Menstrual Stigma, Hygiene, and Human Capital: Experimental Evidence from Madagascar

Karen Macours*, Julieta Vera Rueda†, Duncan Webb*

May 27, 2024

Abstract

This paper presents results from an RCT in 140 schools in Madagascar that targets both hygiene practices and menstrual stigma. First, we show that a bundle of interventions (sanitation infrastructure, menstrual products, and teacher sensitization) leads to substantial (0.15 SD) improvements in learning tests and school marks, without affecting attendance or health. These learning benefits appear to be driven by reduced stress and an improved psychosocial environment in treatment schools, where girls' heart rate at endline is lower (-0.12 SD), severe bullying is less common (-0.08 SD), and a measure of network integration is higher (+0.24 SD). Second, we evaluate the additional effect of nominating and coaching "young girl leaders" - school girls willing to speak out against menstrual stigma - to spread positive messages about hygiene and menstruation. The combined program generates substantial improvements in hygiene knowledge and behavior (0.33-0.56 SD) and in menstrual stigma (0.74 SD), and the Young Girl Leader component significantly increases the impact on all of these dimensions.

*Paris School of Economics. † University College London. We would like to thank our partners, CARE Madagascar and CARE France for their hard work and unwavering commitment to the project – especially Emilie Vigeanel, Marina Ogier, Andrianarilaza Rakotovao, Emma Hanitriniaina, Ifalianiaina Mihantarijaona, and Hasina Razakarivelonirina. We are indebted to Eliane Ralison and the entire survey team from Consultant Associates for countless inputs throughout the survey and lab-in-the-field instrument design, excellent data collection, and for their perseverance and ingenuity in tracking girls that moved. We thank Emanuela Galasso for sharing validated standardized test instruments and data from related work in Madagascar. We would like to thank Oscar Diaz, Mariachiara Iannuzzi, and Alex Verlet for their fantastic research assistance. This work has been generously supported by CARE, the Fonds d'Innovations pour le Developpement, J-PAL's Post-Primary Initiative, CEPREMAP, and the French National Research Agency (ANR), under Grant ANR-17-EURE-0001. This study was pre-registered in the AEA RCT Registry under the unique identifying number AEARCTR-0008972. The study was approved by the Institutional Review Board at the Paris School of Economics.

1 Introduction

Human capital development is strongly affected by psychosocial factors: for example, performance at school can be determined by socioemotional skills (Ashraf et al., 2020; Edmonds et al., 2020; Dinarte & Egana del Sol, 2018), social interactions between students (Rao, 2019; Hu, 2023), or parents' aspirations for their child (Bernard et al., 2019). One potentially important psychosocial barrier to girls' human capital is *menstrual stigma* – the social stigma surrounding menstruation that makes it a taboo topic to discuss, and leads to the marginalization of menstruating women. Such stigma is widespread, with one study estimating that 1 in 3 women across the world who menstruate risk shame and harassment (WaterAid, 2013). Stigma could negatively impact (i) health, by inhibiting demand for menstrual products or inhibiting the flow of information about optimal menstrual hygiene (El-Gilany et al., 2005; Ali & Rizvi, 2010); (ii) school attendance, by discouraging girls from attending school during menstruation; or (iii) learning, by increasing stress or anxiety when menstruating and thereby inhibiting concentration while at school or by affecting the overall school social environment. More broadly, interventions in school – even those targeted on specific behaviors like pedagogy or hygiene – might have important impacts on other aspects of the school psychosocial environment that can have knock-on effects on human capital (Alan et al., 2021; Alan & Kubilay, 2024). For example, hygiene programs might reduce the stress associated with illness, create a collaborative environment that motivates teachers, or change students' gender attitudes by focusing on reproductive health.

Relatively few studies have examined interventions specifically designed to reduce menstrual stigma for girls in school. And while several previous studies have examined interventions aiming to directly improve menstrual *hygiene*, or a broader set of hygiene behaviors, they have generally seen mixed or small impacts on human capital outcomes – perhaps because they do not focus on the psychosocial effects of stigma or on the social environment in school (Oster & Thornton, 2011; Hennegan & Montgomery, 2016; Benshaul-Tolonen et al., 2020; A. Benshaul-Tolonen et al., 2021; Chirgwin et al., 2021).

In this paper, we present results from a randomized controlled trial in 140 schools in rural Madagascar evaluating a program run by our NGO partner, CARE. The program aims to reduce menstrual stigma and improve menstrual hygiene, while also improving a broader set of hygiene behaviors (such as hand-washing and latrine usage) – thereby improving girls' human capital outcomes.

In this context, menstruation is highly stigmatized and hygiene behaviors are typically poor. For example, the taboo that surrounds menstruation means that only 45% of girls in our baseline survey had heard a classmate speak about menstruation (compared to 70% about hand-washing), and 39% say that one *should not* discuss menstruation openly. At baseline, only 55% of girls had received information from their mother (or other caregiver) about menstruation. And only 11% of girls use sanitary pads, while the remainder typically improvise with old cloth or fabric. Knowledge of good hygiene practices is generally mixed at baseline (e.g., 98% know that the place to defecate is not outside, but only 63% can name 2 important moments in the day for hand-washing). But this knowledge does not always translate to correct behavior (e.g., only 32% washed their hands

3 times the last day they were at school, and only 14% cleaned themselves with soap the last time they menstruated at school).

Our experimental design evaluates the effects of two components of the program. First, we evaluate a bundle of interventions that simultaneously (i) engages and trains school teachers to promote hygiene and reduce menstrual stigma in schools, and (ii) relaxes hard constraints to good hygiene practices by building hygiene infrastructure (hand-washing basins and latrines) and distributing free menstrual pads in schools.

Second, we evaluate the additional effect of an innovative approach to addressing the harmful social norm that underlies menstrual stigma, namely, identifying and amplifying the voice of *positive deviants* (Marsh et al., 2004) – people embedded within a social network who are willing to engage in prosocial behaviors in defiance of a harmful social norm. Building on an earlier approach piloted by CARE, we identify around 4 "*Young Girl Leaders*" (YGLs) in each school. Using data from teacher and student surveys, these girls are selected based on their willingness to actively and openly speak about menstruation, despite the taboo nature of the topic, and on their broader leadership skills. YGLs are trained on a curriculum focusing on key hygiene behaviors, the reduction of menstrual stigma, and leadership skills, and are asked to promote hygiene and reduce stigma among their peers and classmates at school. The YGLs are thus intended to act as a prominent example of someone engaging in behaviors that undermine the norm of stigmatization from *within* the social network.

Working with positive deviants in this way may be more likely to lead to norm or behavior change than *individualized* attempts to change stigmatizing attitudes for at least two reasons (Paluck et al., 2021). First, driving norm change through positive deviants leverages a *social multiplier* effect. Rather than changing the attitudes of each person in the network individually, which may be costly and hard to scale, amplifying the voice of a positive deviant can allow them to spread attitude change through peer networks.[1] Second, by leveraging the peer network, it may be possible to not only change attitudes, but also people's beliefs about others' attitudes, thereby coordinating a change in norms.

To evaluate the effects of the program, each of the 140 schools is randomized into one of three conditions: (i) *Control* (35 schools); (ii) *Base only* (35 schools), which receive teacher and parental sensitization, distribution of sanitary pads, and sanitation infrastructure construction; and (iii) *Base + YGL* (70 schools), which receive all the elements of the base program and also have an average of 4 girls nominated to be Young Girl Leaders (YGLs). After approximately 1.5 years of the program, we evaluate the program's effects on human capital, hygiene knowledge and behaviors, and menstrual stigma using an endline that includes an average of 16 girls in each school (N=2,250).

We find that the base program on its own leads to important impacts on human capital outcomes for girls. First, we document that the base program of hygiene infrastructure and school-level interventions leads to substantial improvements in girls' learning, as measured by standardized math

[1]This builds on the insight from the social learning literature that shows that an effective way of spreading *information* through a social network is to "seed" it with influential network members, and applies this insight to norm change (Banerjee et al., 2013; Beaman et al., 2021; Banerjee, Chandrasekhar, et al., 2019).

and language achievement tests at endline and official school grades (0.1–0.2 SD). It also increased the probability of progressing to the next grade-level by 9 p.p. (17%). These improvements are comparable to the effects of interventions specifically designed to improve learning in low-income schools (Muralidharan, 2017), a notable result given that the program did not directly target improvements in learning. Second, there are no measurable effects on school attendance, as measured by the number of students present at school during a series of 5 unannounced spot checks carried out throughout the school year, or by official school enrolment registers and girls' self-reports. This suggests that the improvements in learning were driven by improvements in *quality* of time at school, rather than *quantity* of time at school. Third, we find no measurable effects of the program on self-reported health or psychosocial well-being. A generalized health index, an index of symptoms of urinary tract infections, an index of mental health, and an index of self-esteem all yield null effects for both the *Base only* schools and the *Base+YGL* schools.

Fourth, we find evidence of changes in objectively measured stress levels. Strikingly, girls' *heart rate* during the endline survey is 0.12 SD lower in treatment schools. This reduction in stress is directly welfare-relevant, but also points to an important mechanism that could underlie the effects on learning. Given existing evidence that stress acts as a "bandwidth constraint" that can inhibit cognitive function (Haushofer & Fehr, 2014; Schilbach et al., 2016; Kaur et al., in press), the stress reduction may have directly improved concentration and learning while at school.

The program also generated broader improvements in the schools' psychosocial environment that could drive learning effects. Building on qualitative reports that treatment schools had an improved social environment with more solidarity among students, we find (i) an increase of 0.24 SD in a network connection index, based on questions focusing on girls' network connections with others at school (e.g., their friends, who they do projects with, who they share with); and (ii) a reduction in the girls' reported severity of bullying at school (−0.08 SD). These improvements could enhance learning either by reducing stress, by increasing collaboration and peer learning, or by increasing students' (in line with evidence in Alan et al., 2021 and Alan & Kubilay, 2024). We do not find strong evidence in favor of learning improvements being driven by other channels, such as through better hygiene management or through increased teacher motivation.

These results speak to the importance of evaluating and including psychosocial components in school programs – even ones that do not explicitly target psychosocial outcomes.[2] The pattern of results also indicates that the human capital gains in our context operate through psychosocial channels other than the effects on *health* and *attendance* that are often emphasized in the literature on hygiene. Both qualitative and quantitative evidence suggest that other economic constraints are overriding constraints to attendance in this low-income, rural context: for example, the main reason for school dropout is having to work or not being able to pay school fees. Even in such a context, however, an intervention targeting hygiene, menstrual hygiene and stigma has a substantial impact on learning.

Next, we show that the *Young Girl Leader* intervention combined with the base program is signif-

[2]The results on bullying also add to the relatively scarce literature in economics on the causes and effects of bullying at school, with evidence largely coming from high-income settings, e.g., Brown & Taylor (2008); Wolke et al. (2013); Eriksen et al. (2014); Sarzosa & Urzúa (2021); Hu (2023).

icantly more effective than just the base program at improving several measures along the theory of change. While both the base program with and without the *YGL* program lead to substantial improvements in girls' *knowledge* of hygiene (0.3-0.4 SD), these improvements appear *more quickly* when the *YGL* component is included, showing up as significant improvements even at midline. The *YGL* component also generates significant additional improvements in hygiene *behavior*, suggesting that YGLs are successful at partially closing the knowledge-behavior gap that exists at baseline.

In addition, results show a large reduction in menstrual stigma in schools with the *Base+YGL* compared to the control. A reduction in stigma is an outcome that is both important for welfare in its own right, but could also generate knock-on improvements in hygiene behaviors and human capital. A combined index of stigma indicates a 0.74 SD improvement in *Base+YGL* schools (compared to 0.41 SD in *Base only* schools). For example, girls report being more willing to speak about menstruation openly, having more progressive attitudes about behaviors related to menstruation, believing that others in their network have more progressive attitudes, and being less likely to feel shame in response to vignettes involving menstruation. And when using lab-in-the-field exercises to elicit revealed-preference measures of stigma, we find some indication that these changes are more likely to translate to changes in behavior when the YGL program is included. For example, girls are more likely to select to explain a menstruation-related topic in front of their class in *Base+YGL* schools. Broadly, these results suggest that the YGLs are effective at generating behavior change related to hygiene and menstruation while simultaneously changing norms of behavior among their peers.

The motivating idea behind the YGL program was that *positive deviants* who are willing to engage in prosocial behaviors in defiance of a harmful social norm would be particularly effective. We explore this hypothesis further using a heterogeneity analysis, examining whether schools with YGLs who score particularly highly on an index of positive deviance achieve better outcomes. Here, the results paint a mixed picture. Schools with positively deviant YGLs have better learning test scores, lower rates of severe bullying, and better reported mental health. At the same time, this heterogeneity appears to be strongly linked to the baseline school environment, suggesting that positively deviant YGLs may only be *found* in schools that are already more progressive, or may only be *effective* at changing behavior in more favorable school environments.

Finally, we document suggestive evidence that the base program may generate negative unintended consequences, but that the YGLs may offset these effects. First, there is some evidence that pregnancies increase in *Base only* schools (albeit from a very low baseline, i.e., from 2% to 3.5% of girls who have been pregnant in the last year at endline). Second, mothers report worse psychosocial behavior for their daughters in this same arm, possibly driven by backlash. Such negative effects don't materialize in the *Base+YGL* schools. On the other hand, there is some indication that the YGL program crowds out some of the learning effects of the base program, with generally lower point estimates on learning tests in *Base+YGL* schools compared to *Base only*, and suggestive evidence that the academic performance of girls selected as YGLs worsens. Given this, the relative merits of the base program and the YGL program depend on the importance of each outcome to overall welfare.

Our study contributes by showing that a program focused on reducing menstrual stigma and im-

proving hygiene behaviors can generate substantial learning effects for adolescent girls. We thereby add to the literature on menstrual hygiene (Oster & Thornton, 2011; Hennegan & Montgomery, 2016; Benshaul-Tolonen et al., 2020; A. Benshaul-Tolonen et al., 2021) by studying an intervention that specifically focuses on reducing stigma, by using standardized academic and cognitive tests to document improvements in learning, as well as an objective measure of anxiety/stress. Our learning results directly speak to ongoing policy debates on the learning crisis, and the related academic literature on the effect of various school-based interventions on learning outcomes (Akyeampong et al., 2023) and gender differences in such impacts (Evans & Yuan, 2022). We show that hygiene-focused programs can lead to improvements in learning in schools in low-income settings, a result that appears to be driven by changes in social dynamics at schools. This implies that policies should therefore seek to leverage these social dynamics, crowding in effort and motivation, and increasing between-peer solidarity as part of programs focused on health.

Second, we show that working with positive deviants can be an effective tactic for addressing harmful social norms. This builds on the literature examining the role of positive deviants or "trendsetters" (Bicchieri, 2017; Bicchieri & Funcke, 2018), and more broadly on the literature examining ways of addressing harmful social norms (Jayachandran, 2021; Dhar et al., 2022; Gulesci et al., 2023; Bursztyn et al., 2020; Banerjee, La Ferrara, & Orozco-Olvera, 2019; Webb, 2023). Our results could be applied in other contexts where there are positive deviants whose voices can be amplified, such as against anti-minority discrimination, or against harmful practices like female genital mutilation.

Beyond these domains, the study contributes to the broader literature on water, sanitation and hygiene interventions (see Chirgwin et al., 2021 for a review of systematic reviews), the literatures on interventions targeting adolescent girls in Sub-Saharan Africa (Buehren et al., 2017; Bandiera et al., 2020; Bergstrom & Ozler, 2021; Hamory et al., 2023; Shah et al., 2023), and to the evidence of the effectiveness of using schools as entry points for health interventions (Dupas & Miguel, 2017).

2 Study design

2.1 Interventions

The study evaluated the effect of a bundle of interventions designed by our partner NGO, CARE Madagascar. The programs aimed to spread hygienic behaviors and reduce menstrual stigma, and to thereby improve human capital outcomes for girls in the later grade-levels of primary school, and all grade-levels of secondary schools. The region's secondary schools are mostly *lower* secondary schools, although our sample also includes 6 *upper* secondary schools.

Each of the 140 schools in our sample was randomly allocated into one of the following treatment conditions:

1. *Control* (35 schools), which received no school-level interventions.
2. *Base program only* (35 schools), which received a package of interventions that included teacher training, the construction of sanitation infrastructure, and vouchers for free menstrual pads for female students.

3. *Base program + Young Girl Leaders* (70 schools), which received all the elements of the base program, plus the Young Girl Leaders element of the program.

Randomization was stratified by three school characteristics: (i) whether it was primary or secondary, (ii) whether there was existing functional water access before the program, and (iii) the number of girls above the age of 12.

2.1.1 Base program

The base program aimed to: (i) relax hard constraints that would otherwise prevent girls from engaging in hygienic behaviors, and (ii) promote hygienic behaviors by sensitizing teachers and parents.

Infrastructure. To relax hard constraints related to hygiene, the program financed and built hygiene and sanitation infrastructure in all the base program schools. Specifically, our partner built between 2 and 10 (mean: 3.9) new latrines on the school site, and in some cases rehabilitated existing latrines so that they were functional.[3] The infrastructure also included basins for students to wash their hands.

Vouchers for sanitary pads. To relax hard constraints related to *menstrual* hygiene, the NGO distributed vouchers which could be exchanged for a set of 6 free reusable sanitary pads from local tailors who were trained by the program. The vouchers were distributed in schools to all girls from the targeted grade-levels. Take-up of the vouchers among targeted girls was 91%.

Teacher training & hygiene committee. To sensitize teachers and parents to the hygienic behaviors being promoted, personnel from the local government ministry of education and our NGO partners trained 2 teachers in every school. The curriculum focused on the key hygiene behaviors described above. There were 2 days of training at the start of the program, and shorter refresher trainings that took place every semester. The trained teachers were asked to organize a short (5-minute) weekly session at school to talk about hygiene to the student body. In addition, an 8-member "hygiene committee" composed of teachers, parents, and students was formed in every treatment school. The committee was responsible for promoting hygienic behaviors in school and for maintaining the hygiene infrastructure. Adult members received an initial training session about the roles and responsibilities of the committee.[4]

2.1.2 Young Girl Leaders intervention

The Young Girl Leaders intervention involved selecting girls in each school and designating them as *Young Girl Leaders* (YGLs) who were responsible for peer-to-peer promotion of hygiene and menstrual hygiene behaviors.

An average of 4 YGLs were selected in each treated school.[5] Treatment assignment was fully re-

[3] In all schools, infrastructure was finalized by endline data collection, but was only at very early stages by midline.

[4] Together, the teacher training and the hygiene committee meant that sensitization on hygiene behaviors was not *only* carried out by the young girl leaders (YGLs), described below. This also meant that the activities of YGLs were more likely to be perceived to be endorsed by adults and figures of authority in the school, reducing the risk of backlash against YGLs.

[5] The number selected depended on the number of pupils in the school. There was one YGL for every 60 students in the school, apart from schools with fewer than 120 students who always had 2 YGLs.

spected: the targeted number of YGLs were identified and trained in all 70 treatment schools, and there were no YGLs in control or *Base only* schools.

YGLs were selected based on a survey of teachers and of the students whom those teachers nominated. We selected girls in the same way across all YGL schools, using a predetermined weighting of survey responses that incorporated (i) leadership qualities, (ii) a form of "positive deviance" (i.e., willingness to take prosocial actions in defiance of others, such as speaking openly about menstruation), and (iii) having started menstruating themselves. We also restricted to girls with a sufficiently good academic track record, to avoid hurting the performance of girls who were already struggling academically.

YGLs received an initial intensive training with weekly sessions of 1.5 hours for approximately 3 months. These training sessions were delivered using a cascade model, where technicians from local NGOs were trained by our main partner NGO, and then in turn trained YGLs. The sessions initially focused on enhancing the leadership and sensitization skills of the YGLs, and then moved onto a curriculum of topics on general hygiene and menstrual hygiene and stigma. The key messages and hygiene behaviors being targeted were as follows: (i) menstrual hygiene, including how to properly wash during menstruation, how to use and wash reusable sanitary pads, and basic information about menstruation and the body; (ii) using latrines; (iii) hand-washing with soap; (iv) using and keeping clean water; and (v) food hygiene. After the initial 3-month training period, technicians followed up with YGLs every two weeks to give additional guidance and encouragement. At the start of the second school year, a comprehensive refresher training covering all topics was provided to the YGLs. YGLs who graduated from a school were replaced with new YGLs, who were selected using the same criteria as the first school year and also received the same training.

YGLs were asked to spread positive messages and encourage hygienic behavior among their peers at school, and to openly discuss menstruation in a way that could reduce menstrual stigma. To test whether peer-to-peer sensitization was more effective if it was "formal" or "informal", we randomized the method through which YGLs were asked to sensitize their peers. In a *formal* condition (35 schools), they were told to give formal classroom sessions about hygiene and menstruation. In an *informal* condition (35 schools) they were told to simply talk to their classmates and friends about these topics, but were not instructed to hold classroom sessions. Results from the endline reveal, however, that there was only modest de facto variation in how the two modalities were implemented (Table A1). For example, YGLs in the informal condition often also organized formal sensitization sessions. Possibly as a result, we find few differences in the effects of the two modalities (Table A2), and therefore present the pooled results throughout the paper.

2.1.3 Timeline

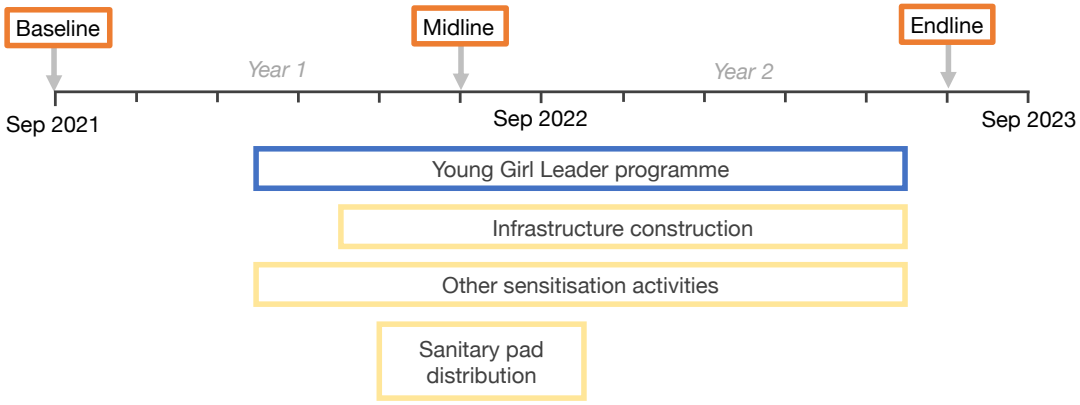
The timeline of the program activities and main data collection exercises are summarized in Figure 1.

2.2 Data collection

2.2.1 Baseline survey and sample selection

140 primary and secondary schools in the rural areas of the Amoron'i Mania department of Madagascar were selected to be part of the study. We selected schools based on three criteria. First, we

Figure 1: *Study timeline*



excluded schools that were deemed be to inaccessible by the operations team (e.g., because they could not be accessed at all by vehicle). Then, we prioritized schools that had the most girls above the age of 12, for whom the intervention would be most relevant. Finally, we used a coarse measure of location (based on the distance to the subdistrict capital) to attempt to exclude schools that were too close together, to minimize spillover effects.

For the baseline survey (September to October 2021), we sampled an average of 17 girls per school to be surveyed, totaling N=2,393. This sampling was based on a listing exercise using official school enrolment registers. 90% of students were between 4th and 9th grade, but the sample included students between 3rd grade and 12th grade.[6] Girls' average age in the sample at baseline was 14 years (Table A3), and about half of the girls had not reached menarche at baseline (with this share naturally declining by endline). The sample is characterized by high levels of poverty, and relatively large accumulated school delays (2.2 years on average).[7] Girls take on average 30 minutes to walk 2.5km to school, while only 2% of their households own a vehicle. Girls live in large households (mean size: 6.5), and 81% live in the same household as their mother. At baseline, we also interviewed the mother or primary female caregiver of all sample girls during the household interview, and, in cases where they existed, a brother in the same household as the girl.

For all relevant treatment comparisons, the treatment groups are well-balanced across key characteristics (Tables A3, A4), with only one statistically significant difference across treatment groups (girls in *Base only* schools are significantly less likely to have a brother in the household, p=0.03). We use LASSO to select all controls that predict both treatment status and outcomes (as per Belloni et al., 2014), to avoid this imbalance affecting the results.

[6]The baseline took place at the end of the summer vacation in 2021, using a sample based on school registers from the previous school year. We therefore avoided sampling grades where girls would be likely to drop out of school. In lower secondary schools, we sampled equally from girls who had just finished 6th, 7th, and 8th grade, and in high schools we sampled equally from girls who had just finished 10th and 11th grade. In primary schools, we sampled girls who were aged 12 and above and prioritized 5th-grade girls who were registered to repeat next year, filling out the remaining sample from other grades and younger ages in smaller schools where necessary.

[7]School dropout at early ages is also frequent: while we cannot observe this at baseline, 17% of girls in control schools drop out between year 1 and year 2 of the study.

2.2.2 Endline survey: girls' survey

Our main endline survey took place between May and July 2023 (i.e., towards the end of the second school year and at the start of the summer vacation). This was approximately 1.5 years after the start of the program's implementation. We surveyed 2255 out of 2393 girls at endline, corresponding to an attrition rate of 5.8%. Attrition at endline is balanced across treatment arms (Table A5, column 1). The majority of girls were interviewed at their homestead. However, many students had moved to a different school, or left school and moved to a different city. For a subset of the 15% of girls who were not found after several attempts in their original location or nearby areas, we carried out additional intensive tracking in October 2023 (see Appendix Section B.5 for more details). This resulted in further in-person interviews at migration destinations, such as the capital Antananarivo and regions next to Amoron'i Mania. A small share of girls (3.7%) were interviewed by phone because they could not be reached in person. The phone survey used a shorter survey instrument that did not include standardized learning tests. Participants in the base program arm were slightly more likely to be surveyed by phone (p=0.03).

2.2.3 Endline survey: other surveys

In addition to the girls' survey, during the endline data collection we also carried out 4 other surveys:

1. *Mothers*. We conducted a home-based survey of the mother or the same female caregiver that had been surveyed in the baseline. 2,287 were found, corresponding to an attrition rate of 4% that was balanced across treatments (Table A5, column 5).[8]
2. *Teachers and directors*. We surveyed all school directors and at least one teacher per grade-level for 3rd grade and above. During this survey, we also elicited the official marks received by all the children in at least one class for every grade-level.
3. *Male classmates*. Using up-to-date school registers, we randomly selected 10 boys in each school to do a short survey centered around a dictator game related to menstruation.
4. *"Willingness to speak" lab-in-the-field*. Using school registers, we randomly sampled 12 girls and 4 boys in each school to take part in an interactive exercise designed to measure girls' willingness to speak about menstruation in front of others. This exercise is described in more detail below.

2.2.4 Midline survey

A midline survey took place in the summer of 2022, after approximately half a year of program implementation. Attrition at midline was 11%, and the combined midline and endline attrition is slightly higher in the treatment villages than in the control (Table A5, column 3). As the midline data is mostly used to provide a broad description of dynamic changes, this attrition is not a primary concern.

[8] For the other samples used, there was (i) no differential attrition for teachers (Table A5, column 7); (ii) a lower probability of successfully eliciting school marks among *Base only* schools (column 8); (iii) no difference in the number of boys successfully interviewed or the probability of having to take boys from a secondary sample list across arms (columns 9 and 10); and (iv) no differential sample selection for girls in the lab-in-the-field (columns 12 and 13).

2.2.5 Spot-check absenteeism and school register data

To measure school absenteeism without having to depend on unreliable official school registers, we carried out 5 waves of unannounced spot checks in all schools. Enumerators visited the schools without notifying beforehand, and took attendance calls for all classes above 3rd grade. They also digitized all the official school enrolment and attendance records from the last 3 months. The first spot check took place in February 2022, before the program had started, acting as a baseline. Subsequent waves took place in June 2022, October 2022, December 2022, January 2023, and June 2023.

2.3 Outcomes

Here, we describe the main outcomes used in the paper (see Appendix Section B.1 for more details).

Because respondents likely knew about the program's objectives to improve hygiene and menstrual hygiene, an important concern when measuring outcomes was the risk of *social desirability bias*, i.e. respondents biasing their responses to be favorable (possibly particularly in the treatment schools). Because of this, a key motivation for the data collection was to create *objective* or behavioral measures of the outcomes of interests. We place most weight on the results on such measures.

