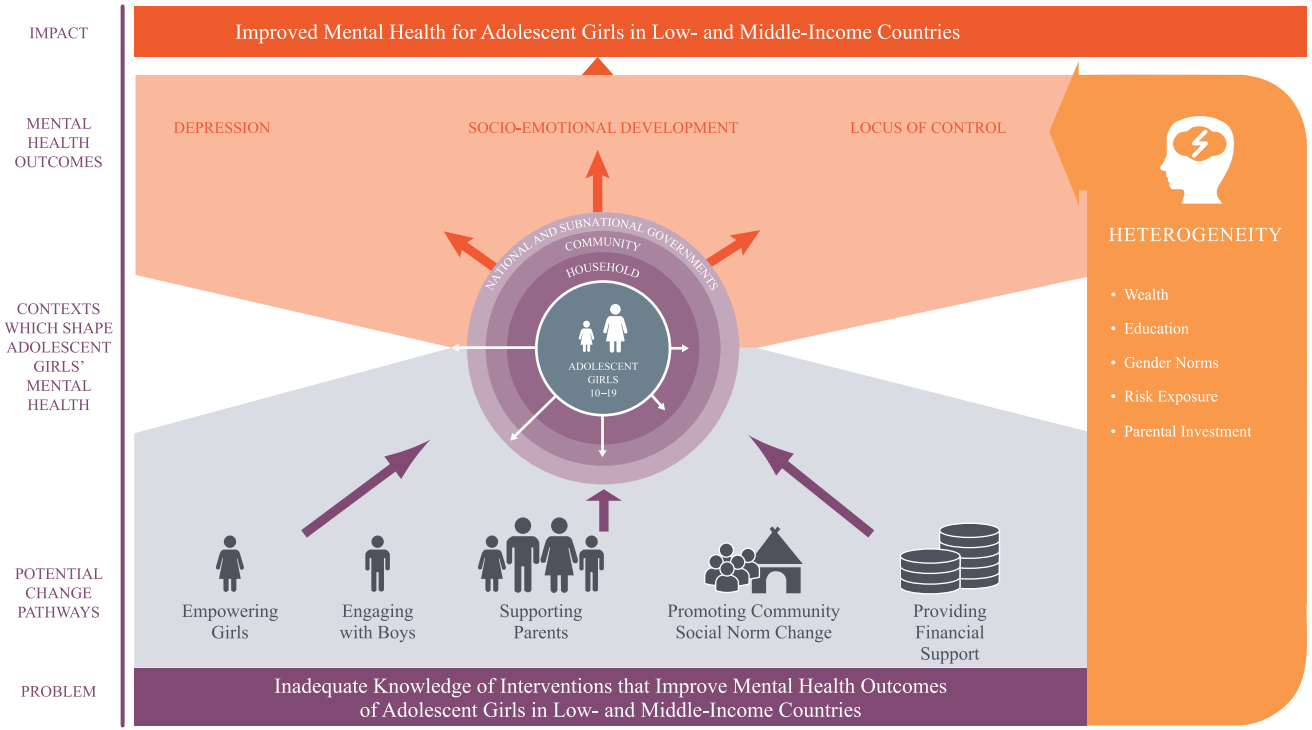

Figure 1

Conceptual Framework

Notes: This diagram summarizes the conceptual framework for our analysis. See the text for more details on this framework.

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

off amongst a lower income population. The machine-learning heterogeneity analysis further suggests that benefits accrue to those who have broader underlying support systems (for example, enrolled in school, parents in the household, etc.).

This work makes three important contributions to the literature. First, we provide causal evidence on the impact of life skills programs on the mental health of adolescent girls using longitudinal data and interventions from three LMICs. While there is some evidence on the efficacy of mental illness–specific interventions for adolescents (Baird et al. 2020a; Layne et al. 2008; Miller-Graff and Cummings 2022), there is limited evidence on the impact of broader adolescent programming (Barry et al. 2013).[2] Given the popularity of life skills programs in LMICs (Haberland, McCarthy, and Brady 2018) and the critical importance of mental health during adolescence, our findings narrow a key evidence gap. Second, by evaluating RCTs with similar goals across three countries and settings, these findings address external validity concerns often directed at small-scale single-context RCTs. Lastly, we investigate heterogeneity of the treatment effects, which allows us to unpack which programs might be more effective and for whom. This heterogeneity analysis provides insight into how and where to scale up similar programs in the future, as well as where additional research is needed.

II. Conceptual Framework

We organize our analysis using the conceptual framework presented in Figure 1, which summarizes how life skills programs could improve adolescent female mental health.[3] Life skills programs can improve adolescent mental health by focusing on adolescent empowerment (for example, Leventhal et al. 2022), providing supportive environments and social networks though club membership and involving boys in the community (for example, Boxho et al. 2023), supporting parents (for example, Falb et al. 2016; Miller-Graff and Cummings 2022), promoting social norm change (for example, Andrew et al. 2022), and providing financial support (for example, Boxho et al. 2023). Our conceptual framework outlines these program features as potential change pathways toward improved adolescent female mental health. We also account for household-, community-, and country-level contexts that shape adolescent girls' experiences.

Each of the identified change pathways has been linked to adolescent female mental health in the literature. *Empowering girls* has been found to be a key component for recovery from mental distress (Grealish et al. 2017; Swift and Levin 1987). Empowerment can be broadly defined as both a psychological sense of personal control and an ability to affect change in one's own life and community, the latter of which relies on the

2. There is a limited evidence base on the impact of cash transfers (Baird, de Hoop, and Özler 2013), as well as for specific vulnerable populations such as adolescents affected by HIV/AIDS (Ssewamala, Han, and Neilands 2009) and children (Fernald, Gertler, and Neufeld 2009) on mental health.

3. This conceptual framework builds on that of the Gender and Adolescence: Global Evidence (GAGE) research program (GAGE Consortium 2019), which uses a capabilities approach as pioneered by Sen (1984, 2004) and further developed by Nussbaum (2011) and Kabeer (2003) to capture the importance of gender dynamics. The capabilities approach recognizes that the ability of individuals to achieve overall well-being, including mental health, is a matter of *what they are able to do* and *who they are able to be*, given the contexts within which they live.

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

malleability of external structures (Swift and Levin 1987). This suggests scope to target broader social and environmental factors in conjunction with girl-focused programming, pointing to the other change pathways.

Peer harassment (for example, sexual harassment and bullying) (Beattie et al. 2019) is negatively associated with the mental health of adolescent girls, suggesting the importance of *engaging with boys* in programming. Weak parental support (for example, death, absenteeism) is associated with worse mental health outcomes among adolescents, and recent systematic reviews in both high-income and LMIC contexts have highlighted the importance of parental involvement in depression interventions for adolescents (Dardas, van de Water, and Simmons 2018; Singla et al. 2020), motivating the role of *supporting parents* in adolescent mental health programming.

At the community level, restrictive gender norms are directly associated with worse mental health outcomes for girls (Baird et al. 2019). In addition, such norms affect mental health indirectly through effects on school dropout and early marriage (Beattie et al. 2019; Guglielmi, Mitu, and Seager 2021). Thus, programs *promoting community social norm change* alongside adolescent-focused programming may be more efficacious at improving female adolescent mental health.

Finally, there is a strong link between poverty and psychological distress (for example, Bøe et al. 2012, 2018), aligning with recent literature on the scope of cash transfers to improve mental health (Baird, de Hoop, and Özler 2013; Zimmerman et al. 2021), motivating the role of *financial support*.

Our conceptual framework also hypothesizes sources of potential heterogeneity in program effectiveness, based on a rapid review of the recent impact evaluation literature of adolescent life skills programming in LMICs. We summarize this literature in Online Appendix Table 1. We compiled studies through a scoping search of several databases for academic publications and working papers from the year 2010 through mid-2023, as well as relevant gray literature (technical reports of impact evaluations).[4] Based on this literature review, we explore possible heterogeneity in program impacts along the dimensions of wealth, education, community gender norms, risk exposure, and parental investment.

III. Study Setting, Sample, Design, and Outcomes

A. Study Settings

This study provides evidence from adolescent girls living in three LMICs—two in Africa (Tanzania and Ethiopia) and one in South Asia (Bangladesh). Online Appendix Table 2 provides a summary of selected country characteristics.

Ethiopia is the poorest of the three countries we study. More than one-quarter of Ethiopians—and more than two-thirds of the rural population—live in extreme poverty,

4. We started with a review of studies included in Temin and Heck (2020). We included all of these studies and also used this review to generate search terms for academic databases. Our database search included ScienceDirect, PubMed, Google Scholar, World Bank, and the National Bureau of Economic Research, and we sought English language articles including the words "adolescents" or "girl groups" and "life skills training." We eliminated any studies that were not conducted in an LMIC or were not randomized. We additionally included research known to the authors related to group therapy and/or growth mindset and reviewed papers included in Singla et al. (2020).

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

in comparison to 45 percent in Tanzania and only 13.5 percent in Bangladesh. All three countries have high youth dependency ratios, ranging from 40 percent in Bangladesh to nearly 83 percent in Tanzania (United Nations, Department of Economic and Social Affairs, Population Division 2022), highlighting the importance of studying interventions that aim at improving outcomes for young people.

Despite high rates of secondary school enrollment (72 percent), the median age of marriage for women aged 25–29 in Bangladesh is the lowest of the three countries at 16.7 years, and rates of adolescent childbearing are highest, with one-third of women giving birth to their first child by age 18 (UNESCO Institute for Statistics 2020; Demographic and Health Survey 2022). In Ethiopia and Tanzania, the median ages of marriage for women aged 25–29 are 18.1 and 19.6, respectively, and a quarter of women in this age group give birth to their first child by age 18 (Demographic and Health Survey 2022).

B. Study Samples

Our sample includes panel data on more than 4,500 adolescent girls aged 10–19 in Tanzania, Bangladesh, and Ethiopia. Table 1 provides a summary of the sample in each setting.

In Tanzania, our sample includes 1,449 adolescent girls aged 10–19 (at 2016 baseline) living across 79 communities in the Dodoma and Iringa regions of Tanzania. These 79 communities were selected due to their inclusion in an evaluation of BRAC's adolescent female clubs, Empowerment and Livelihood for Adolescents (*ELA*), via a cluster-randomized controlled trial (cRCT) in 2009–2011 (Buehren et al. 2017) and where no additional community-level programming was introduced during the subsequent cRCT in 2016–2018 from which this study draws (see Shah et al. 2023 for more details). In communities with *ELA* clubs (32 communities), all members were selected for the survey. In communities without *ELA* clubs (47 communities), adolescent girls were sampled based on a household census and similarity to *ELA* members using propensity score matching based on age, wealth, school enrollment, marital status, and ever-pregnant status.

In Bangladesh, our sample includes 1,200 adolescent girls aged 10–14[5] (at 2020 baseline data collection) attending Grades 7 and 8 across 100 government and semi-private schools in Chittagong and Sylhet divisions in Bangladesh. These divisions were selected by the Government of Bangladesh in partnership with the World Bank for this study due to them being relatively disadvantaged with regards to adolescent health and education outcomes. Data collection covers schools in nine districts of Chittagong (Bandarban, Brahamanbaria, Chandpur, Chittagong, Comilla, Cox's Bazar, Lakshmipur, Noakhali, Rangamati, and Sunamganj) and three districts of Sylhet (Habiganj, Maulvibazar, Sunamganj). Adolescents were randomly sampled from school registration lists to be representative of their gender within their grade and school.

In Ethiopia, our sample includes nearly 1,900 adolescent girls aged 10–12 (at 2017–2018 baseline data collection) living in 126 communities (Enumeration Areas) in two poor, rural zones of central and eastern Ethiopia (South Gondar, Amhara and East Hararghe, Oromia). Ten districts within these two zones were purposely selected for

5. Seven percent of the adolescents were aged 15–18.

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

Table 1
Summary of Interventions, Sample, and Study Design

	Tanzania	Bangladesh	Ethiopia
Panel A: Interventions			
	Goal Setting, implemented in 2017, encourages girls to set a goal to remain healthy and stay STI/HIV free; *Empowerment and Livelihood for Adolescents* (*ELA*), an ongoing intervention, provides a safe social space, life-skills training and community support for girls.	*Growth Mindset* is an intervention aimed at developing "growth mindset," the belief that ability depends on effort and is not constrained by fixed intelligence; *Girl Rising* uses the power of storytelling to change the way the world values girls and their education. Interventions were implemented in 2021.	*Act With Her* is a multilevel intervention that includes a safe social space, life skills training for girls and boys (separately), awareness raising for their caregivers, and (in a random half of communities) community-level work aimed at improved adolescent-focused services; discussion of gender norms and attitudes is interwoven at each level, and goal setting for girls is a focus. *Act With Her + Transfers* additionally includes in-kind school supply and sexual and reproductive health–related *Transfers* for girls. Interventions were implemented during 2019–2020.

(*continued*)

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

Table 1 (*continued*)

	Tanzania	Bangladesh	Ethiopia
Panel B: Sample			
Location	Dodoma and Iringa Regions	Chittagong & Sylhet Divisions	South Gonder Zone (Amhara), East Hararghe Zone (Oromia)
Sublocation/cluster	79 communities	100 schools	126 communities
Urban/rural	Rural/Peri-urban	Urban/Rural	Rural
Baseline age range	10–19-year-olds	10–18-year-olds[a]	10–12-year-olds
Baseline sample size	1,449	1,200	1,863
Baseline survey year(s)	2016	2020	2017/18
Follow-up survey year(s)	2018	2021	2019/20
Panel C: Study Design			
	Cross-cutting design Goal setting: Individually randomized (369 treated, 1,080 control) ELA: Cluster randomized (32 treatment communities, 47 control)	Multi-arm cluster RCT (36 control schools, 36 Growth Mindset, 37 *Growth Mindset + Girl Rising*)	Multi-arm cluster RCT (39 control communities, 58 *Act With Her*, 29 *Act With Her + Transfers*)

[a]93 percent of adolescents are aged 10–14; 61 adolescents are 15 years old, 16 are 16 years old, three are 17 years old, and one is 18 years old.

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

study inclusion (five in each zone), with preference for areas characterized by high levels of food insecurity and high rates of child marriage, while additionally keeping in mind security concerns and programming capacity of the intervention implementation partners (Pathfinder and CARE Ethiopia). Adolescents were randomly sampled from full community lists developed through census-style enumeration by the research team.

Table 2 provides more detail on the study sample within each country, summarizing both adolescent (Panel A) and household (Panel B) characteristics at baseline. Ethiopia focuses on very young adolescents (mean baseline age of 10.9), Bangladesh focuses on adolescents attending Grades 7 and 8 (mean age of 12.7), and Tanzania targets adolescents across a broader age range of 10–22 (mean age of 14.7 in our sample, where we restrict to ages below 19 and younger for this analysis).

Age may be a key factor in driving the success of programs—both due to developmental appropriateness of the intervention and because adolescents may be more susceptible to impacts at certain ages. That said, age is not a perfect marker of development, as factors such as puberty happen over a wide age range and may in fact be more important than biological age (Viner, Allen, and Patton 2017). Understanding the role of age in driving program success among adolescents is part of a much broader research agenda (GAGE Consortium 2019). While this analysis is not well suited to unpack age alongside the many other factors that vary across studies, we control for age in all analyses, include age in our machine-learning heterogeneity analysis, and reflect on the role of age briefly in our discussion.

