report

ADOLESCENT GIRLS EMPOWERMENT PROGRAMME

RESEARCH AND EVALUATION MID-TERM TECHNICAL REPORT

OCTOBER 2016

Corresponding Authors

Karen Austrian, PhD
AGEP Principal Investigator
Associate
Population Council, Kenya
General Accident House – 2nd Floor
Ralph Bunche Road
PO Box 17643-00500
Nairobi, Kenya
Tel: +254-20-2713480
kaustrian@popcouncil.org

Paul C. Hewett, PhD
AGEP Research Lead Investigator
Senior Associate
Population Council, Zambia
4301 Connecticut Avenue
Washington, D.C. 20008
Tel: +1 202 237 9419
phewett@popcouncil.org

Contributing Authors

Erica Soler-Hampejsek, PhD
Associate
Population Council, New York

Fiammetta Bozzani, MSc
Research Fellow
LSHTM, London

Jere R. Behrman, PhD
Professor of Economics
University of Pennsylvania

Jean Digitale, MPH
Data Analyst/Research Coordinator
Population Council, New York

POPULATION COUNCIL
Ideas. Evidence. Impact.

The Population Council confronts critical health and development issues—from stopping the spread of HIV to improving reproductive health and ensuring that young people lead full and productive lives. Through biomedical, social science, and public health research in 50 countries, we work with our partners to deliver solutions that lead to more effective policies, programmes, and technologies that improve lives around the world. Established in 1952 and headquartered in New York, the Council is a nongovernmental, nonprofit organization governed by an international board of trustees.

Population Council, Zambia
Plot 3670, No. 4
Mwaleshi Road, Olympia Park
Lusaka, Zambia
Tel: +260 211 295 925
email: info.zambia@popcouncil.org

© 2016 The Population Council, Inc.

Suggested citation: Austrian, Karen, Paul C. Hewett, Erica Soler-Hampejsek, Fiammetta Bozzani, Jere R. Behrman, and Jean Digitale. 2016. *Adolescent Girls Empowerment Programme: Research and Evaluation Mid-Term Technical Report.* Lusaka, Zambia: Population Council.

For more information about the design of AGEP and the evaluation, please refer to the following:

Hewett, Paul C., Karen Austrian, Erica Soler-Hampejsek, Jere R. Behrman, Christine A. Kelly, Dela Kusi-Appouh, Fiammetta Bozzani, Barbara S. Mensch, and Minyoi Maimbolwa. 2014. *Adolescent Girls Empowerment Programme: Research and Evaluation Baseline Technical Report.* Lusaka, Zambia: Population Council.

Table of Contents

Acknowledgments iii
List of Abbreviations iv
Executive Summary 1
Background 9
Adolescent girls in Zambia 9
AGEP Programme Summary 11
Theory of change 11
AGEP population, sites, and timeline 12
Core components of AGEP in Zambia 13
Safe Spaces implementation 15
Health voucher implementation 24
Savings account implementation 28
AGEP Research and Evaluation 32
Randomised cluster design 32
Adolescent survey 34
Assessed adolescent girl outcomes and impact indicators 36
Statistical analysis and evaluation assessment 39
Economic evaluation 40
AGEP Mid-Term Results 45
Data quality assessments 47
Impact of AGEP on girls' empowerment 53
Impact of AGEP on longer-term outcomes 68
Assessing the impact of the programme by study arm 81
Assessing contamination from spillovers of the programme to controls 82
Economic evaluation 82
Conclusions 89
Annex A: List of Safe Spaces Training Session Topics 92
Annex B: Mentor Evaluation Results 95
Annex C: Intent-to-Treat (ITT) and Treatment-on-Treated (ToT) Summary and Detailed Results
Excel tables available at: http://www.popcouncil.org/uploads/pdfs/2016PGY_AGEPMidTermReport_AnnexC.xlsx 109
Annex D: Summary of Key Lessons Learned 110

Acknowledgments

This mid-term report would not be possible without the hard work of diverse programme and research teams dedicated to improving the lives of vulnerable adolescent girls.

At the Population Council: Natalie Jackson Hachonda and her team of programme officers, programme coordinators, and site coordinators, who successfully implemented an expansive and complex intervention and helped to prepare the communities for research activities as well. We thank Chama Mapulanga, Kevin Kapungu and Mary Namukoko who served as study coordinators, Dominic Nzala and Paul Banda who served as fieldwork coordinators, Minyoi Maimbolwa who served as data manager, Oscar Mbulwe and Jerrix Nkhata who supervised the biomarker collection, Rogers Musonda who served as the study cartographer, Memoona Malik who supervised the anthropometric measurement, and the entire group of team leaders and research assistants who spent so many days in the field collecting data. We would also like to acknowledge the efforts of the finance and administrative team at Population Council Zambia, including Masida Chisambi, Swedish Sianyaka, Khukhie Makukisi, and Memory Sichizuwe. Additional thanks to Samir Souidi and Stan Mierzwa at the Population Council headquarters for the development of the tablet-based electronic data capture tools.

We also thank Zoe Duby and Chipo Zulu for the analysis and write-up of the second round of qualitative data included in this report, Monica Caudillo for the analysis and write-up of the mentor-quality data, and Yana Mayevskaya for the analysis of the health voucher usage. Many thanks to Ann Blanc, Natalie Jackson Hachonda, and Barbara Mensch at the Population Council for feedback on draft versions of the analysis and report and to Barbara Miller for her gracious editing.

We also thank the Ministry of Health and the Ministry of Community Development and Social Welfare for their support of AGEP overall, and the research specifically. Special thanks to Dr. Peter Mwaba, Dr. Elizabeth Chizema, Dr. Maximillian Bweupe, Dr. Caroline Phiri, Dr. Mary Nambao, Dr. Angel Mwiche, and Mrs. Mable Mweemba at the Ministry of Health and Dr. Davy Chikamata and Mr. Yengwe Kakusa at the Ministry of Community Development and Social Welfare.

At the YWCA Zambia, we would like to acknowledge the efforts of Bernard Munkombwe, Debbie Chingobe Kangombe, and all the site coordinators who were also instrumental in the preparation and implementation of research and programme activities at the field level.

The authors thank the UK Department for International Development (DFID) and those from DFID who provided feedback on the draft report. We especially thank Valerie Roberts who has championed the Adolescent Girls Empowerment Programme from its inception. DFID's investment in rigorous research will benefit not only Zambia, but the field of adolescent girls programming as a whole.

Finally, we are grateful to the adolescent girls who shared their thoughts and experiences with us and thank them for participating in both the programme and the research.

List of Abbreviations

ACASI	Audio Computer-Assisted Self-Interviewing
AGEP	Adolescent Girls Empowerment Programme
CAPI	Computer-Assisted Personal Interviewing
CSA	Census Supervisory Area
CSO	Central Statistical Office
DCHO	District Community Health Office
DHS	Demographic and Health Survey
DID	Difference in Differences
HC	Health Centre
HIV	Human Immunodeficiency Virus
HSV-2	Herpes Simplex Virus – type 2
ICERs	Incremental Cost-Effectiveness Ratios
IDI	In-Depth Interviews
ITT	Intent to Treat
MCDMCH	Ministry of Community Development, Mother and Child Health
MCDSW	Ministry of Community Development and Social Welfare
MOH	Ministry of Health
NatSave	National Savings and Credit Bank
NRC	National Registration Card
ODK	Open Data Kit
OLS	Ordinary Least Squares
PCA	Principal Components Analysis
RCD	Randomised Cluster Design
RED	Research Evidence Division
RHC	Rural Health Centre
SEA	Standard Enumeration Area
SRH	Sexual and Reproductive Health
STI	Sexually Transmitted Infection
TOT	Training of Trainers
ToT	Treatment on Treated
UNZA-BREC	University of Zambia, Biomedical Research Ethics Committee
YWCA	Young Women's Christian Association

Executive Summary

The Adolescent Girls Empowerment Programme (AGEP) was a social, health, and economic asset-building programme targeting vulnerable adolescent girls aged 10–19. AGEP was implemented at 10 sites in four of the 10 provinces of Zambia, including Lusaka, Central, Copperbelt, and North-Western. The core of AGEP was weekly "Safe Space" girls' group meetings that were conducted over the course of two years. The girls' groups were comprised of 20 to 30 girls who met with a trained mentor—a young woman from their community; the groups were also segmented by age and marital status. The girls were provided short training sessions on a range of health, life-skills, and financial education topics, as well as a chance to discuss important experiences of the past week. The primary goal of the Safe Space groups was to reduce social isolation and build assets that would facilitate positive change in the intermediate and longer- term for adolescent girls as they transition to adulthood.