Where objective measures were not possible to elicit, we also used self-reported measures. For these, to evaluate the risk of social desirability bias driving our treatment effects, we use a shortened version of the Crowne & Marlowe (1960) social desirability index elicited at baseline, and test whether treatment effects are driven by girls with high social desirability scores (Table A7; see Appendix Section B.4 for more details).

2.3.1 Human capital outcomes

We measure the effect of the base program and the YGL intervention on a set of human capital outcomes.

Learning. We measure how much children have learned in three complementary ways. First, at endline, we administer a set of standardized tests that evaluates academic skills learnt at school and cognitive ability. It includes modules on (i) Malagasy listening; (ii) Malagasy reading fluency; (iii) Malagasy text comprehension; (iv) a digit-span test that measures working memory (forward and backward); and (v) math fluency, based on a timed written maths test involving simple arithmetic calculations. Most modules had been tested and validated earlier by local survey teams on other projects in Madagascar, and all of them elsewhere in Sub-Saharan Africa. In addition, we used pilot data collected before the deadline to validate all modules on a population of the same age and region as our study population. For the main analysis, we combine all 5 modules into a single index using factor analysis. We also consider the results of each test separately, to explore the differences between the academic achievement tests (reading and math fluency) and the other tests that capture attention, memory and cognition.

Second, we ask teachers and directors to give us the official marks received by students for their most recent exams. These are recorded by schools on student-specific paper "bulletins" and on class-specific registers, allowing enumerators to directly verify reported school marks. Students are

given an overall mark out of 20 for the preceding trimester. We use this as an observed measure of academic achievement.

Third, girls report the grade-level they were in at baseline and endline (which were one academic year apart). We can therefore deduce whether a girl progressed to the next grade-level, or did not because they repeated a grade-level.

School attendance. We measure school attendance using the spot check data, first by measuring the number of students physically present at school when the spot check is carried out, and second by recording the number of students enrolled and present according to official school registers. For these outcomes, we include fixed effects for the 4 waves of the spot checks, or day-level fixed effects for school register data, and control for the value of the outcome in the first baseline wave. We also ask girls at endline (i) whether they are currently enrolled in school, and (ii) whether they were absent from school in the last 3 months.

Health. First, we measure generalized health by asking girls the physical functioning sub-scale of the Pediatric Quality of Life Inventory (PedsQL), following A. Benshaul-Tolonen et al. (2021). Girls are asked whether they had difficulty carrying out any daily activities, such as running or completing tasks at home. Second, since improvements in menstrual hygiene are associated with reductions in urinary tract infections (UTIs) (Phillips-Howard et al., 2016), we also asked girls to report symptoms of UTIs, following (Czura et al., 2019).

Psychosocial well-being and mental health. We measure psychosocial well-being by first asking an adapted version of the CESD depression index, in which girls report how many days over the last week they have experienced certain (positive and negative) emotions. Our preferred specification uses an index that combines both positive and negative emotions and is corrected for acquiescence bias, but we also show positive and negative emotions separately.[9] We also elicit a self-esteem module, asking girls a series of questions indicating whether they are confident in themselves and their abilities or not.

Heart rate during endline survey. To act as an objective measure of stress or anxiety, we measured girls' heart rates during the endline survey. This was incorporated to measure a combination of (i) general anxiety levels experienced by girls, and (ii) the anxiety levels experienced specifically when discussing topics related to hygiene and menstruation. We used wristband monitors that included optical heart rate sensors that recorded the girls' heart rate in beats per minute on a second-by-second basis. The monitors we used have been demonstrated to have similar accuracy to industry-standard chest strap monitors (Reece et al., 2021), but were significantly less invasive and more comfortable

[9]Also known as "yay-saying", acquiescence bias is when respondents tend to agree with a statement from the enumerator, even if doing so results in contradictory responses to positive and negative statements intended to measure the same trait. We correct for this bias, following the psychometric literature (Rammstedt & Farmer, 2013; Laajaj & Macours, 2019). For each girl, we calculate an acquiescence score by taking the average difference between the positively-coded questions and the reversed negatively-coded questions, and dividing it by two to retain centering. This score is then added to the negatively-coded questions, and subtracted from the positively-coded questions.

for girls to wear.[10]

2.3.2 Hygiene knowledge and behavior

To understand intermediate outcomes, we elicit a series of measures of knowledge and behavior related to hygiene and menstruation.

Knowledge. We ask girls questions about their knowledge of good hygienic practices. We create one index based on knowledge of overall hygiene practices (e.g., methods for cleaning water and how many times a day they should wash their hands), and another on their knowledge of menstruation and menstrual hygiene (e.g., naming hygienic menstrual products, and naming average length of a period). The sum of correct answers to the questions, normalized to have a mean of 0 and standard deviation of 1 in the control group, is used as the outcome.

Behavior. We ask girls questions related to whether they engage in the types of hygienic behaviors recommended by the YGL curriculum. One index is again related to overall hygiene behaviors (e.g., did they use soap when washing their hands at home last week?; did they use a latrine or defecate openly the last time?), while another is related specifically to menstruation (e.g., did they use a hygienic menstrual product?; did they wash at least 3 times per day the last time they menstruated?).

Observed hygiene. To alleviate concerns about social desirability bias leading to girls *overreporting* hygienic behaviors in treatment schools, enumerators also directly observed levels of cleanliness in three contexts. First, they marked the visible hygiene levels of girls themselves during the endline survey (e.g., was there visible dirt on the girl?; was the girl wearing shoes?). Second, they noted the hygiene levels at the original household of the girl, where the mother resided at endline (e.g., were animals separated from the people's living spaces?; were the dishes in the kitchen area clean?). Finally, in each school, when carrying out the teacher and director survey, enumerators noted features of the environment that indicate engagement in hygienic behaviors, such as whether there was soap available at the basins, whether there was a dedicated dispenser for trash, and whether any latrine had a cover.

2.3.3 Menstrual stigma

To understand whether the constraints imposed by menstrual stigma are relaxed by the program, we elicit an array of complementary measures of stigma. For all measures of stigma, a reduction in stigma is coded positively.

Spread of information. First, we seek to understand whether the taboo around speaking about menstruation is lifted by the program. We measure whether girls have received information about menstruation from others at school or at home using an index of questions focusing on how many times they heard classmates, teachers, and parents speak about menstruation in the last 3 months, and how many times they asked people questions about menstruation. A second series of questions measures how willing girls are to themselves talk about menstruation in front of others.

Attitudes and perceived norms. We measure how progressive girls' attitudes are towards menstrua-

[10] In some cases, heart rate monitors were not fixed properly to girls' wrists, or the battery of monitors ran out on the field. Because of this, heart rate data is available for only 1904 girls. However, the proportion of data available is equal across treatment arms (Table A5, column 6).

tion. We use a series of Likert scale questions asking the level of agreement with statements like "Girls should be allowed to come to school when they menstruate" and "Girls should have to hide the fact that they are on their periods when they are at school". We also measure perceived norms by asking girls to estimate how much people in their social network agree with a subset of these attitude questions. They report how much they think other girls, boys, teachers in their school, and their mothers agree with the statements about menstruation. All four second-order belief questions are combined into a single index of norms.

Shame. To measure shame, we describe hypothetical vignettes about embarrassing scenarios related to menstruation to girls. We then ask them to report how much they feel shame-related emotions in response. To validate these measures, we ask enumerators to observe how visibly uncomfortable girls appear to be when talking about menstruation during the endline survey (e.g., did they look at the floor, did they hesitate, did they seem uncomfortable, did they speak excessively quietly).

"Willingness to speak" lab-in-the-field. This was an interactive exercise designed to measure girls' *revealed-preference* willingness to speak about menstruation in front of others. In each school, this activity was carried out twice, with 6 girls and 2 boys participating in each group. All 8 students sat in a classroom. Each girl was taken out one by one, and had two new topics explained to them by enumerators: one topic related broadly to hygiene/health (e.g., what is pneumonia?), and the other related to menstruation (e.g., what are menstrual cramps?).

Girls then faced two binary choices. In one choice ("*anonymous*"), they were told that a video would be shown to the rest of the class about one of the topics, and they could choose which one — the menstruation-related topic, or the hygiene-related topic. No one would be told which video the girl picked. Since this choice was anonymous, it was designed to minimize concerns about taboo and stigma, and to measure the desire to supply the information to the rest of the class. In the other binary choice ("*explanation*"), by contrast, girls were told that they would have to *themselves* explain the topic to the rest of the class.[11] Since this involved talking openly about a topic in front of others, the choice was likely to be strongly affected by social image concerns and the taboo around menstruation, allowing us to test for changes in the level of taboo driven by the treatments.

Dictator game. A standard prediction of theories of taboo is that those who *break* the taboo should face social sanctions. To test this, we ask endline girls and the 10 randomly selected endline boys to play two dictator games. Students are shown two 1-minute videos in a random order, both depicting a teenage girl from the region explaining a topic. One explains what menstruation is and how to maintain menstrual hygiene, and the other talks about the use of soap in hand-washing and the spread of germs. After each video, the student is shown 5 cookies, and asked how many they want to share with the person in the video. We interpret a decrease in the number of cookies as social sanctioning that may be incurred when breaking a taboo. Just as with the lab-in-the-field, we can test whether the taboo around menstruation has decreased by looking at whether students are more likely to share cookies with the girl speaking about menstruation (relative to the girl speaking about hand-washing).

[11] We randomly varied the order of explanation and anonymous choices, and the topics being proposed to each girl.

Demand for information in envelope. The final revealed-preference measure of stigma we use is whether girls choose to find out information about menstruation or another topic, intended as a measure of the demand for information. In the endline and midline survey, girls were shown a series of pairs of envelopes, each of which showed a question on the outside, some of which related to menstruation (e.g., "Can stress affect the menstrual cycle?") and some which related to broader hygiene and health (e.g., "Are mosquitos more attracted by certain blood types?"). Girls make binary choices over whether they would prefer to open the menstruation-related envelope or the hygiene-related envelope, and get to see the answer to the question inside the envelope.

2.3.4 Psychosocial environment at school

To evaluate the social dynamics at school, we ask girls and teachers to report the severity and frequency of *bullying at school*. We ask them to report separately on cases of light-hearted teasing, severe intimidation or harassment, and incidences of bullying specifically related to menstruation. Girls reported on bullying they experienced themselves, and on a randomly selected subset of 2-3 girls in the same grade-level as them; these are combined into one index. Teachers report on a randomly selected subset of girls.

We also create a *network integration index* by asking girls about their social connections with their peers. For example, they are asked about how many friends they have, how many children they play with, how many they speak with at break, and how often they do homework together. For a subset of these questions, we ask separately about other girls and other boys.

2.3.5 Effects on mothers

To evaluate whether any changes in knowledge and behavior among girls are transmitted vertically to their mothers (e.g., through interactions in the household), we elicit from mothers at endline the same indexes of knowledge and behavior related to general hygiene and menstrual hygiene. We also ask mothers the same set of attitude questions about menstruation as girls. We further ask mothers a set of questions about their perception of the behavior of their daughter, using the standardized Strength and Difficulties (SDQ) instrument. This was intended to evaluate girls' psychosocial well-being and also to understand if mothers had a negative or positive reaction to the changes in girls' behavior induced by the program. Some questions focused on emotional behavior (e.g., whether they are often worried, or often angry), while others focused on their relations with others (e.g., are they often alone, do they have at least one friend?). Finally, we ask all mothers about the fertility status of their daughters (specifically, were they pregnant in the last year or are they pregnant now) to examine potential impacts on fertility.

2.4 Empirical specification

In our main specification, we estimate the following regression:

$$Y_{ij} = \beta_0 + \beta_1 BaseOnly_j + \beta_2 BaseYGL_j + \mathbf{X}'_{ij}\Gamma + \varepsilon_{ij} \quad (1)$$

Y_{ij} is the outcome variable measured at endline for student i in school j. $BaseOnly_j$ is a binary variable, equalling 1 when school j is in the *Base only* treatment and 0 otherwise. $BaseYGL_j$ equals 1 when school j is in the *Base program + YGL* treatment and is 0 otherwise. The main average treatment effects of interest are therefore given by β_1 and β_2. $\mathbf{X}_{ij}$ is a vector of controls, which

always includes stratum fixed effects. In cases where it exists, it also includes Y_{ij}^0, the baseline analog of the outcome. We use double LASSO (Belloni et al., 2014) to select an additional set of controls that predict both the treatment and outcome variables. When further controls are included (e.g., wave fixed effects for spot check outcomes), we specify directly in the table notes. Standard errors are, unless otherwise specified, clustered at the school level.

3 Results

3.1 First stage: program implementation

The program was implemented with high fidelity to the assigned treatment conditions. First, in *Base only* and *Base+YGL* schools, there were improvements in the hygiene infrastructure at endline according to enumerator observations. On average, these treatment schools had 4.5 functional latrines (vs 2.3 in control schools), 85% of them had at least one separate latrine for girls (vs 23% in control schools), and 69% had a functional water basin (vs 37% in control schools). There were no statistically significant differences between *Base only* and *Base+YGL* schools in these outcomes. Second, in *Base only* and *Base+YGL* schools, menstrual pads were successfully distributed, with 100% of directors in these schools reporting that there had been a distribution (vs 0% in control schools), and 72% of girls at endline reporting having received pads (vs 2% in control schools)[12]. Third, in *Base only* and *Base+YGL* schools, directors reported that teachers had received training about hygiene and sanitation in the last 5 years in 99% of cases (vs 31% in control), and teachers sensitized children on hygiene and sanitation issues every week in 93% of schools (vs 46% in control). Finally, the *Young Girl Leader* program was carried out successfully: when asking girls at endline whether certain girls in their school carried out sensitization activities about menstruation and health, in 97% of schools (vs 9% in control schools) at least 20% of the girls said yes.

3.2 Human capital outcomes

In this section, we document the effects of the program on human capital outcomes, including learning, school attendance, health, mental health, and anxiety/stress (as measured by heart rate).

Learning. The base program leads to substantial improvements in girls' academic and cognitive skills, measured through a set of standardized learning tests administered at endline (Table 1). The pooled effect of the *Base only* and *Base+YGL* arms on the combined index (capturing tests of academic skills, memory, and attention) is 0.15 SD (p=0.004, column 1). The effects are strongest for academic achievement tests (language and math fluency) (0.13-0.25 SD, pooled p=0.001, column 2), while effects are smaller and less consistently significant for tests capturing memory, attention and other cognitive skills (0.09-0.14 SD, pooled p=0.05, column 3).[13] The magnitudes of the effects are comparable to the effects of interventions specifically designed to improve learning in schools in low-income settings (Muralidharan, 2017). They are thus substantial given that the program elements did not focus directly on learning.

[12] This figure is lower than the 91% overall take-up rate, because some girls in the sample were too young to be in the grades who received pads.

[13] See Table A8 for results on each individual test.

The improvement in learning translates to improvements in official marks in treatment schools for girls (pooled effect: 0.11 SD, p=0.08; Table 1, column 4). The effect on boys' official marks is positive but not significant (pooled p=0.20, column 5), suggesting that there may be a gender-specific component to the improvements, although we cannot reject that the effect on girls and boys is the same (p=0.70). In addition, the improved marks for girls lead to a substantially increased probability of progressing to the next grade-level, i.e., of not repeating a grade-level (column 6). While in the control group only 51% of girls progressed to the next grade-level, this increases to 60% in the pooled treatment schools (effect: 9p.p., 17%, p=0.01).[14]

The improvements in learning across the different measures appear to be driven by the base program rather than the *YGL* intervention. For all measures, the point estimate is larger for *Base only* schools than *Base+YGL* schools, although the difference between these two arms is only significant for the subset of academic achievement tests (p=0.02). This suggests that relaxing hard constraints by building infrastructure and distributing sanitary pads, along with teacher training and parental involvement together drive the learning effects, while the *YGL* program does not generate additional learning benefits.

School attendance. There are no improvements in school attendance driven by the interventions (Table 2). The program has no significant effect on (i) the number of girls physically present at school during the unannounced spot checks (pooled p=0.96, column 1); (ii) the number of girls physically present in the 2 months preceding the spot checks, according to digitized official school registers (pooled p=0.33, column 2); or the number of students enrolled according to those school registers (pooled p=0.40, column 3).[15] When asking girls at endline, we find no effect on whether girls report being enrolled in school (pooled p=0.33, column 4), or whether they were absent in the last 3 months, conditional on being enrolled (pooled p=0.29, column 5).[16] Together with the results above, these results suggest that the *quality* of learning increases while at school without increasing the *quantity* of time at school.

A plausible explanation for the lack of effects on school attendance is that absenteeism and dropout in this low-income context is driven by other hard economic factors that act as binding constraints. The overall rate of dropout is high (19% over the two-year period between baseline and endline). And in 63% of schools, directors reported that the main reasons for female dropout were economic

[14] Several additional results suggest the learning effects are robust. The randomization inference p-value of the pooled effect is significant for both the learning tests and grade progression (Table A9). The distributions of the learning test results shift fairly uniformly, indicating that the effects are not driven by outliers (Figure A10). We find no evidence of spillovers on learning tests (Table C1), or any other primary human capital outcome. And while we did not elicit learning tests at baseline, the number of years of school delay (a proxy for academic performance) is balanced across arms at baseline (Table A3). Learning results are also robust to including attrition weights (Table A11 and Table A12). Alternatively Lee bounds for the learning results for both treatment groups together show an upper bound of 0.18 SD (p-value <0.001) and a lower bound of 0.08SD (P-value = 0.135). Lee bounds for impacts on achievement tests show 0.35SD (P<0.001) as upper bound and 0.16SD as lower bound (P-value= 0.056).

[15] Our standard errors imply a minimum detectable effect of approximately 8 girls on a given day, corresponding to an absenteeism rate of approximately 12% relative to the control mean. We therefore cannot rule out small effects on school attendance.

[16] There is also no effect on boys' attendance (Table A13).

Table 1: *Education - Learning*

VARIABLES	(1) Achievement + cognitive tests (combined) (Z)	(2) Achievement tests only (Z)	(3) Cognitive tests only (Z)	(4) Official mark (Z): girls	(5) Official mark (Z): boys	(6) Progressed one class (=1)
Base Only	0.208***	0.248***	0.142*	0.160**	0.075	0.112***
	(0.072)	(0.064)	(0.075)	(0.079)	(0.071)	(0.040)
Base + YGL	0.117**	0.131**	0.086	0.086	0.085	0.076**
	(0.053)	(0.051)	(0.058)	(0.064)	(0.069)	(0.036)
Observations	2,167	2,167	2,167	7,586	6,993	2,256
Data source	Girls	Girls	Girls	Marks	Marks	Girls
Control mean	0	0	0	0	0	.515
p: Treated = 0	.0042	.001	.0524	.0803	.1952	.011
p: Base Only = Base + YGL	.1688	.0194	.4324	.2763	.8751	.2056
p: Base Only [Girls] = Base Only [Boys]				.184	.184	
p: Base + YGL [Girls] = Base + YGL [Boys]				.866	.866	
p: Treated [Girls] = Treated [Boys]				.702	.702	

Notes: Standard errors are clustered at the school level and are in parentheses, and controls include stratum fixed effects. Columns (1)-(3) also include the baseline controls selected by double LASSO. The bottom of the table denotes the p-values on the test of being treated *Treated=0* and the equality of the coefficients *Base Only = Base + YGL*. Other p-values are defined analogously. *$* p < 0.1$, $** p < 0.05$, $*** p < 0.01$.* In column (1), all learning tests elicited at endline are combined into one index using factor analysis. In column (2), only reading and math fluency tests are included and are combined using an unweighted sum of Z-scores. In column (3), only listening, comprehension, digit span, and reverse digit span tests are used, and are combined using an unweighted sum of Z-scores. The official school marks (columns 4 and 5) were collected from grades notebooks provided by teachers and include all students in the class. *Progressed one class* = 1 when the girl moved up one grade-level in between the baseline and endline survey, and is coded as 0 if the girl is in the same grade-level at endline or is not enrolled at school at endline.

Table 2: *Education - Attendance*

VARIABLES	(1) # girls present (spot-checks)	(2) # girls present (registers)	(3) # students enrolled (registers)	(4) Enrolled (=1)	(5) Absent in last 3 months (=1)
Base Only	-2.003	-0.849	4.554	-0.019	-0.009
	(3.453)	(2.854)	(4.564)	(0.024)	(0.030)
Base + YGL	0.833	-3.080	2.525	-0.021	-0.038
	(2.702)	(2.715)	(3.918)	(0.022)	(0.029)
Observations	660	34,455	698	2,256	1,824
Data source	Spot checks	Registers	Registers	Girls	Girls
Control mean	56.4	64.8	146.	.826	.442
Control SD	49.1	95.1	103.	.38	.497
p: Treated = 0	.9621	.3323	.4036	.3305	.2854
p: Base Only = Base + YGL	.385	.4206	.5499	.9256	.2561

Notes: The table displays school attendance results from the spot-checks in column (1) and official school registers in columns (2)-(3), collected in June 2022, October 2022, December 2022, January 2023, and June 2023, and the girls survey in columns (4)-(5) at endline. Standard errors are clustered at the school level and are in parentheses, and controls include stratum fixed-effects. Column 1: *# girls present* denotes the number of girls physically present at the school. The data is at the *wave* × *school* level and includes wave fixed-effects and controls for the outcome at baseline. Column 2 is at the *day* × *school* level and includes date (day) fixed effects and controls for the outcome at baseline. *# girls present* denotes how many girls were present according to the registers on that day. Column 3: *# students enrolled* is the number of students (girls and boys) enrolled according to the official registers in that wave of the spot-check. For the last two outcomes, girls were asked about their school attendance at endline: *Enrolled (=1)* refers to whether the girl was enrolled in the 2022-2023 academic year, and *Absent in last 3 months (=1)* is an indicator of whether the girl had been absent in the last 3 months, conditioned on being enrolled. LASSO-selected controls are included for columns 4 and 5. The bottom of the table denotes the p-values on the test of being treated *Treated=0* and the equality of the coefficients *Base Only = Base + YGL*. Other p-values are defined analogously. ** p < 0.1, ** p < 0.05, *** p < 0.01*.

factors, including (i) girls having to find a job (38% of schools), (ii) not being able to pay school fees (44% of schools), or (iii) not having enough food to eat to go to school (13% of schools).

Health. There are no measurable impacts on self-reported health outcomes (Table 3). Our measure of general health based on the Pediatric Quality of Life index shows no significant effect (pooled p=0.90, column 1). There are also no clear decreases in the symptoms of urinary tract infections (UTIs) (pooled p=0.42, column 2). The null effects seen here may be partly driven by measurement error on self-reported health outcomes, given that we do find impacts on objective measures of observed hygiene and heart rate during the endline survey (discussed below).

Table 3: *Health and psychosocial wellbeing*

	(1)	(2)	(3)	(4)	(5)	(6)
			Mental health			
	General	UTI	(Reversed CESD)	Mental health	Mental health	
	health	health	Pos. and neg.	Pos. emotions	Neg. emotions	Self-esteem
VARIABLES	(Z)	(Z)	emotions (Z)	only (Z)	only (Z)	(Z)
Base Only	-0.037	-0.089*	0.012	0.038	0.035	0.017
	(0.060)	(0.054)	(0.062)	(0.071)	(0.059)	(0.067)
Base + YGL	0.009	-0.004	0.051	0.083	0.051	0.010
	(0.048)	(0.043)	(0.060)	(0.066)	(0.051)	(0.052)
Observations	2,256	2,256	2,256	2,256	2,256	2,256
Data source	Girls	Girls	Girls	Girls	Girls	Girls
p: Treated = 0	.9025	.416	.5409	.2925	.2925	.7784
p: Base Only = Base + YGL	.392	.0885	.4053	.3882	.7508	.8991

Notes: The table displays health index results from the girls' surveys at endline. Standard errors are clustered at the school level and are in parentheses; controls include stratum fixed effects and the baseline controls selected by double LASSO. The bottom of the table denotes the p-values on the test of being treated *Treated=0* and the equality of the coefficients *Base Only = Base + YGL*. Other p-values are defined analogously. *$* p < 0.1$, *$** p < 0.05$, *$*** p < 0.01$. *General health* is an index constructed using questions about experiencing symptoms of physical pain and general health during the last month. A higher value indicates *better* health. *UTI health* is constructed using questions specific to experiencing urinary symptoms during the last month (a higher value indicates fewer symptoms, i.e., better health). *Mental health* (column 3) was constructed from questions aimed at measuring depression symptoms during the last week, combining both positive and negative emotions and correcting for acquiescence bias, with a higher value indicating better mental health (i.e., less depression). *Positive emotions* (column 4) and *Negative emotions* (column 5) include only the positive and negative symptoms from this scale, respectively. A higher value for positive emotions indicates more common experiences of positive emotions; a higher value for negative emotions indicates more commonly experiencing negative emotions. *Self-esteem* is constructed from positive and negative questions about how the girl perceived her own abilities.

Self-reported psychosocial well-being. There are no significant effects on self-reported measures of psychosocial wellbeing (Table 3, columns 3-7). The program does not affect a summary index of depression based on the CESD index (p=0.54, column 3), nor does it affect reports of the positive or negative emotions that compose the index when separated out (columns 4 and 5). A related measure of self-esteem, that asks girls questions about their confidence in their abilities and personality, also shows no significant effects (p=0.78, column 6).[17]

[17] At midline, there is some weak evidence of an increased incidence of negative emotions in *YGL* schools (Table A14, column 4), which may be suggestive of an initial "backlash" effect against the program. But this fades out by endline, suggesting that any initially negative effect is mitigated with further program involvement.

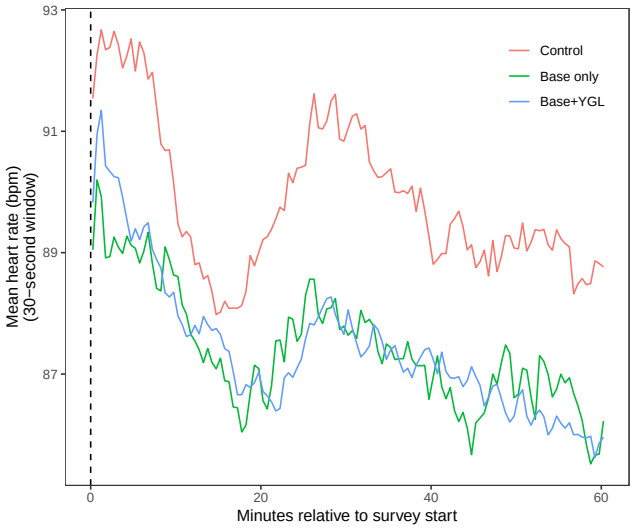
Heart rate. The interventions do, however, lead to a significant *decrease* on an objective biological measure of anxiety or stress: girls' heart rate during the endline survey. While heart rate naturally fluctuates in both treatment and control during the survey, heart rate is on average 2.3 bpm lower in treatment schools (Figure 2; p=0.01; effect size: 0.12 SD when using the between-girl standard deviation). The effects are very similar across *Base only* and *Base+YGL* schools (p of difference: 0.99).[18]

Figure 2: *Stress: heart rate during endline survey*

(a) *Regression*

	Dep var: HR bpm
Base only	−2.346**
	(1.133)
Base+YGL	−2.338**
	(0.998)
Num.Obs.	9 105 420
N girls	1904
Control mean	87.6
Control SD (between)	19.0
p : Treat = 0	0.014
p : Base=Base+YGL	0.993

(b) *Dynamics*



Notes: Panel (a) shows the effect of the intervention on the heart rate beats per minute (bpm) for girls in the endline survey. All girls who were included in the baseline sample and for whom heart rate data is available are included. *Time window*: observations are included until up to 120 minutes after the start of the survey. The unit of observation is the *girl* × *second (relative to the start of the survey)*. Results are robust to using a girl-level regression (Table A16, column 1). The regression controls for 30-second window fixed effects, stratum fixed effects, and variables selected by double LASSO. Standard errors are clustered at the school level and are in parentheses. The bottom of the table denotes the p-values on the test of being treated Treated=0 and the equality of the coefficients Base Only = Base + YGL. * p < 0.1, ** p < 0.05, *** p <0.01. Panel (b) shows the average heart rate in bpm for each 30-second window relative to the start of the survey. The graph ends at 60 minutes, when almost all girls still have not yet finished the survey (see Appendix Section B.2 for more details).