Our sample is broadly reflective of the national averages shown in Online Appendix Table 2. Most of the sample was enrolled in school (including all girls in Bangladesh, where school enrollment was a determinant of study inclusion), but aspirations for a degree beyond secondary school differ enormously between Ethiopia and the two other samples (58 percent of girls report such aspirations in Ethiopia, compared to 81 percent in Tanzania and 85 percent in Bangladesh). For each country, we summarize an index of adolescent perceptions of norms related to gender roles, which includes the roles of men and women in household decision-making and the role of women in caring for her family and home (Online Appendix Table 3 for more details on this index). Higher values of the index (on a scale of 0–2) indicate more traditional gender norms. Girls in Bangladesh perceive gender role norms as the least traditional (with a mean of 0.890), while girls in Ethiopian study sites perceive gender role norms as the most traditional (mean 1.645).

Panel B of Table 2 provides household-level characteristics for sample households in each country. Sample households in Tanzania are somewhat smaller than the corresponding national average (Online Appendix Table 2), while households in Bangladesh and Ethiopia are larger than the respective national averages (at 5.8 and 6.4 household members). Household head school attainment reflects national-level differences, with education levels in Tanzania (8.7 years) and Bangladesh (7.4 years) much higher than in Ethiopia (2.4 years). An indicator of improved flooring further suggests that households in Tanzania and Bangladesh are more advantaged relative to those in Ethiopia. Finally, an indicator for female caregiver participation in paid work indicates the highest rates of mothers working outside of the household in Tanzania (58 percent) relative to Bangladesh (15 percent) and Ethiopia (9 percent).

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

Table 2
Sample Summary Statistics

	Tanzania	Bangladesh	Ethiopia
Panel A: Adolescent Characteristics			
Age at baseline	14.7	12.7	10.9
	(2.5)	(1.1)	(0.8)
Age at follow-up	16.1	14.8	13.2
	(3.0)	(1.1)	(1.1)
=1 if adolescent is enrolled in school	0.721	1.000	0.836
=1 if adolescent aspires to higher than secondary schooling	0.807	0.852	0.581
Index of perceptions of gender stereotypical roles (0–2)	1.009	0.890	1.645
	(0.439)	(0.824)	(0.601)
=1 if adolescent has savings	0.138	0.217	0.023
Panel B: Household Characteristics			
Number of household members	3.3	5.8	6.4
	(1.3)	(2.3)	(1.8)
=1 for mother living in household	0.764	0.947	0.913
=1 for father living in household	0.580	0.862	0.829
Household head highest grade attained	8.7	7.4	2.4
	(3.0)	(5.3)	(3.8)
=1 if improved flooring (wood/tiles/ ceramic/carpet/cement/bamboo)	0.894	0.656	0.017
=1 for primary female caregiver in the household engages in paid work	0.580	0.151	0.089
Number of observations	1,449	1,220	1,863

Notes: This table presents summary statistics for the sample used in each country. All statistics are measured for the entire sample (by country) at baseline, with the exception of "age at follow-up." Means are presented for all measures, and standard deviations are presented (in parentheses) for continuous measures only.

C. Interventions and Study Design

We evaluate two variations of a life skills programs using RCTs in each country. The programs vary in the breadth of their curricula across the change pathways outlined in the conceptual framework (Figure 1). Table 1, Panels B and C provide a summary of the interventions and study designs in each setting. We provide more detail here for each country.

1. Tanzania

In Tanzania, we evaluate a goal-setting activity (*Goal Setting*) that focuses on the first change pathway: empowering girls. Goal setting has been found to be helpful when

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

addressing behavioral or emotional difficulties and has been a valuable tool when used in combination with cognitive-behavioral therapy (Beck 1976; Lochman et al. 2011; Gaudiano 2008). For the activity, facilitators engaged with adolescent girls one-on-one. They asked girls if they were willing to set a goal to remain healthy and stay STI/HIV free for the next year. If they agreed, they went through the process of setting specific, measurable, achievable, relevant, and timely (SMART) goals, as developed by Doran (1981). Girls then identified and committed to up to three specific strategies to achieve the prespecified goal. This initial goal setting activity was implemented in August 2017 and lasted approximately 90 minutes. Facilitators followed up with females four months later in December 2017 to remind them of their goals and discuss successes and challenges in implementing their strategies. This check-in lasted approximately 60 minutes. Follow-up endline data collection took place in mid 2018.

In a subset of communities (32 communities), girls were also members of preexisting *ELA* clubs. *ELA* is an education-based intervention designed to empower adolescent females by providing a safe social space, life-skills training, and support in adolescent development. Participation in *ELA* is voluntary, but members are expected to attend five afternoons per week (for three hours each afternoon). *ELA* club leaders engage parents in sensitization meetings to orient them to the club purpose and activities. Thus, *ELA* has components that address supporting parents and promoting community norm change.

The RCT is an individually randomized controlled trial, with 25 percent of girls randomly selected to be invited to participate in *Goal Setting*, stratified by community (369 treated, 1,080 control). These girls were selected across both *ELA* (32 communities) and control (47 communities). In the present analysis, we evaluate the *Goal Setting* intervention within communities where no other programming is offered and within communities where participants were also exposed to *ELA*. Therefore, we do not evaluate *ELA* directly, but use the context of the preexisting *ELA* clubs to explore differences in the treatment effects of *Goal Setting* when it is layered on top of *ELA* in contrast to when it is delivered as an independent activity. We hypothesize that Goal Setting will be more effective among *ELA* participants, who have been exposed to a broader set of life skills and live in communities that have been more broadly affected by the *ELA* activities. This analysis is on a subset of a broader RCT that includes two additional treatment arms providing sexual and reproductive health programming (see Shah et al. 2023). We exclude these arms from the current study.

2. *Bangladesh*

In Bangladesh, we evaluate two interventions, delivered virtually during COVID-19-related school closures across 100 schools, that span three of the five change pathways: empowering girls, engaging with boys, and promoting community social norm change. The two interventions are (i) a gender-neutral growth mindset program around malleable intelligence (*Growth Mindset*) and (ii) a gender sensitization program that focuses on shifting gender norms around girls' education, as well as sensitizing adolescents to the issues of child marriage and child labor (*Girl Rising*). The *Growth Mindset* programming was offered both alone and in combination with the *Girl Rising* programming. *Girl Rising* was layered on top of *Growth Mindset* in a random subset of schools in order to address the structural constraints upheld by gender norms, so that adolescent girls would be better supported in operationalizing the tenets of the *Growth Mindset* curriculum.

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

The concept of growth mindset is rooted in Carol Dweck's research on malleable intelligence. When children (and adults) have a growth mindset, they believe that their ability depends on effort and are not constrained by fixed intelligence—the belief that talents are innate (Dweck 2006). These interventions encourage adolescents to spend time working to improve their skills and academic performance, essentially setting goals for future improvement. Growth mindset interventions can result in higher motivation, effort, and ultimately increased educational attainment (Paunesku et al. 2015; Yeager and Dweck 2012; Alan, Boneva, and Ertac 2019). In our case, the intervention entailed a facilitated group call around an essay on malleable intelligence, followed by the adolescent writing a letter to a friend explaining the concept and responding to a series of messages with growth mindset content via SMS.

Girl Rising uses storytelling to create awareness of gender-based discrimination, change dominant gendered perceptions, promote gender-equitable attitudes and norms, and provide tools to participants to translate changes in attitudes and greater aspirations into behavior change (Vyas et al. 2020). Adolescents are provided a physical storybook with six stories of six girls from six countries, accompanied by six companion videos, shedding light on a variety of gender-based issues, such as early marriage. Adolescents discuss the stories each week in group calls and, after reading all stories, write and share their own stories.

In Bangladesh the RCT is a multi-arm cluster (school-level) randomized controlled trial. Of the 100 schools included in this sample, 32 schools were randomly assigned to the *Growth Mindset* arm, 34 schools were randomly assigned to the *Growth Mindset + Girl Rising* arm, and 34 schools were randomly assigned to serve as the control, stratified by school type (government and semi-private) and rural/urban status. *Growth Mindset* programming was implemented over the course of eight weeks from April to June 2021. The *Girl Rising* program was implemented over the course of eight weeks from September to November 2021. Girls assigned to receive both *Growth Mindset* and *Girl Rising* programming participated in both interventions in the same groups. Follow-up data were collected from March–April 2022. In the current analysis we compare each treatment arm to the pure control group and to each other. For more details on the Bangladesh sample, data collection, and RCT, see Seager et al. (2022).

3. Ethiopia

In Ethiopia, we evaluate the most expansive programming across the countries we study—programming that touches on all five identified change pathways (Figure 1)—across 126 rural communities. *Act With Her (AWH) Ethiopia* pairs regular curriculum-based group meetings for young female and male adolescents and their caregivers with community-level adolescent-centric systems strengthening activities, offering a more holistic approach to improving adolescent transitions to adulthood.

In the *Act With Her* program sites, participating adolescent girls meet weekly over a period of ten months in life-skills training group sessions led by a female mentor. Several sessions are devoted to aspirations and goal setting, and discussion of norms and attitudes related to gender is interwoven across a range of topics, including interpersonal conflict and negotiation, violence and safety, physical health and nutrition, sexual and reproductive health, and personal finance. Participating adolescent boys of the same age

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

take part in their own curriculum-based sessions approximately twice a month over the same period, with a male mentor covering a subset of similar topics. Four sessions bring the adolescent boys' and adolescent girls' groups together for joint learning and discussion. Over the same ten-month period, caregivers of participating adolescents are invited to six facilitated group sessions of their own, to orient them to the topics covered in the adolescent group meetings and to help them to create a supportive environment for their adolescents. In a random half of these *Act With Her* communities, key community decision-makers and stakeholders are regularly brought together for structured meetings to discuss social norms[6] and adolescent-centric services in order to initiate changes over time. Groups meet monthly to discuss harmful socio-cultural norms relevant to their local community and to devise an action plan as to how they can be tackled.

An additional set of communities in the Ethiopia study, which also received all of the above interventions, were randomly allocated to receive in-kind transfers to participating adolescent girls (*Act With Her + Transfers*) as well. Girls were allowed to choose from three equal-valued (115 USD) packages: school supplies, hygiene supplies, and a combination package. The majority of girls chose the school supplies option, which included a backpack, writing utensils, eraser, ruler, exercise books, solar lantern, compass, and English and math reference books.

The Ethiopia RCT is a multi-arm cluster (community-level) randomized control trial. Communities were first stratified by district (woreda) and marginalization status (lack of programming, isolated from key services and road/transport infrastructure), and then randomly assigned to *Act With Her* (58 communities), *Act With Her + Transfers* (29 communities), or control (39 communities). The program was launched in early 2019, and a follow-up survey was conducted in late 2019 and early 2020, when participants were on average 13 years old. In the present analysis, we compare the two treatment arms to a pure control group. For more details on the Ethiopia sample, data collection, and RCT, see Baird et al. (2020b).

D. Outcome Measures

Our analysis focuses on three dimensions of adolescent mental health that are captured in the WHO definition of mental health—depression, socio-emotional development, and locus of control. We also construct an overall index as a summary measure or a "sufficient statistic" for mental health. Table 3 provides further details and references on measurement by country.

For depression, we focus on two submeasures. First, we utilize a continuous measure of depression from the Patient Health Questionnaire (PHQ) across all samples. The research in Tanzania utilized the PHQ-2 scale (Kroenke, Spitzer, and Williams 2003), while the PHQ-9 scale (Kroenke, Spitzer, and Williams 2001) was administered in both Bangladesh and Ethiopia. These measures have been validated for adolescent samples in LMICs (see, for example, Anum, Adjorlolo, and Kugbey 2019; Mazzuca et al. 2019). The second submeasure of depression is an indicator for moderate or severe depression,

6. Work on catalyzing shifts in social norms is primarily focused on applying CARE's well-known Social Analysis and Action (SAA) approach to gender and social transformation, which seeks to enable communities to identify for themselves the linkages between social factors and well-being, and then determine what actions will help improve them (Mekuria, Sprinkel, and Cowan 2018).

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

Table 3
Outcome Descriptions and Summary Statistics

	Tanzania	Bangladesh	Ethiopia
Construct #1: Depression			
PHQ Depression Score	Patient Health Questionnaire-2 (PHQ- 2) score—a measure of depression, with scores ranging from 0–6. Higher scores indicate higher levels of distress; measure used in regression is standardized within the sample control group.	Patient Health Questionnaire-9 (PHQ 9) score—a measure of depression (Kroenke, Spitzer and Williams 2001), with scores ranging from 0–27. Higher scores indicate higher levels of distress; measure used in regression is standardized within the sample control group.	Patient Health Questionnaire-9 (PHQ 9) score—a measure of depression (Kroenke, Spitzer and Williams 2001), with scores ranging from 0–27. Higher scores indicate higher levels of distress; measure used in regression is standardized within the sample control group.
Summary statistics (mean, standard deviation within control group at follow-up)	1.01 (1.26)	2.43 (2.50)	0.59 (2.09)
Indicator for moderate/ severe depression	Indicator of depression—binary measure taking on a value of 1 if raw PHQ-2 score is ≥3.	Indicator of moderate/severe depression—binary measure taking on a value of 1 if raw PHQ-9 score ≥10.	Indicator of moderate/severe depression—binary measure taking on a value of 1 if raw PHQ-9 score ≥10.
Summary statistics (mean, standard deviation within control group at follow-up)	0.135 (0.342)	0.019 (0.138)	0.01 (0.08)

(*continued*)

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

Table 3 (*continued*)

	Tanzania	Bangladesh	Ethiopia
Construct #2: Socio-Emotional Development			
Confidence/grit/resilience	Confidence—a measure that takes on a value of 1 if the adolescent agrees she is confident she can complete any task set before her. Measure used in regression is standardized within the sample control group.	Grit Scale—a measure of adolescent grit generated from a 7-item scale following Alan, Boneva, and Ertac (2019), with scores ranging from 0 to 4, where higher scores indicate more grit. Measure used in regression is standardized within the sample control group.	Child and Youth Resilience Measure- 12 (CYRM-12)—a measure of youth resilience (Liebenberg, Ungar, and LeBlanc 2013), with scores ranging from 12 to 36, where higher scores indicate more resilience. Measure used in regression is standardized within the sample control group.
Summary statistics (mean, standard deviation within control group at follow-up)	0.529 (0.499)	2.97 (0.457)	31.31 (4.41)

(*continued*)

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

Table 3 (*continued*)

	Tanzania	Bangladesh	Ethiopia
Construct #3: Locus of Control			
Locus of control	Pearlin Mastery Scale items, which is the extent to which the respondent feels in control of the events that influence her life. The Pearlin Mastery Scale (Pearlin and Schooler 1978) is a 7-item scale of seven statements to which respondents indicate if they strongly disagree, disagree, feel neutral, agree, or strongly agree. The responses to each item are coded so that all responses go in the same direction (toward feelings of more control) and summed to an index that takes on values of 12 to 40. We generate a binary indicator = 1 if the score is >28 (median score at baseline). Measure used in regression standardizes this indicator within the sample control group.	Self-reported feeling of control over own life, measured on a scale from 1 to 10, where higher values indicate more control. Adapted from the World Values Survey (Haerpfer et al. 2020). Question asks: "Some people feel that they have a great deal of control over their own lives. Others feel that what they do has very little effect on what happens to them. On a scale from 1 to 10, with 1 being very little and 10 being complete control, how would you classify yourself?" Measure used in regression is standardized within the sample control group.	Self-reported feeling of control over own life, measured on a scale from 1 to 10, where higher values indicate more control. Adapted from the World Values Survey (Haerpfer et al. 2020). Question asks: "Some people feel that they have a great deal of control over their own lives. Others feel that what they do has very little effect on what happens to them. On a scale from 1 to 10, with 1 being very little and 10 being complete control, how would you classify yourself?" Measure used in regression is standardized within the sample control group.
Summary statistics (mean, standard deviation within control group at follow-up)	28.7 (4.66)	6.789 (2.32)	7.05 (3.01)

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

based on the continuous PHQ measure.[7] As the summary statistics in Table 3 show, the measured level of moderate depression is 13.5 percent in the Tanzania sample, but is less than 2 percent in the other two countries, rates consistent with global averages (WHO 2021). These lower rates are partially due to the younger age profile of our sample in Bangladesh and Ethiopia, but also likely reflective of the challenges of measuring depression among adolescents, particularly in LMICs (Carvajal et al. 2022). Our broader measure of mental health is motivated by the fact that diagnosable depression is only one piece of good mental health (Renwick et al. 2022).