Two additional components were added on top of the Safe Spaces meetings in the AGEP programme. The first component of AGEP was that selected girls were provided a health voucher that was redeemable for a package of health services at certain public and private health providers in their communities. The health services covered by the voucher included basic wellness exams and age-appropriate sexual and reproductive health services. The second component was a Girls Dream savings account at the National Savings and Credit Bank (NatSave). Offered to select girls within AGEP, the savings account was tailored to adolescents and the programme facilitated the process of opening bank accounts.

Programme implementation results

The Population Council, in partnership with the Young Women's Christian Association of Zambia (YWCA-Zambia), successfully implemented the AGEP programme from late 2013 to early 2016. Although all 10 AGEP sites completed a full two years of the programme, the sites were staggered in their start and end dates as the programme rolled out sequentially by site. Over 240 mentors in AGEP guided 11,390 adolescent girls aged 10–19 at baseline into the programme and through a planned two years of programming, accumulating to approximately 40,884 meetings and 115,200 hours of mentor effort. While the AGEP Safe Spaces were very successful overall in reaching and providing programming to vulnerable adolescent girls, participation rates in the Safe Space groups were not as high as desired, with only approximately one out of every three girls invited to the programme attending more than half the AGEP group sessions; younger rural adolescents were most likely to attend, while older urban adolescents were less likely to; although the difference is only three percentage points. To address participation, AGEP adapted and responded by adding prizes for attendance, fun days, and increased community sensitization, which was thought to increase excitement for the programme. Initial programme uptake was affected by the recruitment process, as the impact evaluation required household-based recruitment of select girls rather than community-based outreach, with the latter leading to more interested girls self-selecting to participate in the programme, and hence, higher participation rates. Further, due to potential contamination of control sites, which were often located geographically nearby, large community-based events to raise awareness, interest, and excitement for AGEP were not possible, reducing the potential for considerable community support. Ongoing programme participation was affected

particularly by competing interests and activities for the girls, girls migrating out of programme sites, and loss of enthusiasm for the programme.

An evaluation of mentor quality was conducted that defined different dimensions of mentor quality, as well as assessed the impact of mentor quality on programme impacts. Girls with mentors who had positive attitudes towards contraception were less likely to have ever been pregnant; girls with mentors who scored high on "safe-space creation" were less likely to have been married, had sex, had an unwanted pregnancy, or given birth and girls with mentors who scored high on the self-efficacy score were less likely to be HIV positive and have had unwanted sex. Finally, girls who had a mentor who scored high on a "relationship with girls and community" score were also less likely to have had unwanted sex.

In collaboration with the Zambian Ministry of Health, provincial and district health offices, AGEP successfully established and operationalized the AGEP health voucher platform providing adolescent-friendly training to health service providers and facilitating access to general and sexual reproductive health services for adolescent girls in the programme. The AGEP health voucher was rolled out in the AGEP sites approximately one year after the initiation of the programme and will continue through a second year, even after the end of the AGEP Safe Space groups. Hence, the full impact of the voucher cannot be assessed in the mid-term report, as it precedes the end of the health voucher component. While the health voucher was received by 5,789 adolescent girls, this represented approximately three-quarters of those who were eligible to receive the voucher because it was distributed only to girls actively participating in AGEP by the date of rollout. For those girls who did receive the voucher, qualitative data suggest that it was empowering for girls who used it, providing needed confidence in accessing services, paying for medicine, and receiving respect from health service providers. While use of the vouchers was limited to one out of every five girls who received them, approximately one-third of girls who used the voucher did so to receive sexual and reproductive health care services. The modest use of the voucher was attributed to low demand for services, likely due to the relative healthiness of adolescent girls, social norms regarding premarital sex, and the perception that health facilities entail long lines and wait times for services.

The Girls Dream savings account at National Savings and Credit Bank was made available to one-third of AGEP girls who had joined Safe Space groups approximately eight to 10 weeks after group initiation; girls who received the account were randomised to receive it as a requirement of the impact evaluation. Approximately half of all girls who were eligible opened a bank account, although girls who were more highly engaged in AGEP were significantly more likely to do so, with approximately three-quarters of such girls opening an account. Reflecting AGEP programme participation in general, younger rural adolescents were more likely to open an account. Given that the programme facilitated the process and travel requirements, the barrier of distance for rural girls was eliminated, at least for the opening of the account, as rural adolescents had fewer interactions with the account after it had been established. Despite the fact that overall account usage remained low throughout programme implementation, as revealed in the evaluation results, those girls who opened an account were more likely to have saved in the previous year, partially a function of higher informal savings at home. It is hypothesized that access to the bank account reinforces the information and perceived value of the financial education provided with in the AGEP Safe Space groups.

Impact evaluation methods

In addition to programme implementation, a rigorous impact evaluation was embedded whose objective was to assess the impact of AGEP on adolescent girls' outcomes. A randomised cluster design with four study arms was designed and implemented to assess impact. Communities where AGEP was being implemented were randomly assigned to one of three arms of the intervention or to a control arm. The randomised arms of the evaluation included: 1) safe spaces only, 2) safe spaces with a health voucher, 3) safe spaces, the health voucher, and the savings accounts, or 4) a control arm. Analyses were to be conducted immediately after the AGEP programme ended (i.e., after two years) and will be conducted an additional two years later (i.e., after four years). The primary objective of the analysis after two years was to assess the impact of AGEP on the components of girls' empowerment that were believed to be directly influenced by the programme in the shorter term, i.e., the mediating measures of girls' social, economic, and health assets. The primary objective of the final analysis that will be conducted after four years is to assess the impact on adolescent girls' longer-term demographic, reproductive, and health outcomes. The results presented in this mid-term report focus on the first of these analysis, specifically the impact of AGEP on girls' empowerment, although the report also presents preliminary results of the impact on longer-term outcomes after two years.

The statistical analyses of programme impact presented in this report are an "intent-to-treat" (ITT) and a "treatment-on- treated" ( ) analysis. The ITT analysis estimates the average effect of the programme on the adolescents in clusters randomised to AGEP relative to girls in clusters randomised to the control, based on their original randomisation and irrespective of actual participation in the intervention. Randomisation assures that the estimate of impact is unbiased and therefore provides the highest degree of confidence in attributing any differences in observed outcomes specifically to AGEP. As the ITT analysis does not account for the fact that many girls did not choose to participate in AGEP or had different levels of participation, a secondary ToT analysis was conducted using indicators of impact that measured actual programme participation. The ToT analysis conducted here controls for the potential selectivity of girls choosing to participate in AGEP through an instrumental variables estimation approach. The ToT is expected to reveal a stronger programme impact than the ITT if substantial numbers of girls choose not to participate, but the programme has an impact for those who do participate.

In parallel to assessing the impact of AGEP, an evaluation of the cost-effectiveness of the programme was conducted to facilitate comparisons between AGEP and other programmes whose objectives are to improve similar outcomes for adolescent girls. It also allows for an assessment of the incremental cost-effectiveness of each of the components or arms of AGEP. To execute the economic evaluation, direct programme costs were collected from AGEP budgets and financial reports and included both start-up and programme delivery costs for the Population Council and its partners. A decision analytic model was constructed to generate estimates of the incremental costs per negative health outcome averted and positive progress achieved on non-health indicators from participating in AGEP. The focus of the economic evaluation analysis in this mid-term report will centre on the question of cost-effectiveness of all arms of AGEP compared to not implementing any programme. It should be noted, however, that the value of the economic evaluation is dependent upon a significant showing of benefits of the programme for participants.