We see two main interpretations of this reduction in heart rate (which are not mutually exclusive). First, it could capture a *general* improvement in stress or anxiety levels among the girls. Given the evidence that reductions in stress and anxiety can relax cognitive bandwidth constraints and thereby improve decision-making in a variety of domains (Haushofer & Fehr, 2014; Schilbach et al., 2016; Kaur et al., in press), this therefore points to an important potential mechanism that could underly the effects on learning. A reduction in stress or anxiety would also represent an important

[18]The distribution of heart rates shows an overall shift to the left in treatment groups (Figure A15). However, these distributions also show a grouping of large heart rate values at approximately double the magnitude of the mode. These could reflect real episodes of high heart rates (heart rate can typically jump to 150-200 bpm in situations of intense stress). Alternatively, they could reflect measurement error, e.g., driven by the heart rate monitors erroneously counting *both* of the two main peaks in arterial pressure (called the systolic and diastolic peaks) that occur within one heartbeat (see e.g. Paradkar & Chowdhury (2015) for discussion of peak detection issues in arterial pressure). As robustness tests, we drop or halve any observations above 175 bpm, and show that this still yields significant pooled effects (Table A16; p=0.09 and 0.07). Results are also robust to using girl's mean heart rate during the interview rather than all the heart rates in each 30 second window (p=0.02).

improvement in welfare in and of itself.

Second, the reduction in heart rate could be driven by a reduction in stress or anxiety *specifically* when discussing menstruation, thus capturing a reduction in menstrual stigma. We find mixed evidence regarding this channel, since heart rate does not seem to be differentially affected when actually speaking about menstruation during the endline (Table A17, columns 1 and 2), although there is some evidence that the treatment effect grows larger as the survey progresses (Table A17, column 4). On the other hand, girls may simply be correctly anticipating questions about menstruation even before they occur, since they were aware that the survey was about menstruation – so this evidence is not dispositive.

Multiple hypothesis testing. Given that we evaluate multiple primary outcomes, we check robustness to multiple hypothesis testing. We calculate Anderson (2008) sharpened q-values that control for the false discovery rate for our set of primary outcomes that are directly welfare-relevant: learning test scores, school attendance (as measured by spot-checks), general health, heart rate, and menstrual stigma (Table A6). Our substantive conclusions are not affected by this adjustment: the pooled coefficient on learning has a q-value of 0.009 (compared to a p-value of 0.004); the pooled coefficient on heart rate is still significant at the 5% level; and the pooled coefficient on the stigma index is still significant at the 1% level.

3.3 Hygiene knowledge and behavior

In Table 4 we show that the base program led to substantial improvements in hygiene knowledge and behavior, and that the *YGL* intervention led to significant *additional* improvements in these dimensions.

Knowledge. The program led to large improvements of 0.3-0.4 SD in girls' knowledge of both general hygiene (pooled effect: 0.41 SD, $p<0.001$, column 1), and of menstruation and menstrual hygiene (pooled effect 0.33 SD, $p<0.001$, column 2). For both of these knowledge indexes, the point estimates on the *Base+YGL* arm are larger than for the *Base only* arm, hinting that the *YGLs* may have been effective at spreading knowledge about the hygiene-based curriculum. However, the two treatment arms are not significantly different from each other in either case (p of difference: 0.12 and 0.51).

Reported hygiene behavior. Girls in *Base only* schools report engaging in more hygienic behaviors related to menstruation ($p<0.001$, column 4), but not more general hygiene behaviors ($p=0.20$, column 3).[19] By contrast, girls in *Base+YGL* schools report large improvements in both general (0.33 SD, $p<0.001$, column 3) and menstrual hygiene behaviors (0.56 SD, $p<0.001$, column 4). Notably, the coefficients are significantly larger in *Base+YGL* schools than *Base only* schools, suggesting that the *YGLs* play an important role in correcting the knowledge-behavior gap by embedding and promoting correct hygienic behaviors in their social networks.

Observed hygiene. Here we examine whether the reported improvements in hygienic behaviors trans-

[19] While self-reported hygiene behaviors are potentially subject to social desirability bias, we find no evidence that the effects are stronger for girls who are more likely to report socially desirable answers (Table A7).

late to measures of hygiene that are observed directly by enumerators, and are therefore likely to be less subject to reporting biases like social desirability. We find that *home environments* of girls' from *Base+YGL* schools are significantly cleaner (p<0.001, column 6), and treatment schools show very large improvements in observable measures of hygiene in the environment, such as the availability of soap and separation of trash (p<0.001, column 7). The *YGL* intervention leads to additional improvements in this measure (p of difference = 0.05 and 0.02). We do not find significant improvements in the observed hygiene of the girls themselves during the endline survey (pooled p=0.99, column 5), though also on this indicator the *Base+YGL* intervention is significantly better than the *Base* intervention. Overall the significant results on observed measures corroborate the results on reported behavior, namely that the *YGLs* lead to significant improvements in hygienic behaviors in their schools.

Transmission to mothers. We find mixed results on whether the changes in knowledge and behavior seen for the girls are transmitted vertically from girls to their mothers (Table A18). While mothers of girls in *Base + YGL* schools show improvements of 0.22 SD in menstruation-related knowledge (column 2, p<0.01), there are no significant improvements in self-reported behaviors or general hygiene knowledge (columns 1, 3, and 4). There is, however, a large improvement of 0.35 DS in mothers' progressive *attitudes* towards menstruation (column 5, p<0.001).

Table 4: *Knowledge and behavior indexes*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	Hygiene knowl.	Menstr. knowl.	Hygiene behavior	Menstr. behavior	Observed hygiene: Girls	Observed hygiene: Home	Observed hygiene: School
VARIABLES	(Z)	(Z)	(Z)	(Z)	(Z)	(Z)	(Z)
Base Only	0.345***	0.310***	0.110	0.360***	-0.071	0.017	1.917***
	(0.077)	(0.057)	(0.086)	(0.069)	(0.063)	(0.024)	(0.214)
Base + YGL	0.442***	0.339***	0.327***	0.560***	0.044	0.058***	2.353***
	(0.062)	(0.049)	(0.072)	(0.054)	(0.055)	(0.017)	(0.190)
Observations	2,167	2,167	2,167	2,256	2,167	2,287	140
Data source	Girls	Girls	Girls	Girls	Girls	Mothers	Schools
p: Treated = 0	<0.001	<0.001	<0.001	<0.001	.9900	.0095	<0.001
p: Base Only = Base + YGL	.1171	.507	.0022	.0018	.0264	.0493	.0152

Notes: The table displays knowledge and behavior index results from the girl (1)-(6), mother (7), and school (8) surveys in endline. In columns (1)-(7) standard errors are clustered at the school level and are in parentheses; controls include stratum fixed effects and the baseline controls selected by double LASSO. In column (8), standard errors are robust and controls include stratum fixed effects. The bottom of the table denotes the p-values on the test of being treated *Treated=0* and the equality of the coefficients *Base Only = Base + YGL*. Other p-values are defined analogously. ** p < 0.1, ** p < 0.05, *** p < 0.01*.

Hygiene knowledge index combines questions about WASH knowledge related to the program. *Menstruation knowledge index* is constructed from questions that test girls' knowledge about menstruation. *Hygiene behavior index* compiles questions about hygiene behavior during the last week. *Menstruation behavior index* includes questions about girls' behavior during the last period, such as the use of serviettes. *Observed hygiene: Girls* is measured by enumerator observations about the cleanliness of the girl at endline. *Observed hygiene at home* is constructed from enumerators' observations of home status during the mother's interview. *Observed hygiene: School* is constructed from enumerators' observations of the hygiene of the school during the school-level interview of teachers and directors. All outcomes are in control group standard deviations.

3.4 Menstrual stigma

Here, we document the program's effects on measures of menstrual stigma, first concentrating on the effect on survey measures of stigma (Table 5), then on revealed preference lab-in-the-field measures (Table 6).

Survey measures. The base program led to significant improvements in a wide variety of reported measures of menstrual stigma, and the *YGL* intervention led to substantial additional improvements (Table 5). In *Base only* schools, girls reported a 0.41 SD improvement in a combined index of stigma (column 1). This results from 0.20–0.37 SD improvements (all p-values below 0.01) in indexes of (i) how much they had heard people speaking about menstruation (column 2); (ii) how willing they were to speak about menstruation (column 3); (iii) how progressive their attitudes were regarding menstruation (column 4); (iv) how progressive they believed others' attitudes were (i.e., a measure of norms, column 5); and (v) they reported a reduction in emotions of shame and guilt in response to vignettes related to menstruation (column 6).[20] For all of these same measures, the *YGL* intervention generated significantly larger improvements of between 0.45 and 0.73 SD, resulting in an improvement of 0.74 SD on the combined index. These large changes suggest that the program reduced the stigma surrounding menstruation, at least according to girls' perceptions, and did so particularly effectively when *YGLs* were involved in challenging the stigma. This conclusion is reinforced by enumerators' observations of girls during the endline survey (column 7), which indicate that girls in the treatment schools were substantially more comfortable talking about menstruation than girls in control schools (e.g., they were less likely to look at the floor when discussing the subject), though for this last variable there is no significant difference between traetment groups. In addition, while the self-reported measures of stigma are in principle vulnerable to social desirability bias, we find no evidence that the treatment effects are driven by such bias (Table A7).

Revealed-preference lab-in-the-field. A series of revealed-preference measures show suggestive evidence that the reported reductions in menstrual stigma translate into changes in behavior (Table 6). When girls were informed they'd be asked to explain a topic in front of their schoolmates, and were asked whether to choose a menstruation-related topic or a broader hygiene-related topic, they were 4.6 p.p. (22%, 0.11 SD) more likely to choose the menstruation topic in the *Base+YGL* schools (p=0.03, column 1). By contrast, there is no effect on *Base only* schools (p=0.98, column 1; p of difference = 0.06). The *YGLs* in schools are therefore effective at making girls more willing to speak in front of others about menstruation, indicating a reduction in the taboo on the topic. In line with this, the additional effect of *YGLs* is specific to a choice that involves speaking *in front of other people* (when social image concerns are more salient). When girls instead choose *anonymously* which video should be shown to their schoolmates (column 2), the pooled treatment effect is significant (+3.8 p.p., 13%, 0.08 SD, p=0.05), and *Base+YGL* girls do not choose the menstruation topics more often

[20] Girls typically *underestimate* how progressive others' attitudes towards menstruation are in both control and treatment schools (Figure A19), suggesting a form of "pluralistic ignorance" (Katz et al., 1931; Bursztyn et al., 2020) that could be driven by a reluctance to speak openly about menstruation. Their misperceptions about teachers' attitudes are significantly reduced by the program, in line with a reduction in stigma.

Table 5: *Menstrual stigma - Survey measures*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
VARIABLES	Combined index (Z)	Received info. (Z)	Willingness to speak (Z)	Attitudes (Z)	Norms (Z)	Shame response to vignettes (rev.) (Z)	Observed shame (rev.) (Z)
Base Only	0.406***	0.202***	0.252***	0.367***	0.280***	0.308***	0.296***
	(0.065)	(0.078)	(0.076)	(0.064)	(0.060)	(0.059)	(0.061)
Base + YGL	0.740***	0.539***	0.730***	0.567***	0.473***	0.445***	0.229***
	(0.056)	(0.065)	(0.065)	(0.058)	(0.049)	(0.055)	(0.053)
Observations	2,256	2,256	2,250	2,254	2,167	2,165	2,256
Data source	Girls	Girls	Girls	Girls	Girls	Girls	Girls
p: Treated = 0	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
p: Base Only = Base + YGL	<0.001	<0.001	<0.001	<0.001	<0.001	.0081	.1844

Notes: The table displays results from the girls (1)-(7) and the teachers surveys (8) in endline. In all regressions, standard errors are clustered at the school level and are in parentheses, and controls include stratum fixed effects. In addition, columns (1)-(7) include the baseline controls selected by double LASSO. The bottom of the table denotes the p-values on the test of being treated *Treated=0* and the equality of the coefficients *Base Only = Base + YGL*. Other p-values are defined analogously. ** p < 0.1, ** p < 0.05, *** p < 0.01*. *Combined index* is an index that combines all indexes from (2) to (7) into one summary index. *Received info* is a factor analysis index from questions that asks girls how many times they have heard peers, family or teachers speak about menstruation in the last 3 months. *Willingness to speak* is an index constructed from questions that ask girls how likely they are to discuss topics related to menstruation. *Attitudes* is an index of how progressive girls' attitudes are when asked about questions related to menstruation. *Norms* is constructed from questions about the menstruation-related attitudes girls think their teachers and mothers have. *Shame response to vignettes (rev.)* index is constructed from the emotions that girls reported when prompted with situations where others hear them talk about menstruation or realize they had their period. A higher value indicates less shame. *Observed shame (rev.)* index is constructed from enumerator observations about the girls' behaviors during the menstruation survey questions. A higher value indicates less shame in their behavior. All outcomes are in control group standard deviations.

than *Base only* girls (p of difference = 0.52, column 2).[21]

There are no detectable effects on the probability of selecting a menstruation-related topic when presented with envelopes with information about various topics contained inside (pooled p=0.21) at endline. However, at *midline*, we found a stronger pooled effect of 4.1 p.p. on the probability of choosing a menstruation-related envelope (8.6%, 0.14 SD, p=0.003), with no difference between *Base only* and *Base+YGL* schools (p=0.88), suggesting that information demand does go up at least early on in the program.

Since undoing menstrual stigma also requires that girls face fewer social sanctions from an "out-group" when menstruating or discussing menstruation, we examine whether the program affects the behavior of *boys*. We do not find evidence of reduced menstrual stigma when boys play a dictator game with a girl who speaks about either menstruation or hand-washing. There is no effect on how many cookies boys shared with the girl speaking about menstruation (Table 6, column 5, pooled p=0.67), while boys in treatment schools share approximately 6% fewer cookies with the girl who spoke about hand-washing (column 4, pooled p=0.05).[22]

Overall, the set of revealed-preference measures provides suggestive evidence in favor of a reduction in menstrual stigma driven by the program, although the effects are more mixed and generally smaller in magnitude than the large effects on self-reported measures. This suggests that while attitudes and girls' perceptions of stigma changed substantially, this did not consistently translate into shifts in observed behavior. The set of menstrual stigma results also together indicate that the *YGL* component helped improve some of the outcomes more than the base intervention, while for other outcomes the base intervention alone was effective.

3.5 Psychosocial environment at school

We find evidence of generalized improvements in the psychosocial environment at schools driven by the base program (Table 7). Teachers, parents, and students reported changes of these kinds during qualitative work that took place around 9 months after the start of the program, and these findings are corroborated by two outcomes in the endline data.

First, girls report reductions in the *severity* of bullying in schools: while there is no effect on overall bullying levels or light teasing (Table 7, pooled p: 0.93 and 0.35; columns 1 and 2), there is a significant reduction of 0.08 SD in girls' reports of severe intimidation or harassment (pooled p=0.05, column 3).[23] This appears to represent a general improvement in relationships between students, rather than a reduction in menstrual stigma, since it is not driven by reduced bullying *about menstruation*. In fact, such teasing *increases* in treatment schools (Table 7, column 4, p<0.001), an effect

[21] The double difference (comparing the difference between *Base only* and *Base+YGL* across the *explanation* and *anonymous* outcomes) has p-value of 0.06, in line with the claim that the *YGL* intervention's impact is specific to the *explanation* outcome.

[22] There is also no significant effect of any treatment on whether girls themselves share cookies with the other girls (Table A20).

[23] Results on severe intimidation are very similar when girls report about peers or themselves. On the other hand, We do not find significant effects on teachers' reports of bullying, although the point estimate on severe intimidation and harassment is negative and a similar magnitude to the girls' reports (Table A21).

Table 6: *Menstrual stigma - lab measures*

VARIABLES	(1) Lab in field Explanation Chose menstr. (=1)	(2) Lab in field Anonymous Chose menstr. (=1)	(3) Informative envelope: P(chose menstr.)	(4) Prop. cookies shared w/ handwashing girl	(5) Prop. cookies shared w/ menstruation girl
Base Only	-0.001	0.046**	0.012	-0.033*	-0.006
	(0.026)	(0.021)	(0.016)	(0.020)	(0.018)
Base + YGL	0.046**	0.034	0.023	-0.030*	-0.007
	(0.021)	(0.021)	(0.016)	(0.016)	(0.016)
Observations	1,634	1,634	2,161	1,325	1,325
Data source	Lab in the field	Lab in the field	Girls	Boys	Boys
Control mean	.208	.291	.571	.454	.44
Control SD	.406	.455	.274	.206	.199
p: Treated = 0	.1247	.0467	.2135	.0503	.6663
p: Base Only = Base + YGL	.059	.5221	.4198	.8471	.9523

Notes: The table displays results from the revealed-preference lab-in-the-field experiments. In all regressions, standard errors are clustered at the school level and are in parentheses, and controls include stratum fixed effects. Controls selected by double LASSO are included for column (3), where we have baseline measures. The bottom of the table denotes the p-values on the test of being treated *Treated=0* and the equality of the coefficients *Base Only = Base + YGL*. Other p-values are defined analogously. *$* p < 0.1$, *$** p < 0.05$, *$*** p < 0.01$.

Columns (1)-(2) present results from the *lab-in-the-field* experiment where girls were informed they would need to explain or show (anonymously) a video on a topic in front of their schoolmates. They were asked to choose between a menstruation-related topic and a broader hygiene-related topic. *Explanation, chose menstr.* (column 1) indicates whether the girl chose to provide an in-person explanation about a menstruation-related topic to the group rather than another hygiene topic. *Anonymous, chose menstr.* (column 2) indicates that the girl chose to anonymously show a video about a menstruation-related topic instead of a hygiene-related topic. Controls in columns (1)-(2) include fixed effects for choice order (*Explanation first=1*). Column 3 displays results from the envelopes experiment where girls had to choose between envelopes containing multiple topics, including menstruation-related topics. *Informative envelope: P(chose menstr.)* is the proportion of times out of 4 that girls chose to learn about a menstruation-related topic instead of a non-stigmatizing topic (such as whether mosquitoes are equally attracted to all blood types). Columns (4)-(5) show results from a dictator game implemented during the boys' survey. Boys had to decide how many cookies to share with a girl after watching a video of the girl providing a menstruation or hand-washing-related explanation. The outcomes in these columns represent the proportion of cookies (out of 5) that boys agreed to share with the girl in the video. Controls include order of videos fixed effects.

Table 7: *Impacts on the school psychosocial environment*

VARIABLES	(1) Bullying Girls report Combined (Z)	(2) Bullying Girls report Teasing only (Z)	(3) Bullying Girls report Intimidation/ harassment only (Z)	(4) Bullying Girls report About menstr. only (Z)	(5) Bullying Teacher-report Combined (Z)	(6) Network connection index (Z)	(7) Network connection w. girls only (Z)	(8) Network connection w. boys only (Z)
Base Only	0.026	0.106	-0.083*	0.136*	-0.067	0.174**	0.071	0.205**
	(0.062)	(0.067)	(0.046)	(0.070)	(0.056)	(0.086)	(0.076)	(0.084)
Base + YGL	-0.019	0.027	-0.069*	0.270***	-0.028	0.267***	0.183***	0.266***
	(0.054)	(0.059)	(0.040)	(0.061)	(0.052)	(0.074)	(0.062)	(0.076)
Observations	2,167	2,167	2,167	2,167	3,637	2,167	2,167	2,167
Data source	Girls	Girls	Girls	Girls	Teachers	Girls	Girls	Girls
p: Treated = 0	.9284	.3488	.0542	<0.001	.4343	<0.001	.0121	<0.001
p: Base Only = Base + YGL	.3963	.1594	.7384	.0444	.3908	.1859	.0743	.3747
p: Base Only [Girls] = Base Only [Boys]								.21
p: Base + YGL [Girls] = Base + YGL [Boys]								.161
p: Treated [Girls] = Treated [Boys]								.133

Notes: The table displays effects on bullying and network connection between students at endline. Standard errors are clustered at the school level and are in parentheses; controls include stratum fixed effects. All columns except column 5 have baseline controls selected by double LASSO. *Treated=0* and the equality of the coefficients *Base Only = Base + YGL*. Other p-values are defined analogously. ** p < 0.1, ** p < 0.05, *** p < 0.01*.

Columns (1)-(4) use girls' reports about bullying in their school, combining peer reports (reporting about other girls being bullied, using a randomly selected subset of 2 or 3 girls in the same grade-level and removing cases where the respondent did not know the other girl) and self-reports (about themselves being bullied). Column 1 combines both light teasing and severe intimidation and harassment. Column 2 includes only teasing. Column 3 includes only severe intimidation and harassment. Column 4 includes only questions about bullying specifically related to menstruation. Column 5 is analogous to column (1) but uses reports from teachers about bullying in their class. Each teacher reports on a randomly selected sample of 3 girls from their class; the observations are at the *teacher × girl* level. *Network connection index (Z)* combines questions about the girls' awareness, friendship, interaction, cooperation, and communication with school peers. Column 6 refers to interactions with all peers. Column 7 restricts to the subset of questions that were asked specifically about female peers; column 8 is specifically about male peers.

that is driven solely by teasing *from boys* (Table A22, column 5). This suggests that the program was not successful at reducing stigmatizing behaviors among boys, even while attitudes and behaviors strongly improved among girls, and overall severity of bullying decreased.

Second, the program led to a significant increase in a measure of network integration among students (Table 7, column 6, pooled $p<0.001$), based on asking girls how many social connections they have of different types at endline (e.g., how many friends, people they play with, people they share with, people they do projects with). Social connections with both girls and boys increase significantly.[24]

3.6 Dynamics: comparing midline and endline

To understand the dynamic process through which the interventions affected the outcomes, Figure 3 shows the effect sizes of the two treatment arms at both midline (after approximately 4 months of the program) and endline (after approximately 1.5 years of the program).[25] Four key patterns stand out.

[24] Parts of the measured effects are driven by large increases in the number of peers girls speak with specifically about menstruation (Figure A23). But even after removing this question from the index, there are still 0.13-0.17 SD increases in the network integration index for boys (pooled p-value = 0.03), suggesting a more general improvement in social dynamics between girls and boys.

[25] To ensure comparability across waves, we (i) only include individuals found at both midline and endline waves in the samples used for estimates in this particular figures; (ii) construct indexes so that they are only composed of questions that were common across midline and endline waves, and use the same factor loadings for factor analysis-based indexes; and (iii) normalize all outcomes using the control mean and standard deviation *at endline*.

First, the null results seen on school attendance, and self-reported health and mental health also hold at midline, with no significant effects at the 5% level across all measures and treatments (Figure 3, panels 1-7).[26] This suggests that these human capital outcomes are not on average affected by the program at any point. We did not administer learning tests or measure heart rate at midline, so we cannot evaluate the dynamics of those outcomes.

Second, the program reduced *menstrual stigma* relatively rapidly, showing significant effects at midline (using a combined index of all menstrual stigma measures available at both midline and endline; Figure 3, panel 8; p=0.03 for *Base only* and p<0.001 for *Base+YGL*). These effects were consolidated by endline, with slightly larger point estimates (although the endline effects do not differ significantly from the midline ($p \in [0.14, 0.33]$).

Third, *knowledge* about menstruation and hygiene improved more rapidly in the *Base+YGL* schools compared to the *Base only* schools at midline (p of difference = 0.09 for hygiene knowledge; 0.002 for menstruation knowledge). But by endline the differences had diminished, suggesting that *Base only* schools caught up (Figure 3, panels 9-10; p of [*Base only=Base+YGL*] at endline = 0.17 and 0.98). There are also significant improvements in both types of knowledge between the midline and endline (p<0.05 for all comparisons for both treatment groups).

Finally, reported *behavior* was slow to move, with no significant effects at midline except a small effect on menstruation behavior in the *Base+YGL* schools (0.10 SD, p=0.08). Only by endline do we see strong and significant effects (Figure 3, panels 11-12). Combined with the results above, this suggests that behavior takes longer to change than knowledge and reported stigma, in line with a literature showing that generating behavior change can be challenging, especially in the field of hygiene and health (Aboud & Singla, 2012).

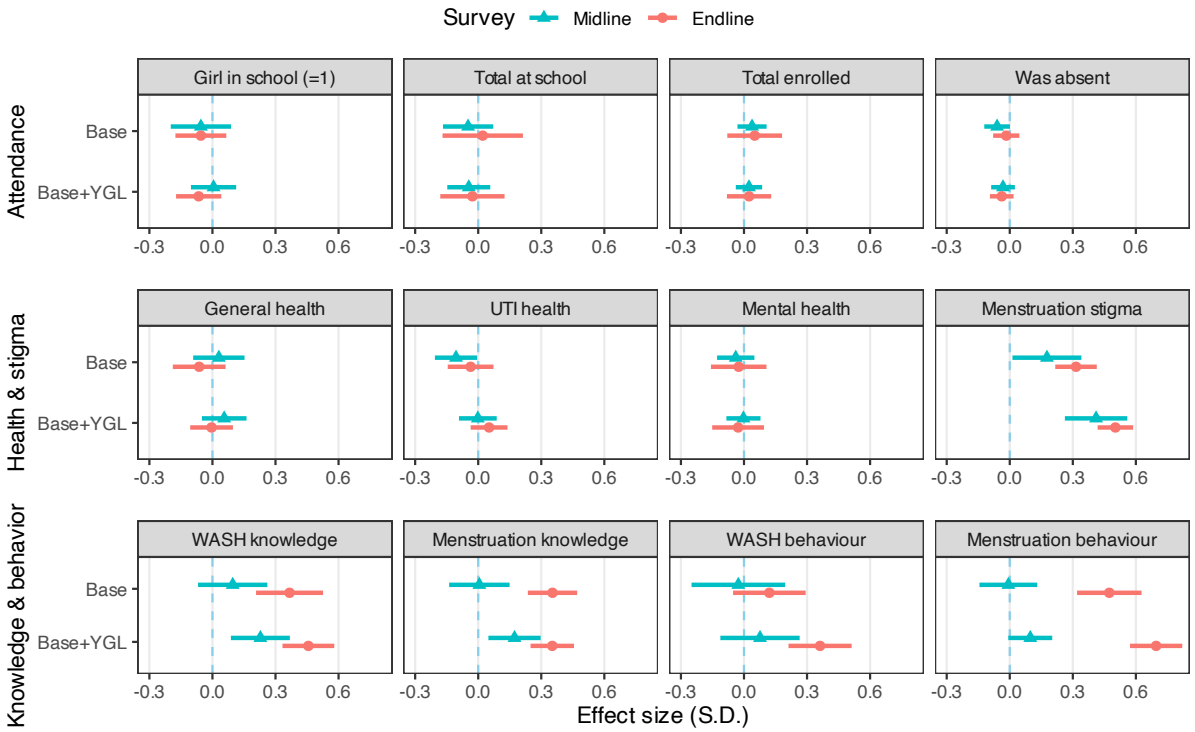
Broadly, the results confirm that outcomes earlier in the theory of change (knowledge and perceptions of stigma) are more responsive at midline, while behavior takes longer to change. The dynamic results further show that the YGLs were instrumental in shifting knowledge, stigma and menstrual behavior more quickly than the base program alone.

4 Mechanisms

In this section, we further investigate the mechanisms behind the program's impacts and the relationships between them. We first note that the relatively large changes in hygiene knowledge, behavior and social stigma in both treatment arms confirm that the program led to meaningful changes along the theory of change. It is thus plausible that they contributed to the observed improvements in final human capital outcomes, in particular in learning. We further argue that the improvements in learning are likely driven by improvements in the psychosocial environment at school, reductions in stress, and associated improvements in quality of learning while at school. We then investigate why the YGL program did not lead to additional impacts on learning or stress, despite generating large additional improvements in hygiene behavior and menstrual stigma. To do so, we explore

[26] One exception, discussed above in section 3.2, is that we find suggestive evidence of an increase in the incidence of negative emotions in *YGL* schools at midline (Table A14).

Figure 3: *Dynamics: comparing effects at midline and endline*



Notes: This figure shows the effects on each outcome at both midline and endline. Indices are re-calculated so that they only use the questions that were asked in both midline and endline. Only girls in both the midline and endline samples are included. All outcomes are normalized using the control mean and standard deviation from the endline. All index outcomes are combined using factor analysis, based on the factor loadings from the endline survey.

heterogeneous impacts of the YGL program to understand how they relate to the effects of the base program. Finally, we examine possible negative consequences of the program by examining effects on fertility and backlash.

4.1 Mechanisms behind learning effects

Why does learning improve in treatment schools? Since the effects on learning are never larger in *Base+YGL* schools than *Base only* schools (Table 1), they are driven by base program. Note that in principle, any component of the base program could have led to reduced stress and improved the psychosocial environment (e.g., sanitary pads could reduce stress associated with menstruating at school; teacher training could increase their motivation and improve classroom dynamics). We cannot experimentally evaluate the relative importance of each component, and instead focus on understanding the key mechanisms.