Our second measure of mental health is socio-emotional development. The measurement of socio-emotional development varies across countries, but, in each country, the measure captures ability to overcome adversity or difficulty while navigating life. In Tanzania, we use an indicator for whether the adolescent is confident that they can complete a task set before them. In Bangladesh, we focus on a measure of grit that is generated from a seven-item scale following Alan, Boneva, and Ertac (2019), with scores ranging from 0 to 4, where higher scores indicate more grit. In Ethiopia, we use the Child and Youth Resilience Measure-12 (CYRM-12), a measure of youth resilience (Liebenberg, Ungar, and LeBlanc 2013), with scores ranging from 12 to 36, where higher scores indicate more resilience.

Our third measure of mental health is locus of control. For the Tanzania sample, we use the Pearlin Mastery Scale, which is a seven-item scale measuring the extent to which the respondent feels in control of the events that influence her life (Pearlin and Schooler 1978). The scale ranges from 12 to 40, with higher values indicating higher locus of control. For both Bangladesh and Ethiopia, we use a measure adapted from the World Values Survey (Haerpfer et al. 2020), which asks the respondent to rate how much control she feels she has over her own life on a scale from 1 to 10, with 10 being complete control. Summary statistics in Table 3 suggest that adolescents in all three samples feel a moderate to high level of control over their lives.

For regression analysis, we standardize each measure, except the indicator for moderate or severe depression, around each country's sample control group mean for consistency of interpretation across countries. In addition, we generate an overall mental health index from the three standardized measures by orienting them all in the same (positive) direction, taking the unweighted mean across the measures, and standardizing around the control group in each country (Kling, Liebman, and Katz 2007). This provides an overall summary statistic across these three diverse constructs of mental health.

IV. Empirical Framework

We estimate intention-to-treat (ITT) regressions, implementing basic reduced form linear regressions as follows:

$$(1) \quad Y_{1iv} = \alpha + \mathbf{T}_v'\beta + X'_{0iv}\gamma + \phi_s + e_{iv}$$

where Y_{1iv} is outcome of interest for adolescent i in village (cluster) v measured at follow-up ($t=1$). $\mathbf{T}$ is a vector of treatment indicators equal to one if the individual/

7. PHQ-9 scores of 10 or higher are classified as "moderate or severe depression" (Kroenke, Spitzer, and Williams 2001). PHQ-2 scores of 3 or higher are classified as likely "major depression" (Kroenke, Spitzer, and Williams 2003).

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

cluster is assigned to a treatment (and zero otherwise). In Bangladesh and Ethiopia, **T** includes an indicator for each of the two treatment arms in the respective country: *Growth Mindset* and *Growth Mindset + Girl Rising* (for Bangladesh) and *Act With Her* and *Act With Her + Transfers* (for Ethiopia). In Tanzania, *T* includes an indicator for being invited to *Goal Setting*, an indicator for being in a community with *ELA*, and the interaction between the two. Each country has a basic set of baseline characteristics in X'_{0iv} to adjust for country-specific sampling design, including adolescent age, with ϕ_s noting randomization strata fixed effects (as defined in Section II.C). Standard errors are clustered at the level of treatment (village in Tanzania, school in Bangladesh, and kebele in Ethiopia) to account for the design effect. We employ sampling weights as appropriate in Bangladesh and Ethiopia.

While the analysis here is motivated by broader preregistered protocols for each of the three country studies, the subset of outcomes and comparisons investigated do not map perfectly to these protocols.[8] Instead, we utilize the conceptual framework (Figure 1) to bring together these three distinct studies and construct outcome measures of mental health that allow for the greatest comparability of measures across studies. To further mitigate against concerns of "*p*-hacking" we implement two additional robustness checks. First, we calculate *q*-values for effect estimates that are adjusted for the false discovery rate (FDR), as described in Anderson (2008), to account for multiple hypothesis testing. We utilize Anderson's Stata code to create FDR-adjusted *q*-values separately by country and intervention across the four main outcomes. Second, we reestimate all regressions using a series of different covariates to ensure that selection of covariates does not influence the findings, essentially validating the randomization process (approach discussed in detail in Section V.B). Details on the construction of these control variables and how they map to the contextual levels in the conceptual framework can be found in Online Appendix Table 3.

We build on our estimates of average treatment effects to explore broader distributional and heterogeneous effects. This analysis is again motivated by our conceptual framework (Figure 1) and rapid scoping review (Online Appendix Table 1) and serves to shed additional light on for whom and where programs might be more effective, allowing for further triangulation of findings across interventions. We approach this in three ways. First, we conduct Kolmogorov–Smirnov (K–S) tests to assess differences in the distributions of outcomes between treatment arms, moving beyond simple differences in means. Second, we explore heterogeneity by including interactions of the treatment indicators in Equation 1 with five key baseline measures identified in the recent impact evaluation literature (see Online Appendix Table 1): wealth, education, community gender norms, risk exposure, and parental investment (discussed in more detail in Section V.C). Finally, we complement this more traditional heterogeneity

8. For Tanzania, the data used for this paper are a subsample of a broader registered RCT (Shah and Seager 2021). While the specific outcomes used in this paper are not prespecified, they fall within "other health behaviors" and "behavioral economic parameters," which were prespecified. The Bangladesh sample is a subset of a broader registered RCT (Seager et al. 2022). The measure of socio-emotional development, grit, is a component of a prespecified primary outcome measuring socio-emotional skills. The Ethiopia sample studied in this paper is also a subset of a broader registered RCT (Baird et al. 2020b). Both the depression and resilience scores were prespecified as primary outcomes for the *Act With Her* impact evaluation, and the depression indicator was prespecified as a secondary outcome of the study. Only the locus of control measure was not prespecified.

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

analysis with machine-learning techniques and utilize the causal forest algorithm developed by Athey and Wager (2019). This method allows flexible, high-dimensional combinations of covariates to identify who gains from the program in a way that researcher-determined interaction effects would typically avoid (Davis and Heller 2020). We use the generalized random forest algorithm by Tibshirani et al. (2023) to estimate conditional average treatment effects (CATEs) and descriptively summarize the characteristics of individuals who benefit most from the life skill programs based on high and low CATE groups. We provide more detail of the methodology in Section V.C.

We note that the evaluations of these interventions have been rigorously designed using detailed power calculations, with sample sizes selected based on standard minimum detectable effect sizes (see, Shah and Seager 2021; Baird et al. 2020b; Seager et al. 2022). We provide evidence on baseline balance for each country's randomization in Online Appendix Tables 4–6 and on attrition in each country's sample in Online Appendix Tables 7–9 (attrition is discussed in Section V.B).

In terms of baseline balance, we see no statistically significant differences for any treatment–country comparison on baseline measures of the outcome of interest (see Online Appendix Tables 4–6). For Tanzania, across 17 comparisons, we find one that is statistically significant at 95 percent confidence, with the joint test not statistically significant. For Bangladesh, across 48 comparisons we find five that are statistically significant at 90 percent confidence, well within the bounds of what would be expected with statistical chance. The joint F-tests on orthogonality across all baseline measures for *Growth Mindset* versus control and *Growth Mindset + Girl Rising* versus control are statistically significant ($p = 0.000$ and $p = 0.066$), so we further check our results through a series of structured covariate controls in the main estimation as robustness (see below). For Ethiopia, we find six of 48 comparisons statistically significant at 90 percent confidence, with all joint F-tests insignificant.

V. Results

A. Main Results

Table 4 and Figure 2 present impact estimates on the four distinct measures of mental health and the overall index of mental health using Equation 1. Table 4, Panel A provides results for Tanzania, Panel B for Bangladesh, and Panel C for Ethiopia. Results presented in Table 4 include basic controls (that is, adolescent age, geographic, and randomization strata fixed effects).[9] Standard errors are presented in parentheses, and p-values from K–S tests of equality of distributions between the treatment and control group are presented in square brackets below the standard errors. Figure 2 provides a graphical representation of the average effects, again with three panels, one per country.[10]

9. We provide results using more extensive control sets in Online Appendix Tables 10–12, and the results are qualitatively the same.

10. Online Appendix Tables 13–15 provide additional detail on program impacts item by item for those interested in impacts on specific questions underlying each measure. We choose to focus the main results on the validated scales for the discussion in the paper.

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

Table 4
Regression Results

	PHQ Depression Score (Standardized) (1)	Indicator for Moderate/Severe Depression (2)	Socio-Emotional Development (Standardized) (3)	Locus of Control (Standardized) (4)	Mental Health Index (Standardized) (5)
Panel A: Tanzania					
Goal Setting	0.150	0.086**	–0.057	–0.112	–0.165
	(0.102)	(0.033)	(0.111)	(0.089)	(0.103)
	[0.237]			[0.355]	[0.155]
Goal Setting × ELA	–0.208	–0.111**	0.058	0.096	0.199
	(0.131)	(0.043)	(0.137)	(0.137)	(0.129)
	[0.746]			[0.345]	[0.121]
ELA only	0.136	0.054	–0.135	–0.079	–0.174
	(0.102)	(0.037)	(0.105)	(0.118)	(0.123)
	[0.384]			[0.310]	[0.104]
Number of observations	1,167	1,167	1,176	1,176	1,167
p-value on K–S (GS + ELA vs. ELA)	[0.736]			[0.470]	[0.271]
Panel B: Bangladesh					
Growth Mindset	–0.018	–0.002	0.317***	0.175	0.286**
	(0.090)	(0.009)	(0.111)	(0.110)	(0.125)
	[0.220]		[0.128]	[0.469]	[0.054]
Growth Mindset + Girl Rising	–0.001	–0.003	0.140	0.027	0.090
	(0.107)	(0.011)	(0.103)	(0.101)	(0.095)
	[0.276]		[0.258]	[0.534]	[0.275]

(*continued*)

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

Table 4 (*continued*)

	PHQ Depression Score (Standardized) (1)	Indicator for Moderate/Severe Depression (2)	Socio-Emotional Development (Standardized) (3)	Locus of Control (Standardized) (4)	Mental Health Index (Standardized) (5)
Number of observations	1,096	1,096	1,100	1,104	1,092
p-value on (1st row – 2nd row)	[0.877]	[0.960]	[0.126]	[0.272]	[0.137]
p-value on K–S (GM vs. GM + GR)	[0.471]		[0.579]	[0.429]	[0.130]
Panel C: Ethiopia					
Act With Her	–0.113	–0.006	0.158**	0.104	0.172**
	(0.071)	(0.007)	(0.067)	(0.072)	(0.071)
	[0.767]		[0.000]	[0.301]	[0.007]
Act With Her + Transfers	0.011	0.007	0.087	0.056	0.063
	(0.091)	(0.010)	(0.072)	(0.075)	(0.076)
	[0.897]		[0.041]	[0.483]	[0.125]
Number of observations	1,518	1,518	1,405	1,617	1,288
p-value on (1st row – 2nd row)	[0.091]	[0.130]	[0.329]	[0.490]	[0.116]
p-value on K–S (AWH vs. AWH + Transfers)	[0.780]		[0.086]	[0.412]	[0.031]

Notes: Each column and panel presents results from a separate regression. The column title indicates the outcome measure (construct) of interest. Regression controls include only those necessary to account for the study sampling design by country (basic controls); more specific information on the control sets for each country can be found in Online Appendix Table 3. For regressions with other control specifications, see Online Appendix Tables 10–12. Each regression is properly weighted and clustered according to the country's sample design. Results presented for each treatment indicator include the coefficient estimates, standard errors in parentheses, and in brackets the *p*-values from Kolmogorov–Smirnov (K–S) tests of equality of distributions between the treatment group and control group. Significance: $*p<0.1$, $**p<0.05$, $***p<0.001$. The results presented in brackets in the final two rows of each panel are (i) *p*-values on the test of equality of the treatment indicator coefficients and (ii) *p*-values from Kolmogorov–Smirnov (K–S) tests of equality of distributions between the two treatment groups.

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

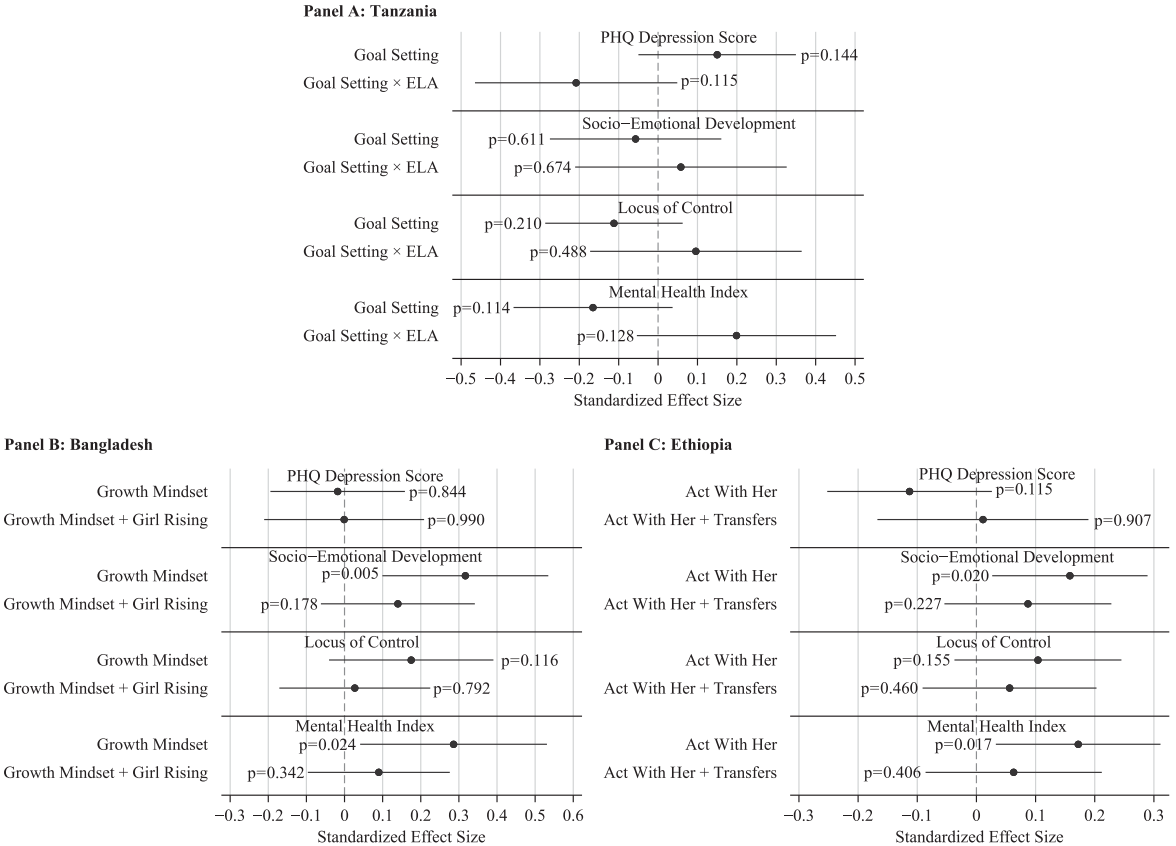

Figure 2

Main Regression Results

Notes: Each subpanel of the figure displays the point estimates (dots) and 95 percent confidence intervals (lines) on the treatment group terms from the regressions presented in Table 4 for the respective country (panel title) and outcome (subpanel title). Outcome measures for each country are constructed and standardized as described in Online Appendix Table 3. Regressions are performed as described in the notes to Table 4. The *p*-values listed next to each treatment group test the null hypothesis of no program impact (versus control) for that intervention and outcome and are drawn from the regressions presented in Table 4.