Impact evaluation results

A total of 5,235 respondents completed the AGEP baseline (Round 1) survey. Approximately, 90% of the adolescents interviewed at baseline were re-interviewed a year later in Round 2, and 89% of Round 1 participants were re-interviewed in Round 3. The baseline characteristics of girls, as well as the loss-to-follow-up over time were well balanced across study intervention arms and controls, meaning that the randomisation by study arm was effective at evenly distributing the characteristics of the sample and that the programme did not differentially affect the study's ability to track study participants. The impact results presented in the mid-report focus on the change that occurred between Round 1 and Round 2 (after the first year of the programme), between Round 2 and Round 3 (after the second year of the programme), and between Round 1 and Round 3 (cumulative effect of the programme). The results are also provided by the various age and residential groupings, including younger urban girls, young rural girls, older urban girls, and older rural girls; as delineated by their ages (10–14, 15–19) and residencies (urban, rural) at baseline. Additional results are presented for each of the three intervention arms separately, as compared to the control cases and each other. The mid-term results are first discussed in reference to their impact on girls' empowerment and then turn towards a preliminary assessment of impact on what are considered longer-term outcomes for girls.

Empowerment is defined within AGEP as the condition of possessing the assets and capacities that allow adolescent girls to maximize the opportunities they might encounter during the early and later adolescent years. As reflected by AGEP's theory of change, the programme was designed to build upon and expand these assets through the Safe Spaces meetings and provision of a health voucher and a bank account. Empowerment was also seen as a key ingredient to improving the longer-term adolescent outcomes. The indicators used to measure empowerment across the social, economic, and health domains are presented in the report in Table 12 below. While not exhaustive of every possible dimensions of the multifaceted nature of girls' empowerment, the AGEP research instrument was designed to capture what were thought to be the key mediating factors underlying the theory of change.

In measuring girls' empowerment in the three rounds of data that were analysed for the mid-term report, three themes emerge. The first theme was that adolescent girls on average were not absent of assets prior to the initiation of the AGEP and that, in fact, on some asset indicators could be said to have possessed higher levels of empowerment than their vulnerability would have suggested. For instance, girls on average had relatively high levels of self-efficacy at baseline, agreeing to statements that indicated that they felt they were able to manage problems faced, overcome difficulties, find solutions, and accomplish goals. Also, girls expressed confidence in their ability to change outcomes based on any plans they make and that they were proficient in making good decisions regarding the use of money. Girls also were shown to possess basic financial literacy skills at baseline. Of course, it should be noted that averages hide potentially large numbers and percentages of girls who do not possess such assets.

A second theme of assessing the measures of empowerment across the survey rounds in the evaluation is that a notable degree of change was observed in the assets over time, but that the change was not solely due to the impact of AGEP. There may be a range of factors driving change in girls' empowerment. For instance, as girls age they may acquire new capabilities, acquire new information, or hone the skills they already possess. Adolescents may also be

exposed in their communities to other programmes, interventions, and/or information similar in nature to that provided through AGEP. Further, as they become older, girls may be provided or take on new responsibilities and challenges that naturally enhance their sense of efficacy, skills, autonomy, and control. Across many empowerment indicators, the AGEP programme enhanced the development of girls' social, economic, and health assets, although in many cases, despite being positive, they did not reach the level of statistical significance.

A third observation was that a set indicators remained relatively impervious to any change over time, specifically those regarding the perceptions of gender equality and norms regarding gender-based violence against women and girls. For instance, the notion that boys and girls, men and women are similar in their innate characteristics, such as intelligence, or in norms concerning who should attend school or make decisions in the household remained unchanged, whether among AGEP or among control girls. This was the case for AGEP participants even though the health and life-skills curriculum specifically addressed gender roles, gender-based violence, and human and children's rights. The lack of change in perceptions of gender, gender roles, and violence against women points to the deeply ingrained nature of these normative attitudes and beliefs and the need to potentially complement Safe Space group trainings with additional community-based interventions that can serve to reinforce girls' own assessments. It is possible, for example, that while many internalized assets are more malleable to a direct girl-based asset-building approach, perceptions of what is normative or external to the girls themselves in the enabling environment require appropriate messaging to come from actors in that environment, e.g., boys, men, families, and communities.

The primary impact results for the empowerment indicators were summarized in the mid-term report in Tables 18 (ITT) and 19 (ToT), while the detailed results, including the impact coefficients and p-values are provided in Tables C-2 and C-4; in most cases, the results of the ITT and ToT reinforce one another providing confidence in the robustness of the statistical assessment and conclusions drawn overall. Of the range of indicators that were used to measure girls' empowerment in the evaluation, a total of six indicators (two in each of the three asset domains) in the ITT and ToT revealed statistically significant differences at $p < .05$ among girls participating in AGEP (all arms) and girls in the control clusters. All of these six indicators show, as hypothesized, greater positive improvement, whether among girls who were randomised to receive AGEP or among those who had actually participated intensely in the programme. For instance, there was an increase in girls' access to a place in the community where they felt safe to meet their friends apart from school and home. Also, girls in AGEP had made significantly greater improvements in financial literacy over time. One additional indicator (self-efficacy) was statistically significant different between Round 1 and 2, but control girls had caught up by Round 3. Overall, the impact results on girls' empowerment due to the programme were modest as measured immediately after the end of two-year AGEP program period. It is possible that AGEP has, however, set girls on a different trajectory of further acquisition of assets and hence the cumulative impact of AGEP on empowerment may be revealed in later rounds of observation.

For the longer-term outcomes, in both the ITT and ToT analyses the impact of AGEP is already apparent in two of the indicators. These results are presented in summary form in Tables 20 and 21 and in detail in Tables C-3 and C-4. There was no significant difference between the AGEP and control girls on educational attainment, timing of marriage, pregnancy and birth,

experience of violence, or HIV/HSV-2 prevalence. AGEP girls were, however, significantly less likely to engage in transactional sex and more likely to use a condom at first sex than girls who were not exposed to the programme. Between Rounds 1 and 3, the percentage of girls in AGEP who had engaged in transactional sex was 13% less than girls in the control group; notably, nearly half the girls in the sample who had initiated sexual activity had reported transactional sex. AGEP girls between Round 1 and Round 3 were also more likely to use a condom at first sex, by approximately five percentage points. Although lower at 4% points and not statistically significant overall between Round 1 and 3, condom use at last sex with a nonmarital partner was also higher among AGEP girls compared to control girls. Additional years of observation will reveal if these impacts are able to translate into delayed pregnancy and STI acquisition. While the majority of longer-term indicators were not statistically significant after two years, it should be noted that the study was statistically powered based on the full four years of observation that will be made on these indicators, as such it would be inappropriate to conclude that AGEP had no effect on these indicators because of their lack of statistical significance in these analyses.

Analysis of the impact of AGEP on girls' empowerment was also conducted for each intervention arm separately against the control arm. These results are presented in summary form in Table 22 and in detail in Table C-5. The study was powered to assess differences between the intervention and control arms, as well as between each intervention arm separately. This latter fact provides an opportunity for the cost-effectiveness evaluation to offer an assessment regarding the cost-effectiveness of the AGEP components if scaling or expanding the programme to other settings is an objective. The study results by arm parallel the overall impact of the programme, in that most of the statistically significant findings are consistently significant across intervention arms relative to the control arm, suggesting that the impact is attributable to the safe space component. There are two indicators (% having saved in the past year, % having saved more than 20 kwacha) in which statistical significance is observed only in the arm that includes the additional bank account component. Hence, one may conclude that the addition of having a bank account, whether or not it is used very much, increases the propensity to save; this result is back up by the observation from the data that girls with the bank account are accumulating their savings informally as much as they are formally through the use of the account.

The economic evaluation looked at the implementation cost per beneficiary (Table 24), which totalled $394 per girl for the Safe Space groups, an additional $293 per girl for the health voucher, and an additional $551 per girl for the savings account (adjusted for inflation to 2016 $US). A large component of the Safe Space group costs were staff costs of implementing and monitoring the groups and the per diems that were paid to mentors over the two years of the programme. The additional costs associated with the bank account were largely related to transporting all of the girls to the NatSave branches to receive an introduction and to open the bank accounts. It is expected that a scaled-up version of the programme would be cheaper as approximately 20% of the costs were start-up costs, as well as added expenses incurred to implement the programme in the context of a randomised controlled trial. The average cost-effectiveness ratios were also presented in Table 29 and reveal that given the limited impact on the longer-term outcomes for adolescents that it is extremely expensive for unit changes in the outcomes within the first two years of the programme. In many ways, these results are an artifact of the lack of impact in the short-term for outcomes that are expected to change over a longer period of time.