Health and attendance. The null results on other human capital outcomes suggest that the effect on learning does not occur through measurable improvements in physical health outcomes or through spending an increased *quantity* of time at school. Instead, the *quality* of time at school seems to have improved.

Stress. The reduction in heart rate discussed above points to an important potential channel for this improvement in quality: reducing girls' anxiety and stress could allow them to concentrate more effectively while at school. This is in line with a literature on "bandwidth constraints" (Haushofer & Fehr, 2014; Mullainathan & Shafir, 2013; Kaur et al., in press) that suggests that stress can inhibit decision-making ability and productivity.

Psychosocial environment. The reductions in severe bullying in treatment schools and increases in network integration could also underly the learning effects. This channel could operate through the reductions in stress measured as reductions in heart rate (e.g., by reducing the stress associated with bullying or with social isolation). Additionally, the increased social integration could directly improve learning by encouraging peer-to-peer motivation or improving students' motivation.

Hygiene behaviors and time in class. The interventions focused on menstrual management and hygiene behaviors could have increased learning while at school by directly impacting how much time girls spent in class during a school day, even conditional on attending school.[27] For example, access to private toilets, sanitary pads, or improved knowledge of good menstrual hygiene may have reduced the likelihood of leaving class due to leaks during menstruation. We find some evidence that the program may indeed affect this "intensive margin" of school attendance. Of the 1053 girls who had started menstruating at baseline, girls in treatment schools were by endline 3.1 p.p. (27%) less likely to report having been absent from school because of menstruation (Table A24, column 2, pooled p=0.11); 3.8 p.p. (35%) less likely to have left school during the day because of menstruation (column 4, pooled p=0.05); and 3.8 p.p. (17%) less likely to report not being able to fully

[27] While the unannounced spot checks at school could in theory pick up this intensive margin of school attendance, we are likely underpowered to detect such changes, especially since girls may be marked as attending school even if they are (for example) in the toilet during the rollcall.

participate in school activities because of menstruation (column 5, pooled p=0.10).[28]

However, it is not clear that improved menstrual management allowing more time in the classroom is a key mechanism behind the learning effects. First, these mechanisms only affect those who have started menstruating by endline, i.e., around 67% of the sample. Second, the effects of menstruation on classroom time seem relatively small, even in the control group. For example, only 11% of menstruating girls in control schools report *ever* having had to leave class because of menstruation. Third, learning effects are not larger for girls who had reached menarche at baseline (Table A25) or for girls who started menstruating between baseline and endline (Table A26). The treatment effects are therefore not driven by girls who we would expect to be able to change their menstrual management behaviors the most.[29] These results together suggest that the learning effects are unlikely to be *primarily* driven by menstrual management behaviors or associated changes in time spent in class. However, this does not rule out that they could contribute to the program's effects, or that the "hard" components of the program could drive *psychosocial* changes that generate learning effects. For example, access to sanitary pads and good menstrual management may reduce the stress associated with girls fearing leaks or smell, enabling better concentration in class.

Teacher behavior. We do not find evidence that the learning effects are driven by increases in teacher effort in response to the program, since teacher absenteeism rates are unaffected by the program (Table A27, column 4, p=0.44).[30] However, the program generates large improvements in teachers' progressive attitudes towards menstruation (Table A27, column 1, pooled effect: 0.67 SD; p<0.001), and smaller but significant improvements in knowledge about menstruation (column 2, pooled p=0.08). These changes may make teachers more attentive to *girls* in classrooms, or more responsive when they are being bullied about menstruation, possibly leading to improved psychosocial outcomes or learning improvements for girls.

Crowd out in the YGL program. The effects on learning (Table 1) suggest that the base program drives the improvements, while including Young Girl Leaders in the intervention generates no further improvements. There may even be a negative effect of the Young Girl Leaders: the point estimates are lower for *Base+YGL* schools than for *Base Only* schools across all measures, and is significantly lower for the *achievement tests* (column 2; p=0.02). There is some suggestive evidence that this could be because the time and effort YGLs dedicate to sensitizing others about hygiene and stigma comes at the cost of academic performance, perhaps by crowding out traditional study time. In particular, we find a negative trend in grades of girls nominated as YGLs: between baseline and endline, YGLs' average academic score decreases by 1.2 (on a 20-point scale) and their rank within

[28] We use the baseline measure of menarche, because at endline the reduction in stigma in treatment schools appears to have made girls *more* likely to reveal that they had started menstruating than in control schools, making it an endogenous outcome. Note that we cannot rule out that lower menstruation-driven absenteeism is simply due to fewer girls being *enrolled* in treatment schools (given the negative, although insignificant, point estimate on enrollment in Table 2).

[29] In addition, the timing of impacts suggests that infrastructure access is unlikely to be a key driver of learning. Girls were more likely to progress to the next grade-level during the study period (Table 1, column 6), implying that learning had improved *by the end of the first school year*, when infrastructure construction had not been completed in the majority of treatment schools.

[30] There is also no effect on *female* teacher absenteeism, that could plausibly have been driven by improved access to hygiene infrastructure (p=0.64).

the class worsens by 6.4 places (Table A28). These results should be interpreted with caution, since (i) YGLs were not chosen randomly, preventing us from making causal inferences about their performance compared to non-YGLs, and (ii) some of the time-trend could partly capture regression to the mean, since YGLs were selected to have above-average academic scores (and at endline they still have higher scores than non-YGLs; +0.34 points, $p<0.01$). Nevertheless, this pattern may explain overall weaker learning results in YGL schools. If YGLs influence others' behavior by acting as exemplars, their focus on hygiene and stigma may shift other students' attention away from traditional academic learning.[31] There is also a mechanical effect on school-level learning, since YGLs in the sample will reduce the measured mean.

4.2 Which YGLs were effective?

The design of the YGL program was motivated by the idea that one could identify *positively deviant* girls in each school, i.e., those willing to speak out in defiance of a harmful norm, and that those girls would be especially effective at changing others' behavior. To examine this hypothesis, we analyze heterogeneity with respect to the selected YGLs' positive deviance in each school using data from the YGL selection process.[32] We find that schools with more positively deviant YGLs indeed exhibit better human capital and behavior outcomes, suggesting that selecting the "right" YGLs can improve outcomes significantly. However, these improvements are strongly conditional on the baseline school environment, either because positively deviant YGLs can only be *found* in already-progressive school environments, or because positively deviant YGLs can only have positive effects if they are working within a progressive school environment.

We first motivate our analysis by using machine learning to test whether there is heterogeneity in the effect of the YGL program, based on a large set of baseline characteristics (following Chernozhukov et al., 2018). We find evidence of significant heterogeneity on the *Base + YGL* treatment for 3 outcomes (Table A29): (i) whether girls are enrolled in school at endline; (ii) menstruation knowledge; and (iii) the index of diffusion of information about menstruation. The heterogeneity parameter also has a p-value of 0.105 for the reading fluency measure.

Motivated by this evidence, we then analyze heterogeneity with respect to positive deviance. To do this, we construct an index of each YGLs' *positive deviance*, using data collected during the YGL selection process on their their willingness to engage in prosocial behaviors in defiance of existing

[31] Girls in *Base+YGL* schools are significantly less likely to progress to the next grade-level when the *YGLs* were asked to carry out *formal* classroom sessions, compared to when they were asked to carry out *informal* sensitization, e.g., during recreation (p=0.05, Table A2). While a similar difference is not observed in the learning tests (p=0.70), it nevertheless suggests that using classroom time to learn about hygiene and menstrual hygiene may crowd out academic learning.

[32] In the design phase, we considered randomizing the YGL selection criteria - selecting positive deviants in some schools and *popular* (network-central) girls in others. We stepped away from this design after piloting, because in many schools there were not enough candidates who satisfied other key criteria (having started menarche and having good school grades) to generate variation in the type of selected YGL *within* a school. Given that there was, however, variation *between* schools in positive deviance, we instead decided at the design stage to test for heterogeneity with respect to the selected YGLs' positive deviance.

norms (see Appendix Section B.4 for the questions used).[33] We then average across the YGL of a school to obtain a school-level indicator of positive deviance of the selected YGLs. Importantly, the variation in this index does not only capture individual YGLs' characteristics, but is also likely to capture features of the baseline school environment, since the probability of finding more positive deviant girls does not vary randomly across schools.

We then parametrically analyze heterogeneity using this school-level indicator of YGL positive deviance (Table A30). Schools with more positively deviant YGLs have significantly better learning test scores, lower rates of severe bullying, and better reported mental health. They also generate strong improvements in UTI health. This suggests that schools with more positively deviant YGLs were more effective at generating learning, health, and psychosocial improvements. Plausibly, girls who are willing and experienced in speaking out in defiance of social norms are better able to speak openly about menstruation without damaging the overall psychosocial environment of the school. Conversely, schools where these positively deviant YGLs were *not* selected have significantly weaker human capital outcomes, and are likely to drive the overall lower point estimates on learning found in *Base+YGL* schools compared to *Base only* schools.

That said, the improved outcomes in schools with positively deviant YGLs appear to be strongly linked to the baseline school environment. First, the positive deviance of selected YGLs in a school at baseline is positively correlated with mothers' education and being a secondary school, and negatively correlated with girls having a brother at baseline (all p-values below 0.01). This suggests that the ability to find positively deviant YGLs is strongly related to the school environment.[34] Second, when examining the positive deviance of selected YGLs *relative* to the wider pool of girls who were considered as candidates for the YGL position in a given school, the pattern of heterogeneity does not persist: for example, relatively positively deviant YGLs are not associated with higher learning scores, and are even associated with less progression to the next grade (Table A31). This implies that the earlier heterogeneity results are primarily driven by some schools having more positive deviants among their possible candidate YGLs. Together these results suggest that either (i) positively deviant girls who are willing to defy a harmful norm tend to be mainly *found* in school environments that are already more educated or progresive, or that (ii) positively deviant girls are only *effective* when the school environment is already educated or progressive. In either case, the results re-emphasize the importance of the school environment for girls' human capital outcomes.

4.3 Possible negative consequences of the program

Fertility. We find some evidence that the base program increased the probability of pregnancy (Table A32, column 2). It is difficult to assess the robustness of this result, because the overall rates of pregnancy in the sample are very low: they move from 2% in the control group to 3.5% in the *Base only* group. However, the p-value is robust to using randomization inference (Table A9), the effects

[33] Our measure of positive deviance does not simply pick up the overall "quality" of the YGL, e.g., in terms of leaderships skills, confidence, popularity, or intelligence. When we construct a quality measure based on the first factor from an exploratory factor analysis of the YGL selection data, we do not find similar heterogeneity results.

[34] This could explain the negative heterogeneity result on "Received info.": in places where girls already knew more at baseline, YGLs plausibly focused less on spreading information.

are not driven by a few outlier schools with many pregnancies (Figure A33), and the results are corroborated by more reports of girls dropping out of school due to pregnancy (Table A32, column 3). This result therefore suggests an important cautionary note for programs that focus on reducing stigma regarding menstrual and reproductive health. Opening up discussion about such topics may also decrease stigma regarding sexual interactions and could lead to early pregnancies for a subset of girls.

Backlash. The program encourages girls to speak out against the harmful taboo surrounding menstruation, which could expose them to backlash from teachers, parents, or other students. The overall pattern of results suggests that such backlash was not common, for two reasons. First, the program generated improvements in psychosocial environment and large reductions in menstrual stigma. Second, *YGLs*, who are plausibly most vulnerable to backlash, appear to be well supported by the school community members: 92% of teachers and directors say that YGLs had a positive impact on the school atmosphere, and 80% of girls said that they *admire* the YGLs in an open-ended question. Nevertheless, as discussed above, teasing about menstruation does increase in treatment schools (Table 7, column 4, $p<0.001$), which could have negative effects on a subset of girls. Since this was driven by teasing from boys, it suggests that a program that more explicitly targeted boys' attitudes may be helpful in mitigating backlash.

We also find some evidence that mothers report worsened psychosocial behavior for their daughters in *Base only* schools (effect: -0.14 SD, Table A18, column 5, $p=0.06$), although not for *Base+YGL schools* (p of difference: 0.02). The fact that mothers are unhappy with their daughters' behavior may be an indication that they do not agree with the messages of the program. It may also be related to the increases in pregnancy rates in *Base only* schools.

4.4 Welfare implications of the YGL program

Despite potential costs to academic learning, the overall welfare implications of the YGL program can still be positive, given the countervailing improvements in hygiene and menstrual hygiene-related knowledge, behavior and stigma documented in Sections 3.3 and 3.4, and the fact that the YGLs offset the negative impacts of the base program on fertility and mothers' perceptions of behavior. Broadly, girls in YGL schools may gain less in academic skills like reading and math, but they learn other skills that can improve their health and well-being. We cannot adjudicate which is more beneficial over the long run. The results do show, however, that these welfare trade-offs are weaker in schools with a more favorable initial environment.

5 Discussion

Motivated by the evident social stigma that surrounds menstruation in many contexts, policymakers have developed interventions that target menstrual hygiene in schools in low- and middle-income countries. While some interventions focus mainly on relaxing hard constraints to menstrual hygiene management, others also target the menstrual stigma that limits the flow of information and worsens the psychosocial environment for menstruating girls more broadly. This paper provides empirical evidence on the effects of such a broader intervention, which combined sanitary infrastructure and

access to sanitary pads with efforts to destigmatize menstruation and its discussion. A unique feature of the approach was to identify and train "Young Girl Leaders" in each school to spread positive messages about hygiene and menstruation, and to encourage healthy behaviors.

After 1.5 years, the program led to large shifts in knowledge and reported hygiene and menstrual hygiene behaviors, as well as improvements in several measures of menstrual stigma, including impacts on objective and observed measures. The results demonstrate that targeted interventions in school can tackle social taboos. While we observe important impacts along the causal chain for the base program, impacts on behavior and stigma were larger (and started earlier) in schools with *Young Girl Leaders* - suggesting that this component was effective. This confirms that identifying and supporting positive deviants has the potential to be an effective way of addressing harmful social norms in school environments. At the same time, the attitudes and behaviors of girls' mothers and male classmates do not seem to be shifted by the program, and positively deviant YGLs only seem to be effective in already-favorable school environments. This points to the need for future research on interventions that target a broader set of actors (e.g., teachers, parents, and boys) to help YGLs be effective in more challenging school environments.

The base intervention itself led to substantial improvements in learning, despite being designed to primarily target hygiene and menstrual hygiene, and having no apparent impact on physical health outcomes and school attendance. It also led to reductions in stress (as measured by reduced heart rate at endline) that could generate improved learning and concentration, along with improvements in the school environment (as measured by increases in social ties students and reductions in severe bullying). This suggests that the focus on health and attendance (in both policy discussions and the academic literature) may be ignoring effects on important welfare-relevant outcomes like learning and stress, and may be neglecting the psychosocial channels that may underlie them. More broadly, the results are therefore highly relevant for discussions on girls' education and the learning crisis in low- and middle-income countries, and point to the importance of the school environment in affecting girls' human capital outcomes.

References

Aboud, F. E., & Singla, D. R. (2012, August). Challenges to changing health behaviours in developing countries: A critical overview. *Social Science & Medicine (1982)*, *75*(4), 589–594. doi: 10.1016/j.socscimed.2012.04.009

Akyeampong, K., Andrabi, T., Banerjee, A., Banerji, R., Dynarski, S., Glennerster, R., ... Yoshikawa, H. (2023). *Cost-Effective Approaches to Improve Global Learning - What does recent evidence tell us are "Smart Buys" for improving learning in low- and middle-income countries?* London, Washington D.C., New York: FCDO, the World Bank, UNICEF, and USAID.

Alan, S., Baysan, C., Gumren, M., & Kubilay, E. (2021, November). Building Social Cohesion in Ethnically Mixed Schools: An Intervention on Perspective Taking. *The Quarterly Journal of Economics*, *136*(4), 2147–2194. doi: 10.1093/qje/qjab009

Alan, S., & Kubilay, E. (2024). Empowering Adolescents to Transform Schools: Lessons from a Behavioral Targeting. *European University Institute, mimeo*.

Ali, T. S., & Rizvi, S. N. (2010, August). Menstrual knowledge and practices of female adolescents in urban Karachi, Pakistan. *Journal of Adolescence*, *33*(4), 531–541. doi: 10.1016/j.adolescence.2009.05.013

Anderson, M. L. (2008, December). Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects. *Journal of the American Statistical Association*, *103*(484), 1481–1495. doi: 10.1198/016214508000000841

Ashraf, N., Bau, N., Low, C., & McGinn, K. (2020). Negotiating a Better Future: How Interpersonal Skills Facilitate Intergenerational Investment*. *The Quarterly Journal of Economics*. doi: 10.1093/qje/qjz039

Bandiera, O., Buehren, N., Burgess, R., Goldstein, M., Gulesci, S., Rasul, I., & Sulaiman, M. (2020, January). Women's Empowerment in Action: Evidence from a Randomized Control Trial in Africa. *American Economic Journal: Applied Economics*, *12*(1), 210–259. doi: 10.1257/app.20170416

Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2013, July). The Diffusion of Microfinance. *Science*, *341*(6144), 1236498–1236498. doi: 10.1126/science.1236498

Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2019, November). Using Gossips to Spread Information: Theory and Evidence from Two Randomized Controlled Trials. *The Review of Economic Studies*, *86*(6), 2453–2490. doi: 10.1093/restud/rdz008

Banerjee, A., La Ferrara, E., & Orozco-Olvera, V. H. (2019). The Entertaining Way to Behavioral Change: Fighting HIV with MTV. *NBER Working Paper*(26096).

Beaman, L., BenYishay, A., Magruder, J., & Mobarak, A. M. (2021, June). Can Network Theory-Based Targeting Increase Technology Adoption? *American Economic Review*, *111*(6), 1918–1943. doi: 10.1257/aer.20200295

Belloni, A., Chernozhukov, V., & Hansen, C. (2014, April). Inference on Treatment Effects after Selection among High-Dimensional Controls. *The Review of Economic Studies*, *81*(2), 608–650. doi: 10.1093/restud/rdt044

Benshaul-Tolonen, Anja, Aja, Zulaika, G., Sommer, M., & Phillips-Howard, P. A. (2020). Measuring menstruation-related absenteeism among adolescents in low-income countries. In *The Palgrave Handbook of Critical Menstruation Studies* (pp. 705–723). Palgrave Macmillan.

Benshaul-Tolonen, A., Zulaika, G., Nyothach, E., Odour, C., Mason, L., Obor, D., & Laserson, K. F. (2021). *Sanitary products, absenteeism and psychosocial well-being: Evidence from a three-arm cluster randomized controlled feasibility study in Western Kenya* (Working Paper No. WP No.93). Columbia University, SIPA, CDEP-CGEG.

Bergstrom, K., & Ozler, B. (2021). *Improving the Well-Being of Adolescent Girls in Developing Countries*. The World Bank. doi: 10.1596/1813-9450-9827

Bernard, T., Dercon, S., Orkin, K., & Taffesse, A. S. (2019, May). Parental Aspirations for Children's Education: Is There a "Girl Effect"? Experimental Evidence from Rural Ethiopia. *AEA Papers and Proceedings*, *109*, 127–132. doi: 10.1257/pandp.20191072

Bicchieri, C. (2017). *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*. Oxford University Press. doi: 10.1093/acprof:oso/9780190622046.001.0001

Bicchieri, C., & Funcke, A. (2018, March). Norm Change: Trendsetters and Social Structure. *Social Research: An International Quarterly*, *85*(1), 1–21. doi: 10.1353/sor.2018.0002

Brown, S., & Taylor, K. (2008). Bullying, education and earnings: Evidence from the National Child Development Study. *Economics of Education Review*, *27*(4), 387–401.

Buehren, N., Goldstein, M. P., Gulesci, S., Sulaiman, M., & Yam, V. (2017, February). Evaluation of an adolescent development program for girls in Tanzania. *Policy Research Working Paper Series*(7961).

Bursztyn, L., González, A. L., & Yanagizawa-Drott, D. (2020, October). Misperceived Social Norms: Women Working Outside the Home in Saudi Arabia. *American Economic Review*, *110*(10), 2997–3029. doi: 10.1257/aer.20180975

Chernozhukov, V., Demirer, M., Duflo, E., & Fernández-Val, I. (2018, June). *Generic Machine Learning Inference on Heterogeneous Treatment Effects in Randomized Experiments, with an Application to Immunization in India* (Working Paper No. 24678). National Bureau of Economic Research. doi: 10.3386/w24678

Chirgwin, H., Cairncross, S., Zehra, D., & Sharma Waddington, H. (2021, December). Interventions promoting uptake of water, sanitation and hygiene (WASH) technologies in low- and middle-income countries: An evidence and gap map of effectiveness studies. *Campbell Systematic Reviews*, *17*(4), e1194. doi: 10.1002/cl2.1194

Crowne, D. P., & Marlowe, D. (1960). Marlowe-Crowne social desirability scale. *Journal of Consulting Psychology*.

Czura, K., Menzel, A., & Miotto, M. (2019). Menstrual Health, Worker Productivity and Well-being among Female Bangladeshi Garment Workers. *SSRN Electronic Journal*. doi: 10.2139/ssrn.3519895

Dhar, D., Jain, T., & Jayachandran, S. (2022, March). Reshaping Adolescents' Gender Attitudes: Evidence from a School-Based Experiment in India. *American Economic Review*, *112*(3), 899–927. doi: 10.1257/aer.20201112

Dinarte, L., & Egana del Sol, P. (2018). Preventing Violence in the Most Violent Contexts. Behavioral and Neurophysiological Evidence. (May).

Dupas, P., & Miguel, E. (2017). Impacts and determinants of health levels in low-income countries. In *Handbook of economic field experiments* (Vol. 2, pp. 3–93). Elsevier.

Edmonds, E., Feigenberg, B., & Leight, J. (2020, July). *Advancing the Agency of Adolescent Girls* (Tech. Rep. No. w27513). Cambridge, MA: National Bureau of Economic Research. doi: 10.3386/w27513

El-Gilany, A.-H., Badawi, K., & El-Fedawy, S. (2005, January). Menstrual Hygiene among Adolescent Schoolgirls in Mansoura, Egypt. *Reproductive Health Matters*, *13*(26), 147–152. doi: 10.1016/S0968-8080(05)26191-8

Eriksen, T. L. M., Nielsen, H. S., & Simonsen, M. (2014). Bullying in Elementary School. *Journal of Human Resources*, *49*(4), 839–871.

Evans, D. K., & Yuan, F. (2022, February). What We Learn about Girls' Education from Interventions That Do Not Focus on Girls. *The World Bank Economic Review*, *36*(1), 244–267. doi: 10.1093/wber/lhab007

Gulesci, S., Jindani, S., La Ferrara, E., Smerdon, D., Sulaiman, M., & Young, H. (2023). A Stepping Stone Approach to Norm Transitions. *SSRN Electronic Journal*. doi: 10.2139/ssrn.4425503

Hamory, J., Baird, S., Das, S., Jones, N., Woldehanna, T., & Yadete, W. (2023). Do layered adolescent-centric interventions improve girls' capabilities? Evidence from a mixed-methods cluster randomised controlled trial in Ethiopia.

Haushofer, J., & Fehr, E. (2014, May). On the psychology of poverty. *Science*, *344*(6186), 862–867. doi: 10.1126/science.1232491

Hennegan, J., & Montgomery, P. (2016, February). Do Menstrual Hygiene Management Interventions Improve Education and Psychosocial Outcomes for Women and Girls in Low and Middle Income Countries? A Systematic Review. *PLOS ONE*, *11*(2), e0146985. doi: 10.1371/journal.pone.0146985

Hu, Q. (2023). Breaking Down Bullying: Empathy, Social Networks, and Adolescents.

Jayachandran, S. (2021, September). Social Norms as a Barrier to Women's Employment in Developing Countries. *IMF Economic Review*, *69*(3), 576–595. doi: 10.1057/s41308-021-00140-w

Katz, D., Allport, F. H., & Jenness, M. B. (1931). *Students' attitudes; a report of the Syracuse University reaction study*. Oxford, England: Craftsman Press.

Kaur, S., Mullainathan, S., Oh, S., & Schilbach, F. (in press). Do financial concerns make workers less productive? *The Quarterly Journal of Economics*.

Laajaj, R., & Macours, K. (2019, October). Measuring Skills in Developing Countries. *Journal of Human Resources*, 1018-9805R1. doi: 10.3368/jhr.56.4.1018-9805R1

Marsh, D. R., Schroeder, D. G., Dearden, K. A., Sternin, J., & Sternin, M. (2004, November). The power of positive deviance. *BMJ*, *329*(7475), 1177–1179. doi: 10.1136/bmj.329.7475.1177

Mullainathan, S., & Shafir, E. (2013). *Scarcity: Why Having Too Little Means So Much*. Henry Holt and Company.

Muralidharan, K. (2017). Field Experiments in Education in Developing Countries. In A. Banerjee & E. Duflo (Eds.), *Handbook of Field Experiments* (pp. 323–385). doi: 10.1016/bs.hefe.2016.09.004

Orr, L., Ferris, J. D., Jacob, R., Beecroft, E., Sanbonmatsu, L., Katz, L. F., ... Kling, J. R. (2003). *Moving to opportunity: Interim impacts evaluation* (Report). Washington D.C: U.S. Dept. of HUD.

Oster, E., & Thornton, R. (2011, January). Menstruation, Sanitary Products, and School Attendance: Evidence from a Randomized Evaluation. *American Economic Journal: Applied Economics*, *3*(1), 91–100. doi: 10.1257/app.3.1.91

Paluck, E. L., Porat, R., Clark, C. S., & Green, D. P. (2021, January). Prejudice Reduction: Progress and Challenges. *Annual Review of Psychology*, *72*(1), 533–560. doi: 10.1146/annurev-psych-071620-030619

Paradkar, N., & Chowdhury, S. R. (2015). Primary study for detection of arterial blood pressure waveform components. *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference*, *2015*, 1959–1962. doi: 10.1109/EMBC.2015.7318768

Phillips-Howard, P. A., Nyothach, E., Ter Kuile, F. O., Omoto, J., Wang, D., Zeh, C., ... Laserson, K. F. (2016, November). Menstrual cups and sanitary pads to reduce school attrition, and sexually transmitted and reproductive tract infections: A cluster randomised controlled feasibility study in rural Western Kenya. *BMJ open*, *6*(11), e013229. doi: 10.1136/bmjopen-2016-013229

Radloff, L. S. (1977, June). The CES-D Scale: A Self-Report Depression Scale for Research in the General Population. *Applied Psychological Measurement*, *1*(3), 385–401. doi: 10.1177/014662167700100306

Rammstedt, B., & Farmer, R. F. (2013). The impact of acquiescence on the evaluation of personality structure. *Psychological Assessment*, *25*(4), 1137–1145. doi: 10.1037/a0033323

Rao, G. (2019). Familiarity does not breed contempt: Generosity, discrimination, and diversity in Delhi schools. *American Economic Review*. doi: 10.1257/aer.20180044

Reece, J. D., Bunn, J. A., Choi, M., & Navalta, J. W. (2021, June). Assessing Heart Rate Using Consumer Technology Association Standards. *Technologies*, *9*(3), 46. doi: 10.3390/technologies9030046

Sarzosa, M., & Urzúa, S. (2021). Bullying among adolescents: The role of skills. *Quantitative Economics*, *12*(3), 945–980. doi: 10.3982/QE1215

Schilbach, F., Schofield, H., & Mullainathan, S. (2016, May). The Psychological Lives of the Poor. *American Economic Review*, *106*(5), 435–440. doi: 10.1257/aer.p20161101

Shah, M., Seager, J., Montalvao, J., & Goldstein, M. (2023, August). *Sex, Power, and Adolescence: Intimate Partner Violence and Sexual Behaviors* (Working Paper No. 31624). National Bureau of Economic Research. doi: 10.3386/w31624

WaterAid. (2013). *We can't wait: A report on sanitation and hygiene for women and girls.* WaterAid International. Retrieved from: http://www.zaragoza.es/contenidos/medioambiente/onu/1325-eng_We_cant_wait_sanitation_and_hygiene_for_women_and%20girls.pdf.

Webb, D. (2023). Silence to Solidarity: Using Group Dynamics to Reduce Anti-Transgender Discrimination in India. *mimeo, Paris School of Economics*.

Welz, M., Alfons, A., Demirer, M., & Chernozhukov, V. (2022). GenericML: Generic machine learning inference. *R package version*.