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

Panel A of Table 4 and Figure 2 summarize the results from the evaluation of the *Goal Setting* intervention in Tanzania with and without the presence of *ELA*. We find no evidence that goal setting alone improves mental health, and it may, in fact, worsen it. In communities without *ELA* clubs, girls who were invited to *Goal Setting* are 8.6 percentage points more likely to exhibit symptoms of depression ($p = 0.011$, Column 2). We also observe a sizeable positive coefficient for the standardized PHQ-2 depression score (0.15 standard deviations), though the coefficient is not statistically significant at conventional levels ($p = 0.144$, Column 1). However, girls who were invited to *Goal Setting* in *ELA* club communities appear to be better off—they are 11 percentage points less likely to exhibit symptoms of moderate/severe depression compared to girls who were invited to *Goal Setting* in non-*ELA* communities ($p = 0.012$, Column 2). It appears that *ELA* provides a base level of life skills training that might be necessary to make *Goal Setting* successful in terms of improving mental health.

The mental health index estimates corroborate this with a negative and large coefficient on *Goal Setting* (–0.165, $p = 0.114$) and a positive and large coefficient on *Goal Setting* in *ELA* communities (0.199, $p = 0.128$). Figure 3 (Panel A) suggests that this difference in means is driven by distributional differences at the bottom of the distribution.

In Bangladesh, we find no evidence of impacts on depression for either *Growth Mindset* or *Growth Mindset + Girl Rising* (Panel B of Table 4 and Figure 2). This is perhaps not surprising given the very low reported rates of moderate/severe depression in this sample (1.9 percent, Table 3). In terms of socio-emotional development and locus of control, however, we find large positive effects from *Growth Mindset*. In particular, girls assigned to the *Growth Mindset* arm exhibit 0.317 standard deviations higher socio-emotional development ($p = 0.005$, Column 3) and 0.175 standard deviations higher locus of control ($p = 0.116$, Column 4) than girls in the control group. There is also suggestive evidence that differences go beyond mean differences to broader distributional differences, particularly with socio-emotional development ($p = 0.128$, Column 3). Online Appendix Figure 1 further illustrates this point, showing a clear distributional positive shift for GM compared to the control.

While the coefficients for *Growth Mindset + Girl Rising* also go in the expected direction, they are small and not statistically significant at conventional levels. For socio-emotional development, the impact of the *Growth Mindset + Girl Rising* arm is noticeably lower than the *Growth Mindset* arm (0.140 vs. 0.317), with a *p*-value on their equality of 0.126 (Column 3). The difference between *Growth Mindset* and *Growth Mindset + Girl Rising* is the inclusion of gender-transformative messaging, delivered through stories of girls around the world overcoming obstacles to attain education. Qualitative evidence suggests that many of the girls identified with the stories of other girls' hardships, which may have highlighted external barriers to achieving goals and muted the salience of the growth mindset messaging. The *Girl Rising* stories discuss external factors that can prevent girls from completing their education and the role of social support in overcoming these challenges, potentially contradicting the *Growth Mindset* messaging that promotes internal locus of control and power to reshape outcomes. These factors may have contributed to the small and statistically insignificant impacts of *Growth Mindset + Girl Rising*. This highlights the need to consider how messaging across programs may complement or undermine each other prior to bundling.

The overall mental health index findings (Table 4, Panel B, Column 5 and Figure 2, Panel B) in Bangladesh summarize these conclusions, with large and significant positive

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

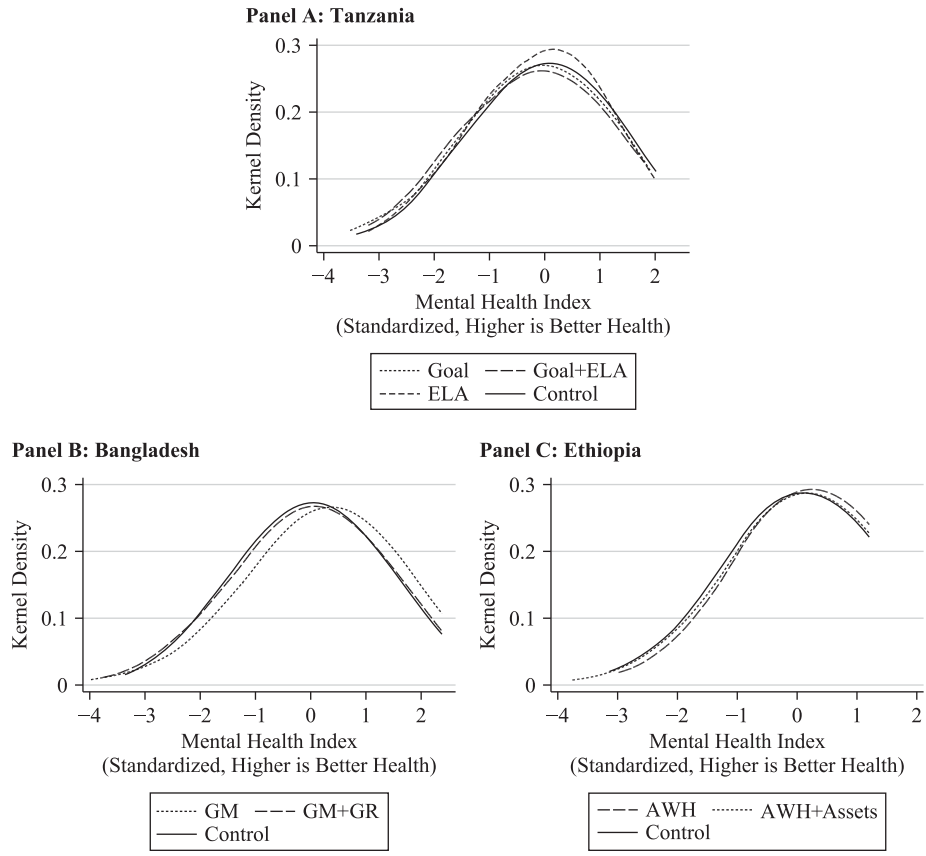

Figure 3

K-Densities for Mental Health Index by Country

Notes: Each subpanel of this figure displays, separately by country, the kernel densities for the mental health index by treatment group. The kernel densities are weighted appropriately according to each country's study design. The mental health measure is constructed and standardized as described in the text. In Panel B for Bangladesh, the data was trimmed below −4 ($n = 3$). GM and GR are abbreviations for *Growth Mindset* and *Girl Rising*, respectively. In Panel C for Ethiopia, the control group data was trimmed below −5 ($n = 3$). AWH is an abbreviation for *Act With Her*.

impacts in the *Growth Mindset* arm (0.286, $p = 0.024$) and null effects in the *Growth Mindset + Girl Rising* arm (0.090, $p = 0.342$). Figure 3, Panel B also highlights that this shift occurs across the distribution, showing a clear distributional positive shift in the mental health index for girls assigned to the *Growth Mindset* arm compared to girls assigned either to the control or to *Growth Mindset + Girl Rising*.

Panel C of Table 4 and Figure 2 present results from Ethiopia, where the most expansive programming was evaluated. We find evidence consistent with *Act With Her* improving mental health for adolescents across outcomes, with the strongest impacts on socio-emotional development. Assignment to *Act With Her* improved resilience by

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

0.158 standard deviations ($p = 0.020$, Column 3). Similar to Bangladesh, there is also strong evidence of a broader distributional shift beyond the difference in means ($p < 0.001$, Column 3). Online Appendix Figure 1 shows positive shifts in resilience at the upper end of the distribution. There is also suggestive evidence that *Act With Her* improves mental health, reducing the standardized PHQ-9 index by 0.113 standard deviations ($p = 0.115$, Column 1), and it improves locus of control by 0.104 standard deviations ($p = 0.155$, Column 4). Again, we see no impact on moderate/severe depression, but as in Bangladesh, baseline levels were low at 1 percent.

As in Bangladesh, we find that the expanded version of the program in Ethiopia, which offered asset transfers, mutes overall impacts. While coefficients are still in the desired direction, they are no longer statistically significant at conventional levels. However, there is evidence of a distributional shift for socio-emotional development between *Act With Her + Transfers* ($p = 0.041$, Column 3) versus control and evidence that the distributions for the two interventions arms differ ($p = 0.086$, Column 3; see also Online Appendix Figure 2).

One hypothesis for the cause of this muted and uneven effect is that asset transfers were only offered to adolescent girls, not adolescent boys. Qualitative data suggest there was backlash from boys because of this asymmetry that negatively influenced impacts for a subset of girls. Some of the boys reported feeling angry that girls received things like solar lamps while they did not. This result points to the importance of thinking carefully about whether programs should focus specifically on one gender, with potential unexpected adverse effects.

The impacts on the standardized mental health index once again provide a useful summary statistic for mental health, with large, positive impacts for *Act With Her* (0.172, $p = 0.017$) and muted effects for *Act With Her + Transfers* (0.063, $p = 0.406$), further exemplified by the clear shift in the distribution of mental health, particularly at the upper end of the distribution (Figure 3, Panel C).

In summary, these findings point to the potential of life skills programming to improve mental health of adolescent girls. Across 30 comparisons in Table 4, 23 go in the direction of improved mental health, something unlikely to happen by chance. Of those that go in the undesired direction, two are precisely measured null effects (*Act With Her + Transfers*), and the remaining five are all for *Goal Setting* in the absence of *ELA*—the one intervention that appears to have had an adverse effect on mental health (see Table 4).

The impacts on individual measures further highlight the importance of program design and context in driving impact. In all three countries, two variations of programming were evaluated, and, in each country, there is evidence that one version of the program improved mental health, while there are null effects (and possibly negative impacts) from the other. Impacts in Tanzania highlight the importance of understanding the underlying systems and skills that are needed to support potential program curricula. Reduced treatment effects in Bangladesh when layering gender sensitization programming suggest programs addressing gendered social norms may make salient the limitations they impose. *Girl Rising* seemed to remind girls of all the gender norms in their communities, which make their lives more difficult. Muted programming impacts from an asset transfer in Ethiopia remind us of the potential perverse effects of offering programming to particular groups, but not to their peers. We further unpack heterogeneity in Section V.C and return to the role of program design in the final section.

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

B. Robustness

1. Attrition

Overall, there is little evidence that attrition is driving our findings. In Tanzania there is no evidence of differential attrition, either overall or by baseline covariates or outcome measures (Online Appendix Table 7). For Bangladesh we are more likely to survey (by approximately 3.6 percentage points) those in the *Growth Mindset + Girl Rising* group compared to *Growth Mindset* ($p = 0.031$), with some evidence of differential attrition by baseline covariates for *Growth Mindset* versus control ($p = 0.043$, Online Appendix Table 8). In Bangladesh, follow-up survey rates are high at 92.6 percent, so these small differences in follow-up rates are likely inconsequential. For Ethiopia, there is no overall differential attrition, but some evidence of differential attrition between treatment and the control communities based on baseline covariates ($p = 0.067$ for *Act With Her* interactions, and $p = 0.084$ for *Act With Her + Transfers* interactions; Online Appendix Table 9). Given that follow-up survey rates are slightly lower in Ethiopia (86.1 percent), this further motivates our robustness check to ensure that choice of covariates is not impacting outcomes.

2. Control sets

While the randomized design should limit the effect of covariate choice on the estimated treatment impacts, we subject our main results to a battery of different covariate adjustment strategies to address any potential concerns about baseline imbalance, differential attrition, and the fact that our specific analysis in this paper was not prespecified. We take a structured approach starting with Equation 1, then add sets of adolescent and household covariates sequentially, guided by our conceptual framework (Figure 1). Online Appendix Table 3 describes these covariates in detail. We estimate our empirical models using traditional methods of controlling for covariates and double-lasso regression for covariate selection (Urminsky, Hansen, and Chernozhukov 2016). Online Appendix Tables 10–12 present the results from this estimation.

Column 1 of Online Appendix Tables 10–12 replicates the results presented in the main text, using a sparse control set for each country, based on the randomization design in that location. Column 2 then adds a set of household- and adolescent-level controls that are common across countries, what we call the "consistent" control set, as we can measure them across all three countries. Column 3 implements the double-lasso regression using the previously described consistent controls, and Column 4 uses the same method but adds an additional set of country-specific household and adolescent controls, which we call the "expanded controls." Finally, Column 5 implements the same specification as Column 4 but with the addition of the baseline outcome where available. The estimation results are consistent across specifications, supporting that randomization was implemented successfully and that any small baseline imbalances are likely inconsequential.

3. Multiple hypothesis testing

Finally, Online Appendix Table 16 provides the original *p*-values alongside the FDR-adjusted *q*-values to assess whether statistically significant findings are likely due to

multiple hypothesis testing, as opposed to real impacts. Reassuringly, across all comparisons, results are qualitatively the same once the FDR adjustments are made. Ultimately, this battery of robustness checks further confirms our core findings.

4. Heterogeneity

While our overall findings suggest that life skills programs have the potential to improve mental health for adolescent girls in LMICs, they also highlight important variation in impacts across countries, interventions, and the pre-intervention distribution of the outcome of interest. It is possible, and in fact likely, that there is further heterogeneity in treatment effects based on sample baseline characteristics. To try to further unpack these findings and assess for whom the interventions are working—in the hope of providing some generalizable insights for future programming—we conduct within-country heterogeneity analysis. We tackle this both through traditional interactions and through machine-learning techniques.

Heterogeneity Interactions. The conceptual framework in Figure 1 outlines five key dimensions along which we test for heterogeneity, which we derived from a rapid review of the impact evaluation literature on adolescent life skills programming (Online Appendix Table 1). We identified 34 randomized evaluations of adolescent life skills programming in LMICs between 2010 and mid-2023. Fifteen of these studies conducted heterogeneity analysis (summarized in Online Appendix Table 1). While our samples do not allow us to study heterogeneity across dimensions such as ethnicity, age, marital status, and rural–urban residential location, we do provide analysis along the following dimensions: (i) household wealth, which captures baseline financial support; (ii) access to education, which captures aspects of both parental support and community norms; (iii) community gender norms, which captures how supportive environments in the community might be toward young women; (iv) baseline depressive symptoms, which captures an element of adolescent risk exposure; and (vi) an indicator for whether the adolescent's mother lives in the household, which captures parental support.

Table 5 displays results from regressions that interact treatment indicators and baseline measures of each aforementioned dimension: (i) an indicator for above-median assets in the country sample,[11] (ii) an indicator for enrolled in school in Tanzania and Ethiopia and for aspired to university degree in Bangladesh,[12] (iii) an indicator for community gender norms greater than the median in the country sample,[13] (iv) an indicator for PHQ score greater than median value within the country sample, and (v) an indicator of mother living in the household at baseline. Given this analysis is exploratory, we focus this discussion on the overall index of mental health but include results for heterogeneity for each of the other four mental health outcomes in Online Appendix Tables 17–19.