Considerations

The results presented in the mid-term report have implications for recommendations on future programming for adolescent girls in Zambia and elsewhere and should be coupled with burgeoning evidence from AGEP and the literature to adapt programming for vulnerable adolescent girls in order to improve impact. In many ways, the longer-term impact of AGEP will not be known until the final round of data collection in 2017. While the impact of AGEP on empowerment was not as strong as expected immediately after the end of the programme, it is possible that the measures of assets did not comprehensively capture all aspects of girls' empowerment that were changed by AGEP and, in turn, influence longer-term adolescent outcomes. Further, it is also possible that AGEP has set girls on a different trajectory of further acquisition of assets and hence the cumulative impact of AGEP on empowerment and subsequent outcomes may be revealed in later years of observation. Finally, it is also possible that the AGEP programme will have a direct effect on longer-term adolescent outcomes, independent of its indirect impact expected to occur through empowerment. For these reasons and others, the AGEP study provides a wealth of data that will be explored for further understanding of the adolescent transitions in Zambia, as well as application to adolescent programming.

It should be noted that the AGEP programme implementation was, in many ways, constrained by the need to integrate a highly rigorous evaluation. In particular, due to fears of contaminating the control areas and undermining the evaluation, community engagement and involvement was necessarily limited. On the positive side, this allowed for a rigorous assessment of the Safe Space girls' group model by isolating its activities. The mid-term results, however, are suggestive that an adolescent girls' asset-building programme may not be sufficient to lead to immediate and substantial change on its own. This may particularly be the case given the entrenched nature of traditional social gender norms, attitudes towards adolescent sexuality, and use of contraceptives, to name a few. A more comprehensive ecological theory would dictate that complementary work is needed in the enabling environment, particularly at the family and community level. A promising approach that has been found effective elsewhere would be to engage the boys, men, adults, and other key stakeholders in girls' lives, addressing norms at household and community levels in order to benefit girls. This may be an important way to help girls leverage the assets they are building in the safe spaces.

AGEP was also an ambitious project directed towards changing girls' lives in a significant and meaningful way across areas of education, sexual and reproductive health, marriage and fertility, and experience of violence. While the programme length of two years was sufficient to cover these areas in the Safe Spaces group curricula, it is possible that depth of the programme was foregone in place of breadth. While the underlying root causes of girls' vulnerabilities are interrelated, it is possible that a more direct focus on a particular outcome, driven by a more targeted intervention, would have led to greater impact in the shorter term. A domain-specific conceptualization of change and related theory of change for particular outcomes would be a natural first step in this process. For adolescents, it may imply a direct focus for older adolescents on livelihoods and entrepreneurship and for younger adolescents may need more focus on educational support. Also, providing direct resources through incentivised activities may be a constructive approach to increasing engagement with the programme.

Overall, the AGEP cohort data, and lessons it has generated from the Adolescent Girls Empowerment Programme are rich, nuanced, and important for informing the next generation of programmes for adolescents in Zambia and elsewhere. Even though the study is still underway, and the full longer-term effects of AGEP remain to be seen, the information presented in this report can be used to guide programmes and policymakers on programme areas of promise, gaps that need to be filled, and a range of questions about how to best serve this population that still need to be answered.

Background

Adolescent girls in Zambia

Adolescent girls[1] in Zambia face a range of risks and vulnerabilities that challenge their healthy development from girls into young women and they often lack the social, health, and economic assets to mitigate these risks. The issues that confront vulnerable girls — high rates of gender-based violence, unsafe sex that puts girls at risk for unwanted pregnancy and HIV infection, school dropout, lack of economic resources and income-generating options, lack of agency and participation — are intertwined and have similar causes.

Early marriage: Married adolescents have limited social, health, and economic assets. Child brides often experience a sudden shrinking of their social networks as they move to their husbands' residences, leaving them with few, if any, friends and peers. While parents may want their daughters to marry at an early age for apparent economic stability, anticipating that the new husband will cater to the needs of the girl, in reality early marriage often leaves girls with limited control over resources, restricted mobility, and little or no power in their new households.[2] Thus, married girls may face significant challenges in negotiating safe sexual relations. In Zambia, according to the 2013–14 DHS,[3] 7% of girls aged 15–17 were currently married and 2% of girls had been married by age 15. Among 20–24-year-old females, 6% were married by age 15–19 and 31% by age 18. Girls living in this context are vulnerable, sexually active, with a unique set of needs, and are often missed by general youth programmes, as well as by general sexual and reproductive health programmes intended for women.

Education: It is known that education is a critical component of a healthy transition to adulthood. For example, being in school has been associated with delays in the age at first sex, marriage, and childbearing. However, in both rural and urban settings in Zambia, 41% of girls ages 10–14 in Zambia are already off course with their schooling—meaning that they have either never attended primary school, are not currently in school, or are in school but are two years or more behind grade-for-age. When comparing girls' and boys' non-enrolment, levels are similar for the 10–14-year-old group (approximately 10%), but the gender disparity is clear among older adolescents with 44% of girls aged 15–19 not in school as compared to 31.1% of boys. In other words, girls leave school at significantly higher rates than boys. Interventions that are school-based may, therefore, miss large segments of the adolescent girl population—often including those who are most vulnerable.

Gender-based Violence (GBV): Lack of social protection and voice leaves girls exposed to violence, both physical and sexual. More than 35% of 15–24-year-old females have experienced physical violence, and 12% have already experienced sexual violence. A study done by the Population Council[4] among girls aged 15–24 in urban slums in Lusaka found pervasive levels of violence for adolescent girls at home and in the community. School, a place supposed to be protective, is not necessarily safe for girls. Seventy percent of girls in that study reported having been teased or sexually harassed in school and 53% reported that girls in their

---

[1] For this document, adolescent girls are defined as girls ages 10–19.
[2] Erulkar, A. 2013. "Adolescence lost: The realities of child marriage," *Journal of Adolescent Health* 52(5): 513–514.
[3] In this Background section, except in the case of sources otherwise noted, the data discussed come from the *2013–2014 Zambia Demographic and Health Survey report or data analysis conducted*. Central Statistical Office (CSO) [Zambia], Ministry of Health (MOH) [Zambia], and ICF International. 2014. *Zambia Demographic and Health Survey 2013--14.* Rockville, MD: Central Statistical Office, Ministry of Health, and ICF International.
[4] Brady, Martha et. al. 2009. *Understanding Adolescent Girls' Protection Strategies Against HIV: An Exploratory Study in Zambia.* New York: Population Council.

school were molested. Eighty-six percent reported that girls their age were pressured to do things they did not want to do in exchange for money. These issues go beyond experience of violence, to include acceptance of it. Data from the 2013–14 DHS confirm that in Zambia harmful gender norms are formed by an early age. Among 15–24-year-old females, 48% believe that wife beating is justified in at least one circumstance. Therefore, it is critical to reach girls early, addressing gender norms and rights, in an effort to reduce violence against girls.

HIV and other health risks: HIV is a significant health risk to girls in Zambia. The threat is particularly acute among young women and girls with limited social and economic assets who are unable to avoid, manage, or leave unsafe sexual relationships, whether inside or outside of marriage[5]. In Zambia, the 2014 DHS indicates that HIV prevalence is 15% for females and 11% for males. Among 15–19-year-olds, 4.8% of females are infected as compared to 4.1% of males. The gender disparity increases in the 20–24-year-old group with 11.2% of females infected as compared to 7.3% of males. Due to the gender-based and economic inequalities present in many marriages and sexual relationships, women, especially girls, may find it difficult to communicate about HIV and negotiate condom use with their partners. Therefore, while knowledge of and access to HIV testing and other services are important, they are not sufficient to ensure HIV prevention. Girls must have strong social, health, and economic assets in order to gain more control within their sexual relationships.

Additional health risks—especially early pregnancy, unsafe abortion, and STIs—result from high levels of unprotected sex. In Zambia, among 15–19-year old never-married girls, 30.9% in urban settings and 44.5% in rural settings report having experienced sex. Overall, 12% of women 20–24 report first sex before age 15 and 54% before age 18. Married and unmarried sexually active adolescents need information, social support for informed decisionmaking, and access to services and technologies to effectively maintain their sexual and reproductive health. The common large age gap between girls and their partners, which can result in financial dependency, often puts these girls at a social and economic disadvantage, which compromises their ability to avoid violence and negotiate healthy sexual behaviour. In Zambia, there is a large unmet need for contraception among adolescents and young women; only 18% of never-married sexually active girls aged 15–19, and 47% of young women 20–24, have ever used a modern method of contraception.