Wolke, D., Copeland, W. E., Angold, A., & Costello, E. J. (2013). Impact of Bullying in Childhood on Adult Health, Wealth, Crime, and Social Outcomes. *Psychological Science*, *24*(10), 1958–1970.

A Supplementary tables and figures

Table A1: *Implementation methods used by YGLs*

	Base + YGLs Informal	**Base + YGLs Formal**
Informal implementation	52%	24%
Outside the classroom	23%	9%
Without specific slots	20%	10%
Unannounced	7%	4%
Without adult help	2%	1%
Formal implementation	48%	76%
With the help of teachers	11%	18%
In a classroom	15%	26%
With specific time slots	22%	32%
N	35	35

Notes: The table shows results from the school survey. Directors were asked about the way the YGLs implemented their sensitization. Each row shows the proportion of each response. Column 1 includes the 35 *Base + YGL* schools where YGLs were asked to carry out informal discussions with their peers. Column 2 includes the 35 *Base + YGL* schools where YGLs were asked to carry out formal sensitization sessions.

Table A2: *Effects of formal vs informal YGL program*

Outcome	(1) Base + YGL (Informal)	(2) Base + YGL (Formal)	(3) *p*-val. of diff.	(4) N
Learning test (Z)	0.105*	0.129**	.7021	2167
	(0.059) [0.076]	(0.063) [0.040]		
Progressed one class (=1)	0.108***	0.044	.0488	2256
	(0.040) [0.006]	(0.038) [0.240]		
Enrolled (=1)	–0.011	–0.031	.4464	2256
	(0.026) [0.686]	(0.024) [0.208]		
Absent in last 3 months (=1)	–0.040	–0.036	.8881	1824
	(0.032) [0.206]	(0.033) [0.279]		
General health (Z)	0.015	0.003	.8308	2256
	(0.053) [0.781]	(0.058) [0.962]		
UTI health (Z)	–0.048	0.039	.1155	2256
	(0.053) [0.369]	(0.048) [0.422]		
Mental health (Z)	0.030	0.070	.509	2256
	(0.067) [0.647]	(0.068) [0.305]		
Self-esteem (Z)	0.046	–0.025	.2473	2256
	(0.059) [0.441]	(0.061) [0.685]		
Heart rate (Z)	–0.179***	–0.078	.0655	1904
	(0.056) [0.001]	(0.058) [0.180]		
Hygiene knowl. (Z)	0.407***	0.477***	.2451	2167
	(0.062) [0.000]	(0.075) [0.000]		
Menstr. knowl. (Z)	0.349***	0.330***	.6987	2167
	(0.052) [0.000]	(0.056) [0.000]		
Hygiene behavior (Z)	0.376***	0.277***	.186	2167
	(0.080) [0.000]	(0.083) [0.001]		
Menstr. behavior (Z)	0.569***	0.553***	.8143	2256
	(0.065) [0.000]	(0.062) [0.000]		
Observed hygiene: Girls (Z)	0.082	0.007	.2102	2167
	(0.064) [0.200]	(0.061) [0.913]		
Combined stigma index (Z)	0.682***	0.797***	.079	2256
	(0.062) [0.000]	(0.067) [0.000]		
Received info. (Z)	0.504***	0.571***	.3725	2256
	(0.077) [0.000]	(0.071) [0.000]		
Willingness to speak (Z)	0.614***	0.843***	.0024	2250
	(0.072) [0.000]	(0.076) [0.000]		
Attitudes (Z)	0.539***	0.595***	.4506	2254
	(0.064) [0.000]	(0.073) [0.000]		
Norms (Z)	0.431***	0.516***	.1259	2167
	(0.053) [0.000]	(0.059) [0.000]		
Shame response to vignettes (rev.) (Z)	0.436***	0.455***	.7521	2165
	(0.060) [0.000]	(0.065) [0.000]		
Observed shame (rev.) (Z)	0.243***	0.216***	.6147	2256
	(0.061) [0.000]	(0.059) [0.000]		
Bullying: teasing (Z)	0.082	–0.041	.0449	2167
	(0.059) [0.163]	(0.071) [0.560]		
Bullying: intimidation/harassment (Z)	–0.022	–0.116**	.035	2167
	(0.044) [0.612]	(0.046) [0.011]		
Network (Z)	0.275***	0.258***	.8443	2167
	(0.083) [0.001]	(0.087) [0.003]		
Official mark (Z)	0.056	0.107	.4375	14579
	(0.074) [0.452]	(0.067) [0.115]		

Notes: The table compares the effects of the *YGL* informal and formal modalities using girls' outcomes at endline. Each row represents a regression of the outcome on *Base Only*, *Base + YGL (Informal)* and *Base + YGL (Formal)*. Column (1) shows the coefficient on *Base + YGL (Informal)*; column (2) shows the coefficient on *Base + YGL (Formal)*; column (3) shows the *p*-value for the test of *Base + YGL (Informal)* = textitBase + YGL (Formal). Standard errors are clustered at the school level and are in parentheses; coefficient p-values are in brackets; controls include stratum fixed effects and the baseline controls selected by double LASSO. ** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.*

Table A3: *Baseline balance: covariates*

Variable	(1) Total Mean/(SD)	(2) Control Mean/(SD)	(3) Base + YGL Mean/(SD)	(4) Base Only Mean/(SD)	(3)-(2) Pairwise t-test P-value	(4)-(2) Pairwise t-test P-value
Girl's age (years)	14.000	13.741	14.104	14.033	0.292	0.483
	(6.798)	(6.904)	(6.535)	(7.356)		
Girl already had her period (=1)	0.471	0.447	0.483	0.468	0.606	0.807
	(1.387)	(1.392)	(1.340)	(1.509)		
Girl has a brother (=1)	0.548	0.574	0.556	0.508	0.512	0.033**
	(0.548)	(0.531)	(0.567)	(0.511)		
Distance to school (km)	2.320	2.260	2.379	2.259	0.518	0.996
	(3.521)	(3.546)	(3.732)	(3.094)		
Time to school (min)	30.167	29.722	30.755	29.404	0.637	0.896
	(40.650)	(44.587)	(40.191)	(38.511)		
Was absent last 3 months (=1)	0.552	0.546	0.573	0.518	0.497	0.517
	(0.798)	(0.739)	(0.858)	(0.718)		
Days absent last 3 months	1.412	1.391	1.322	1.613	0.671	0.289
	(3.633)	(2.988)	(3.637)	(4.106)		
Father passed away (=1)	0.085	0.077	0.083	0.096	0.687	0.256
	(0.319)	(0.225)	(0.357)	(0.319)		
Father in household (=1)	0.696	0.711	0.696	0.682	0.592	0.401
	(0.604)	(0.556)	(0.605)	(0.657)		
Mother passed away (=1)	0.049	0.040	0.050	0.053	0.331	0.282
	(0.210)	(0.183)	(0.223)	(0.209)		
Mother in household (=1)	0.816	0.835	0.805	0.820	0.208	0.595
	(0.450)	(0.469)	(0.439)	(0.459)		
Guardian knows how to read and write (=1)	0.828	0.829	0.836	0.810	0.799	0.474
	(0.485)	(0.479)	(0.514)	(0.428)		
Guardian has no education (=1)	0.095	0.085	0.096	0.102	0.465	0.310
	(0.365)	(0.255)	(0.418)	(0.342)		
Guardian attended secondary or higher education (=1)	0.303	0.308	0.293	0.318	0.648	0.794
	(0.648)	(0.681)	(0.623)	(0.685)		
Guardian occupation: Agriculture (=1)	0.773	0.768	0.781	0.762	0.696	0.899
	(0.666)	(0.716)	(0.632)	(0.697)		
Observations	2390	568	1216	606		
Schools	140	35	70	35		

Notes: Sample includes all girls in baseline. Columns (1)-(4) show means and standard deviations of covariates from the girls' baseline survey. Columns (5)-(6) show the p-value of a pairwise test comparing *Base Only* and *Base + YGL* with *control* , respectively. Standard errors cluster at the school level. *$* p < 0.1$, **$ p < 0.05$, ***$ p < 0.01$.

Table A4: *Baseline balance: outcome indexes*

Variable	(1) Total Mean/(SD)	(2) Control Mean/(SD)	(3) Base + YGL Mean/(SD)	(4) Base Only Mean/(SD)	(3)-(2) Pairwise t-test P-value	(4)-(2) Pairwise t-test P-value
Wealth index (Z)	0.000	–0.136	0.083	–0.039	0.208	0.632
	(3.487)	(3.397)	(3.480)	(3.578)		
Years of school delay	2.224	2.160	2.233	2.264	0.628	0.598
	(3.278)	(2.881)	(3.191)	(3.830)		
WASH knowledge index (Z)	0.093	0.000	0.107	0.151	0.317	0.206
	(1.988)	(2.178)	(1.960)	(1.883)		
WASH behaviour index (Z)	0.008	–0.000	–0.011	0.054	0.906	0.606
	(1.678)	(1.901)	(1.596)	(1.658)		
Menstruation knowledge index (Z)	0.047	–0.000	0.099	–0.014	0.504	0.937
	(2.999)	(2.890)	(2.978)	(3.195)		
Menstruation behaviour index (Z)	0.070	–0.000	0.094	0.086	0.475	0.591
	(2.700)	(2.555)	(2.677)	(2.937)		
Mental health index (Z)	0.020	–0.000	0.061	–0.042	0.412	0.604
	(1.373)	(1.488)	(1.383)	(1.234)		
Menstrual health problems index (Z)	0.018	0.000	0.042	–0.013	0.699	0.919
	(2.110)	(2.187)	(2.055)	(2.203)		
Menstruation inhibiting school activity index (Z)	0.032	0.000	0.042	0.041	0.316	0.330
	(0.489)	(1.006)	(0.021)	(0.017)		
Received info. on hygiene (Z)	0.037	–0.000	0.055	0.036	0.455	0.687
	(1.311)	(1.604)	(1.107)	(1.418)		
Received info. on menstruation (Z)	0.021	–0.000	0.066	–0.049	0.520	0.694
	(2.096)	(1.999)	(2.077)	(2.243)		
Willingness to speak about hygiene (Z)	–0.006	–0.000	–0.011	–0.001	0.899	0.989
	(1.701)	(1.752)	(1.776)	(1.540)		
Willingness to speak about menstruation (Z)	0.040	–0.000	0.058	0.043	0.415	0.648
	(1.571)	(1.344)	(1.527)	(1.857)		
Exposure to bullying (self-reported) (Z)	0.032	0.000	0.021	0.082	0.757	0.357
	(1.491)	(1.308)	(1.453)	(1.729)		
Attitudes towards menstruation (Z)	–0.072	0.000	–0.134	–0.014	0.213	0.907
	(2.057)	(2.163)	(1.997)	(2.078)		
Shame response to vignettes (Z)	0.020	–0.000	0.015	0.048	0.880	0.692
	(2.096)	(1.898)	(2.107)	(2.292)		
Perceived norms related to menstruation (Z)	0.076	–0.000	0.089	0.120	0.277	0.222
	(1.748)	(1.516)	(1.813)	(1.841)		
Social desirability index (Z)	0.006	0.000	0.037	–0.051	0.538	0.456
	(1.165)	(1.218)	(1.166)	(1.105)		
Observed hygiene: home (Z)	–0.024	0.000	–0.065	0.035	0.407	0.721
	(1.606)	(1.571)	(1.515)	(1.789)		
Observations	2390	568	1216	606		
Schools	140	35	70	35		

Notes: Sample includes all girls in baseline. Columns (1)-(4) show means and standard deviations of indexes calculated using the girls' baseline survey. Columns (5)-(6) show the p-value of a pairwise test comparing *Base Only* and *Base + YGL* with *control*, respectively. Standard errors cluster at the school level. *$* p < 0.1$, **$ p < 0.05$, ***$ p < 0.01$.

Table A5: *Probability of attrition*

VARIABLES	(1) In Endline (=1)	(2) Endline in person (=1)	(3) In Midline and Endline (=1)	(4) Mother in Endline (=1)	(5) Heart rate available (=1)	(6) Class with marks (=1)	(7) Class with teacher int. (=1)	(8) Boy from main sample list (=1)	(9) Girl from main sample list (=1)	(10) P(register available)
Base Only	-0.028	-0.045**	-0.059**	-0.018	-0.048	-0.093**	-0.019	0.034	-0.021	-0.007
	(0.019)	(0.020)	(0.029)	(0.012)	(0.030)	(0.046)	(0.048)	(0.022)	(0.025)	(0.013)
Base + YGL	0.001	-0.020	-0.044*	-0.013	-0.026	-0.028	-0.004	-0.016	-0.012	0.006
	(0.016)	(0.017)	(0.024)	(0.011)	(0.026)	(0.034)	(0.040)	(0.023)	(0.019)	(0.011)
Observations	2,390	2,390	2,390	2,390	2,390	500	500	1,325	1,627	3,626
Data source	Girls	Girls	Girls	Girls	Heart Rate	Teachers	Teachers	Boys	Lab in field	Registers
Control mean	.951	.928	.856	.963	.802	.919	.887	.837	.853	.901
p: Treated = 0	.5705	.0789	.0383	.1835	.1783	.1487	.8183	.9634	.4041	.8938
p: Base Only = Base + YGL	.0902	.1579	.5183	.5264	.3869	.1012	.7259	.0119	.7112	.2559

Notes: The table displays the probability of each observation from the sampling frame being in the data. *Outcomes:* Column (1): a girl from the baseline sample was interviewed at endline. Column (2): girl from the baseline sample was interviewed in person (rather than through phone). Column (3): girl was interviewed both at midline and endline. Column (4): the girl's mother or caretaker from the baseline sample was also interviewed at endline. Column (5): heart rate data was available for the girl's interview. Columns (6)-(7): we aimed to digitize the official school marks and interview teachers from a sample of 500 classes that included all target grade-levels (from 3rd grade to 12th grade), randomly selecting one class per grade-level if there were multiple. Column (8) shows probability of collecting school marks for each class; column (9) shows probability of interviewing the teacher. Column (10): the proportion of boys interviewed from the main sampling frame of 10 randomly selected boys based on the school registers (rather than a randomly selected replacement list). Column (11): for the lab in the field, the proportion of girls interviewed based on a sampling frame of 12 randomly selected girls based on school registers (instead of a randomly selected replacement list). Column (12): at the *class* × *wave* level, the proportion of measured days for which official absenteeism data was available. Standard errors are clustered at the school level and are in parentheses; controls include stratum fixed effects.

Table A6: *Multiple hypothesis testing adjustment*

Outcome	Pooled coeff.	p-val.	q-val.
Learning test (Z)	0.147	0.004	0.009
# girls present (spot-checks)	-0.121	0.332	0.200
General health (Z)	-0.006	0.902	0.565
Heart rate (bpm)	-2.341	0.014	0.015
Stigma index (Z)	0.628	<0.001	<0.001

Notes: *q*-values are calculated using the Anderson (2008) sharpened process that controls for the false discovery rate across the 5 outcome variables shown. All specifications follow the corresponding specification in the main results, but with a pooled treatment that includes both *Base only* and *Base+YGL* schools. All models cluster at the school level. All models control for baseline analogues of the outcome variable, and all models apart from the spot checks (row 2) use LASSO-selected controls. The spot checks model controls for wave fixed effects; the heart rate model controls for time-window fixed effects and includes all observations less than 120 minutes after the start of the survey.

Table A7: *Are effects driven by participants with high social desirability bias?*

Outcome	Base Only	Base + YGL	High SDB	Base × High SDB	(Base+YGL) × High SDB	N
Learning test (Z)	0.207***	0.168***	0.055	0.003	–0.092	2167
	(0.077) [0.007]	(0.063) [0.008]	(0.093) [0.552]	(0.090) [0.978]	(0.088) [0.295]	
Progressed one class (=1)	0.130***	0.105**	–0.005	–0.036	–0.058	2256
	(0.049) [0.008]	(0.048) [0.030]	(0.057) [0.933]	(0.058) [0.533]	(0.056) [0.300]	
Enrolled (=1)	0.000	–0.016	–0.003	–0.038	–0.011	2256
	(0.034) [0.998]	(0.034) [0.644]	(0.049) [0.947]	(0.050) [0.454]	(0.046) [0.821]	
Absent in last 3 months (=1)	0.065	–0.061	0.036	–0.155**	0.036	1824
	(0.043) [0.135]	(0.039) [0.122]	(0.062) [0.565]	(0.062) [0.012]	(0.059) [0.540]	
General health (Z)	0.001	–0.007	0.030	–0.077	0.034	2256
	(0.080) [0.989]	(0.076) [0.927]	(0.122) [0.808]	(0.144) [0.593]	(0.120) [0.776]	
UTI health (Z)	–0.119	0.014	0.088	0.055	–0.039	2256
	(0.086) [0.169]	(0.061) [0.819]	(0.097) [0.364]	(0.112) [0.622]	(0.085) [0.648]	
Mental health (Z)	0.102	0.065	0.128	–0.183	–0.029	2256
	(0.073) [0.164]	(0.069) [0.350]	(0.107) [0.233]	(0.122) [0.132]	(0.111) [0.791]	
Self-esteem (Z)	–0.004	–0.040	0.042	0.048	0.104	2256
	(0.088) [0.961]	(0.068) [0.554]	(0.094) [0.658]	(0.109) [0.655]	(0.096) [0.275]	
Heart rate (Z)	–0.091	–0.150*	–0.031	–0.081	0.044	1904
	(0.099) [0.358]	(0.080) [0.059]	(0.120) [0.797]	(0.147) [0.583]	(0.116) [0.705]	
Hygiene knowl. (Z)	0.376***	0.469***	–0.104	–0.061	–0.047	2167
	(0.085) [0.000]	(0.071) [0.000]	(0.099) [0.291]	(0.102) [0.550]	(0.089) [0.594]	
Menstr. knowl. (Z)	0.328***	0.366***	–0.001	–0.035	–0.047	2167
	(0.067) [0.000]	(0.060) [0.000]	(0.086) [0.994]	(0.095) [0.709]	(0.080) [0.562]	
Hygiene behavior (Z)	0.142	0.347***	0.065	–0.064	–0.044	2167
	(0.093) [0.130]	(0.085) [0.000]	(0.120) [0.591]	(0.120) [0.595]	(0.115) [0.699]	
Menstr. behavior (Z)	0.280***	0.605***	–0.096	0.165	–0.084	2256
	(0.084) [0.001]	(0.062) [0.000]	(0.085) [0.258]	(0.103) [0.111]	(0.083) [0.307]	
Combined stigma index (Z)	0.468***	0.810***	–0.003	–0.123	–0.128	2256
	(0.075) [0.000]	(0.063) [0.000]	(0.081) [0.974]	(0.082) [0.135]	(0.081) [0.113]	
Received info. (Z)	0.307***	0.618***	0.098	–0.212	–0.157	2256
	(0.103) [0.003]	(0.080) [0.000]	(0.118) [0.406]	(0.146) [0.146]	(0.109) [0.150]	
Willingness to speak (Z)	0.304***	0.765***	–0.083	–0.105	–0.065	2250
	(0.092) [0.001]	(0.076) [0.000]	(0.113) [0.459]	(0.110) [0.339]	(0.106) [0.544]	
Attitudes (Z)	0.412***	0.613***	–0.048	–0.086	–0.079	2254
	(0.075) [0.000]	(0.061) [0.000]	(0.080) [0.548]	(0.080) [0.282]	(0.080) [0.325]	
Norms (Z)	0.249***	0.475***	–0.055	0.069	0.010	2167
	(0.065) [0.000]	(0.054) [0.000]	(0.100) [0.583]	(0.102) [0.502]	(0.091) [0.909]	
Shame response to vignettes (rev.) (Z)	0.353***	0.510***	0.088	–0.096	–0.124	2165
	(0.072) [0.000]	(0.065) [0.000]	(0.086) [0.305]	(0.097) [0.323]	(0.082) [0.129]	
Bullying: teasing (Z)	0.073	0.062	0.037	0.067	–0.093	2167
	(0.085) [0.391]	(0.079) [0.436]	(0.100) [0.709]	(0.108) [0.535]	(0.099) [0.347]	
Bullying: intimidation/harassment (Z)	–0.093	–0.063	–0.076	0.028	–0.017	2167
	(0.077) [0.223]	(0.065) [0.329]	(0.107) [0.477]	(0.123) [0.820]	(0.103) [0.867]	
Network (Z)	0.238**	0.340***	0.055	–0.129	–0.139	2167
	(0.113) [0.034]	(0.096) [0.000]	(0.121) [0.650]	(0.129) [0.315]	(0.102) [0.173]	

Notes: Each line is one regression, and each column is a coefficient from that regression *High SDB* = 1 when a girl's Crowne-Marlowe social desirability score elicited at baseline is above the median, and is 0 otherwise. Controls include stratum fixed effects and the baseline controls selected by double LASSO. Standard errors are clustered at the school level and are in parentheses; *p*-values are in brackets. Only self-reported outcomes from the girls' survey are included.

Table A8: *Effects on each individual learning test*

VARIABLES	(1) Math fluency	(2) Reading fluency	(3) Listening	(4) Comprehension	(5) Digit span	(6) Reverse digit span
Base Only	0.254***	0.159***	0.110	0.080	0.114*	0.048
	(0.082)	(0.049)	(0.080)	(0.076)	(0.062)	(0.068)
Base + YGL	0.150**	0.073*	0.044	0.041	0.033	0.107**
	(0.062)	(0.041)	(0.062)	(0.058)	(0.051)	(0.050)
Observations	2,167	2,167	2,167	2,167	2,167	2,167
Data source	Girls	Girls	Girls	Girls	Girls	Girls
p: Treated = 0	.0036	.0106	.2596	.3388	.2386	.0575
p: Base Only = Base + YGL	.1093	.0372	.3621	.5816	.1534	.3625

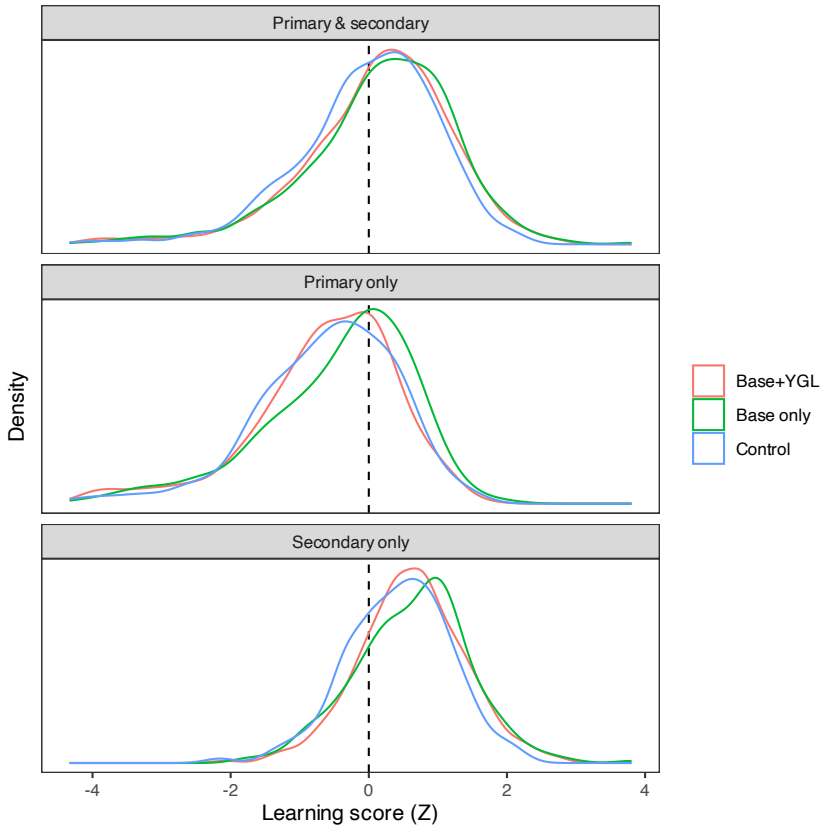
Notes: * p< 0.1, ** p < 0.05, *** p < 0.01. Standard errors are clustered at the school level and are in parentheses. Sample includes all girls in endline interviewed in person. Controls include stratum fixed effects and the baseline controls selected by double LASSO. The outcomes are the individual components of the learning tests implemented during the girls' endline survey. All scores are in control group standard deviations.

Table A9: *Randomization inference p-values*

	Base+YGL	Base Only	Pooled
Learning test (Z)	.043	.023	.015
Progressed one class (=1)	.050	.030	.015
Enrolled (=1)	.418	.517	.405
Absent in last 3 months (=1)	.268	.790	.364
General health (Z)	.878	.619	.911
UTI health (Z)	.940	.123	.524
Mental health (Z)	.470	.871	.550
Heart rate (Z)	.030	.067	.011
Combined stigma index (Z)	<0.001	<0.001	<0.001
Self-esteem (Z)	.857	.855	.825
Hygiene knowl. (Z)	<0.001	.003	<0.001
Menstr. knowl. (Z)	<0.001	<0.001	<0.001
Hygiene behavior (Z)	<0.001	.275	.003
Menstr. behavior (Z)	<0.001	<0.001	<0.001
Pregnant now or last year (=1)	.309	.024	.158

Notes: Table displays the *p*-values using randomization inference from 1000 replications. There is one row per outcome. Columns (1) and (2) result from a regression of the outcome on *Base Only* and *Base+YGL*. Column (3) results from a regression of the outcome on a pooled treatment variable. All regressions control for stratum fixed effects and LASSO controls.

Figure A10: *Distribution of learning test scores at endline*



Notes: Density plot of learning test scores (academic achievement and cognitive tests combined) in each treatment arm. First panel includes both primary and secondary, second panel is only primary schools, and third panel is only secondary schools.

Table A11: *Effects on girls with attrition weights*

	(1) Base Only	(2) Base + YGL	(3) p: Treated=0	(4) p:Base Only=Base+YGL	(5) N
Learning test (Z)	0.213***	0.132**	.0022	.2155	2167
	(0.072) [0.003]	(0.053) [0.013]			
Progressed one class (=1)	0.102***	0.064*	.0249	.1678	2256
	(0.039) [0.008]	(0.035) [0.068]			
Enrolled (=1)	–0.026	–0.028	.2222	.9406	2256
	(0.024) [0.284]	(0.023) [0.235]			
Absent in last 3 months (=1)	0.000	–0.030	.4464	.2234	1824
	(0.030) [0.990]	(0.028) [0.286]			
General health (Z)	–0.058	–0.010	.5919	.3799	2256
	(0.060) [0.334]	(0.048) [0.843]			
UTI health (Z)	–0.098*	–0.007	.3538	.0722	2256
	(0.053) [0.068]	(0.044) [0.872]			
Mental health (Z)	0.005	0.059	.507	.2756	2256
	(0.064) [0.933]	(0.061) [0.334]			
Self-esteem (Z)	–0.020	–0.008	.8405	.8494	2256
	(0.068) [0.771]	(0.052) [0.872]			
Heart rate (Z)	–0.107*	–0.110**	.0214	.9511	1904
	(0.057) [0.061]	(0.050) [0.029]			
Hygiene knowl. (Z)	0.332***	0.444***	<0.001	.0741	2167
	(0.077) [0.000]	(0.063) [0.000]			
Menstr. knowl. (Z)	0.311***	0.334***	<0.001	.6032	2167
	(0.056) [0.000]	(0.048) [0.000]			
Hygiene behavior (Z)	0.105	0.315***	<0.001	.003	2167
	(0.084) [0.211]	(0.070) [0.000]			
Menstr. behavior (Z)	0.359***	0.551***	<0.001	.0041	2256
	(0.070) [0.000]	(0.053) [0.000]			
Observed hygiene: Girls (Z)	–0.062	0.053	.8544	.0314	2167
	(0.064) [0.336]	(0.056) [0.343]			
Combined stigma index (Z)	0.411***	0.746***	<0.001	<0.001	2256
	(0.066) [0.000]	(0.058) [0.000]			
Received info. (Z)	0.192**	0.524***	<0.001	<0.001	2256
	(0.078) [0.013]	(0.063) [0.000]			
Willingness to speak (Z)	0.255***	0.746***	<0.001	<0.001	2250
	(0.076) [0.001]	(0.065) [0.000]			
Attitudes (Z)	0.354***	0.564***	<0.001	<0.001	2254
	(0.065) [0.000]	(0.061) [0.000]			
Norms (Z)	0.277***	0.464***	<0.001	<0.001	2167
	(0.061) [0.000]	(0.051) [0.000]			
Shame response to vignettes (rev.) (Z)	0.296***	0.425***	<0.001	.0165	2165
	(0.059) [0.000]	(0.055) [0.000]			
Observed shame (rev.) (Z)	0.278***	0.231***	<0.001	.3591	2256
	(0.063) [0.000]	(0.056) [0.000]			
Bullying: teasing (Z)	0.110*	0.024	.3486	.1142	2167
	(0.065) [0.091]	(0.058) [0.672]			
Bullying: intimidation/harassment (Z)	–0.059	–0.060	.1449	.9922	2167
	(0.049) [0.226]	(0.041) [0.150]			
Network (Z)	0.179**	0.269***	<0.001	.1974	2167
	(0.085) [0.036]	(0.074) [0.000]			

Notes: The table displays results for Base Only (1) and Base+YGL (2) from the girls survey in endline. The estimations are adjusted by attrition weights; standard errors are clustered at the school level and are in parentheses; coefficient p-values are in brackets; controls include stratum fixed effects and controls selected by LASSO. Columns (3) and (4) denotes the p-values on the test of being treated *Treated=0* and the equality of the coefficients *Base Only = Base + YGL*. Column (5) shows the number of observations. *$* p < 0.1$, **$p < 0.05$, ***$p < 0.01$.