11. We construct an asset index using principal components analysis following the methods of Filmer and Pritchett (2001). A household is defined as having above-median assets according to their asset score.

12. In Bangladesh, school enrollment was the basis for study enrollment. Thus, we instead use aspirations to university degree in place of enrolled in school.

13. To generate a measure of community gender norms, we calculate the average of the Index of adolescent gendered attitudes related to gender roles (defined in Online Appendix Table 3) at the community level (village in Tanzania, school in Bangladesh, kebele in Ethiopia). We use sample weights in Bangladesh and Ethiopia. We then generate an indicator that the community-level gendered attitudes are above the median within each countries study sample, meaning norms around gender are more conservative.

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

Table 5
Heterogeneity Regressions for Mental Health Index by Country

	× Indicator for Assets >Median (1)	× Indicator for Enrolled in School[a] (2)	× Indicator for Perceptions of Gendered Norms in Community >Median (3)	× Indicator for PHQ Score > Median[b] (4)	× Indicator for Mother Lives in Household (5)
Panel A: Tanzania					
Goal Setting	–0.243*	–0.209	–0.212*	–0.157	–0.227
	(0.131)	(0.161)	(0.119)	(0.131)	(0.193)
Goal Setting × ELA	0.254	0.078	0.158	0.112	0.229
	(0.166)	(0.269)	(0.144)	(0.175)	(0.287)
GS × Heterogeneity	0.168	0.056	0.125	–0.024	0.076
	(0.142)	(0.204)	(0.227)	(0.164)	(0.221)
(GS × ELA) × Heterogeneity	–0.114	0.162	0.012	0.157	–0.036
	(0.193)	(0.345)	(0.277)	(0.293)	(0.335)
Number of observations	1,167	1,167	1,167	1,139	1,167
Panel B: Bangladesh					
Growth Mindset	0.010	0.076	0.388**	0.337**	0.201
	(0.129)	(0.242)	(0.155)	(0.135)	(0.417)
Growth Mindset + Girl Rising	0.137	0.227	0.054	0.005	0.042
	(0.116)	(0.212)	(0.118)	(0.127)	(0.272)
GM × Heterogeneity	0.456***	0.242	–0.318	–0.110	0.089
	(0.167)	(0.244)	(0.224)	(0.216)	(0.421)

(*continued*)

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

Table 5 (*continued*)

	× Indicator for Assets >Median (1)	× Indicator for Enrolled in School[a] (2)	× Indicator for Perceptions of Gendered Norms in Community >Median (3)	× Indicator for PHQ Score > Median[b] (4)	× Indicator for Mother Lives in Household (5)
(GM + GR) × Heterogeneity	–0.104	–0.164	0.055	0.173	0.051
	(0.133)	(0.233)	(0.195)	(0.226)	(0.275)
Number of observations	1,092	1,092	1,092	1,092	1,092
Panel C: Ethiopia					
Act with her	0.029	0.402*	0.132**	0.227**	–0.015
	(0.079)	(0.224)	(0.066)	(0.091)	(0.126)
Act with her + Transfers	–0.013	0.403	0.017	0.118	–0.117
	(0.086)	(0.246)	(0.076)	(0.096)	(0.164)
AWH × Heterogeneity	0.342***	–0.261	0.127	–0.143	0.209
	(0.115)	(0.222)	(0.187)	(0.132)	(0.153)
(AWH + T) × Heterogeneity	0.166	–0.373	0.147	–0.133	0.198
	(0.138)	(0.252)	(0.191)	(0.154)	(0.194)
Number of observations	1,288	1,288	1,288	1,270	1,288

Notes: Each column and panel presents results from a separate regression. The column title indicates the heterogeneity metric of interest, and the panel title indicates the country sample. The outcome measure is the mental health index. Regression controls include only those necessary to account for the study sampling design (basic controls); more specific information on the control sets can be found in Online Appendix Table 3. Each regression is properly weighted and clustered according to the country's sample design. Results presented include the coefficient (standard error) on a treatment indicator or the interaction between the heterogeneity metric and a treatment indicator. Significance: *$p < 0.1$, **$p < 0.05$, ***$p < 0.001$.

[a]For Bangladesh, the educational heterogeneity term is a measure of education aspirations, as school enrollment was a requirement for study eligibility.

[b]For Ethiopia, the depression heterogeneity term is the General Health Questionnaire (GHQ-12) measure, another validated measure of mental health (as the PHQ was not collected at baseline in Ethiopia).

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

Focusing first on wealth in Column 1, Table 5, we find strong evidence in Bangladesh and Ethiopia that positive impacts on mental health are significantly larger for girls in relatively wealthier households for the programs that had statistically significant overall effects (*Growth Mindset* in Bangladesh and *Act With Her* in Ethiopia). We find no evidence of heterogeneity for the two interventions that had overall null effects (*Growth Mindset + Girl Rising* and *Act With Her + Transfers*). In Bangladesh, *Growth Mindset* has a 0.456 standard deviation ($p = 0.008$) larger effect for adolescents from households with above-median assets. Similarly, in Ethiopia, *Act With Her* had a 0.342 standard deviation ($p < 0.001$) larger effect in wealthier households. In Tanzania, though not statistically significant at conventional levels, there is also evidence that girls in households with above-median assets in the *Goal Setting* only communities fare better than those in households with at- or below-median assets (0.168, $p = 0.242$), corroborating the hypothesis that a base level of resources is necessary for *Goal Setting* programming to be effective. In *ELA* communities, where all girls are exposed to supportive programming, the coefficient on the interaction is negative but with a larger p-value ($p = 0.558$), suggesting that variation in wealth is less important when supportive programming is present. This finding is similar in spirit to findings from the evaluation of graduation programs on financial inclusion where benefits accrue to top quantiles given necessary base levels of access to savings/credit to benefit from the program (Banerjee et al. 2015).

Moving to Column 2 in Table 5, there does not appear to be statistically significant heterogeneity according to school enrollment status, though interactions are positive and sizeable for Tanzania (*Goal Setting* in *ELA* communities) and Bangladesh (both treatment arms). This is perhaps not surprising since enrollment rates are high across settings. In Bangladesh, 81 percent aspired to university education. The Ethiopia and Tanzania sample have similarly high rates of school enrollment, at 84 percent and 72 percent. Thus, the estimation of these interaction terms is based on a relatively small number of girls in each treatment arm.

Columns 3–5 show that the heterogeneity findings on the overall wellness index are not statistically significant at conventional for community gender norms, baseline mental health, or presence of mother in the household. That said, they do suggest some interesting dynamics worth further exploration. We highlight a few suggestive findings here.

Looking first at Tanzania, findings suggest that *Goal Setting* was more effective in communities with restrictive gender norms. Online Appendix Table A17, Panel B, shows that in *ELA* communities, the *Goal Setting* intervention is associated with larger decreases in moderate/severe depression in communities with more restrictive gender norms (–0.186, $p = 0.034$). There is also evidence that *Goal Setting* had more positive effects on mental health for those with higher baseline adverse mental health; adolescents with worse baseline mental health experienced greater improvements in locus of control (0.292, $p = 0.089$; Online Appendix Table A17, Panel D).

Turning to Bangladesh, findings suggest that impacts of *Growth Mindset* are muted in contexts with more conservative gender norms, while the opposite is true for *Growth Mindset + Girl Rising*, suggesting that the *Girl Rising* curriculum may be important in contexts where gender norms are particularly restrictive. We see a similar pattern in Bangladesh with regards to baseline mental health, with *Growth Mindset* once again less effective for adolescents with higher baseline adverse mental health, with the opposite for

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

Growth Mindset + Girl Rising. This is further supported by Online Appendix Table A18, which shows a significant negative interaction of *Growth Mindset* with above-median gender norms (–0.450, $p < 0.027$) and a positive and significant interaction of *Growth Mindset + Girl Rising* with above-median PHQ score at baseline (0.415, $p < 0.001$). This points to the potential importance of more comprehensive programming for those that are more at risk.

In Ethiopia, both programs appear to be most effective in contexts with restrictive gender norms, a result that likely reflects the importance of the gender transformative curriculum. This is further supported in Online Appendix Table 19, which shows that the interaction of *Act With Her + Transfers* with more restrictive gendered norms is negative and significantly associated with lower PHQ scores, indicating the program is more effective at reducing depression in contexts with more restrictive gender norms. On the other hand, both variations of *Act With Her* were less effective for those with worse baseline mental health, pointing to the importance of a more direct focus of mental health in the curricula. Finally, across all contexts, having a mother in the household generally appears to bolster program effectiveness, illustrating that broader support systems may be important for successful programming.

Overall, our heterogeneity findings using traditional interaction methods provide some evidence of important heterogeneity, particularly in terms of household wealth. We summarize the findings as follows: (i) programs appear to be more effective when some other type of support system exists (for example, wealth, education, mother in the household) and (ii) programs that focus specifically on a certain constraint (for example, gender norms) do work better for subgroups that face this constraint (for example, *Act With Her* is more effective in communities with more restrictive gender norms). However, this appears to come at a cost, as these targeted components are also counterproductive for those who do not face the constraint. This points to the importance of both identifying the constraints that are preventing adolescent girls from achieving optimal mental health and then exploring targeting of the program. We expand on this further in our discussion.

Conditional Average Treatment Effects. Next, we estimate conditional average treatment effects (CATEs) to shed further light on which adolescent girls are benefiting from these programs. Specifically, we estimate CATEs nonparametrically using the causal forest function developed by Tibshirani et al. (2023). This technique allows us to be largely "agnostic" about the set of possible dimensions for heterogeneous treatment effects and uses a machine-learning approach to assess differential effects by population subgroups. It has become increasingly common in empirical economics (for example, see Athey and Wager 2019; Davis and Heller 2020; Abaluck et al. 2020) to avoid ad hoc heterogeneity analysis. See the Appendix at the end of this paper for more description on the estimation process. Based on the estimated CATEs, we then categorize adolescents into high and low average treatment effect (ATE) groups by creating below- and above-median CATE subgroups and estimate the ATE for each group. The final step in the process is to calculate and compare the mean values of our covariates of interest across the two CATE subgroups. This approach enables us to present descriptively the characteristics of adolescents who are more likely to benefit from treatment and those who are least likely to benefit.

For power reasons, we apply the causal forest analysis to data comprising adolescent girls pooled across all three countries by assessing heterogeneity for each of our five

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

main outcomes mapped to a multidimensional vector space of common covariates observed across all countries (see Online Appendix Table 1 for full list of covariates).[14] We present results of the primary heterogeneity analysis via augmented inverse propensity scores (AIPW) based on random forests fitted to pooled data across study countries for all combinations of treatment arm comparisons.[15] Table 6 shows whether certain adolescents benefitted significantly more than others from the life skills programming—that is, is there any evidence of significant heterogeneity? If there is a significant difference in the effect between those who benefited the most (above- or equal-to-median CATE) and those who benefited the least (below-median CATE), we can then further unpack the characteristics of adolescent girls that are more likely to benefit.

Table 6 provides estimates of the CATEs across all outcomes and treatment arms. Across the 30 estimates, 50 percent have the opposite sign on the effect size between the above- and below-median CATE, and seven out of 30 show statistically significant differences between the two groups. For example, returning to our overall findings presented in Table 4, in Bangladesh there was no overall impact of *Growth Mindset* on the standardized PHQ depression score (–0.018, $p = 0.844$), but when we look at Table 6, we see that this masks considerable heterogeneity. For those with above-median CATEs (that is, more negative impacts), we see a –0.166 ($p < 0.10$) standardized effect size, while for those with below-median CATEs (that is, fewer negative impacts), we observe a positive coefficient (0.129, $p = 0.196$). The difference between these two is 0.295 and significant at the 95 percent level. Similarly, in Tanzania, we see no overall impact of *Goal Setting + ELA* on locus of control (0.096, $p = 0.488$), but we see substantial differences between those with above-median CATEs (0.106, $p = 0.418$) and those with below-median CATEs (–0.261, $p < 0.05$). Once again, the difference between these two is large (0.367 standard deviations) and significant at the 95 percent level. This finding suggests important underlying heterogeneity that drives some adolescent girls to benefit, while others are not impacted or, in some cases, made worse off.

In Table 7, we further explore differences in the characteristics of adolescent girls and their households, comparing those who benefited more (above- or equal-to-median CATE) to those who benefited less (below-median CATE). Looking at the types of adolescents who are more likely to benefit from programming across a wider set of characteristics than our empirical based heterogeneity analysis allows us to further unpack the types of adolescent girls likely to benefit from programming to inform future program design. We focus this analysis on the seven comparisons where there is a statistically significant difference between the two CATE groups in Table 6, as this is where there are potentially important differential treatment effects to unpack further. We focus on covariates that map to the conceptual framework from Figure 1.

14. Although the causal forest algorithm can handle complex relationships across our various covariates, it performs best in large sample cases. We pool our data to overcome some of the challenges of power to detect HTEs and unstable HTE estimates due to overfitting when sample sizes are relatively small.

15. The results we present use the robust augmented inverse probability weights (AIPW) approach to estimate CATEs and ATEs within CATE subgroups. The AIPW approach leverages information about adolescents' characteristics to predict the likelihood of receiving treatment. It then adjusts the data to give more weights to adolescents who received the treatments but were less likely to and less weight to those who did not receive the treatments but more likely to (Athey and Wager 2019).

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

Table 6
Conditional Average Treatment Effect Regressions

	Mental Health Index			PHQ Depression Score			Indicator for Moderate/ Severe Depression		
	Above-Median CATE (1)	Below-Median CATE (2)	Difference (3)	Above-Median CATE[a] (4)	Below-Median CATE (5)	Difference (6)	Above-Median CATE[a] (7)	Below-Median CATE (8)	Difference (9)
Panel A: Tanzania									
Goal Setting	–0.037	–0.142	0.105	0.006	0.123	–0.117	0.022	0.055	–0.033
	(0.101)	(0.099)	[0.458]	(0.108)	(0.107)	[0.442]	(0.035)	(0.039)	[0.529]
Goal Setting × ELA	–0.042	–0.059	0.017	–0.092	0.097	–0.189	–0.02	0.005	–0.025
	(0.135)	(0.145)	[0.931]	(0.135)	(0.144)	[0.338]	(0.052)	(0.046)	[0.719]
Panel B: Bangladesh									
Growth Mindset	0.106	0.08	0.025	–0.166	0.129	–0.295**	–0.02	0.026	–0.046**
	(0.106)	(0.101)	[0.864]	(0.080)	(0.100)	[0.021]	(0.010)	(0.018)	[0.025]
Growth Mindset + Girl Rising	0.177	–0.091	0.268**	0.022	0.070	–0.048	–0.017	0.016	–0.033
	(0.096)	(0.093)	[0.045]	(0.111)	(0.094)	[0.741]	(0.016)	(0.013)	[0.109]
Panel C: Ethiopia									
Act With Her only	0.159	0.146	0.013	–0.088	–0.01	–0.078	–0.016	0.003	–0.019***
	(0.075)	(0.070)	[0.899]	(0.077)	(0.056)	[0.413]	(0.007)	(0.002)	[0.009]
Act With Her + Transfers	0.074	–0.169	0.243**	0.023	0.038	–0.015	0.006	0.015	–0.009
	(0.082)	(0.088)	[0.043]	(0.085)	(0.100)	[0.909]	(0.008)	(0.011)	[0.508]

Notes: This table presents results of average treatment effects (ATEs) for adolescents based on groups defined by adolescents with high and low conditional average treatment effects (CATEs, estimated using grf model in R) for individuals in the three study settings. Above- (below-) median groups represent adolescents whose CATEs are greater than (less than) the median estimated CATEs. Standard errors are in parentheses. Each panel of the table shows the treatment effects for each CATE subgroup derived from country-specific treatment arms and the respective set of outcomes. Significance: *$p < 0.1$, **$p < 0.05$, ***$p < 0.001$.