Assets that prevent pregnancy among girls are critical because early motherhood is linked to school dropout, as well as social isolation and early marriage. According to the 2013–14 Zambia DHS, 29% of all 15–19-year-olds have been pregnant or had a child (15% among those who have never been married). Despite the high levels of risk for HIV and unwanted pregnancy, however, less than 50% of sexually active girls aged 15–19 have had an HIV test in the last 12 months, and only 36% of sexually active, unmarried 15–19-year-olds used a condom at last sex.

The situation described above is a challenging one for adolescent girls in Zambia. Across a wide range of issues—education, social support, safety, and sexual and reproductive health—they are at a disadvantage compared to their male peers. The added vulnerabilities confronting Zambian adolescent girls formed the basis for designing an intervention that a) focuses on adolescent girls aged 10–19 and b) selects participants for the programme with the direct intention of capturing those girls who face the most vulnerabilities described above among their female counterparts.

---

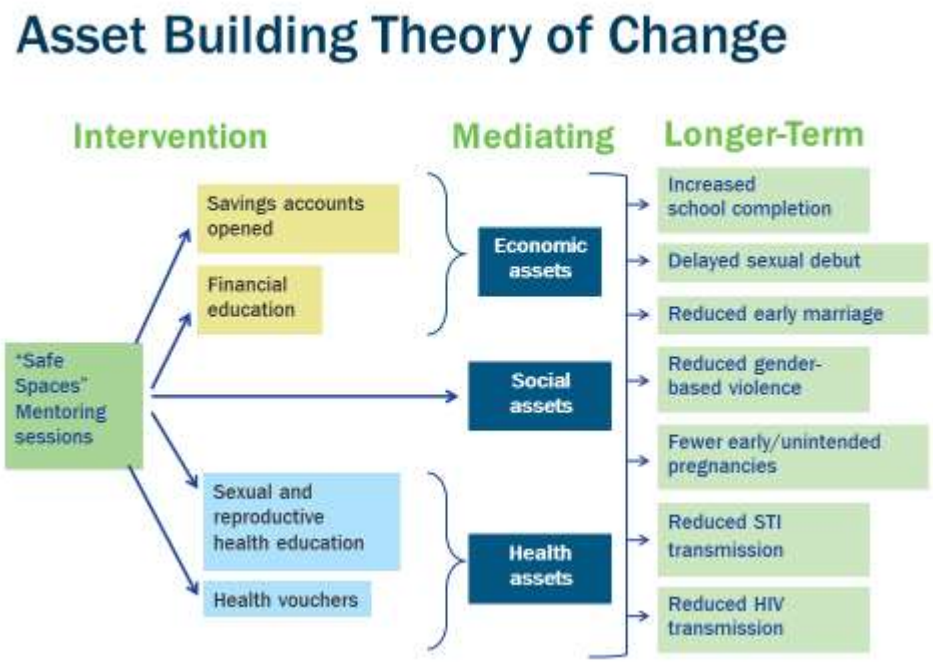
[5] Bruce, Judith. 2007. "Girls left behind: Redirecting HIV interventions towards the most vulnerable." Transitions to Adulthood Brief no. 23. New York: Population Council.

AGEP Programme Summary

Theory of change

The theory of change behind AGEP posited that adolescent girls are empowered (the desired outcome) by acquiring social, health, and economic assets that they can then draw on to reduce vulnerabilities and expand opportunities, thereby increasing their likelihood of completing school and delaying sexual debut, and reducing the risks of early marriage, unintended pregnancy, acquisition of HIV, and so on (the impact).

Figure 1. AGEP Theory of Change



The literature shows an association of social and economic assets to key longer-term health outcomes. For example, data among adolescent girls in South Africa showed that an increase in social isolation and increases in economic vulnerability have all been independently linked to an increase in having experienced sexual coercion and having engaged in transactional sex.[6] There is a growing literature that the Safe Spaces model—girls meeting regularly for sessions facilitated by a female mentor—has been successful in several countries in sub-Saharan Africa on longer-term health and economic outcomes[7,8,9].

---

[6] Hallman, K. 2005. "Gendered socioeconomic Cconditions and HIV risk behaviours among young people in South Africa," *Africa Journal of AIDS Research* 4(1): 37–50.

[7] Austrian, K. and E. Muthengi. 2013. "Safe and Smart Savings Products for Vulnerable Adolescent Girls in Kenya and Uganda: Evaluation Report." Nairobi: Population Council. http://www.popcouncil.org/uploads/pdfs/2013PGY_SafeSmartSavingsEvalReport.pdf

[8] Erulkar, A. and E. Muthengi. 2009. "Evaluation of Berhane Hewan: A program to delay child marriage in rural Ethiopia," *International Perspectives on Sexual and Reproductive Health* 35(1): 6–14.

[9] Bandiera, O., N. Buehren, R. Burgess, M. Goldstein, S. Gulesci, I. Rasul, and M. Sulaiman. 2012. *Empowering Adolescent Girls: Evidence from a Randomized Control Trial in Uganda.* Washington, DC : World Bank Group. http://documents.worldbank.org/curated/en/543691468309340444/Empowering-adolescent-girls-evidence-from-a-randomized-control-trial-in-Uganda

AGEP population, sites, and timeline

To target those adolescents considered most at risk in Zambia, AGEP drew participants from lower-income backgrounds and living with multiple levels of vulnerability, e.g., physical and social isolation, without parents, in low-income households, and not attending school. In service to the AGEP goal of reaching the most vulnerable girls, while also creating conditions for conducting a rigorous cluster randomised evaluation, eligible girls were identified through a household survey for recruitment in the programme. Information collected from the household allowed for targeted invitations for the most vulnerable adolescent girls. This method of invitation differs from recruitment typically used by community-based programmes.

A one-year pilot of AGEP was conducted in Matero (a high-density housing compound in Lusaka) and Chibombo (a rural district in Central Province) with 1,200 girls. The purpose of the pilot was to test the intervention components, as well as the vulnerability selection process. Key adaptations were made prior to the rollout of the main intervention and presented in a "'lessons learned" report.[10]

After pilot implementation, AGEP was rolled out in 10 sites in four provinces in Zambia, five sites in urban and five sites in rural areas (see Figure 2). AGEP aimed to reach 1,000 girls per site, for a total of 10,000 girls who would be recruited into the programme. There were two urban sites in Lusaka province (Misisi/Chawama and Chipata/Chazanga); two rural sites (Mumbwa and Kapiri Mposhi), and one urban site (Kabwe) in Central Province; two urban sites (Ndola and Kitwe) and two rural sites (both in Masaiti) in Copperbelt Province; and one rural site (Solwezi) in Northwestern Province. Study provinces and the number of sites per province were selected purposefully, on the basis of feasibility of operating the AGEP programme while also conducting a research evaluation, as well as through discussions with the donor regarding the target populations. A site in a rural area contained multiple contiguous or proximal villages or chiefdoms, while in urban sites the programme was implemented within high-density housing compounds. The programme communities in urban areas were directly proximal to participating banks and health centres; in rural areas there were on average greater distances between households and health facilities and participating banks.

To select sites within the study provinces, a sampling frame of sites containing two or three public health facilities proximal to each other was generated. Urban and rural areas were treated separately. For urban AGEP sites, adjacent high-density compounds that included the necessary two or more health facilities were considered a single unit for sampling. This was done to achieve a sufficient number of randomisation units (clusters) for the second stage of selection.[11] For rural AGEP sites, all clusters within a 15km radius of the health centres were included in the site.

---

[10] For more detail about the lessons learned from the pilot and the adaptations made for the rollout, please see Austrian, Karen, Natalie Jackson-Hachonda, and Paul C, Hewett. 2013. *The Adolescent Girls Empowerment Programme: Lessons Learned from the Pilot Program*. Lusaka: Population Council. http://www.popcouncil.org/uploads/pdfs/2013PGY_AGEP-PilotReport.pdf.