Table A12: *Effects on school marks and teacher outcomes with attrition weights*

	(1) Base Only	(2) Base + YGL	(3) p: Treated=0	(4) p:Base Only=Base+YGL	(5) N
Grades:					
Official mark (Z)	0.125*	0.085	.0965	.4915	14579
	(0.069) [0.072]	(0.061) [0.165]			
Official mark (Z): girls	0.168**	0.084	.0653	.2077	7586
	(0.077) [0.031]	(0.060) [0.167]			
Official mark (Z): boys	0.081	0.092	.1695	.8626	6993
	(0.072) [0.264]	(0.069) [0.188]			
Teachers:					
Positive Menstruation Beliefs (Z)	0.694***	0.685***	<0.001	.8881	743
	(0.090) [0.000]	(0.087) [0.000]			
Mens. Knowledge Index (Z)	0.166	0.228**	.0294	.5859	743
	(0.123) [0.179]	(0.101) [0.026]			
Positive Gender Views (Z)	0.108	0.134	.2521	.8076	779
	(0.128) [0.401]	(0.117) [0.255]			
Absent Last Month (=1)	–0.101*	–0.036	.2127	.2254	743
	(0.056) [0.075]	(0.048) [0.463]			

Notes: The table displays results for Base Only (1) and Base+YGL (2) from the teacher survey in endline. The estimations are adjusted by attrition weights; standard errors are clustered at the school level and are in parentheses; coefficient p-values are in brackets; controls include stratum fixed effects. Columns (3) and (4) denotes the p-values on the test of being treated *Treated=0* and the equality of the coefficients *Base Only = Base + YGL*. Column (5) shows the number of observations. *$* p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.*

Table A13: *Effects on attendance for boys*

VARIABLES	(1) # boys present (spot-checks)	(2) # boys present (registers)
Base Only	-2.444	1.906
	(4.002)	(3.103)
Base + YGL	-1.746	-0.057
	(3.057)	(2.615)
Observations	660	34,455
Data source	Spot checks	Spot checks
Control mean	52.5	57.3
Control SD	42.5	78.7
p: Treated = 0	.513	.8138
p: Base Only = Base + YGL	.8338	.5323

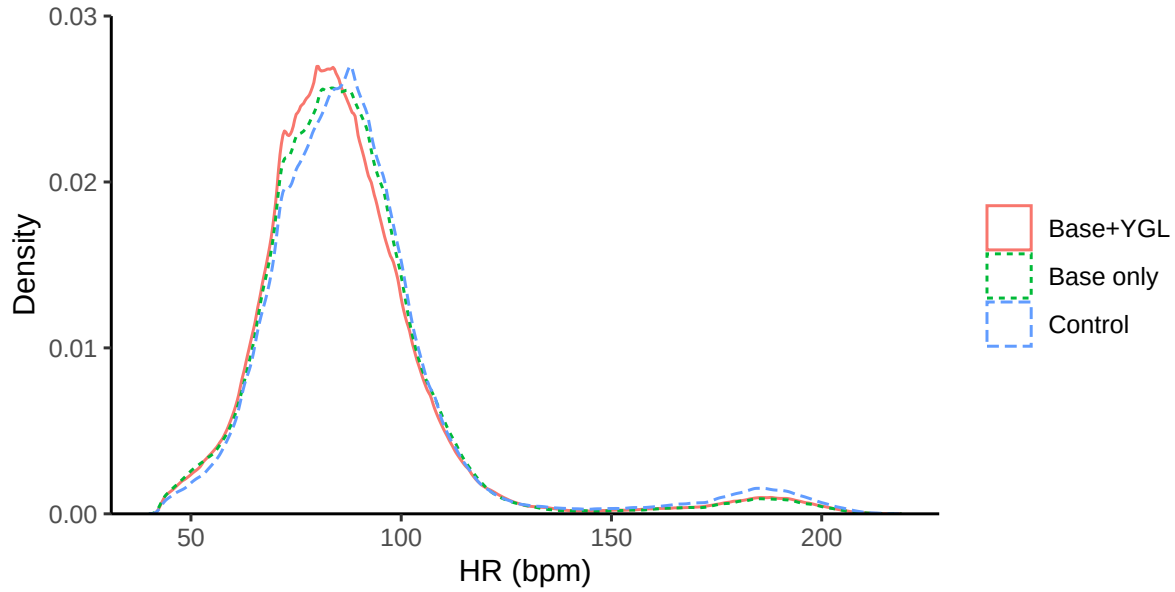
Notes: Column 1: outcome is the number of boys physically present during the spot checks. Observations are at the school × wave level. Specification controls for spot check wave fixed effects. Column 2: outcome is the number of boys present on a given day according to official school registers. Observations are at the school × day level. Specification controls for day fixed effects. Both columns control for stratum fixed effects and cluster standard errors at the school level.

Table A14: *Mental health (depression) at midline and endline*

VARIABLES	(1) Midline Index Factor Analysis (Z)	(2) Midline Index Raw Sum (Z)	(3) Midline Positive Emotions (Z)	(4) Midline Negative Emotions (Z)	(5) Endline Index Factor Analysis (Z)	(6) Endline Index Raw Sum (Z)	(7) Endline Positive Emotions (Z)	(8) Endline Negative Emotions (Z)
Base Only	-0.039	-0.022	0.035	-0.018	-0.025	0.012	0.060	0.101
	(0.045)	(0.052)	(0.049)	(0.068)	(0.067)	(0.061)	(0.073)	(0.064)
Base + YGL	-0.003	0.062	0.051	0.140**	-0.028	0.004	0.052	0.074
	(0.041)	(0.047)	(0.041)	(0.057)	(0.063)	(0.058)	(0.066)	(0.054)
Observations	1,963	1,963	1,963	1,963	1,964	1,964	1,964	1,964
Data source	Girls	Girls	Girls	Girls	Girls	Girls	Girls	Girls
p: Treated = 0	.6995	.4181	.2336	.1006	.6492	.8998	.3847	.107
p: Base Only = Base + YGL	.3904	.0834	.7313	.0132	.9586	.8705	.8885	.624

Notes: The table shows the disaggregated results on mental health at midline and endline. Only girls who were interviewed at *both* midline and endline are included. All outcomes are normalized by the mean and standard deviation of the control group at endline. Only questions that were included in both midline and endline are included. *Index: Factor Analysis* is an index constructed using factor-analysis-based weightings on both "positive" and "negative" emotion questions, and correcting for acquiescence bias. *Index: Raw Sum* simply sums the values of the positive and (reverse-coded) negative emotion questions. For both of these measures, a higher index indicates improved mental health. *Positive emotions* is a sum of the value of only the positive emotions. *Negative emotions* is a sum of the value of only negative emotions (a higher score indicating more frequent negative emotions).

Figure A15: *Distribution of heart rate at the girl × second level*



Notes: Shows the distribution of heart rate observations in each treatment group, measured in beats per minute. Unit of observation is the *girl* × *second*. Sample includes all girls with heart rate data at endline.

Table A16: *Robustness of effects on heart rate to dealing with outliers*

	HR (bpm)		
	Individual-level	Drop outliers	Divide outliers
Base only	−2.212*	−1.007	−1.068
	(1.169)	(0.800)	(0.789)
Base+YGL	−2.261**	−1.202*	−1.262*
	(1.018)	(0.695)	(0.690)
Num.Obs.	1904	8 877 622	9 105 420
N girls	1904	1904	1904
Control mean	87.9	85	85.3
Control SD (between)	19.0	14.6	12.2
p : Treat = 0	0.021	0.087	0.069
p : Base=Base+YGL	0.961	0.77	0.768

Notes: Outcomes are heart rate measured in beats per minute. Column (1): Observations are amalgamated to the *girl* level by taking the mean heart rate throughout the first 120 minutes of the endline survey. Column (2): observations are at the *girl* × *second* level, with observations with heart rate above 175 bpm are dropped from the sample. Column (3): observations are at the *girl* × *second* level, with observations with heart rate above 175 bpm are divided by 2 (based on the assumption that the heart rate monitor may be erroneously double counting both the systolic and diastolic peaks within a single heartbeat). This division creates a unimodal distribution of heart rate observations. Observations are at the girl × second level. * p< 0.1, ** p < 0.05, *** p < 0.01. Standard errors are clustered at the school level and are in parentheses. Sample includes all girls in endline interviewed with heart rate data. Controls include stratum fixed effects and the baseline controls selected by double LASSO. Columns 2 and 3 also include 30-second time-window fixed effects.

Table A17: *Heart rate: dynamics*

	(1)	(2)	(3)	(4)
Treated (=1)	−1.705	−1.720	−1.372	−1.374
	(1.094)	(1.093)	(1.103)	(1.103)
After first question about menstr. (=1)	−1.219*	1.520*		
	(0.667)	(0.893)		
Treated * After first question about menstr.	−0.026	−0.016		
	(0.758)	(0.772)		
Time since start of survey (m)			−0.038*	
			(0.020)	
Treated * Time since start of survey			−0.012	−0.077***
			(0.022)	(0.022)
FEs: 30-second window		X		X
Num. obs.	6 678 438	6 678 438	6 678 438	6 678 438

Notes: All specifications use stratum fixed effects and LASSO-selected controls. Standard errors are clustered at the school level. Observations are at the second × girl level. Only observations in the first 60 minutes of the survey are included, since girls' interviews typically last between 60 and 90 minutes and so the number of observations begins to drop rapidly and become a selected sample. *Treated*=1 when a girl is in a *Base only* or *Base+YGL* school, and 0 otherwise. Columns (1) and (2) examine the role of the first question about menstruation ("Have you ever had your period?"). Column (1) does not control for 30-second time-window fixed effects, but column (2) does. Columns (3) and (4) examine a linear time trend (measured in minutes). Column (4) also controls for time-window fixed effects. Since *Time since start of survey* is perfectly correlated with the time-window fixed effects, it is not included in the model in column (4).

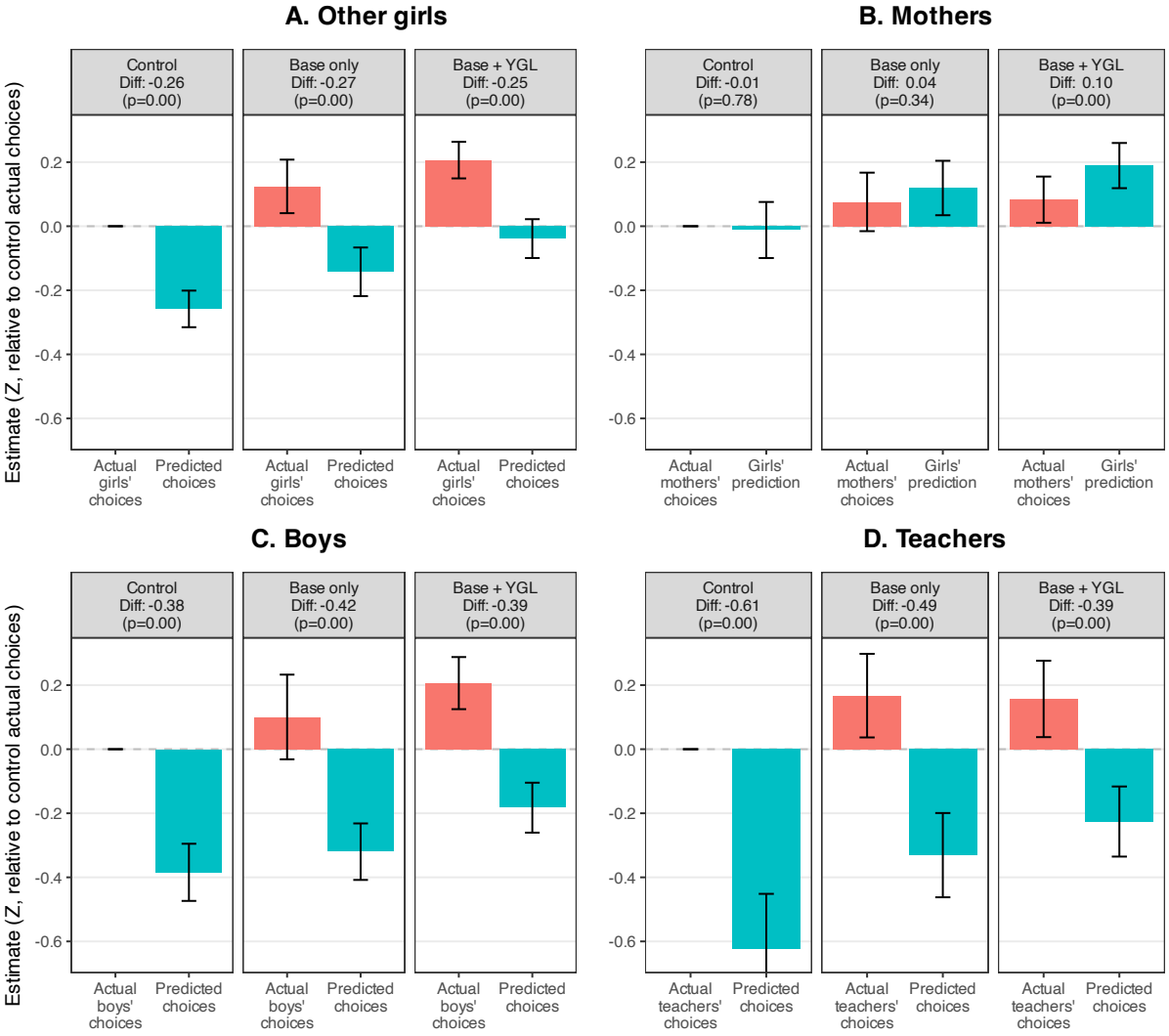
Table A18: *Effects on mothers' knowledge, behavior, and perceptions of girls' behavior*

VARIABLES	(1) Hygiene knowl. (Z)	(2) Menstr. knowl. (Z)	(3) Hygiene behavior (Z)	(4) Menstr. behavior (Z)	(5) Attitudes (Z)	(6) Girls' behavior (Z)
Base Only	0.010	0.093	-0.104	0.028	0.333***	-0.143*
	(0.085)	(0.073)	(0.076)	(0.061)	(0.065)	(0.077)
Base + YGL	0.029	0.220***	-0.046	0.007	0.350***	0.017
	(0.073)	(0.059)	(0.064)	(0.057)	(0.055)	(0.061)
Observations	2,287	2,286	2,287	2,287	2,287	2,260
Data source	Mothers	Mothers	Mothers	Mothers	Mothers	Mothers
p: Treated = 0	.7483	.0018	.2798	.8039	<0.001	.5778
p: Base Only = Base + YGL	.7853	.0463	.3866	.6174	.7849	.0176

Notes: The table displays knowledge and behavior index results from the mother survey in endline. Standard errors are clustered at the school level and are in parentheses; controls include stratum fixed effects and the baseline controls selected by double LASSO. The bottom of the table denotes the p-values on the test of being treated *Treated=0* and the equality of the coefficients *Base Only = Base + YGL*. Other p-values are defined analogously. ** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.*

Hygiene knowledge index combines questions about WASH knowledge related to the program. *Menstruation knowledge index* is constructed from questions that tests mothers' knowledge about menstruation. *Hygiene behavior index* compiles questions about hygiene behavior during the last week. *Menstruation behavior index* includes questions about mothers' behavior during the last period, such as the use of serviettes. *Attitudes* includes questions about attitudes towards menstruation, where a higher value indicates more progressive attitudes. *Girls' behavior* includes mothers' reports about the psychosocial behavior of their daughters.

Figure A19: *Misperceptions about others' attitudes*



Notes: Panel **A** shows predictions about other girls at school. Panel **B** shows girls' predictions about their mother. Panel **C** shows girls' predictions about boys at school. Panel **D** shows predictions about teachers at school. All outcomes are always normalized to the actual choices in the control group. Pooled treatment effect of [*Base only* and *Base + YGL*] on the average *misperception* regarding teachers' attitudes (predicted - actual) is 0.20 SD (p=0.03). All other treatment effects on average misperceptions are not significant at the 10% level.

Table A20: *Dictator game: girls*

VARIABLES	(1) Prop. cookies shared w/ handwashing girl	(2) Prop. cookies shared w/ menstruation girl
Base Only	-0.005	0.019
	(0.014)	(0.013)
Base + YGL	0.019	0.009
	(0.013)	(0.012)
Observations	2,161	2,161
Data source	Girls	Girls
Control mean	.42	.422
Control SD	.221	.206
p: Treated = 0	.3435	.2569
p: Base Only = Base + YGL	.0444	.3966

Notes: This shows the results of the girls' dictator game. In all regressions, standard errors are clustered at the school level and are in parentheses, and controls include stratum fixed effects. The bottom of the table denotes the p-values on the test of being treated *Treated=0* and the equality of the coefficients *Base Only = Base + YGL*. Other p-values are defined analogously. ** $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$*. Columns (4)-(5) show results from a dictator game implemented during the girls' survey. Respondents had to decide how many cookies to share with another girl after watching a video of that girl providing a menstruation- or hand-washing-related explanation. The outcomes in these columns represent the proportion of cookies (out of 5) that respondents agreed to share with the girl in the video. Controls include order of videos fixed effects.

Table A21: *Effects on bullying*

VARIABLES	(1) Self-report Combined (Z)	(2) Self-report Teasing only (Z)	(3) Self-report Intimidation/ harassment only (Z)	(4) Peer-report Combined (Z)	(5) Peer-report Teasing only (Z)	(6) Peer-report Intimidation/ harassment only (Z)	(7) Teacher-report Combined (Z)	(8) Teacher-report Teasing only (Z)	(9) Teacher-report Intimidation/ harassment only (Z)
Base Only	0.029	0.108	-0.083*	-0.074	-0.029	-0.081*	-0.067	-0.023	-0.074
	(0.059)	(0.066)	(0.046)	(0.052)	(0.061)	(0.046)	(0.056)	(0.067)	(0.049)
Base + YGL	-0.026	0.020	-0.069*	-0.061	-0.006	-0.070	-0.028	0.032	-0.052
	(0.053)	(0.058)	(0.040)	(0.050)	(0.059)	(0.044)	(0.052)	(0.061)	(0.047)
Observations	2,167	2,167	2,167	3,042	3,042	3,042	3,637	3,637	3,637
Data source	Girls	Girls	Girls	Girls	Girls	Girls	Teachers	Teachers	Teachers
p: Treated = 0	.8933	.3858	.0624	.1459	.7864	.0652	.4343	.7762	.1917
p: Base Only = Base + YGL	.2927	.1093	.7349	.7788	.6944	.7869	.3908	.3266	.5793

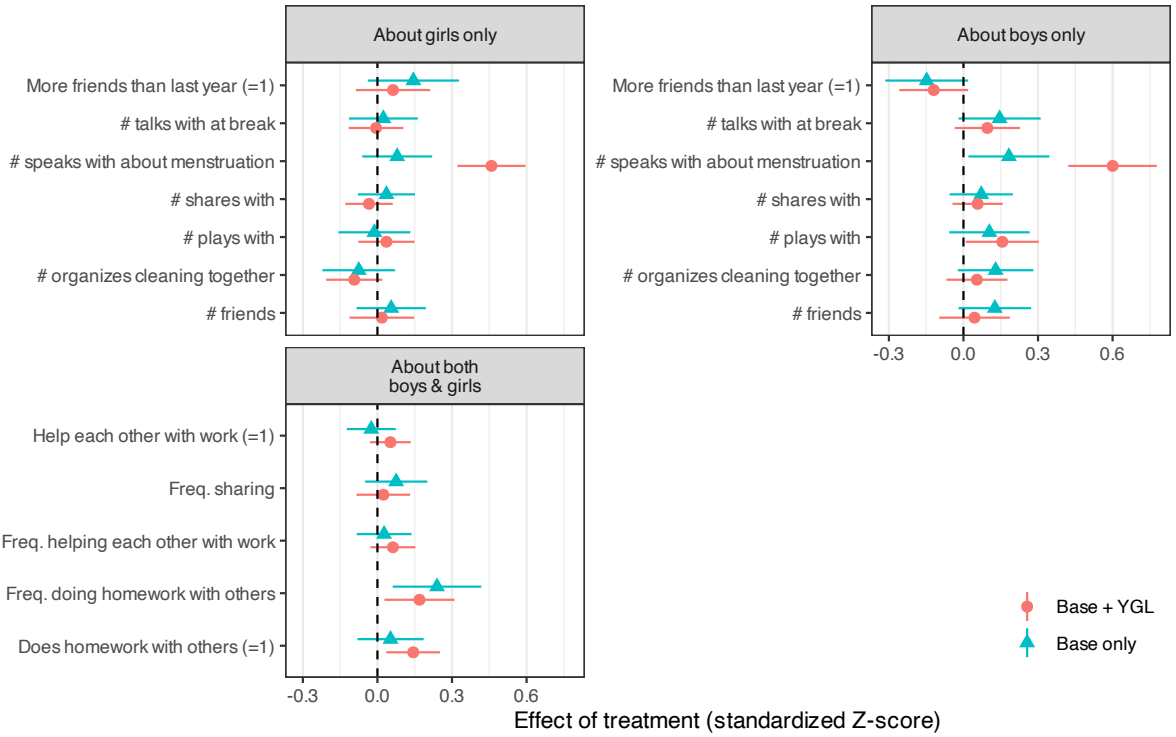
Notes: Table shows girls' and teachers' reports of bullying. Columns (1)-(3) are girls' reports about bullying towards themselves. Columns (4)-(6) are girls' reports about bullying towards other girls, using a randomly selected subset of 2 or 3 girls in their grade-level (and removing cases where the respondent did not know the other girl). Columns (7)-(9) are teachers' reports about randomly selected girls in their class. Each teacher reports on a randomly selected sample of 3 girls from their class; the observations are at the *teacher* × *girl* level. Columns (1), (4) and (7) combine reports of both light teasing and severe intimidation/harassment. Columns (2), (5), and (8) include only light teasing. Columns (3), (6) and (9) include only severe intimidation/harassment. Combined indexes are constructed using loadings from exploratory factor analysis. The same relative loadings are used for teasing and intimidation/harassment. Standard errors are clustered at the school level and are in parentheses. Sample includes all girls in endline interviewed in person. Controls include stratum fixed effects and the baseline controls selected by double LASSO. All outcomes are in control group standard deviations.

Table A22: *Effects on bullying by boys or by other girls*

VARIABLES	(1) Teased by boy(s) (Self-reported) (=1)	(2) Teased by girls(s) (Self-reported) (=1)	(3) Teased by boy(s) (Peer-reported) (=1)	(4) Teased by girls(s) (Peer-reported) (=1)	(5) Teased about menstr. by boy(s) (=1)	(6) Teased about menstr. by girls(s) (=1)
Base Only	0.024	0.022	-0.009	-0.013	0.034**	-0.003
	(0.026)	(0.022)	(0.026)	(0.020)	(0.016)	(0.012)
Base + YGL	0.006	0.009	0.013	0.007	0.068***	0.007
	(0.022)	(0.020)	(0.024)	(0.018)	(0.014)	(0.011)
Observations	2,167	2,167	3,042	3,042	2,256	2,256
Data source	Girls	Girls	Girls	Girls	Girls	Girls
Control mean	.241	.222	.252	.157	.05	.048
Control SD	.428	.416	.435	.364	.218	.214
p: Treated = 0	.5892	.4896	.8341	.9663	<0.001	.6125
p: Base Only = Base + YGL	.419	.5294	.382	.3633	.0241	.3718

Notes: Columns (1), (3), (5): outcome = 1 when girl was teased by a boy. Columns (2), (4), (6): outcome = 1 when girl was teased by a girl. Columns (1) and (2) are girls' reports about bullying towards themselves. Columns (3) and (4) are girls reports about randomly selected girls in their class. Columns (5) and (6) are self-reports about bullying specifically about menstruation. Sample includes all girls in endline interviewed in person. Controls include stratum fixed effects and the baseline controls selected by double LASSO. Standard errors are clustered at the school level and are in parentheses.

Figure A23: *Effect on individual network questions*



Notes: This figure shows the effects on each outcome that is used in the network index in Table 7. Only girls included in the baseline sample were used. All outcomes are normalized using the control mean and standard deviation from the endline. All regressions control for stratum fixed effects and confidence intervals are calculated based on standard errors clustered at the school level.

Table A24: *Effects on restrictions to girls' activities due to menstruation (among girls who reached menarche at baseline)*

VARIABLES	(1) Any (=1)	(2) Absent (=1)	(3) Absent because ill (=1)	(4) Left school during day (=1)	(5) Couldn't participate fully (=1)	(6) Missed social activities (=1)	(7) Other activities (=1)
Base Only	-0.005	-0.021	-0.000	-0.047*	0.002	0.001	-0.008
	(0.046)	(0.027)	(0.019)	(0.025)	(0.035)	(0.025)	(0.026)
Base + YGL	-0.074**	-0.036*	-0.017	-0.033*	-0.059**	-0.029	0.005
	(0.034)	(0.019)	(0.017)	(0.020)	(0.023)	(0.022)	(0.025)
Observations	1,055	1,055	1,052	1,055	1,055	1,055	1,055
Data source	Girls	Girls	Girls	Girls	Girls	Girls	Girls
Control mean	.375	.117	.05	.108	.225	.096	.113
Control SD	.485	.322	.219	.311	.418	.295	.317
p: Treated = 0	.1174	.1069	.5066	.0541	.1049	.3843	.9719
p: Base Only = Base + YGL	.1234	.5282	.2358	.4908	.0647	.1067	.5638

Notes: Shows girls' reports of restrictions to their activities because of menstruation at endline. Sample only includes 1,053 girls who reported having started menstruating at *baseline* and interviewed in person at endline. We use baseline menarche rather than endline because at endline there is a small (5%) but significant difference in the share of girls report having started menarche, possibly related to reduced menstrual stigma leading to less underreporting in treatment schools. Standard errors are clustered at the school level and are in parentheses. Controls include stratum fixed effects and the baseline controls selected by double LASSO. Column (1): outcome is whether any of columns (2)-(7) are true. Column (2): has the girl ever been absent from school because of menstruation? Column (3): has the girl ever been absent because they were ill because of menstruation? Column (4): has the girl ever left school during the school day because of menstruation? Column (5): does the girl report not being able to participate fully in school activities because of menstruation? Column (6): has the girl missed any social activities because of menstruation? Column (7): has the girl been restricted from any other activities due to menstruation? The recall period used is whether these restrictions have *ever* occurred to them, so the questions were asked to all girls including those who had dropped out of school in the last year.