[a]For PHQ Depression Score and Indicator for Moderate/Severe Depression, the "above-median CATE" indicates a more negative impact on the measure.

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

Table 6
Conditional Average Treatment Effect Regressions (continued)

	Socio-Emotional Development			Locus of Control		
	Above-Median CATE (10)	Below-Median CATE (11)	Difference (12)	Above-Median CATE (13)	Below-Median CATE (14)	Difference (15)
Panel A: Tanzania						
Goal Setting	0.006	–0.091	0.097	0.033	–0.115	0.148
	(0.097)	(0.094)	[0.472]	(0.095)	(0.094)	[0.268]
Goal Setting × ELA	–0.004	–0.116	0.112	0.106	–0.261	0.367**
	(0.134)	(0.129)	[0.547]	(0.131)	(0.129)	[0.046]
Panel B: Bangladesh						
Growth Mindset	0.196	0.083	0.113	0.142	0.021	0.121
	(0.097)	(0.083)	[0.376]	(0.096)	(0.103)	[0.390]
Growth Mindset + Girl Rising	0.16	0.118	0.042	0.105	–0.15	0.255*
	(0.083)	(0.089)	[0.730]	(0.097)	(0.093)	[0.058]
Panel C: Ethiopia						
Act With Her only	0.167	0.13	0.037	0.132	–0.008	0.14
	(0.066)	(0.084)	[0.729]	(0.072)	(0.068)	[0.157]
Act With Her + Assets	0.061	–0.092	0.153	0.056	–0.074	0.13
	(0.083)	(0.098)	[0.234]	(0.081)	(0.083)	[0.262]

Notes: This table presents results of average treatment effects (ATEs) for adolescents based on groups defined by adolescents with high and low conditional average treatment effects (CATEs, estimated using grf model in R) for individuals in the three study settings. Above- (below-) median groups represent adolescents whose CATEs are greater than (less than) the median estimated CATEs. Standard errors are in parentheses. Each panel of the table shows the treatment effects for each CATE subgroup derived from country-specific treatment arms and the respective set of outcomes. Significance: *$p < 0.1$, **$p < 0.05$, ***$p < 0.001$.

Table 7
Average Covariate per CATE Subgroup

	Tanzania			Bangladesh			Bangladesh		
	Locus of Control Goal Setting × ELA			PHQ Depression Score Growth Mindset			= 1 Moderate/Severe Depression Growth Mindset		
	Above-Median CATE (1)	Below-Median CATE (2)	Difference (3)	Above-Median CATE[a] (4)	Below-Median CATE (5)	Difference (6)	Above-Median CATE[a] (7)	Below-Median CATE (8)	Difference (9)
Age	14.027 (0.092)	15.391 (0.105)	−1.364*** [0.000]	14.844 (0.043)	14.853 (0.047)	−0.009 [0.888]	14.889 (0.045)	14.816 (0.045)	0.073 [0.251]
Indicator for enrolled in school	0.886 (0.013)	0.561 (0.020)	0.325*** [0.000]	0.005 (0.003)	0.002 (0.001)	0.003 [0.956]	0.481 (0.149)	0.682 (0.153)	−0.201 [0.347]
Household size	3.422 (0.056)	3.054 (0.049)	0.368*** [0.000]	5.942 (0.043)	5.631 (0.094)	0.311** [0.034]	5.819 (0.106)	5.754 (0.103)	0.065 [0.660]
Indicator for mother lives in HH	0.736 (0.018)	0.782 (0.017)	−0.046 [0.632]	0.94 (0.010)	0.944 (0.010)	−0.004 [0.777]	0.947 (0.010)	0.936 (0.010)	0.011 [0.437]
Indicator for father lives in HH	0.612 (0.127)	0.524 (0.021)	0.088*** [0.002]	0.849 (0.015)	0.862 (0.015)	−0.013 [0.540]	0.862 (0.015)	0.848 (0.015)	0.014 [0.509]
HH head years of education	8.586 (0.092)	8.836 (0.125)	−0.25 [0.161]	7.528 (0.223)	6.48 (0.207)	1.048*** [0.000]	6.054 (0.204)	7.934 (0.221)	−1.88*** [0.000]
Indicator for parent participation in paid work	0.526 (0.021)	0.658 (0.020)	−0.132*** [0.000]	0.162 (0.016)	0.118 (0.014)	0.044** [0.039]	0.127 (0.014)	0.153 (0.015)	−0.026 [0.205]
Asset count	5.556 (0.125)	4.578 (0.110)	0.978*** [0.000]	6.027 (0.137)	4.871 (0.099)	1.156*** [0.000]	4.912 (0.127)	5.974 (0.114)	−1.062*** [0.000]
Index of perceptions of gender stereotypical roles	−0.002 (0.044)	−0.01 (0.043)	0.008 [0.897]	0.106 (0.040)	0.038 (0.042)	0.068 [0.241]	0.094 (0.041)	0.051 (0.042)	0.043 [0.463]
Indicator for aspires to earn bachelor's degree	0.838 (0.015)	0.791 (0.017)	0.047** [0.038]	0.837 (0.016)	0.807 (0.017)	0.03 [0.199]	0.813 (0.017)	0.831 (0.016)	−0.018 [0.441]

Notes: This table presents country-level covariate means estimates for conditional average treatment effect (CATE) subgroups (above-median and below-median CATE) for selected outcomes and treatments. In particular, the outcome–treatment combinations shown are those that displayed statistically significant heterogeneous treatment effects across CATE subgroup in Table 6. Significance: *$p < 0.1$, **$p < 0.05$, ***$p < 0.001$.
[a]For PHQ Depression Score and Indicator for Moderate/Severe Depression, the "above-median CATE" indicates a more negative impact on the measure.

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

Table 7
Average Covariate per CATE Subgroup (continued)

	Bangladesh			Bangladesh			Ethiopia		
	Locus of Control Growth Mindset + Girl Rising			Mental Health Index Growth Mindset + Girl Rising			= 1 Moderate/Severe Depression Act with Her		
	Above-Median CATE (10)	Below-Median CATE (11)	Difference (12)	Above-Median CATE (13)	Below-Median CATE (14)	Difference (15)	Above-Median CATE[a] (16)	Below-Median CATE (17)	Difference (18)
Age	14.712 (0.042)	14.987 (0.046)	–0.275* [0.062]	14.94 (0.043)	14.758 (0.047)	0.182*** [0.004]	10.85 (0.028)	10.968 (0.029)	–0.118*** [0.003]
Indicator for enrolled in school	0.002 (0.002)	0.004 (0.003)	–0.002 [0.579]	0.571 (0.021)	0.421 (0.021)	0.150*** [0.000]	0.879 (0.012)	0.816 (0.014)	0.063*** [0.000]
Household size	5.895 (0.095)	5.69 (0.113)	0.205 [0.165]	5.799 (0.109)	5.769 (0.100)	0.03 [0.839]	6.642 (0.060)	6.124 (0.068)	0.518*** [0.000]
Indicator for mother lives in HH	0.944 (0.010)	0.94 (0.010)	0.004 [0.777]	0.934 (0.011)	0.95 (0.009)	–0.016 [0.260]	0.974 (0.006)	0.871 (0.012)	0.103*** [0.000]
Indicator for father lives in HH	0.848 (0.015)	0.863 (0.015)	–0.015 [0.800]	0.849 (0.015)	0.862 (0.015)	–0.013 [0.540]	0.904 (0.011)	0.771 (0.015)	0.133*** [0.000]
HH head years of education	9.571 (0.199)	4.425 (0.172)	5.146*** [0.000]	7.362 (0.228)	6.648 (0.202)	0.714** [0.019]	2.397 (0.147)	2.473 (0.139)	–0.076 [0.707]
Indicator for parent participation in paid work	0.132 (0.014)	0.147 (0.015)	–0.015 [0.465]	0.138 (0.015)	0.139 (0.015)	–0.001 [0.962]	0.024 (0.006)	0.157 (0.013)	–0.133*** [0.000]
Asset count	6.824 (0.111)	4.066 (0.103)	2.758*** [0.000]	5.505 (0.128)	5.378 (0.117)	0.127 [0.464]	5.241 (0.108)	5.881 (0.103)	–0.64*** [0.000]
Index of perceptions of gender stereotypical roles	–0.026 (0.043)	0.172 (0.039)	–0.198*** [0.001]	0.009 (0.041)	0.13 (0.042)	–0.121** [0.039]	0.011 (0.038)	–0.019 (0.036)	0.03 [0.567]
Indicator for aspires to earn bachelor's degree	0.9 (0.013)	0.744 (0.019)	0.156*** [0.000]	0.836 (0.016)	0.809 (0.017)	0.027 [0.247]	0.696 (0.017)	0.523 (0.018)	0.173*** [0.000]

Notes: This table presents country-level covariate means estimates for conditional average treatment effect (CATE) subgroups (above-median and below-median CATE) for selected outcomes and treatments. In particular, the outcome–treatment combinations shown are those that displayed statistically significant heterogeneous treatment effects across CATE subgroup in Table 6.

[a]For Indicator for Moderate/Severe Depression, the "above-median CATE" indicates a more negative impact on the measure.

Table 7
Average Covariate per CATE Subgroup (continued)

	Ethiopia		
	Mental Health Index Act With Her + Assets		
	Above-Median CATE (19)	Below-Median CATE (20)	Difference (21)
Age	10.873 (0.032)	10.849 (0.031)	0.024 [0.590]
Indicator for enrolled in school	0.86 (0.014)	0.885 (0.013)	–0.025 [0.191]
Household size	6.627 (0.070)	6.006 (0.067)	0.621*** [0.000]
Indicator for mother lives in HH	0.935 (0.010)	0.91 (0.011)	0.025* [0.093]
Indicator for father lives in HH	0.862 (0.014)	0.815 (0.015)	0.047** [0.022]
HH head years of education	2.315 (0.143)	2.695 (0.169)	–0.38* [0.086]
Indicator for parent participation in paid work	0.095 (0.012)	0.09 (0.011)	0.005 [0.759]
Asset count	5.602 (0.113)	5.507 (0.119)	0.095 [0.563]
Index of perceptions of gender stereotypical roles	0.102 (0.037)	–0.125 (0.043)	0.227*** [0.000]
Indicator for aspires to earn bachelor's degree	0.475 (0.020)	0.731 (0.017)	–0.256*** [0.000]

Notes: This table presents country-level covariate means estimates for conditional average treatment effect (CATE) subgroups (above-median and below-median CATE) for selected outcomes and treatments. In particular, the outcome–treatment combinations shown are those that displayed statistically significant heterogeneous treatment effects across CATE subgroup in Table 6.

Table 7 paints a complex picture of the underlying heterogeneity. We highlight a few findings worthy of further exploration in future studies. First, similar to our traditional heterogeneity findings in Table 5, benefits appear to accrue to the wealthier households. For example, in Tanzania (*Goal Setting* in *ELA* communities) and Bangladesh (*Growth Mindset and Girl Rising*), we find that above-median CATEs on locus of control are concentrated among the wealthiest. In Tanzania, the asset index was on average 5.56 (compared to 4.58) for adolescent girls who experienced above-median CATEs, and in Bangladesh, it was 6.82 (compared to 4.07). Again building on our traditional heterogeneity findings, we see that effects in Tanzania are also concentrated

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

among those enrolled in school—88.6 percent of adolescents who experienced above-median CATEs were in school versus 56.1 percent of adolescents who had below-median CATEs. In Bangladesh, where all girls were enrolled at school at baseline, we similarly see that adolescent girls who had aspirations to earn a bachelor's degree are more likely to have improved locus of control as a result of *Growth Mindset + Girl Rising*; in the above-median CATE group 90 percent aspire versus 74 percent in the below-median CATE group.

We are also able to explore age in this analysis. In Tanzania, where we have the widest age range, we do see larger treatment effects of *Goal Setting* in *ELA* communities on locus of control for younger adolescents (14.0 vs. 15.4). This finding points to the potentially important role of intervening early and supports a current push in the adolescent research field on building the evidence base on interventions for very young adolescents (Blum et al. 2014).

More generally, Table 7 suggests that benefits accrue to those that have broader underlying support systems, a finding that is also supported in our traditional heterogeneity analysis (Table 5). For example, in Tanzania, *Goal Setting* in *ELA* communities is more effective at improving locus of control for adolescents who are enrolled in school, have a father in the household, are wealthier, and have higher education aspirations at baseline. In Bangladesh, *Growth Mindset + Girl Rising* has larger effects on locus of control once again for wealthier households, those where parental education is higher, and those where adolescents had higher education aspirations. Finally, in Ethiopia, we observe larger reductions in depression for adolescents in *Act With Her* where the father and mother are in the household and the adolescent is enrolled in school and aspires to more education.

Taken together, the machine-learning findings largely align with the findings from the traditional heterogeneity analysis and, in this way, serve as a useful robustness check. They do, however, add some additional valuable nuance. In particular, these findings highlight that existing support goes beyond household wealth and mother in the household to household head's education and presence of a father in the household. In a sense, this helps create a broader understanding of what a supportive environment consists of, providing valuable insight for future programming.

Ultimately, this exploratory heterogeneity analysis underscores that programming is most effective when household and community contexts are conducive to adolescents exercising agency and when adolescents have some preexisting underlying capabilities. Alongside girl-focused adolescent programming, it is likely necessary to incorporate intervention components specifically targeting supportive household and community environments to compensate when they are absent (for example, transformative community gender norm programming that is included with *Act With Her* and *ELA* with *Goal Setting*).

VI. Discussion and Conclusion

Our findings provide novel evidence that life skills programs in LMICs can promote the mental health of adolescent girls. In Tanzania, Bangladesh, and Ethiopia, we see positive movement on at least one dimension of mental health. The size

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

of these effects is on par with effects of life skills programming on other outcomes of interest for adolescent girls.[16]

While the life skills interventions were successful at promoting mental health across all contexts, program design mattered. Effects appear larger in contexts where programs included adolescent boys alongside adolescent girls (Bangladesh and Ethiopia). This aligns with findings from Evans and Yuan (2022) and Shah et al. (2023), who show that programming that targets both genders is more effective at improving outcomes for females. This point is further supported by evidence from Ethiopia, where providing asset transfers to adolescent girls but not adolescent boys muted effects, indicating that favoring one sex over another could have detrimental impacts for all. Adding further nuance to the importance of program design, in Bangladesh, adding gender-focused programming (*Girl Rising*) to the gender-neutral intervention (*Growth Mindset*) muted the large and positive impacts of *Growth Mindset*. Qualitative evidence suggests that *Girl Rising* served as a reminder of the challenges that adolescent girls need to overcome to achieve their goals. The potential negative impact of making gender norms salient aligns with some of the suggestive evidence from heterogeneity analysis that effect sizes are smaller in communities with more gendered norms.