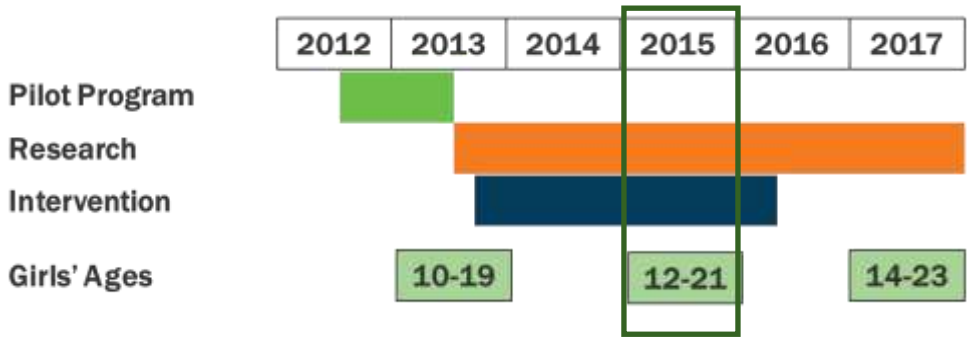
[11] For more detail on the sampling frame for site selection, please see Hewett, Paul C., Karen Austrian, Erica Soler-Hampejsek, Jere R. Behrman, Christine A. Kelly, Dela Kusi-Appouh, Fiammetta Bozzani, Barbara S. Mensch, and Minyoi Maimbolwa. 2014. *Adolescent Girls Empowerment Programme: Research and Evaluation Baseline Technical Report.* Lusaka, Zambia: Population Council.

Figure 2. AGEP Sites

One urban site Two urban sites One rural site

AGEP started in a phased approach across the 10 sites, approximately one month after the baseline data collection was completed. Therefore, the first site started in August 2013 and ended in August 2015 and the last site started in March 2014 and ended in March 2016 (see Figure 3). This report reflects the programme learning and data collected between 2013 and 2015 (outlined by rectangle). A final report will be published in 2018 that will reflect the full five rounds of data collection (see Table 10).

Figure 3. AGEP Programme and Research Timeline



Core components of AGEP in Zambia

Safe spaces: Implemented in partnership with YWCA Zambia, Safe Spaces were weekly girls' group meetings in which 20 to 30 girls met with a mentor—a young woman from their community—for short training sessions on a variety of topics as well as an opportunity to discuss their experiences in the past week. AGEP developed three curricula used in the context

of these meetings: 1) a health and life-skills curriculum, 2) a financial education curriculum, and 3) a nutrition curriculum[12] for adolescent girls (see Annex A for a full list of training sessions). Each trained mentor used the same curricula and was instructed on the order in which the sessions should be delivered to ensure standardisation across all groups. The groups met weekly over the course of two years. At the start of the programme, girls were assigned based on their age to a group for 10–14-year-olds or 15–19-year-olds. The groups themselves chose the meeting date, time, and location based on when the girls were available and where in the community they deemed it safe to meet. See below for more details on mentors.

Health vouchers: In partnership with the Ministry of Community Development, Mother and Child Health (MCDMCH[13]), participants received a health voucher redeemable for a package of general wellness and sexual and reproductive health services at partner public and private healthcare providers. During the weekly meetings, mentors taught the girls in the group about the voucher services and informed them of participating clinics where the voucher could be used. For private and NGO providers, payment was made on a "fee for service" basis with pre-approved reimbursement rates, whereas for the public facilities an incentive for each service was paid to the District Community Health Office (DCHO) and then distributed between the district health office and the clinic in previously agreed-upon percentages.

Savings accounts: The Population Council worked in partnership with the National Savings and Credit Bank (NatSave) and Making Cents International to develop the Girls Dream savings account for AGEP girls. The NatSave account had a very low minimum opening balance of ZMW 2.5 (US $0.50) and any amount could be deposited or withdrawn with no fee. Mentors of girls opening savings accounts were trained by AGEP staff in the savings account features, and in turn the mentors conducted an orientation session with the girls and their co-signatory prior to account opening, to instruct them how to use the account and begin the account-opening process. A field trip to the branch was also organized for girls and their co-signatory to complete the account-opening process. At the time the project began, there were approximately 32 NatSave branches throughout Zambia, predominantly located in urban areas.

Programme implementation vis-à-vis research design: In order to assess the effect of the safe spaces alone, as well as the added effect of the "add-on components," AGEP was implemented in three versions across randomised study arms: Arm 1 included safe spaces only, Arm 2 included safe spaces and the health voucher, and Arm 3 included safe spaces, the health voucher, and the savings account (Table 1). Clusters were randomly assigned to receive one of the arms (or to be in the control). Further details of the evaluation design are discussed in the Research and Evaluation section below.

[12] Half of all mentors in each site, stratified by arm, were randomly selected to be trained on and implement the nutrition curriculum in their safe space groups. This embedded sub-study will assess the impact of nutrition education integrated into the safe spaces curriculum on anaemia, dietary diversity, and anthropometric growth of adolescent girls and their children. The sub-study findings will be published in 2018 in a separate report.

[13] Since the intervention, the Mother and Child Health unit has moved to the Ministry of Health and the current ministry is now called Ministry of Community Development and Social Welfare.

Table 1. Randomisation arms of AGEP

AGEP Randomisation Arms			Control
Arm 1 (40 clusters)	Arm 2 (40 clusters)	Arm 3 (40 clusters)	Arm 4 (60 clusters)†
 Safe Space Groups	 + Health Voucher	 + + Bank Account	No activities

† Includes 40 internal control clusters and 20 external control clusters (urban areas only).

Safe Spaces implementation

Using household-listed data to select girls based on vulnerability (see below for details) between 1,200 and 2,000 invitations were delivered in each AGEP site. Mentors, who had already been selected and participated in an eight-day training, worked in pairs to deliver invitations to each selected household. Up to three attempts were made to deliver the invitations. The first preference was to give the invitation to the girl herself, but if she was not present, it was delivered to a guardian. The invitation provided information for an introductory meeting in the community at which a girl and her guardian would learn more about the programme and be assigned to a Safe Spaces group. A total of 16,867 invitations were delivered, among those, 3,515 were part of the research sample.[14]

Girls' level of programme participation was assessed from the research sample using three categories: never attended, attended less than 52 meetings, and attended 52 or more meetings; 52 meetings represented roughly half of the number of possible sessions. As is noted in Table 2, just over a quarter of girls invited to the programme never joined a Safe Spaces group; just under 50% of girls joined a group, but participated in less than half of the sessions, while approximately 30% participated in more than half the group sessions. Data collected in 2014 indicated that the main reasons why girls never joined the programme were: a) they were not aware of the programme or how to join and b) they were too busy or had other commitments/chores to complete in that time. However, for girls who joined and then stopped attending sessions within in the first year, the key drivers of dropout were relocation to an area outside of the AGEP programme and losing interest in the programme activities.

[14] Programme targets called for at least 1,000 girls participating in each site, or 10,000 in total. The research sample targets, determined by power calculations, were just over one-third.

Assessing differential programme participation among each of the study arms, the mean number of sessions attended was roughly the same (see Table 3); hence, programme arm assignment or the added components of the intervention did not affect programme participation.

Table 2. Programme uptake (among research sample)

	Sample	%
Invited never attended	921	26.2
Invited attended <52 meetings	1,559	44.4
Invited attended 52+ meetings	1,035	29.4
All research girls	3,515	100.0

Table 3. Meetings attended, by arm

	Sample	Mean	SD	Min	Max
Arm 1 – Safe spaces only	1,186	33.6	33.7	0	106
Arm 2 – Safe spaces + health voucher	1,169	32.6	33.5	0	113
Arm 3 – Safe spaces + health voucher + savings account	1,160	32.1	32.7	0	107
All arms	3,515	32.8	33.3	0	113

In order to understand the dynamics of programme participation, an analysis of who was more likely to participate was conducted. As observed in Table 4 and Table 5, there was differential uptake of programme participation on several key indicators. Overall, girls who were younger, living in rural areas, in school, with higher numeracy skills, and the biological daughter of the head of household were the most likely to participate in a greater number of sessions. One hypothesis for why younger and rural girls were more likely to participate is that there were fewer competing interests, i.e., fewer additional programmes being offered and fewer social activities that girls could engage in during their free time. Especially in the remote rural areas, AGEP was "the only show in town," and therefore girls were more eager to join and participate. For rural girls, this is of note given the greater distances needed to attend meetings.