Table A25: *Heterogeneity by whether girls were menstruating at baseline*

	Base Only	Base + YGL	Menstr. at baseline	Base × Menstr. at base.	Base+YGL × Menstr. at base.	N
Learning test (Z)	0.200**	0.056	0.007	–0.128	0.024	2165
	(0.094) [0.033]	(0.067) [0.398]	(0.140) [0.962]	(0.126) [0.309]	(0.088) [0.785]	
Progressed one class (=1)	0.145***	0.105**	–0.050	–0.082	–0.072	2254
	(0.047) [0.002]	(0.042) [0.014]	(0.085) [0.557]	(0.065) [0.213]	(0.059) [0.222]	
Enrolled (=1)	–0.020	–0.023	–0.072	–0.008	0.001	2254
	(0.032) [0.526]	(0.029) [0.429]	(0.069) [0.298]	(0.043) [0.852]	(0.042) [0.978]	
Absent in last 3 months (=1)	–0.026	–0.038	0.004	0.034	–0.008	1822
	(0.046) [0.581]	(0.042) [0.359]	(0.093) [0.964]	(0.065) [0.602]	(0.057) [0.893]	
General health (Z)	–0.037	0.012	0.333**	0.019	0.017	2254
	(0.072) [0.611]	(0.064) [0.845]	(0.154) [0.030]	(0.109) [0.862]	(0.085) [0.838]	
UTI health (Z)	–0.102	–0.026	0.224	0.034	0.056	2254
	(0.077) [0.185]	(0.058) [0.650]	(0.179) [0.210]	(0.115) [0.767]	(0.093) [0.544]	
Mental health (Z)	0.027	0.081	–0.001	–0.050	–0.078	2254
	(0.088) [0.762]	(0.079) [0.304]	(0.175) [0.998]	(0.123) [0.683]	(0.118) [0.509]	
Self-esteem (Z)	0.052	0.112	0.356*	–0.072	–0.209*	2254
	(0.103) [0.610]	(0.074) [0.131]	(0.192) [0.063]	(0.142) [0.611]	(0.118) [0.076]	
Heart rate (Z)	–0.066	–0.050	0.231	–0.103	–0.135	1902
	(0.079) [0.407]	(0.068) [0.462]	(0.168) [0.168]	(0.113) [0.363]	(0.103) [0.190]	
Hygiene knowl. (Z)	0.339***	0.497***	0.241	–0.028	–0.148	2165
	(0.104) [0.001]	(0.075) [0.000]	(0.174) [0.166]	(0.137) [0.838]	(0.109) [0.175]	
Menstr. knowl. (Z)	0.398***	0.469***	0.153	–0.214**	–0.294***	2165
	(0.085) [0.000]	(0.069) [0.000]	(0.118) [0.196]	(0.104) [0.039]	(0.087) [0.001]	
Hygiene behavior (Z)	0.063	0.320***	–0.201	0.134	0.028	2165
	(0.114) [0.578]	(0.095) [0.001]	(0.172) [0.244]	(0.154) [0.383]	(0.117) [0.808]	
Menstr. behavior (Z)	0.266***	0.444***	0.408***	0.223*	0.273***	2254
	(0.080) [0.001]	(0.065) [0.000]	(0.138) [0.003]	(0.119) [0.060]	(0.091) [0.003]	
Observed hygiene: Girls (Z)	–0.054	0.101	0.178	–0.048	–0.139	2165
	(0.076) [0.478]	(0.068) [0.135]	(0.175) [0.308]	(0.136) [0.724]	(0.115) [0.226]	
Combined stigma index (Z)	0.470***	0.833***	0.205	–0.170	–0.220**	2254
	(0.085) [0.000]	(0.061) [0.000]	(0.144) [0.155]	(0.112) [0.129]	(0.091) [0.016]	
Received info. (Z)	0.121	0.598***	0.050	0.163	–0.128	2254
	(0.092) [0.190]	(0.080) [0.000]	(0.238) [0.835]	(0.142) [0.251]	(0.111) [0.250]	
Willingness to speak (Z)	0.294***	0.658***	0.206	–0.114	0.146	2248
	(0.084) [0.000]	(0.070) [0.000]	(0.178) [0.248]	(0.130) [0.383]	(0.116) [0.209]	
Attitudes (Z)	0.435***	0.689***	0.287*	–0.179*	–0.292***	2252
	(0.083) [0.000]	(0.067) [0.000]	(0.167) [0.085]	(0.109) [0.099]	(0.095) [0.002]	
Norms (Z)	0.301***	0.569***	0.036	–0.065	–0.224**	2165
	(0.082) [0.000]	(0.052) [0.000]	(0.168) [0.830]	(0.109) [0.552]	(0.089) [0.012]	
Shame response to vignettes (rev.) (Z)	0.394***	0.540***	0.031	–0.213**	–0.225***	2163
	(0.082) [0.000]	(0.069) [0.000]	(0.138) [0.823]	(0.100) [0.034]	(0.082) [0.006]	
Observed shame (rev.) (Z)	0.411***	0.314***	0.176	–0.269***	–0.198**	2254
	(0.081) [0.000]	(0.068) [0.000]	(0.155) [0.256]	(0.098) [0.006]	(0.085) [0.019]	
Bullying: teasing (Z)	0.098	0.070	–0.153	0.049	–0.105	2165
	(0.095) [0.307]	(0.083) [0.401]	(0.163) [0.349]	(0.133) [0.712]	(0.124) [0.396]	
Bullying: intimidation/harassment (Z)	–0.083	–0.047	–0.143	0.002	–0.047	2165
	(0.063) [0.185]	(0.056) [0.400]	(0.118) [0.227]	(0.095) [0.981]	(0.082) [0.570]	
Network (Z)	–0.059	0.167**	–0.191	0.490***	0.204*	2165
	(0.086) [0.494]	(0.079) [0.034]	(0.210) [0.364]	(0.144) [0.001]	(0.122) [0.094]	

Notes: Each row represents a regression of a different outcome. Each column represents the coefficient on each term. *Menstr. at base.* is 1 when a girl was menstruating at baseline. All regressions include stratum fixed effects and the baseline controls selected by double LASSO, and cluster standard errors at the school level.

Table A26: *Heterogeneity by whether girls started menstruating between baseline and endline*

	Base Only	Base + YGL	Newly menstr.	Base × Newly menstr.	Base+YGL × Newly menstr.	N
Learning test (Z)	0.135*	0.060	0.016	0.072	0.070	2167
	(0.071) [0.057]	(0.055) [0.279]	(0.094) [0.862]	(0.147) [0.626]	(0.100) [0.481]	
Progressed one class (=1)	0.120***	0.090**	0.071	–0.014	–0.056	2256
	(0.044) [0.007]	(0.040) [0.025]	(0.068) [0.293]	(0.075) [0.854]	(0.069) [0.419]	
Enrolled (=1)	–0.020	–0.020	–0.064*	0.017	0.004	2256
	(0.024) [0.402]	(0.024) [0.406]	(0.039) [0.097]	(0.052) [0.742]	(0.047) [0.924]	
Absent in last 3 months (=1)	0.000	–0.021	0.013	–0.047	–0.070	1824
	(0.037) [0.994]	(0.033) [0.537]	(0.070) [0.855]	(0.087) [0.586]	(0.072) [0.333]	
General health (Z)	–0.022	–0.007	–0.108	–0.015	0.060	2256
	(0.067) [0.744]	(0.050) [0.882]	(0.124) [0.383]	(0.159) [0.922]	(0.131) [0.643]	
UTI health (Z)	–0.123*	–0.038	–0.214*	0.136	0.140	2256
	(0.063) [0.051]	(0.054) [0.475]	(0.111) [0.055]	(0.168) [0.416]	(0.125) [0.261]	
Mental health (Z)	–0.016	0.011	–0.101	0.060	0.067	2256
	(0.066) [0.809]	(0.065) [0.860]	(0.142) [0.477]	(0.154) [0.698]	(0.141) [0.636]	
Self-esteem (Z)	–0.003	–0.038	–0.143	0.012	0.155	2256
	(0.072) [0.969]	(0.063) [0.543]	(0.135) [0.290]	(0.159) [0.938]	(0.146) [0.291]	
Heart rate (Z)	–0.102*	–0.108**	–0.058	–0.040	–0.021	1904
	(0.060) [0.089]	(0.055) [0.048]	(0.106) [0.586]	(0.129) [0.755]	(0.115) [0.856]	
Hygiene knowl. (Z)	0.318***	0.399***	0.033	0.003	0.074	2167
	(0.080) [0.000]	(0.067) [0.000]	(0.121) [0.787]	(0.143) [0.984]	(0.122) [0.546]	
Menstr. knowl. (Z)	0.293***	0.281***	0.443***	–0.072	0.028	2167
	(0.060) [0.000]	(0.053) [0.000]	(0.072) [0.000]	(0.089) [0.421]	(0.075) [0.712]	
Hygiene behavior (Z)	0.099	0.297***	–0.030	0.004	0.118	2167
	(0.087) [0.258]	(0.071) [0.000]	(0.097) [0.756]	(0.139) [0.978]	(0.112) [0.296]	
Menstr. behavior (Z)	0.273***	0.405***	1.662***	0.292**	0.334***	2256
	(0.061) [0.000]	(0.052) [0.000]	(0.085) [0.000]	(0.113) [0.010]	(0.089) [0.000]	
Observed hygiene: Girls (Z)	–0.177**	–0.039	0.127	0.252*	0.164	2167
	(0.069) [0.010]	(0.060) [0.519]	(0.104) [0.222]	(0.135) [0.062]	(0.116) [0.157]	
Combined stigma index (Z)	0.343***	0.666***	0.095	0.043	0.095	2256
	(0.071) [0.000]	(0.058) [0.000]	(0.089) [0.285]	(0.111) [0.696]	(0.100) [0.340]	
Received info. (Z)	0.227***	0.502***	0.117	–0.167	0.024	2256
	(0.086) [0.008]	(0.068) [0.000]	(0.107) [0.273]	(0.128) [0.191]	(0.123) [0.848]	
Willingness to speak (Z)	0.179**	0.666***	–0.001	0.079	0.101	2250
	(0.073) [0.014]	(0.060) [0.000]	(0.087) [0.993]	(0.118) [0.505]	(0.114) [0.373]	
Attitudes (Z)	0.281***	0.479***	0.027	0.181	0.168	2254
	(0.066) [0.000]	(0.064) [0.000]	(0.107) [0.803]	(0.126) [0.151]	(0.117) [0.149]	
Norms (Z)	0.237***	0.422***	0.108	0.012	0.052	2167
	(0.070) [0.001]	(0.060) [0.000]	(0.107) [0.310]	(0.126) [0.927]	(0.113) [0.648]	
Shame response to vignettes (rev.) (Z)	0.284***	0.428***	0.191*	–0.071	–0.092	2165
	(0.063) [0.000]	(0.055) [0.000]	(0.104) [0.066]	(0.124) [0.565]	(0.106) [0.384]	
Observed shame (rev.) (Z)	0.275***	0.204***	0.059	–0.007	0.019	2256
	(0.067) [0.000]	(0.056) [0.000]	(0.116) [0.613]	(0.135) [0.960]	(0.118) [0.874]	
Bullying: teasing (Z)	0.112	0.042	0.084	–0.010	–0.110	2167
	(0.074) [0.131]	(0.065) [0.519]	(0.132) [0.525]	(0.164) [0.953]	(0.147) [0.455]	
Bullying: intimidation/harassment (Z)	–0.062	–0.025	–0.024	0.006	–0.113	2167
	(0.062) [0.316]	(0.055) [0.648]	(0.130) [0.851]	(0.155) [0.967]	(0.127) [0.374]	
Network (Z)	0.202**	0.246***	0.035	–0.192	0.055	2167
	(0.092) [0.028]	(0.079) [0.002]	(0.131) [0.788]	(0.171) [0.261]	(0.152) [0.717]	

Notes: Each row represents a regression of a different outcome. Each column represents the coefficient on each term. *Newly menstr.* takes the value 1 when a girl was not menstruating at baseline, but was menstruating by endline, and 0 otherwise. All regressions include stratum fixed effects and the baseline controls selected by double LASSO, and cluster standard errors at the school level.

Table A27: *Effects on teachers*

VARIABLES	(1) Progressive Menstruation Attitudes (Z)	(2) Mens. Knowledge Index (Z)	(3) Progressive Gender Attitudes (Z)	(4) Absent Last Month (=1)
Base Only	0.671***	0.155	0.113	-0.076
	(0.086)	(0.114)	(0.122)	(0.055)
Base + YGL	0.684***	0.186**	0.182	0.005
	(0.088)	(0.093)	(0.112)	(0.048)
Observations	739	739	775	739
Data source	Teachers	Teachers	Teachers	Teachers
Control mean	0	0	0	.497
Control SD	1	1	1	.501
p: Treated = 0	<0.001	.048	.1298	.6225
p: Base Only = Base + YGL	.8156	.7598	.5048	.1091

Notes: The table displays results from the teachers' survey in endline. Standard errors are clustered at the school level and are in parentheses; controls include stratum fixed effects and a dummy for a primary schools. *Progressive menstruation attitudes* combines questions about teachers' attitudes towards menstruation, e.g., should girls be allowed to go to school when menstruating. *Mens. knowledge index*: index of menstruation knowledge using a series of true or false questions. *Progressive gender attitudes*: index of progressive gender attitudes, e.g., using agreement with questions like "Boys are better leaders than girls". *Absent last month*=1 when the teacher reported being absent at least once last month, and is 0 otherwise.

Table A28: *Changes in academic performance of nominated YGLs*

	Change in grade (/20)	Change in class rank
Intercept	−1.173***	6.429***
	(0.205)	(0.820)
Num.Obs.	191	191
Grade/rank at baseline	11.7	4.8

Notes: Sample includes all girls who were selected as YGLs (in January 2022) *and* interviewed at endline (in summer 2023) and who were in school at endline. 255 YGLs were selected, 46 were not found at endline, and 18 of those found were no longer in school and thus did not have school marks to elicit. During YGL nomination and at endline, girls were asked their school marks (out of a maximum of 20) and their rank in class. Column (1) shows the change in school marks from nomination to endline. Column (2) shows the change in class rank from nomination to endline (with a higher rank indicating relatively worse academic performance). Regressions control for stratum fixed effects, and standard errors are clustered at the school level. Standard errors cluster at the school level. ** p < 0.1, ** p < 0.05, *** p < 0.01.*

Table A29: *Heterogeneity as predicted by generic machine learning: Best Linear Predictor*

	(1)		(2)	
	Treated pooled		Base + YGL	
	ATE (β_1)	HTE (β_2)	ATE (β_1)	HTE (β_2)
Learning test (Z)	.037**	.803	.118	.241
Achievement tests only (Z)	.018**	.353	.086*	.661
Cognitive tests only (Z)	.134	.203	.302	.138
Math fluency (Z)	.060*	.600	.136	.433
Reading fluency (Z)	.093*	.944	.239	.105
Listening (Z)	.518	.617	.695	.679
Comprehension (Z)	.501	.422	.678	.569
Digit span (Z)	.409	.494	.659	.194
Reverse digit span (Z)	.229	.971	.148	.201
Progressed one class (=1)	.070*	.462	.117	.418
Enrolled (=1)	.633	.361	.977	.006***
Absent in last 3 months (=1)	.577	.293	.358	.342
Hygiene knowl. (Z)	<0.001***	.488	<0.001***	.777
Menstr. knowl. (Z)	<0.001***	.033**	<0.001***	.011**
Hygiene behavior (Z)	.003***	.553	<0.001***	.935
Menstr. behavior (Z)	<0.001***	.083*	<0.001***	.397
Observed hygiene: Girls (Z)	.671	.124	.508	.587
Received info. (Z)	<0.001***	.102	<0.001***	.030**
Willingness to speak (Z)	<0.001***	.886	<0.001***	.253
Attitudes (Z)	<0.001***	.324	<0.001***	.434
Norms (Z)	<0.001***	.681	<0.001***	.407
Shame response to vignettes (rev.) (Z)	<0.001***	.326	<0.001***	.821
Observed shame (Z)	<0.001***	.206	.001***	.514
Bullying: Self-report (Z)	.980	.235	.728	.953
Heart rate (Z)	.048**	.352	.038**	.890
General health (Z)	.949	.375	.962	.145
UTI health (Z)	.613	.419	.809	.960
Mental health (Z)	.595	.901	.631	.289
Self-esteem (Z)	.722	.246	.826	.215
Network (Z)	.008***	.475	.002***	.359

Notes: The table presents the results of the Best Linear Predictor of the Conditional Average Treatment Effect (CATE) on the ML proxies following Welz et al. (2022) over 1000 splits ($\alpha = .05$). We report the p-value for the hypothesis that the parameter is equal to zero against the two-sided alternative of the coefficients β_1 and β_2, which correspond to the ATE and heterogeneity loading (HTE) parameters in the *Best Linear Predictor*. * p< 0.1, ** p < 0.05, *** p < 0.01. The first column shows the outcomes from the girls' endline survey, and each row represents one estimation. In (1) we test heterogeneity comparing all the treated girls against the control, and in (2) we test heterogeneity comparing the girls in Base+YGL schools against the control.

Table A30: *Heterogeneous effects by YGL positive deviance*

	Base Only	Base + YGL	Base + YGL*Pos. Deviance Index	N
Learning test (Z)	0.207***	0.111**	0.095**	2167
	(0.071) [0.004]	(0.052) [0.032]	(0.038) [0.013]	
Progressed one class (=1)	0.113***	0.078**	–0.030*	2256
	(0.039) [0.004]	(0.035) [0.026]	(0.018) [0.099]	
Enrolled (=1)	–0.019	–0.021	–0.005	2256
	(0.024) [0.422]	(0.022) [0.347]	(0.016) [0.784]	
Absent in last 3 months (=1)	–0.009	–0.037	–0.024	1824
	(0.030) [0.757]	(0.029) [0.201]	(0.020) [0.243]	
General health (Z)	–0.037	0.009	–0.005	2256
	(0.060) [0.532]	(0.048) [0.851]	(0.034) [0.877]	
UTI health (Z)	–0.090*	–0.008	0.067**	2256
	(0.053) [0.091]	(0.042) [0.847]	(0.034) [0.048]	
Mental health (Z)	0.012	0.045	0.081*	2256
	(0.062) [0.847]	(0.059) [0.447]	(0.046) [0.079]	
Self-esteem (Z)	0.017	0.009	0.015	2256
	(0.067) [0.799]	(0.052) [0.867]	(0.036) [0.674]	
Heart rate (Z)	–0.129**	–0.126**	–0.044	1904
	(0.056) [0.022]	(0.050) [0.012]	(0.033) [0.178]	
Hygiene knowl. (Z)	0.345***	0.442***	–0.002	2167
	(0.077) [0.000]	(0.062) [0.000]	(0.041) [0.956]	
Menstr. knowl. (Z)	0.310***	0.341***	–0.018	2167
	(0.057) [0.000]	(0.049) [0.000]	(0.035) [0.594]	
Hygiene behavior (Z)	0.110	0.323***	0.053	2167
	(0.086) [0.200]	(0.072) [0.000]	(0.040) [0.184]	
Menstr. behavior (Z)	0.360***	0.559***	0.028	2256
	(0.068) [0.000]	(0.053) [0.000]	(0.039) [0.466]	
Observed hygiene: Girls (Z)	–0.070	0.048	–0.061	2167
	(0.062) [0.260]	(0.055) [0.382]	(0.038) [0.108]	
Combined stigma index (Z)	0.406***	0.741***	–0.016	2256
	(0.065) [0.000]	(0.056) [0.000]	(0.041) [0.695]	
Received info. (Z)	0.202***	0.543***	–0.073*	2256
	(0.077) [0.009]	(0.064) [0.000]	(0.039) [0.061]	
Willingness to speak (Z)	0.252***	0.729***	0.023	2250
	(0.076) [0.001]	(0.065) [0.000]	(0.048) [0.630]	
Attitudes (Z)	0.367***	0.567***	0.005	2254
	(0.064) [0.000]	(0.058) [0.000]	(0.045) [0.913]	
Norms (Z)	0.280***	0.474***	–0.012	2167
	(0.060) [0.000]	(0.049) [0.000]	(0.039) [0.760]	
Shame response to vignettes (rev.) (Z)	0.308***	0.448***	–0.036	2165
	(0.058) [0.000]	(0.054) [0.000]	(0.038) [0.344]	
Observed shame (rev.) (Z)	0.296***	0.230***	–0.016	2256
	(0.060) [0.000]	(0.053) [0.000]	(0.044) [0.710]	
Bullying: teasing (Z)	0.108*	0.024	–0.053	2167
	(0.065) [0.098]	(0.058) [0.685]	(0.040) [0.179]	
Bullying: intimidation/harassment (Z)	–0.082*	–0.064	–0.078***	2167
	(0.045) [0.068]	(0.039) [0.103]	(0.027) [0.003]	
Network (Z)	0.174**	0.266***	0.004	2167
	(0.086) [0.043]	(0.074) [0.000]	(0.038) [0.924]	

Notes: Each row is one regression for a given outcome. *Positive deviance index (Z)* is a school-level index constructed based on how high the selected *YGLs* score on an index of positive deviance, composed of willingness to adopt new behaviors, tell others about new things, discuss sensitive topics, convince others of new behaviors, and ease at explaining how to use a sanitary pad. The first column shows the coefficient on *Base Only*, the second column shows the coefficient on *Base+YGL*, and the third column shows the coefficient on (*Base+YGL* × *Positive Deviance Index*). All regressions control for stratum fixed effects, include LASSO selected controls, and cluster standard errors at the school level. Standard errors are in parentheses, and *p*-values are in brackets. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table A31: *Heterogeneous effects by YGL positive deviance, relative to positive deviance of other candidates in the same school*

	Base Only	Base + YGL	Base + YGL × Rel. Pos. Deviance	N
Learning test (Z)	0.208***	0.116**	0.021	2167
	(0.072) [0.004]	(0.053) [0.029]	(0.061) [0.737]	
Progressed one class (=1)	0.114***	0.084**	–0.087***	2256
	(0.039) [0.004]	(0.035) [0.018]	(0.025) [0.000]	
Enrolled (=1)	–0.018	–0.017	–0.043*	2256
	(0.024) [0.441]	(0.022) [0.435]	(0.023) [0.057]	
Absent in last 3 months (=1)	–0.009	–0.038	–0.013	1824
	(0.030) [0.759]	(0.029) [0.194]	(0.031) [0.677]	
General health (Z)	–0.036	0.017	–0.099***	2256
	(0.060) [0.547]	(0.047) [0.724]	(0.037) [0.007]	
UTI health (Z)	–0.089*	–0.004	0.000	2256
	(0.054) [0.096]	(0.043) [0.929]	(0.058) [0.999]	
Mental health (Z)	0.012	0.048	0.039	2256
	(0.062) [0.852]	(0.061) [0.433]	(0.061) [0.523]	
Self-esteem (Z)	0.017	0.009	0.015	2256
	(0.067) [0.801]	(0.053) [0.871]	(0.050) [0.762]	
Heart rate (Z)	–0.131**	–0.135***	0.081*	1904
	(0.057) [0.020]	(0.049) [0.006]	(0.046) [0.080]	
Hygiene knowl. (Z)	0.344***	0.435***	0.084*	2167
	(0.077) [0.000]	(0.062) [0.000]	(0.045) [0.063]	
Menstr. knowl. (Z)	0.308***	0.331***	0.103***	2167
	(0.057) [0.000]	(0.049) [0.000]	(0.037) [0.005]	
Hygiene behavior (Z)	0.109	0.321***	0.071	2167
	(0.086) [0.203]	(0.073) [0.000]	(0.077) [0.358]	
Menstr. behavior (Z)	0.364***	0.574***	–0.156***	2256
	(0.069) [0.000]	(0.054) [0.000]	(0.050) [0.002]	
Observed hygiene: Girls (Z)	–0.072	0.037	0.088*	2167
	(0.063) [0.256]	(0.056) [0.505]	(0.046) [0.055]	
Combined stigma index (Z)	0.405***	0.738***	0.028	2256
	(0.065) [0.000]	(0.055) [0.000]	(0.072) [0.698]	
Received info. (Z)	0.203***	0.541***	–0.027	2256
	(0.078) [0.009]	(0.064) [0.000]	(0.056) [0.626]	
Willingness to speak (Z)	0.251***	0.725***	0.067	2250
	(0.076) [0.001]	(0.064) [0.000]	(0.083) [0.422]	
Attitudes (Z)	0.367***	0.568***	–0.010	2254
	(0.064) [0.000]	(0.057) [0.000]	(0.076) [0.897]	
Norms (Z)	0.280***	0.475***	–0.026	2167
	(0.060) [0.000]	(0.049) [0.000]	(0.055) [0.640]	
Shame response to vignettes (rev.) (Z)	0.308***	0.446***	–0.010	2165
	(0.059) [0.000]	(0.056) [0.000]	(0.083) [0.902]	
Observed shame (rev.) (Z)	0.295***	0.224***	0.069	2256
	(0.061) [0.000]	(0.054) [0.000]	(0.045) [0.128]	
Bullying: teasing (Z)	0.108	0.019	0.013	2167
	(0.066) [0.101]	(0.058) [0.742]	(0.053) [0.800]	
Bullying: intimidation/harassment (Z)	–0.082*	–0.062	–0.086**	2167
	(0.045) [0.069]	(0.039) [0.110]	(0.040) [0.032]	
Network (Z)	0.175**	0.274***	–0.097*	2167
	(0.086) [0.041]	(0.074) [0.000]	(0.057) [0.085]	

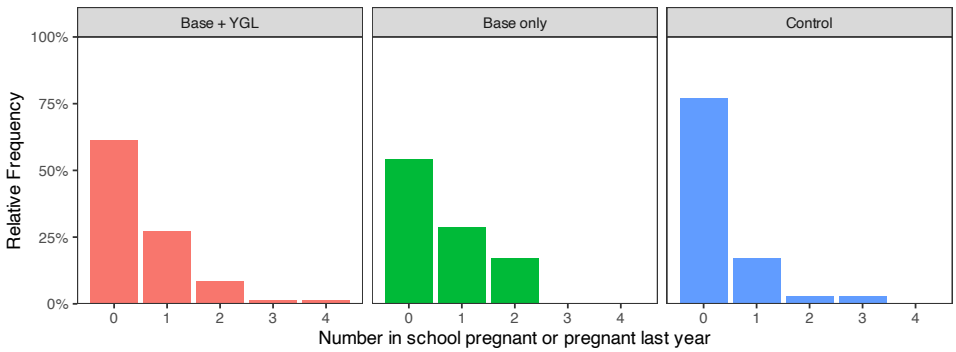
Notes: Each row is one regression for a given outcome. *Relative Positive Deviance Index (Z)* is a school-level index constructed based on how high the selected *YGLs* score on an index of positive deviance, *relative to* the full pool of candidate YGLs in that same school. The first column shows the coefficient on *Base Only*, the second column shows the coefficient on *Base+YGL*, and the third column shows the coefficient on (*Base+YGL* × *Positive deviance index*). All regressions control for stratum fixed effects, include LASSO selected controls, and cluster standard errors at the school level. Standard errors are in parentheses, and *p*-values are in brackets. ** p < 0.1, ** p < 0.05, *** p < 0.01*.

Table A32: *Effects on marriage and fertility*

VARIABLES	(1) Has been married (=1)	(2) Pregnant now or last year (=1)	(3) Dropout because of baby (=1)
Base Only	0.009	0.016**	0.019***
	(0.008)	(0.007)	(0.007)
Base + YGL	0.002	0.009	-0.001
	(0.007)	(0.007)	(0.005)
Observations	2,287	2,287	2,256
Data source	Mothers	Mothers	Girls
Control mean	.024	.02	.013
Control SD	.152	.141	.113
p: Treated = 0	.5022	.0663	.3231
p: Base Only = Base + YGL	.4096	.3156	<0.001

Notes: Table shows effects on marriage, pregnancy, and dropout due to pregnancy at endline. Columns (1) and (2) come from mothers' reports of the marriage and pregnancy status of their daughters. Column (3) comes from girls' own reports. Column (1) outcome is 1 if the girl has been married by endline, 0 otherwise. Column (2) is 1 if the girl has been pregnant in the last year or is currently pregnant, 0 otherwise. Column (3) is 1 if a girl reported dropping out of school because they had to look after a baby or because they were pregnant, 0 otherwise. All specifications include stratum fixed effects, baseline controls selected by double LASSO, and cluster standard errors at the school level.

Figure A33: *School-level relative frequency of number of pregnant girls in the sample*



Notes: Observation is at the school level. Shows the histogram of the number of pregnant girls in a school at endline, divided by the total number of schools in each treatment group in order to maintain comparability between *Base+YGL* (70 schools), *Base only* (35 schools) and *Control* schools (35 schools). Pregnancy = 1 if mothers responded that their daughter was currently pregnant or had been pregnant in the last year.

B Data

B.1 Construction of outcomes

This section documents the questions used to construct the main outcome indexes used in the paper.

Knowledge and behavior indexes are created by summing the number of correct answers. All other indexes are constructed by conducting an exploratory factor analysis of all questions in a module, and combining using factor loadings based on a single factor.

General health. Girls are asked how much the following things have been difficult for them over the last month, with possible answers as (i) never; (ii) almost never; (iii) sometimes; (iv) often; or (v) almost always:

- It's hard for me to walk more than a block.
- It's hard for me to run.
- It is difficult for me to play sports or exercise.
- It's hard for me to lift something heavy.
- It's difficult for me to take a bath or shower by myself.
- It is difficult for me to do household chores.
- I have pain or aches.
- I have little energy.