The existence of basic support structures seems like a necessary condition for program effectiveness. In Tanzania, the introduction of *Goal Setting* was harmful in the absence of supportive environments (*ELA*). Without the life skills provided by *ELA*, goal setting likely appeared unrealistic, thus creating undue stress for recipients. Both traditional interaction methods and machine-learning techniques highlight the importance of the existence of household wealth (measured by assets) in promoting better outcomes. While it is important to keep in mind that these measures are all relative—many of the better off households are still poor by standard definitions—this finding suggests that broader systems to promote adolescent well-being are necessary at the household and community level for life skills to bring about change, and the absence of them may in fact turn impacts negative. This is also why interventions that tackle constraints at multiple levels are likely to be particularly important when contexts do not have the broader systems in place. At the same time, our heterogeneity analyses point to the importance of tailoring programming to the specific constraints that adolescents are facing to maximize program benefit.

Finally, our results suggest that life skills interventions targeted at younger adolescents may be more efficacious at improving mental health more broadly. In Bangladesh and Ethiopia, where interventions focused on younger adolescents (mostly 10–14 years old in Bangladesh and 10–12 years old in Ethiopia), one intervention in each setting improved socio-emotional development, and there were suggestive improvements in locus of control. In addition, in our exploratory heterogeneity analysis, there is evidence of improvement in locus of control in Tanzania among younger adolescents. These results point to intervening with adolescents early. At the same time, Tanzania, where 50 percent of the sample is older than 15 years, is the only country where there was evidence of reductions in depression, as measured by the PHQ, suggesting that impacts on depression manifest at older ages during adolescence, when prevalence is higher

16. For example, Dhar, Jain, and Jayachandran (2022) find effect sizes on gender norms of 0.18 standard deviations of gender-focused life skills programming. Bandiera et al. (2020a) find *ELA* in Uganda has effect sizes of 0.12 standard deviations for gender empowerment and 0.370 standard deviations for HIV knowledge.

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

(WHO 2021). This aligns with previous research that has found that for depressive disorders, 3.1 percent of diagnosed cases are onset by age 14 years and 13.2 percent of cases are onset by age 18 (Solmi et al. 2022) and points to the importance of continued intervention over the course of adolescence to address changing needs.

Our findings, drawn from diverse populations exposed to six interventions across three countries, begin to unpack the conditions under which life skills interventions are likely to promote mental health for adolescent girls. As life skills programs remain a feature of adolescent programming globally, future research is needed. Ultimately, life skills programming is unlikely to be the golden ticket to improved adolescent mental health without improving gender norms and the education, health, and financial systems that underpin the broader society where LIMC adolescents reside.

Appendix: Heterogeneous Treatment Effects Estimation via Causal Forest

We present a step-by-step process of how we estimate heterogeneous treatment effects (HTEs) in this section. We use the grf package developed by Tibshirani et al. (2023) to estimate all our causal forest models and implement these stages in R. In general terms, this package combines random forests, a popular machine-learning (ML) algorithm, with causal inference techniques. In our case, we use this to explore and understand how the mental health effects of life skills programming varies across observed characteristics of adolescents. The estimation algorithm is able to handle complex relationships across our various covariates without imposing any strict assumptions about our data generating process. The algorithm performs best in large sample cases, due to potential issues of reduced power to detect HTEs and risk of overfitting the models resulting in unstable HTE estimates. We circumvent some of these challenges by pooling our data across our study settings and conducting standardized checks to validate the robustness of our estimates. Our estimation process comprises the following stages: training, evaluation, and exploring heterogeneity stages. We discuss details of these three stages below.

1. **Training stage.** The first step in building our causal forest model to assess HTEs is the training stage. In this stage, the causal forest algorithm develops several decision trees, and each is trained on a random sample of observations and all covariates across study countries to predict the treatment effect for each adolescent girl. We utilize a doubly robust (DR) estimator in the training stage to ensure that we achieve the best model performance. The DR estimator first predicts both treatment propensities and outcome values for each observation. We then combine the predicted treatment propensities and outcome values as inputs for our main causal forest model to minimize bias in the estimated conductional average treatment effects (CATEs). The DR approach uses the two models to construct a consistent estimate for the CATEs even if either one of the models are misspecified.

   Other inputs for the trained model include the treatment and outcome values, our set of covariates, and other arguments like "tune.parameters" that ensure

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

optimal model performance.[17] For instance, by specifying the "tune.parameters = TRUE" in our case, the model considers the total number of observations in each tree node split trees. This ensures all characteristics of observations in each split node are considered in the classification process for estimating the treatment effect for each observation. We then use the causal forest model to estimate CATEs for each adolescent observation. We repeat the estimation for various treatment and outcome combinations in the pooled data. In total, we estimated 30 different CATE models that comprised a blend of treatment indicators for the different arms of setting-specific treatment for each outcome of interest.

2. **Evaluation.** After estimating our CATEs in the training phase, we evaluate the model's performance regarding potential overfitting issues. This is to assess whether our models have captured the true underlying patterns of heterogeneity in our data set and to ensure that CATEs are not biased due to noise or random fluctuations of signal in the data set. Following Chernozhukov and Semenova (2018), we use the best linear predictor (BLP) regression to evaluate the performance of CATE predictions on the held-out data. The BLP estimation validates the performance of our trained causal forest model on held-out trials (that is, predicting the values of the outcome for subpopulations not used to train the models). Overall, the BLP analysis showed that our models performed well and were efficiently calibrated. Thus, we did not have to do additional hyperparameter tuning. The mean prediction (coefficient that shows the presence of heterogeneity) form the BLP analysis, range from 0.83 to 2.53. The closer a coefficient is to one, the stronger the presence of heterogeneous effects in the data. We test the statistical significance of the models with BLP mean prediction coefficients between 0.95 and 1.2 to assess the presence of heterogeneity.
3. **Exploring heterogeneity.** After estimating the CATEs and evaluating the model performance, the next step is to assess whether there is variation in the adolescent level treatment effects across the different treatments and outcomes in each study country. To do this we first disaggregate the pooled data to the country level. We then group adolescent girls in each country according to whether their CATE estimates for each outcome are above (or equal) or below the median treatment effect (that is, groups girls into high or low treatment effects). We then estimate the average treatment effects in these two subgroups across each outcome and country-specific treatment separately using the augmented inverse probability weights (AIPWs) predicted from the propensity model during the training stage. We then make inferences on the evidence of heterogeneity by testing if the identified treatment effects in the high and low CATE subgroups are statistically different from each (that is, test null of no heterogeneity in treatment effects). This procedure is heuristic, as the "high" and "low" subgroups only provide qualitative insights about the strength of heterogeneity in treatment effects. Overall, we find that only seven out of the 30

17. When the "tune.parameters" is turned on, our models identify the ideal values for parameters such as num.trees, min.node.size, and splitrule for which our models performs best.

models we estimated showed strong evidence of heterogeneous effects. We then select these models for the final stages of the analysis. In the final step, we determine the important dimensions of heterogeneity in our treatment effects by calculating the sample means of the covariates used to train the model for the CATE subgroups and test whether the sample means are statistically different when comparing adolescents with above-median CATE to those below. Through this, we are able to identify qualitatively the characteristics of young girls/women who benefit most (and least) from the life skills programs.

References

Abaluck, J., L. Agha, D.C. Chan Jr., D. Singer, and D. Zhu. 2020. "Fixed Misallocation with Guidelines: Awareness vs. Adherence." NBER Working Paper 27467. Cambridge, MA: NBER.

Alan, S., T. Boneva, and S. Ertac. 2019. "Ever Failed, Try Again, Succeed Better: Results from a Randomized Educational Intervention on Grit." *Quarterly Journal of Economics* 134(3): 1121–62.

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):1481–95.

Andrew, A., S. Krutikova, G. Smarrelli, and H. Verma. 2022. "Gender Norms, Violence and Adolescent Girls' Trajectories: Evidence from a Field Experiment in India." Working Paper 22/41. London: Institute for Fiscal Studies.

Anum, A., S. Adjorlolo, and N. Kugbey. 2019. "Depressive Symptomatology in Adolescents in Ghana: Examination of Psychometric Properties of the Patient Health Questionnaire-9." *Journal of Affective Disorders* 256:213–18.

Ashraf, N., N. Bau, C. Low, and K. McGinn. 2020. "Negotiating a Better Future: How Interpersonal Skills Facilitate Intergenerational Investment." *Quarterly Journal of Economics* 135 (2):1095–151.

Athey, S., and S. Wager. 2019. "Estimating Treatment Effects with Causal Forests: An Application." *Observational Studies* 5(2):37–51.

Azzopardi, P., Z. Hijazi, N. Wulan, J. Requejo, J. Lai, L. Carvajal-Velez, and G. Patton. 2023. "Bringing a Wider Lens to Adolescent Mental Health: Aligning Measurement Frameworks with Multisectoral Actions." *Journal of Adolescent Health* 72:S9–S11.

Baird, S., Z. Bhutta, B. Abu Hamad, J. Hamory Hicks, N. Jones, and J. Muz. 2019. "Do Restrictive Gender Attitudes and Norms Influence Physical and Mental Health during Very Young Adolescence? Evidence from Bangladesh and Ethiopia." *SSM-Population Health* 9:100480.

Baird, S., J. de Hoop, and B. Özler. 2013. "Income Shocks and Adolescent Mental Health." *Journal of Human Resources* 48(2):370–403.

Baird, S., J. Hicks, N. Jones, T. Woldehanna, and W. Yadete. 2020b. "Gender Transformative Programming to Improve Capabilities of Young Adolescents in Ethiopia: A Cluster Randomized Control Trial." AEA RCT Registry. https://doi.org/10.1257/rct.4024-1.1

Baird, S., B. Özler, C. Dell'Aira, and D. Us Salam. 2020a. "Using Group Interpersonal Psychotherapy to Improve the Well-Being of Adolescent Girls." *Journal of Development Economics*, Registered Report Phase 1. https://drive.google.com/file/d/1tc9_52GtcKhc1VZWvRd9Iina QmDvJRrH/view (accessed October 19, 2023).

Bandiera, O., N. Buehren, R. Burgess, M. Goldstein, S. Gulesci, I. Rasul, and M. Sulaiman. 2020a. "Women's Empowerment in Action: Evidence from a Randomized Control Trial in Africa." *American Economic Journal: Applied Economics* 12(1):210–59.

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

Bandiera, O., N. Buehren, M. Goldstein, I. Rasul, and A. Smurra. 2019. "The Economic Lives of Young Women in the Time of Ebola: Lessons from an Empowerment Program." World Bank Policy Research Working Paper 8760. Washington, DC: World Bank.

Bandiera O., Buehren, N., Goldstein, M., Rasul, I., Smurra, A. 2020b. "Do School Closures during an Epidemic Have Persistent Effects? Evidence from Sierra Leone in the Time of Ebola." Working Paper.

Banerjee, A., E. Duflo, N. Goldberg, D. Karlan, R. Osei, W. Pariente, J. Shapiro, B. Thuysbaert, and C. Udry. 2015. "A Multifaceted Program Causes Lasting Progress for the Very Poor: Evidence from Six Countries." *Science* 348(6236):1260799.

Barry, M.M., A.M. Clarke, R. Jenkins, and V. Patel. 2013. "A Systematic Review of the Effectiveness of Mental Health Promotion Interventions for Young People in Low and Middle Income Countries." *BMC Public Health* 13(1):835.

Beattie, T.S., R. Prakash, A. Mazzuca, L. Kelly, P. Javalkar, T. Raghavendra, S. Ramanaik, M. Collumbien, S. Moses, L. Heise, S. Isac, and C. Watts. 2019. "Prevalence and Correlates of Psychological Distress among 13–14 Year Old Adolescent Girls in North Karnataka, South India: A Cross-Sectional Study." *BMC Public Health* 19(1):48.

Beck, A.T. 1976. *Cognitive Therapy and the Emotional Disorders.* New York: International Universities Press.

Blum, R.W., N.M. Astone, M.R. Decker, and V.C. Mouli. 2014. "A Conceptual Framework for Early Adolescence: A Platform for Research." *International Journal of Adolescent Medicine and Health* 26(3):321–31.

Bøe, T., S. Øverland, A. Lundervold, and M. Hysinng. 2012. "Socioeconomic Status and Children's Mental Health: Results from the Bergen Child Study." *Social Psychiatry Epidemiology* 47:1557–66.

Bøe, T., E. Dearing, K.M. Stormark, and H.D. Zachrisson. 2018. "Subjective Economic Status in Adolescence and Associations with Mental Health in the Hordaland Study." *Journal of Family and Economic Issues* 39:232–336.

Boxho, C., D. Kanga, E. Koussoubé, L. Rouanet, and O. Boulhane. 2023. "Empowering Adolescent Girls through Safe Spaces and Accompanying Measures in Cote d'Ivoire." Unpublished.

Buchman, N., E. Field, R. Glennerster, N. Shahana, S. Pimkina, and I. Sen. 2022. "Power or Money: Alternative Approaches to Reducing Child Marriage in Bangladesh, a Randomized Control Trial." Unpublished.

Buehren, N., M. Goldstein, S. Gulesci, M. Sulaiman, and V. Yam. 2017. "Evaluation of an Adolescent Development Program for Girls in Tanzania." World Bank Policy Research Working Paper 7961. Washington, DC: World Bank.

Campbell, O.L.K., D. Bann, and P. Patalay. 2021. "The Gender Gap in Adolescent Mental Health: A Cross-National Investigation of 566,829 Adolescents across 73 Countries." *SSM – Population Health* 13:100742.

Carvajal, L., J.W. Ahs, J.H. Requejo, C. Kieling, A. Lundin, M. Kumar, N.P. Luitel, M. Marlow, S. Skeen, M. Tomlinson, and B.A. Kohrt. 2022. "Measurement of Mental Health among Adolescents at the Population Level: A Multicountry Protocol for Adaptation and Validation of Mental Health Measures." *Journal of Adolescent Health* 72(1):S27–S33.

Chernozhukov, V., and V. Semenova. 2018. "Simultaneous Inference for Best Linear Predictor of the Conditional Average Treatment Effect and other Structural Functions." CeMMAP Working Paper CWP40/18. London: Centre for Microdata Methods and Practice, Institute for Fiscal Studies.

Cobb-Clark, D.A. 2015. "Locus of Control and the Labor Market." *IZA Journal of Labor Economics* 4:3.

Currie, Janet, and Mark Stabile. 2006. "Child Mental Health and Human Capital Accumulation: The Case of ADHD." *Journal of Health Economics* 25(6):1094–118.

Currie, Janet, Mark Stabile, Phongsack Manivong, and Leslie Roos. 2010. "Child Health and Young Adult Outcomes." *Journal of Human Resources* 45(3):517–48.

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

Dardas, L.A., B. van de Water, and L.A. Simmons. 2018. "Parental Involvement in Adolescent Depression Interventions: A Systematic Review of Randomized Clinical Trials." *International Journal of Mental Health Nursing* 27:550–70.

Davis, J.M.V., and S.B. Heller. 2020. "Rethinking the Benefits of Youth Employment Programs: The Heterogeneous Effects of Summer Jobs." *Review of Economics and Statistics* 102(4): 664–77.

Demographic and Health Survey. 2022. STAT Compiler Database. https://www.statcompiler.com/en/ (accessed October 19, 2023).

Dhar, D., T. Jain, and S. Jayachandran. 2022. "Reshaping Adolescents' Gender Attitudes: Evidence from a School-Based Experiment in India." *American Economic Review* 112(3): 899–927.

Doran, G.T. 1981. "There's a S.M.A.R.T. Way to Write Management's Goals and Objectives." *Management Review* 70(11):35–36.

Dweck, C.S. 2006. *Mindset: The New Psychology of Success*. New York: Random House.