Measures that directly capture vulnerability suggest that girls in less stable living situations were less likely to participate. Although programme staff made a tremendous effort to recruit every girl on the invitation list, it is possible that reluctant guardians served as gatekeepers who did not want the girls to participate. From the qualitative research it was clear that at least some guardians were highly suspicious of the intentions of the programme. It is also possible that, confirmed by reflections of community leaders, the intervention was not of direct interest to the most vulnerable girls, after all a choice still had to be made by the girl to participate. It is possible that such girls are in need of interventions that are directed more to their immediate needs, including nutrition, economic resources, and direct support of formal education. If the

programme were to be replicated, adaptations would need to be made to focus additional recruitment methods and intervention approaches to reach and attract those who are most isolated and vulnerable.

> *If they would include skill training…they would use it to get a job even after the programme is finished. That would make the programme much more meaningful than it already is*. Community Leader, Ndola

> *Education-wise, you need to sponsor these girls…to help financially, buy them books and uniforms so that they can get an education and help out parents who are struggling to meet daily needs.* Community Leader, Solwezi

Table 4. Mean number of meetings attended, by age and geography

	Sample	Mean	SD	Min	Max
Younger rural	904	40.8	34.6	0	110
Younger urban	911	35.6	34.1	0	113
Older rural	804	28.3	30.8	0	108
Older urban	896	25.7	31.1	0	107

Table 5. Characteristics by attendance, results from multivariate probit model predicting intense programme participation

	0–51 meetings		52+ meetings		Bivariate Diff	Multivariate model*	
	N	%/Mean	N	%/Mean	p-val	Coef	p-val
Urban (vs rural)	**2,480**	**53%**	**1,035**	**48%**	**0.004**	**-0.190**	**0.016**
Age at baseline	**2,480**	**14.5**	**1,035**	**13. 6**	**0.000**	**-0.065**	**0.000**
Grade attainment at baseline	2,480	5.5	1,035	5.1	0.000	-0.026	0.193
Attended school at baseline	**2,480**	**76%**	**1,035**	**88%**	**0.000**	**0.205**	**0.010**
Literate in any language at baseline	2,480	42%	1,035	40%	0.316	0.011	0.877
Numeracy score at baseline (0–15)	**2,437**	**6.7**	**1,024**	**6.9**	**0.123**	**0.018**	**0.037**
Cognitive score at baseline (0–16)	2,442	6.5	1,025	6.3	0.163	-0.006	0.485
Mother alive at baseline	2,480	88%	1,035	90%	0.026	-0.008	0.924
Father alive at baseline	2,479	76%	1,034	82%	0.001	0.087	0.110
Mother completed primary school	2,478	44%	1,034	46%	0.449	0.036	0.501
Father completed primary school	2,479	50%	1,035	50%	0.857	-0.034	0.470
Is biological daughter of HH head	**2,480**	**59%**	**1,035**	**66%**	**0.000**	**0.139**	**0.009**
Vulnerability quintiles (1st – lowest)	2,480	23%	1,035	24%	0.343	Ref	Ref
Vulnerability quintiles (2nd)	2,480	19%	1,035	20%	0.313	0.019	0.806
Vulnerability quintiles (3rd)	2,480	20%	1,035	21%	0.309	-0.002	0.984
Vulnerability quintiles (4th)	2,480	19%	1,035	19%	0.855	0.005	0.942
Vulnerability quintiles (5th)	2,480	20%	1,035	16%	0.004	-0.104	0.259
Ever had sex at baseline[a]	1,312	44%	388	35%	0.002		

Ever been pregnant at baseline[a]	1,312	17%	388	13%	0.061		
Ever given birth at baseline[a]	1,308	11%	388	9%	0.178		
HIV positive at baseline[a]	1,251	3%	372	3%	0.567		
HSV-2 positive at baseline[a]	1,223	8%	363	6%	0.186		

*Multivariate probit model controls for arms, number of household visits conducted by the mentor, household assets, household savings, vulnerability scores, and tribe, and accounts for clustering at CSA level.

[a] Only asked of 15–19-year-olds.

Mentors: Mentors were key, if not *the* key, to the Safe Spaces implementation. They were the main point of contact with the girls, as well as their families and community. All of the content in the training curricula was delivered through the mentor, so it is logical to hypothesize that the success and quality of a mentor would have an effect on programme retention and impact. Therefore, more detail is provided on the mentor process, and on an evaluation undertaken to understand the dimensions of mentor quality. The mediating effect of mentor quality on programme impact will also be discussed.

AGEP Mentor Profile. AGEP mentors were trusted female adults between the ages of 20 and 40 living in the same community as the girls in the programme. In addition to facilitating the weekly group meetings, they served as role models to girls in their groups. Mentors had to be literate, speak English and the local language fluently, have completed grade 12, and have facilitation and/or training experience, as well as basic sexual and reproductive health knowledge and a commitment to improving the situation of girls in their community.

Recruitment. In order to recruit mentors with the required background, skills, and commitment, AGEP staff in each site shared the mentor job description with key leaders and posted the job advertisement in public spaces (clinics, schools, council offices, etc.). Interested candidates completed an application letter indicating their personal details and relevant work experience, attaching a curriculum vitae when available. From the applications received, programme staff shortlisted candidates in each site for interviews. As for a professional interview, staff assessed and scored the following criteria for each candidate: appropriate age and residential location, education and communication skills, relevant knowledge and experience, adaptability and social outlook, and personality. Additionally, interviewers asked candidates to share how their experiences had prepared them to be a mentor, assessed candidates' values on sensitive topics such as sexual and reproductive health, and gauged their interest in working with adolescent girls in the community. After scoring the interviews, the top 35 candidates at each site received an invitation to attend an eight-day mentor training of trainers (TOT).

Training. During the initial AGEP mentor TOT, programme staff focused on building the knowledge and facilitation skills needed by mentors to successfully implement the programme. Simultaneously, programme staff used the eight days to assess candidates' language capabilities, knowledge, skills, and values related to working with adolescent girls. The training sessions covered a range of critical topics including: programme overview and mentor roles; the AGEP training methodology and curricula; information sessions on topics such as puberty, contraceptives, self-esteem, counseling, working with parents and adults, conflict management, and HIV/AIDS; practical sessions focused on facilitating AGEP sessions followed by a peer review; values clarification sessions about sexual and reproductive health issues for adolescent girls; and the introduction of monitoring and evaluation tools such as attendance-

keeping, monthly reports, and lesson plans for weekly group meetings. Also the team prepared participants for the programme launch, practising the facilitation of the introductory meeting session where girls were registered and put into groups, forming AGEP safe spaces.

Based on performance during the TOT, programme staff confirmed 24 mentors per site and identified an additional four to six alternate mentors per site. Confirmed mentors signed contract agreements which laid out the contract period, monthly stipend rate ($50 USD per month), working hours (up to 40 hours per month), reporting (monthly activity reports, weekly attendance), roles and responsibilities, penalties for missing sessions ($5 USD per session missed) or falsifying attendance records, and a termination clause.

Midway through the programme, AGEP mentors attended a 5-day refresher TOT workshop to continue to build their capacity to implement Safe Space meetings effectively. Mentors provided feedback about their experiences delivering the Safe Space sessions, working with parents/guardians and communities, and keeping girls engaged in the programme, raising challenges encountered and suggesting solutions to improve the programme. Sessions aimed to build upon the lessons learned during the first year of the programme while also working to enhance mentor facilitation skills and continuing to build their knowledge of key AGEP topics, particularly those they found challenging during the first year. This also allowed the training of new mentors since over time some mentors dropped out of the programme or had their contracts terminated due to poor performance.

Beyond the two off-site TOTs, mentors attended monthly meetings with the site coordinators. These meetings provided a time to collect attendance data, pay the mentors, address any challenges that arose during the month, and have short refresher trainings on key topics, facilitation skills, home visits, and how to keep attendance rates up. Also, programme staff conducted quarterly monitoring visits to each mentor as a spot check to assess mentor performance and any challenges faced during Safe Space implementation.

Roles and Responsibilities. Mentors were responsible for managing two to three weekly Safe Space groups which lasted one to two hours each and were made up of 20 to 35 girls. Mentors also played a critical role in recruiting AGEP girls into the programme, delivering personalized invitations to the girls' homes at the outset of the programme.