The index is constructed so that a higher value indicates *better* health (i.e., fewer difficulties).

UTI health. Girls are asked: "Please indicate if you have had the following symptoms for a few days in the last month." (Answer options: (i) No, (ii) Yes, light symptoms, (iii) Yes, moderate symptoms, (iv) Yes, serious symptoms.):

- Frequent urination of small volumes of urine (going to the toilet very often)
- Urgent urination (strong, uncontrollable urge to urinate)
- Painful or burning sensation when urinating
- Sensation of incomplete emptying of the bladder after urination
- Feeling of pain or discomfort in the lower abdomen (suprapubic region)
- Presence of blood in the urine

They are also asked "Did these symptoms occur at the same time (on the same days)?". The index is constructed so that a higher value indicates *better* health, i.e., fewer symptoms of urinary tract infections.

Mental health. Girls are asked how many times they felt the following emotions over the last 7 days, based on an adapted version of the CESD index (Radloff, 1977):

- How many days have you been bothered by things that don't usually bother you?
- How many days have you not wanted to eat as usual?
- How many days did manage to shake sadness off, even with the help of family and friends?

- How many days were you happy? (rev.)
- How many days you could concentrate on the things you were doing (rev.)
- How many days did you feel discouraged?
- How many days did you feel that everything you did was difficult?
- How many days have you been full of hope for the future? (rev.)
- How many days have you felt so moody that you didn't want to talk to anyone?
- How many days were you afraid?
- How many days was your sleep restless?
- How many days have you been happy? (rev.)
- How many days have you spoken less than usual?
- How many days have you felt alone?
- How many days did others look out for you to chat? (rev.)
- How many days did you want to do your daily activities? (rev.)
- How many days have you cried so hard you couldn't stop?
- How many days have you felt sad?
- How many days have you told yourself that other girls don't like you?
- How many days have you lacked the desire to do things?

The index is constructed so that a higher value indicates *better* mental health (i.e., fewer symptoms of depression).

Self esteem index. Girls are asked "For each description, indicate how much it is true for you." Possible answers: Strongly disagree; disagree; agree; strongly agree. Questions are:

- Do you think you have a certain number of good qualities?
- All things considered, you are inclined to consider yourself a failure (rev.)
- you have a positive attitude towards yourself
- Overall, you are satisfied with yourself
- You wish you had more respect for yourself
- Sometimes you feel really useless (rev.)
- Sometimes you think you're good for nothing (rev.)

The index is constructed so that a higher value indicates *better* self esteem, i.e., more belief in oneself.

Menstruation-related knowledge. Index is constructed as the sum of correct answers to:

- How often does a girl usually get her period? (Ans: once per month / 28 days)
- How many days does the period last on average? (Ans: 3-7 days)
- At what age do women stop having their periods? (Ans: 40-60 years)
- What are the signs before or during your period? (Ans: at least 2 correct)
- What are healthy behaviors to adopt during your period? (Ans: at least 3 correct)
- T/F: When girls reach menarche, they can become pregnant from sexual intercourse?

- T/F: when blood flows during menstruation, it usually means that a woman is pregnant?
- T/F: is it common to experience pain and discomfort during menstruation?
- T/F: Menstrual blood is the destruction of tissue surrounding the uterus that occurs when fertilization does not occur.
- T/F: Is it the first sexual intercourse that causes a period for young girls?
- T/F: You should not shower during your period
- T/F: periods are symptoms of an illness
- List the different types of menstrual products that girls can use. (Ans: names at least 3)
- In your opinion, what are the most hygienic products to use during your period? (Ans: at least 2 correct)
- How many times a day should girls change/wash cloths or other menstrual products when they are on their period? (Ans: at least 3)
- How many times a day should girls change reusable pads when they are on their period? (Ans: at least 2)
- How often should girls wash when they are on their period? (Ans: at least 3x per day)
- What should you wash with during your period? (Ans: with soap)
- How much does one reusable sanitary napkin (made by a seamstress) cost? (Ans: between 5000 and 10,000 Ary)

Hygiene knowledge. Index is constructed as the sum of correct answers to:

- Name 3 good habits you can adopt that can help ensure you don't get sick.
- Name all the things you need to do to ensure you don't contaminate the water in a water container (Ans: at least 2 correct)
- Name all the things you can do to improve the hygiene of latrine use.
- To your knowledge, how many times a day should we wash our hands? (Ans: at least 3)
- To your knowledge, when are the most important times to wash your hands? (Ans: at least 2 correct)
- Can you name two methods for making water drinkable?

Menstruation-related behaviour. Index is constructed as the sum of "correct" answers to:

- During your last period, were you able to wash and change in private at home?
- During your last period, were you able to wash and change in private at school?
- Have you used reusable sanitary pads (made by a seamstress) in the last 3 months?
- In general, how often do you change your sanitary pads or other sanitary products when you have your period? (Ans: at least 3 times per day)
- The last time you used sanitary pads, where you wash them? (Ans: not in the river)
- The last time you used sanitary pads, where did you dry them? (Ans: outside in the sun)
- Where do you wash your underwear? (Ans: not in the river)
- Where do you dry your underwear? (Ans: outside in the sunshine)

- The last time you had your period, how many times did you wash during a typical day at home? (Ans: at least 3)
- The last time you had your period, how many times did you wash with soap on a typical day at home? (Ans: at least 3)
- The last time you had your period during term time, how many times did you wash during a typical day at school? (Ans: at least 3)
- The last time you had your period during school term, how many times did you wash with soap on a typical day at school? (Ans: at least 3)
- Generally, where do you dry your sanitary products? (Ans: outside in the sun)

Hygiene behaviour. Index is constructed as the sum of "correct" answers to:

- Over the past week, where did the water you used to wash your hands at home come from? (Ans: from clean source; from tap; from cleaned source; from village water system)
- During the last week, what did you wash your hands with at home? (Ans: with soap)
- During the last week, when did you wash your hands at home? (Ans: at least 3 moments)
- Think about the last time you went to urinate or defecate when you were at home. Where did you go? (Ans: in a toilet)
- How many times during the day did you wash your hands the last time you spent a day at home? (Ans: at least 3 times)
- During the last week you were at school, where did the water you used to wash your hands come from? (Ans: from clean source; from tap; from cleaned source; from village water system)
- During the last week you were at school, what did you wash your hands with? (Ans: soap)
- During the last week you were at school, when did you wash your hands? (Ans: at least 3 moments)
- Think about the last time you went to urinate or defecate at school. Where did you go? (Ans: in a toilet)
- How many times during the day did you wash your hands the last time you spent a day at school? (Ans: at least 3)
- During the last week, how many times have you gone swimming? Think about the last time you went swimming. Where did you go? (Ans: not in the river)

Observed hygiene (girls). For the first 5 questions, the enumerator observes whether the girl is (i) very clean; (ii) quite clean; (iii) slightly dirty; or (iv) very dirty for the following: (i) hands, (ii) face, (iii) neck, (iv) clothes, (v) hair. The 6th question is whether the girl is wearing shoes. The index is constructed so that a higher value indicates *better* hygiene.

Observed hygiene (home). Enumerator observes the following about the girl's homestead, with (+) indicating a positive hygienic observation, (-) indicating a negative reverse coded observation:

- Observe the presence of water at the place of hand washing. Check that the tap/pump, basin/bucket, water tank, or similar objects have/contain water. (+)

- Is there soap, detergent or ash at the hand-washing area? (+)
- Human and animal feces in and around the house (-)
- Waste accumulated around the house (-)
- Stagnant water around the house (-)
- Stables, pigs or kennels below or to the side of the house (-)
- Lack of ventilation inside the house (-)
- The terrace and the garden are clean (+)
- In the house there are separate spaces for cooking, sleeping and socializing (+)
- The house is neat (+)
- The kitchen area is neat (+)
- Animals have access to areas where food is prepared (-)
- There is water served in the kitchen (+)
- The dishes in the kitchen are clean (+)

The index is constructed so that a higher value indicates *better* hygiene.

Observed hygiene (school). Enumerators observe the following about the school environment, all of which indicate more hygienic environment:

- Is there a cover for the holes in the toilet?
- Is there soap available to use at the water point?
- Are there garbage holes in the ground?
- Is the garbage separated?
- Are there trash collections in classrooms?
- Is there a water point for hand washing with soap?

The index is constructed so that a higher value indicates *better* hygiene.

Received information about menstrual hygiene. Girls are asked the following questions:

- In the past 3 months, how many times have you heard a classmate, or another student at school, talk about menstrual hygiene?
- In the past 3 months, how many times have you heard someone in your family talk about menstrual hygiene?
- In the past 3 months, how many times have you heard a teacher talk about menstrual hygiene?
- In the past 3 months, have you asked anyone a question about menstrual hygiene?
- In the past 3 months, has anyone given you information about menstrual hygiene, whether you asked for it or not?
- In the past 3 months, has [MOTHER] given you information about menstrual hygiene?

Willingness to speak about menstruation. Answer options: (i) strongly disagree; (ii) disagree; (iii) neither agree nor disagree; (iv) agree; (v) strongly agree. Girls are asked whether they would be willing to discuss...

- the importance of menstrual hygiene in a casual chat with friends, if only girls were present.
- the importance of menstrual hygiene in a casual chat with friends, if both boys and girls were present.
- the importance of menstrual hygiene in front of my class in a classroom, if only girls were present.
- the importance of menstrual hygiene in front of my class, in a classroom, if boys and girls were present.

Attitudes towards menstruation. Each question is a Likert scale, with possible answers: (i) strongly disagree; (ii) disagree; (iii) neither agree nor disagree; (iv) agree; (v) strongly agree.

- "Girls should be allowed to go to school when they are on their period"?
- "Girls should be required to hide the fact that they have their period while at school." ? (rev)
- "People should discuss menstruation openly with others"?
- "Mothers should broach the subject of menstruation with daughters only after their first period" (rev)
- "It is prohibited to dry washable towels in a public place" (rev)
- "Girls should be punished in case their brothers discover its intimate effects during the period of menstruation" (rev)
- "Girls should feel embarrassed or ashamed when they have their period in public." (rev)
- "Boys should know about menstruation."
- "Boys should be allowed to laugh at topics related to menstruation" (rev)
- "Periods are natural phenomena."

The index is constructed so that a higher value indicates *more progressive* attitudes towards menstruation (e.g., being more likely to say that girls should be allowed to go to school during menstruation).

Norms. Girls were asked how many other girls and then other boys in their school would agree with the following three statements:

- Girls should be obliged to hide the fact that they are on their periods when they are at school.
- People should be able to openly discuss menstruation with others.
- Boys should be allowed to laugh about things related to menstruation.

To aid with the questions, we asked how many girls (or boys) out of 10 would agree with the statement, using 10 small stones and asking girls to move stones to the right to represent people agreeing and to the left to represent people disagreeing.

Girls were then asked about their teachers' opinions of the same statements (answer options: all teachers; most teachers; half of teachers; only a few teachers; no teachers), and then their mother's opinion for the same statements (answer options: strongly disagree; disagree; neither disagree nor agree; agree; strongly agree). For teachers and mothers, we also asked perceived agreement with a fourth statement: "Girls should be allowed to come to school when they have their periods."

The index is constructed so that a higher value indicates that girls believe that others have *more progressive* attitudes towards menstruation (e.g., being more likely to say that girls should be allowed to go to school during menstruation).

Shame in response to vignettes. Enumerators describe 2 vignettes to girls:

1. Imagine that you went to school and you were speaking about periods with a friend. You turn around and realise that someone you don't know heard the whole conversation.
2. Imagine that you go to school and that someone who you don't know realises that you have started your periods, because he has seen a reusable sanitary pad in your bag.

After each vignette, girls are asked how much they feel the following emotions in response, with answer options: (i) I feel it strongly, (ii) I feel it a bit, (iii) I don't feel it at all

- Ashamed
- Indifferent (rev.)
- Humiliated
- Guilty
- That it's funny (rev.)
- Ridiculous
- That it's not a problem (rev.)

When used in the paper, the score is reversed so that a higher value indicates *less* shame in response to vignettes.

Observed shame. Enumerator is asked (after asking questions about menstruation) about the girl's behavior over the last 5 minutes of the survey:

- Was the girl looking at the floor when she spoke?
- Did the girl speak: (Quietly or very quietly)
- Did the girl laugh while answering or listening to the questions?
- How did the girl answer the questions? (Hesitated a bit or lots)
- Did the girl seem: (Uneasy or slightly uneasy)
- Did the girl seem: (More uneasy when menstruation was mentioned than other subjects)

When used in the paper, the index is constructed so that a higher value indicates *fewer* observable markers of shame.

Network. Girls are asked the following questions:

- Name your friends (how many girls; how many boys)
- Compared to the last school year, this year did you have [more friends vs less friends; girls and boys separately]
- Who are the girls or boys you play with at school? (how many girls; how many boys)
- Who are the girls or boys with whom you share the way to school? (how many girls; how many boys)

- Who are the girls or boys you talk to during recess? (how many girls; how many boys)
- Are there any girls or boys with whom you organize to clean or do the garden at school? (If yes: how many girls; how many boys)
- Do you sometimes share things (snacks, bikes, money to buy a snack, etc.) with your friends? (If yes: how many girls; how many boys)
- How often do you share with these people? (Every day, several times a week, once a week, once every 2 weeks, once a month, less than once a month)
- Do you sometimes do homework or school projects together with your classmates?
- Do your classmates help each other with school work in other ways?
- How often? (Every day, several times a week, once a week, once every 2 weeks, once a month, less than once a month)
- Who are the girls or boys you talk to about menstruation? (how many girls; how many boys)

Bullying. Girls are asked about a random subset of 2-3 other girls in the same grade-level as them in their school. If the respondent doesn't know the other girl being asked about, we ask about additional girls (up to a maximum of 5). Only girls whom the respondent knows are included in the results. Girls are also asked about bullying towards themselves. Teachers are asked about a random subset of 3 girls in their class.

The indexes are constructed so that higher values indicate *more* or more severe bullying.

Light teasing questions include:

- How often has at least one other classmate laughed at [NAME] in the past month? (Never, rarely, some days, most days, almost every day)
- How often has at least one other classmate teased [NAME] in the past month? (Never, rarely, some days, most days, almost every day)
- Please indicate how many other classmates tease [NAME]?
- Please indicate the severity of the teasing she receives from class members? (No teasing, light teasing, moderate teasing, serious teasing)

Severe intimidation/harassment questions include:

- How often has at least one other classmate intimidated/harassed [NAME] in the past month? (Never, rarely, some days, most days, almost every day)
- Please indicate how many other classmates intimidate/harass [NAME]? (Never, rarely, some days, most days, almost every day)
- Please indicate the severity of the intimidation/harassment inflicted on her by class members? (No intimidation, light teasing, light bullying, moderate intimidation, serious intimidation such as violence)

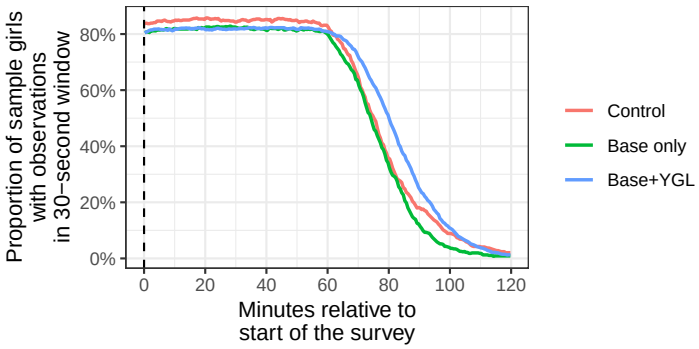
B.2 Heart rate data

We use the *Scosche Rhythm 24* heart rate wristband monitor to measure girls' heart rate during the endline survey. This has been shown to have comparable accuracy to chest-strap monitors (Reece et al., 2021), but has the advantage of being significantly less invasive. Qualitative reports from enumerators indicate that girls frequently forgot that they were wearing the wristband by the end of the survey, suggesting that girls were not overly distracted or stressed by wearing the wristband itself.

Sample. In some cases, heart rate monitors were not fixed properly to girls' wrists, or the battery of monitors ran out on the field. Because of this, heart rate data is available for only 1904 girls. However, the proportion of data available is equal across treatment arms (Table A5, column 6).

Time-window restrictions. For the main results in Figure 2, panel (a), we include all observations less than 120 minutes after the start of the survey. This includes almost all observations across all arms (see Figure B34). We do not cut off at an earlier window in the main specification, because the average length of the survey is different across different treatments (e.g., because girls in *Base+YGL* schools are asked about the YGLs in their schools in a separate module), so any earlier cutoff could generate composition effects in the heart rate data. However, for tests of the *dynamics* over time (Figure 2 panel (b), and Table A17), we use a 60 minute cutoff window. This avoids highly noisy time periods with few observations driving the results, since the number of girls with data available drops rapidly after 60 minutes because many girls finish the survey (see Figure B34).

Figure B34: *Number of girls with heart rate observations available in each 30-second window*



Notes: Shows the proportion of sample girls in each treatment arm who have heart rate data available in each 30-second window of the survey.

B.3 Baseline variable definitions for LASSO

- *Girl's age (years):* Age of the girl in years.
- *Girl has a brother (=1):* Indicates whether the girl has a brother.
- *Distance to school (km):* Self-reported distance from the girl's home to her school in kilometers.
- *Time to school (min):* Self-reported time it takes for the girl to travel from home to school in minutes.
- *Was absent last 3 months (=1):* Indicates whether the girl was absent from school in the last three months (self-reported).
- *Days absent last 3 months:* The total number of days the girl was absent from school in the last three months (self-reported).
- *Years of school delay:* The number of years the girl is behind in her schooling compared to the typical age for her grade-level.
- *Father in household (=1):* Indicates whether the girl's father lives in the household.
- *Mother in household (=1):* Indicates whether the girl's mother lives in the household.
- *Knows how to read and write (=1) (guardian):* Indicates whether the girl's guardian can read and write.
- *No education (guardian) (=1):* Indicates whether the girl's guardian has no formal education.
- *Secondary or higher (guardian) (=1):* Indicates whether the girl's guardian has at least a secondary education.
- *Occupation: Agriculture (guardian) (=1):* Indicates whether the girl's guardian is employed in agriculture.
- *Household size:* The total number of people living in the girl's household.
- *Menstruation inhibiting school activity index (Z):* The sum of positive responses to whether the girl reported (over the last 3 months): (i) having missed school because of her period; (ii) having left school during the day because of her period; (iii) having not fully participated in school because of her period; (iv) having missed a social activity because of her period.
- *Menstrual health problems index (Z):* The sum of positive responses to whether the girl reported having had (over the last 3 months) (i) menstrual cramps, (ii) a urinal infection, and (iii) other health problems linked to menstruation.
- *Wealth index (Z):* is constructed from household module on asset ownership during the mother's interview.

Other indexes are constructed analogously to the endline outcomes documented in Section B.1.

B.4 Other variable definitions

Social desirability bias. At baseline, we elicited a shortened version of the Crowne & Marlowe (1960) social desirability scale for the sample of girls, translated into the local language, Malagasy. The questions ask whether the respondent has a number of "too good to be true" traits (or the reverse), so that answering in the socially desirable way is likely to indicate a desire to please the interviewer. The questions used were:

1. It is sometimes difficult for me to continue my work if I am not encouraged. (-)
2. I sometimes feel resentful when I don't get what I want. (-)
3. On a few occasions, I have given up on doing something because I thought too little of my abilities. (-)
4. There were times when I wanted to rebel against the authorities, even though I knew they were right. (-)
5. No matter who I talk to, I always know how to listen.
6. There have been occasions when I have taken advantage of someone. (-)
7. I'm always willing to admit when I make a mistake.
8. I sometimes try to take revenge rather than forgive and forget. (-)
9. I am always polite, even to unpleasant people.
10. I have never been irritated when people expressed ideas very different from mine.
11. There were times when I was quite jealous of other people's good fortune. (-)
12. I sometimes get irritated by people who ask me for favors. (-)
13. I have deliberately said something that hurt someone. (-)

We construct an index of each girl's social desirability score using factor loadings, and then construct *High SDB*, which indicates whether a girl has an above-median social desirability score. This is the variable we use as an interaction in Table A7.

YGL positive deviance index. During the YGL selection process, all girls nominated by teachers as potential candidates were asked a series of survey questions to evaluate their suitability for the role. The positive deviance index is constructed using a factor-analysis combination of the following questions:

1. How often do you tell your friends about new things you have learned and new experiences?
2. How much do you agree with the following statement: "If I learn new knowledge that I believe is good for me, I am ready to adopt it even if it is very different from my habits and those around me (for example, your friends, your sister, your classmates)."
3. If someone talks to you about a sensitive subject (like menstruation or sexual health), do you feel comfortable and open?
4. During the previous year, have you discussed and expressed your opinions in front of those around you about a sensitive subject such as menstruation or sexual health?
5. During these moments of discussion, overall, did you feel comfortable discussing the sensitive subject?
6. Imagine that you are discussing and expressing your opinions to those around you on a sensitive subject such as menstruation or sexual health. If the people around you feel uncomfortable or show embarrassment, frustration or anger, what action(s) would you take? (Options: I would insist / try to convince them more; I would ask them to talk to someone else; I would suggest another time to speak; I would give up)

7. Would you feel comfortable explaining how to use a sanitary pad in front of a group of *all the girls in your class*?
    - (If no) Would you feel comfortable explaining how to use a sanitary pad in front of a smaller group of girls?
    - (If yes) Would you feel comfortable explaining how to use a sanitary pad in front of a larger group of girls that also includes girls in other classes?

To construct a school-level indicator of YGL's positive deviance, we take the average positive deviance of all YGLs selected in the school in the first year of the program. To construct the *relative* measure of YGL's positive deviance, we take this measure and subtract the average positive deviance of all the candidates who were in the potential pool of YGLs (whether they were selected or not).

B.5 Tracking and attrition weights

Tracking strategy. Regular data collection occurred between May and July 2023. This phase included repeated visits to absent girls, mothers, and teachers. An intensive tracking phase was then organized in October 2023 following a multi-pronged approach to minimize attrition.

We revisited all schools where there was still a relatively large share of girls to track after the regular data collection for any of the following reasons: (i) schools where many girls had been absent during regular tracking because of seasonal work in mines, and where girls were expected to return to their homestead in October; (ii) schools where there were many refusals because the regular survey visit was too close to exams, and girls asked the team to return later; (iii) schools for which the survey team did not have enough time for the girls' survey during the regular phase; and (iv) schools where fewer than two teachers were surveyed in the regular tracking phase.

In addition to this set of schools, we randomly selected 19 of the remaining 62 schools where at least one girl was missing, and included these 19 schools for intensive tracking. This resulted in a total sample of 83 schools to be revisited.

In the revisited schools, all remaining girl, mother, and teacher surveys were targeted for completion. Some teams targeted the original school locality, while in parallel, other teams looked for girls who had moved in their migration destination. This latter step involved in-person interviews in common migration destinations (e.g., the capital city Antananarivo, and regions nearby the study region of Amoron'i Mania), along with phone surveys (for 3.7% of the sample) in the remaining destinations.

Attrition weights. Given the random selection of a subset of schools for intensive tracking, we can calculate *effective tracking rates* (Orr et al., 2003). Intensive tracking was successful in finding a very large share of targeted girls, leading to an effective tracking rate of 99%.

We calculate attrition weights that account for the random selection. Attrition weights are 1 for all girls found during regular tracking, or in schools that were selected for revisits during intensive tracking with probability 1 (as described above). Attrition weights are 62/19 for the girls in the additional schools randomly selected for intensive tracking.

Among the schools randomly selected for intensive tracking, 9 had a missing teacher, among the

total of 38 remaining schools with missing teachers, so the attrition weight for teachers is set at 38/9 for teachers from these schools, and 1 for other teachers.

The results adjusting for attrition weights are shown in Table A11 and Table A12.

C Spillovers

Our sample of 140 schools was selected to avoid cases where schools were very close to each other. Nevertheless, since these schools were selected from only 3 districts in the region of Amoron'i Mania, schools and households are in some cases close enough that there is a risk of spillover effects from treatment schools to control schools.

To evaluate whether such spillovers threaten our main results, we estimate whether control girls who live within a certain radius of other sampled treatment girls have different outcomes. We use data on the GPS coordinates of the girl's survey at either baseline or endline.[35] We then use the following specification for girl i in school j:

$$Y_{ij} = \beta_0 + \beta_1 BaseOnly_j + \beta_2 BaseYGL_j + \delta(Control_j \times AboveMedTreatedWithin2km_{ij}) + \mathbf{X}'_{ij}\Gamma + \varepsilon_{ij}$$

where $Control_j = 1$ for girls in control schools, and 0 otherwise; $AboveMedTreatedWithin2km_{ij}$=1 if girl i has more than the median number of treated girls in our sample who were interviewed within 2km of i's own interview, and 0 otherwise.[36] The coefficient δ tests whether outcomes are different for control girls with more than 3 treated girls in the sample within a 2km radius, where 3 is the median number of treated girls within 2km. $AboveMedTreatedWithin2km_{ij}$=1 for 273/540 (51%) of control girls.

We do not randomly vary the proportion of treated schools in a given locality, so this specification does not generate *causal* estimates of treatment spillovers. For example, $AboveMedTreatedWithin2km_{ij}$ is likely to be positively correlated with population density in an area, so our estimates of δ may partially pick up the effects of such density.

We find no consistent evidence in favour of spillovers that would undermine our main results (Table C1). There are significant negative coefficients on grade progression and on observed hygiene for girls, but not on other related proxies of these measures (learning, hygiene knowledge, or hygiene behavior).

[35] Due to errors with the survey tablets, only 56% girls had GPS coordinates available at baseline. We use these where available, and use endline coordinates if they are not available.

[36] All other variables are defined analogously to Equation 1. We use the *number* rather than the *proportion* of treated girls, because there are many girls who have 0 girls from other schools within 2km, meaning that the denominator would be 0.

Table C1: *Spillovers: heterogeneity by whether control group have above-median treated individuals living within 2km*

	Control & above med. treated in 2km	N
Learning test (Z)	–0.008	2167
	(0.093) [0.932]	
Progressed one class (=1)	–0.112*	2256
	(0.061) [0.064]	
Enrolled (=1)	0.019	2256
	(0.035) [0.591]	
Absent in last 3 months (=1)	–0.053	1824
	(0.055) [0.331]	
General health (Z)	–0.128	2256
	(0.087) [0.140]	
UTI health (Z)	–0.078	2256
	(0.069) [0.258]	
Mental health (Z)	–0.175	2256
	(0.132) [0.184]	
Self-esteem (Z)	–0.140	2256
	(0.098) [0.152]	
Heart rate (Z)	0.010	1904
	(0.098) [0.918]	
Hygiene knowl. (Z)	0.098	2167
	(0.098) [0.318]	
Menstr. knowl. (Z)	0.013	2167
	(0.103) [0.898]	
Hygiene behavior (Z)	–0.086	2167
	(0.127) [0.497]	
Menstr. behavior (Z)	0.108	2256
	(0.083) [0.193]	
Observed hygiene: Girls (Z)	–0.285***	2167
	(0.091) [0.002]	
Combined stigma index (Z)	–0.036	2256
	(0.089) [0.682]	
Received info. (Z)	–0.124	2256
	(0.111) [0.264]	
Willingness to speak (Z)	–0.137	2250
	(0.101) [0.173]	
Attitudes (Z)	0.029	2254
	(0.094) [0.760]	
Norms (Z)	0.032	2167
	(0.083) [0.699]	
Shame response to vignettes (rev.) (Z)	0.079	2165
	(0.093) [0.397]	
Observed shame (rev.) (Z)	–0.117	2256
	(0.091) [0.201]	
Bullying: teasing (Z)	0.113	2167
	(0.112) [0.311]	
Bullying: intimidation/harassment (Z)	0.084	2167
	(0.079) [0.293]	
Network (Z)	–0.123	2167
	(0.139) [0.376]	

Notes: Shows indications of spillovers within 2km of girls' homes. Each row represents a regression of the outcome on *Base Only*, *Base+YGL*, and *Control & above med. treated in 2km*. Coefficients on *Base Only* and *Base+YGL* are not shown. *Control & above med. treated in 2km* = 1 when a girl is in the control group, and the number of girls from a treated school interviewed within 2km was above median. The median is 3. *Control & above med. treated in 2km* = 1 for 273/540 (51%) control girls.