Edmonds, E., B. Feigenberg, and J. Leight. 2023. "Advancing the Agency of Adolescent Girls." *Review of Economics and Statistics* 105(4):1–15.

Evans, D.K., and F. Yuan. 2022. "What We Learn about Girls' Education from Interventions that Do Not Focus on Girls." *World Bank Economic Review* 36(1):244–67.

Falb K.L., S. Tanner, L. Ward, D. Erksine, E. Noble, A. Asham, T. Bakomere, E. Graybill, C. Lowry, P. Mallinga, A. Neiman, C. Poulton, K. Robinette, M. Sommer, and L. Stark. 2016. "Creating Opportunities through Mentorship, Parental Involvement, and Safe Spaces (COMPASS) Program: Multi-Country Study Protocol to Protect Girls from Violence in Humanitarian Settings." *BMC Public Health* 16(1):231.

Fernald, L.C., P.J. Gertler, and L.M. Neufeld. 2009. "10-Year Effect of Oportunidades, Mexico's Conditional Cash Transfer Programme, on Child Growth, Cognition, Language, and Behaviour: A Longitudinal Follow-Up Study." *Lancet* 374(9706):1997–2005.

Filmer, D., and L.H. Pritchett. 2001. "Estimating Wealth Effects without Expenditure Data-or Tears: An Application to Educational Enrollments in States of India." *Demography* 38(1):115–32.

GAGE Consortium 2019. *Gender and Adolescence. Why Understanding Adolescent Capabilities, Change Strategies and Contexts Matters*. Second edition. London: Gender and Adolescence: Global Evidence.

Gaudiano, B.A. 2008. "Cognitive-Behavioral Therapies: Achievements and Challenges." *Evidence Based Mental Health* 11(1):5–7.

Grealish, A., S. Tai, A. Hunter, R. Emsley, T., Murrells, and A.P. Morrison. 2017. "Does Empowerment Mediate the Effects of Psychological Factors on Mental Health, Well-Being, and Recovery in Young People?" *Psychology and Psychotherapy* 90(3):314–35.

Guglielmi, S., K. Mitu, and J. Seager. 2021. "'I Just Keep Quiet': Addressing the Challenges of Married Rohingya Girls and Creating Opportunities for Change." *European Journal of Development Research* 33:1232–51.

Gureje, O., and B.O. Oladeji. 2022. "Quality Care for People with Severe Mental Disorders in Low-Resource Settings." *Lancet Psychiatry* 9(1):3–5.

Haberland, N.A., K.J. McCarthy, and M. Brady. 2018. "A Systematic Review of Adolescent Girl Program Implementation in Low- and Middle-Income Countries: Evidence Gaps and Insights." *Journal of Adolescent Health* 63(1):18–31.

Haerpfer, C., R. Inglehart, A. Moreno, C. Welzel, K. Kizilova, J. Diez-Medrano, M. Lagos, P. Norris, E. Ponarin, and B. Puranen (eds.) 2022. "World Values Survey: Round Seven—Country-Pooled Datafile Version 5.0.0." Madrid, Spain & Vienna, Austria: JD Systems Institute & WVSA Secretariat. https://doi.org/10.14281/18241.20

Hale, D.R., L. Bevilacqua, and R.M. Viner. 2015. "Adolescent Health and Adult Education and Employment: A Systematic Review." *Pediatrics* 136(1):128–40.

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

Heckman, J.J., J. Stixrud, and S. Urzua. 2006. "The Effects of Cognitive and Noncognitive abilities on Labor Market Outcomes and Social Behavior." *Journal of Labor Economics* 24(3):411–82.

Kabeer, N. 2003. "Making Rights Work for the Poor: Nijera Kori and the Construction of 'Collective Capabilities' in Rural Bangladesh." Working Paper 200. Brighton: Institute of Development Studies.

Kapungu, C., and S. Petroni. 2017. "Understanding and Tackling the Gendered Drivers of Poor Adolescent Mental Health." Washington, DC: International Center for Research on Women.

Kling, J.R., J.B. Liebman, and L.F. Katz. 2007. "Experimental Analysis of Neighborhood Effects." *Econometrica* 75(1):83–119.

Koenig, L., R. Blum, D. Shervington, J. Green, M. Li, H. Tabana, and C. Moreau. 2021. "Unequal Gender Norms Are Related to Symptoms of Depression among Young Adolescents: A Cross-Sectional, Cross-Cultural Study." *Journal of Adolescent Health* 69(1):S47–S55.

Krishnan, P., and S. Krutikova. 2012. "Noncognitive Skill Formation in Poor Neighborhoods of Urban India." Cambridge Working Papers in Economics 1010.

Kroenke, K., R.L. Spitzer, and J.B. Williams. 2001. "The PHQ-9: Validity of a Brief Depression Severity Measure." *Journal of General Internal Medicine* 16:606–13.

———. 2003. "The Patient Health Questionnaire-2: Validity of a Two-Item Depression Screener." *Medical Care* 41(11):1284–92.

Layne, C.M., W.R, Saltzman, L. Poppleton, G.M. Burlingame, A. Pašalić, E. Duraković, M. Mušić, N. Ćampara, N. Dapo, B. Arslanagić, A.M. Steinberg, and R.S. Pynoos. 2008. "Effectiveness of a School-Based Group Psychotherapy Program for War-Exposed Adolescents: A Randomized Controlled Trial." *Journal of the American Academy of Child & Adolescent Psychiatry* 47(9):1048–62.

Leventhal, K.S., P.L. Cooper, L.M. DeMaria, P. Priyam, H. Shanker, G. Andrew, and S. Leventhal. 2022. "Promoting Wellbeing and Empowerment via Youth First: Exploring Psychosocial Outcomes of a School-Based Resilience Intervention in Bihar, India." *Frontiers in Psychiatry* 13. https://doi.org/10.3389/fpsyt.2022.1021892

Liebenberg, L., M. Ungar, and J.C. LeBlanc. 2013. "The CYRM-12: A Brief Measure of Resilience." *Canadian Journal of Public Health* 104(2):e131–e135.

Lochman, J.E., N.P. Powell, C.L. Boxmeyer, and L. Jimenez-Camargo. 2011. "Cognitive-Behavioral Therapy for Externalizing Disorders in Children and Adolescents." *Child and Adolescent Psychiatric Clinics of North America* 20(2):305–18.

Manwell, L.A., S.P. Barbic, K. Roberts, Z. Durisko, C. Lee, E. Ware, and K. McKenzie. 2015. "What Is Mental Health? Evidence towards a New Definition from a Mixed Methods Multidisciplinary International Survey." *BMJ Open* 5(6):e007079.

Mazzuca, A., D. Nagarchi, S. Ramanaik, T. Raghavendra, P. Javalkar, S. Rotti, P. Bhattarcharjee, S. Isac, A. Cohen, and T. Beattie. 2019. "Developing a Mental Health Measurement Strategy to Capture Psychological Problems among Lower Caste Adolescent Girls in Rural, South India." *Transcultural Psychiatry* 56(1):24–47.

Mekuria, F., A. Sprinkel, and E. Cowan. 2018. *Social Analysis and Action Global Implementation Manual*. Atlanta, GA: Cooperative for Assistance and Relief Everywhere, Inc. (CARE).

Miller-Graff, L.E., and E.M. Cummings. 2022. "Supporting Youth and Families in Gaza: A Randomized Controlled Trial of a Family-Based Intervention Program." *IJERPH 19*(14):8337.

Musumari, P.M., A. Tangmunkongvorakul, K. Srithanaviboonchai, T. Techasrivichien, S.P. Suguimoto, M. Ono-Kihara, and M. Kihara. 2018. "Grit Is Associated with Lower Level of Depression and Anxiety among University Students in Chiang Mai, Thailand: A Cross-Sectional Study." *PLoS One* 13(12):e0209121.

Nussbaum, M. 2011. *Creating Capabilities: The Human Development Approach*. Cambridge, MA: Harvard University Press.

Orth, Z., F. Moosajee, and B. Van Wyk. 2022. "Measuring Mental Wellness of Adolescents: A Systematic Review of Instruments." *Frontiers in Psychology* 13. https://doi.org/10.3389/fpsyg.2022.835601

Patel, V. 2013. "Reducing the Burden of Depression in Youth: What Are the Implications of Neuroscience and Genetics on Policies and Programs?" *Journal of Adolescent Health* 52(2): S36–S38.

———. 2018. "Acting Early: The Key to Preventing Mental Health Problems." *Journal of the Royal Society of Medicine* 111(5):153–57.

Paunesku, D., G. Walton, C. Romero, E. Smith, D. Yeager, and C. Dweck. 2015. "Mind-Set Interventions are a Scalable Treatment for Academic Underachievement." *Psychology Science* 26(6):784–93.

Pearlin, Leonard I., and Carmi Schooler. 1978. "The Structure of Coping." *Journal of Health and Social Behavior* 19(1):2–21.

Pollard, E.L., and P.D. Lee. 2003. "Child Well-Being: A Systematic Review of the Literature." *Social Indicators Research* 61:59–78.

Reiss, F. 2013. "Socioeconomic Inequalities and Mental Health Problems in Children and Adolescents: A Systematic Review." *Social Science and Medicine* 90:24–31.

Renwick, L., R. Pedley, I. Johnson, V. Bell, K. Lovell, P. Bee, and H. Brooks. 2022. "Conceptualisations of Positive Mental Health and Wellbeing among Children and Adolescents in Low- and Middle-Income Countries: A Systematic Review and Narrative Synthesis." *Health Expectations* 25(1):61–79.

Richter, L.M., M.N. Ahun, S. Besharati, S.N. Naicker, and M. Orri. 2021. "Adolescent Mental Health Problems and Adult Human Capital: Findings from the South African Birth to Twenty Plus Cohort at 28 Years of Age." *Journal of Adolescent Health* 69(5):782–89.

Seager, J., T.M. Asaduzzaman, S. Baird, S. Guglielmi, S. Faiz Rashid, N. Jones, S. Sabarwal, M. Sultan, and A. Vyas. 2022. "Improving Education and Gender Attitudes: A Cluster RCT of Growth Mindset and Girl Rising among Adolescents in Bangladesh." AEA RCT Registry. https://doi.org/10.1257/rct.9214-1.1

Sen, A. 1984. Commodities and Capabilities. Development. Oxford: Oxford University Press.

———. 2004. "Capabilities, Lists, and Public Reason: Continuing the Conversation." *Feminist Economics* 10(3):77–80.

Shah, M., and J. Seager. 2021. "Promoting Safe Sex among Adolescents in Tanzania." AEA RCT Registry. https://doi.org/10.1257/rct.1305-6.3

Shah, M., J. Seager, J. Montalvao, and M. Goldstein. 2023. "Sex, Power, and Adolescence: Intimate Partner Violence and Sexual Behaviors." NBER Working Paper 31624. Cambridge, MA: NBER.

Singla, D.R., A. Waqas, S.U. Hamdani, N. Suleman, S.W. Zafar, Zill-e-Huma, K. Saeed, C. Servili, and A. Rahman. 2020. "Implementation and Effectiveness of Adolescent Life Skills Programs in Low- and Middle-Income Countries: A Critical Review and Meta-Analysis." *Behaviour Research and Therapy* 130:103402.

Smith, J.P., and G.C. Smith. 2010. "Long-Term Economic Costs of Psychological Problems during Childhood." *Social Science and Medicine* 71(1):110–15.

Solmi, M., J. Radua, M. Olivola, E. Croce, L. Soardo, G.S. de Pablo, J. Shin, J.B. Kirkbride, P. Jones, J.H. Kim, J.Y. Kim, A.F. Carvalho, M.V. Seeman, C.U. Correll, and P. Fusar-Poli. 2022. "Age at Onset of Mental Disorders Worldwide: Large-Scale Meta-Analysis of 192 Epidemiological Studies." *Molecular Psychiatry* 27:281–95.

Ssewamala F.M., C. Han, and T.B. Neilands. 2009. "Asset Ownership and Health and Mental Health Functioning among AIDS-Orphaned Adolescents: Findings from a Randomized Clinical Trial in Rural Uganda." *Social Science and Medicine* 69(2):191–8.

Swift, C., and G. Levin. 1987. "Empowerment: An Emerging Mental Health Technology." *Journal of Primary Prevention* 8(1&2):71–94.

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023

Tait, R.J., D.J. French, and G.K. Hulse. 2003. "Validity and Psychometric Properties of the General Health Questionnaire-12 in Young Australian Adolescents." *Australian and New Zealand Journal of Psychiatry* 37(3):374–381.

Temin, M., and C.J. Heck. 2020. "Close to Home: Evidence on the Impact of Community-Based Girl Groups." *Global Health: Science and Practice* 8(2):300–324.

Tibshirani, J., S. Athey, R. Friedberg, V. Hadad, D. Hirshberg, L. Miner, E. Sverdrup, S. Wager, and M. Wright. 2023. "Generalized Random Forests: Package 'grf'." https://github.com/grf-labs/grf (accessed October 17, 2023).

UNESCO Institute for Statistics. 2020. "Data for the Sustainable Development Goals." http://uis.unesco.org/ (accessed October 19, 2023).

United Nations, Department of Economic and Social Affairs, Population Division. 2022. "World Population Prospects 2022." https://population.un.org/wpp/ (accessed October 19, 2023).

Urminsky, O., C. Hansen, and V. Chernozhukov. 2016. "Using Double-Lasso Regression for Principled Variable Selection." http://dx.doi.org/10.2139/ssrn.2733374

Viner, R.M., N.B. Allen, and G.C. Patton. 2017. "Puberty, Developmental Processes, and Health Interventions." In *Child and Adolescent Health and Development*, 3rd edition, ed. D.A.P. Bundy, N.D. Silva, S. Horton, D.T. Jamison, and G.C. Patton. Washington, DC: The International Bank for Reconstruction and Development/The World Bank.

Vyas, A.N., N. Nagaraj, J. Genovese, G. Malhotra, N. Dubey, R. Hingorani, and L. Manning. 2020. "The Girl Rising 'We Dream, We Rise' Social Media Campaign in India: Reach, Engagement and Impact." *Journal of Creative Communications* 15(1):106–24.

WHO. 2021. "Adolescent Mental Health. https://www.who.int/news-room/fact-sheets/detail/adolescent-mental-health (accessed October 19, 2023).

———. 2022. "World Mental Health Report: Transforming Mental Health for All." Geneva: World Health Organization.

Yeager, D.S., and C.S. Dweck. 2012. "Mindsets that Promote Resilience: When Students Believe that Personal Characteristics Can Be Developed." *Educational Psychologist* 47:302–14.

Zhou, W., F. Ouyang, O.E. Nergui, J.B. Bangura, K. Acheampong, I.Y. Massey, and S. Xiao. 2020. "Child and Adolescent Mental Health Policy in Low- and Middle-Income Countries: Challenges and Lessons for Policy Development and Implementation." *Frontiers in Psychiatry* 11:150.

Zimmerman, A., E. Garman, M. Avendano-Pabon, R. Araya, S. Evans-Lacko, D. McDaid, A. Park, P. Hessel, Y. Dias, A. Matijasevich, C. Ziebold, A. Bauer, C. Silvestre Paula, and C. Lund. 2021. "The Impact of Cash Transfers on Mental Health in Children and Young People in Low-Income and Middle-Income Countries: A Systematic Review and Meta-Analysis." *BMJ Global Health* 6:e004661.

Downloaded from at BIOSCIENCE & NATURAL RES on February 26, 2025. Copyright 2023