At AGEP introductory meetings, mentors and programme staff split the girls into 10–14 and 15–19-year-old age groups, then facilitated fun games, songs, and dances. Then they worked with the girls to conduct safety mapping to identify safe and unsafe areas in the community to select a safe meeting place for the weekly group meetings. Together with the girls, they decided on the meeting day and time that was most convenient.

The group meetings were meant to be engaging, interactive, and dynamic, avoiding a classroom feel and a teacher/pupil relationship. Prior to each weekly group meeting, mentors were responsible for preparing the training session from the AGEP curricula, covering a range of topics such as sexual and reproductive health, life skills, gender-based violence, nutrition, and financial education. Each curriculum laid out a set order of session topics for mentors to follow over the two-year period.

When an AGEP girl missed two consecutive Safe Space meetings, the mentor conducted a home visit to meet with the girl and/or her parent or guardian to determine why the girl had stopped attending sessions. This provided a critical opportunity for the mentor to connect with

the girl and her family, assess if the girl was facing any problems, and encourage her to return to the weekly group meetings.

Beyond the two site coordinators based in each site, mentors were the face of AGEP in the community. Mentors worked closely with programme staff to organize parent/husband and community meetings. They encouraged parents to allow their daughters to participate in the programme, addressed any myths that came up about the programme's aim, and fostered support for the programme within the community.

Mentor Quality Evaluation: A set of tools to assess mentor quality were developed to illuminate the link between different dimensions of mentor quality and girls' outcomes. The evaluation included three tools: 1) a self-administered mentor survey that captured socio-demographics, knowledge and attitudes on AGEP-related topics, as well as the mentor's attitude towards her work; 2) observation by an external interviewer of an entire Safe Space session conducted by each mentor; and 3) a module added to the adolescent survey that included questions concerning the girls' views of their mentor and their relationship to her.

Mentor Self-Administered Survey. Completed during the mentor refresher TOT, midway through the programme, the self-administered mentor survey captured socio-demographics (age, education, marital status, age at first pregnancy, and work/volunteer experience); knowledge (literacy, numeracy, HIV/AIDS, contraceptives, STIs, human rights and financial literacy); attitudes about family planning, gender norms, and gender-based violence; mentor job satisfaction; and self-perceived performance. At the second-to-last mentor meeting before sites closed, the mentors completed the same self-administered survey, providing an opportunity to assess any changes in the indicators listed above.

Mentor Observation. The mentor observation tool measured 21 dimensions of mentor quality in six domains including: facilitation, creating a supportive environment, communication, relationship-building, knowledge, and leadership. The observation was conducted by trained external research assistants (RAs). The RAs conducted surprise visits to each AGEP mentor over a 3.5-month period. Mentors were observed when facilitating one Safe Space meeting for each age group, 10–14 and 15–19 years. They were not informed in advance when an observer would come to observe, which prevented mentors from preparing the Safe Space session more thoroughly than usual.

Each dimension of the observation tool had a level 1, 2, or 3 score, with 1 being low and 3 being high. For example, under the communication domain, one dimension addressed "communication: use of body language to enhance girls' engagement." The scoring options were as follows:

- Level 1: mentor made no adjustments to her body language;
- Level 2: mentor used engaging facial expressions and modified her posture; and
- Level 3: mentor adjusted facial expressions, posture, hand gestures, and tone of voice—all to enhance girls' engagement.

Girls' Survey. A set of 18 questions was added to the AGEP Round 3 quantitative questionnaire which was administered to the girls in the research sample at the end of the two-year intervention period. Questions focused on the girl's relationship with and perceptions of her mentor. Questions captured different dimensions of mentor quality, including perceived knowledge, fairness, and trust.

Dimensions of Mentor Quality.[15] We applied Principal Components Analysis (PCA) as a data-reduction strategy to build eight mentorship quality scores based on variables collected in the mentor self-administered surveys, girl surveys, and mentor observation questionnaires. The dimension-specific scores represented different aspects of mentorship quality: 1) knowledge, 2) safe-space creation, 3) relationship with girls and community, 4) self-efficacy, and 5) global gender attitudes; in addition to this gender attitudes score, we calculated three sub-scores that captured attitudes on three different topics: 6) gender norms, 7) contraception, and 8) gender-based violence. Finally, we averaged all[16] of these standardized, dimension-specific scores to estimate an overall measure of mentorship quality.

We evaluated the reliability, as well as the external validity of the scores to make sure they were adequate measures of the quality of mentorship received by the girls. We used Cronbach's alpha to measure the reliability of each of the scores. In Annex B, all the detailed results for the mentor evaluation are shown. Table B-1 shows the variables pre-selected to build each of the scores. We dropped variables with low (or negative) correlation to the rest of the indicators. If the Cronbach's alpha significantly increased after omitting a variable from the set, we dropped the variable. In Table B-2, the items in italics represent excluded variables. We followed this procedure until reliability was maximized for each of the variable sets, while at the same time retaining enough theoretically relevant variables to keep the score meaningful. Then, we analysed each of the variable groups using PCA. We then generated three ordered categories for each of the scores, defined by the distributions' tertiles: Poor, Regular, and Good.

Table B-3 shows a matrix of correlations among the eight dimension-specific mentorship quality scores and the overall score. With the exception of the score on relationship with girls and community, the other scores were positively correlated among themselves, which suggests they had common causal factors and measure-related concepts.

As a check for external validity, Tables B-4 and B-5 show the distribution of the eight mentorship quality scales across a series of demographic indicators, including mentor's socio-economic status (SES). With the exception of "relationship with girls and community," mentors' completion of secondary school was positively and consistently related to good mentorship quality across dimensions. Apart from their scores on "relationship with girls and community," "self-efficacy," and "safe-space creation," mentors' SES level was positively related to mentorship quality as well. Mentor's age was positively related with scores on "knowledge," "Safe Space creation," and "relationship with girls and community." In contrast, age at first pregnancy and age at first marriage did not show a clear relation with the mentorship scores in these bivariate comparisons. Table B-6 shows the same bivariate exercise, but looking at the overall mentorship quality score. In summary, the table suggests the best mentors were older, more educated, with higher SES, and never married.

While descriptive statistics are helpful to assess the external validity of the scores, they do not fully answer what the determinants of good mentorship are. To evaluate this question, Table B-7 shows a series of logistic regressions that control for multiple mentor characteristics at once. Each of the nine models in this table explains a different mentorship quality score, while

[15] A separate paper is available that describes the process to develop measures of mentor quality in much greater detail.
[16] With the exception of the global gender attitudes score, which was substituted by the three topic-specific scores on attitudes.

controlling for mentor's education, age, SES level, age at first pregnancy, and whether the mentor was never married. With the exception of "relationship with girls and community" and "attitudes on gender-based violence," completed secondary school had positive and significant effects on the probability of good mentorship quality. In addition, older mentors had a higher and significant probability of providing good mentorship, particularly along the lines of "knowledge," "safe-space creation," and "gender attitudes." With the exception of "relationship with girls and community," high SES was positively and significantly related to good mentorship. Although age at first pregnancy seems to have been irrelevant in explaining mentorship quality, a "never married" status had positive and significant effects on mentor's self-efficacy and liberality of gender attitudes. In sum, older mentors who had completed secondary school, with a relatively high SES background, and who had never been married, had the greatest probability of providing good mentorship.

We also want to understand if the key dimensions of mentor quality affect the program's impact on the girls. In order to do this we restricted the analysis to girls who attended at least half of the programme sessions and found that girls with mentors who had positive attitudes towards contraception were less likely to have ever been pregnant; girls with mentors who scored high on "safe-space creation" were less likely to have been married, had sex, had an unwanted pregnancy, or given birth; and girls with mentors who scored high on the self-efficacy score were less likely to be HIV-positive and have had unwanted sex. Finally, girls who had a mentor who scored high on the "relationship with girls and community" score were also less likely to have had unwanted sex. See Tables A-7 to A-10 in Annex A for the full results.
The varied experiences with mentors were also reflected in the qualitative data. For example, some girls described mentors encouraging contraceptives and others the reverse.

> *Our mentor told us that we shouldn't use contraceptives before we have children because you will have problems conceiving…they say if you don't have a child it's not good to use (contraceptives)….* (Girl, 18-21, SS+HV+SA, Ndola)

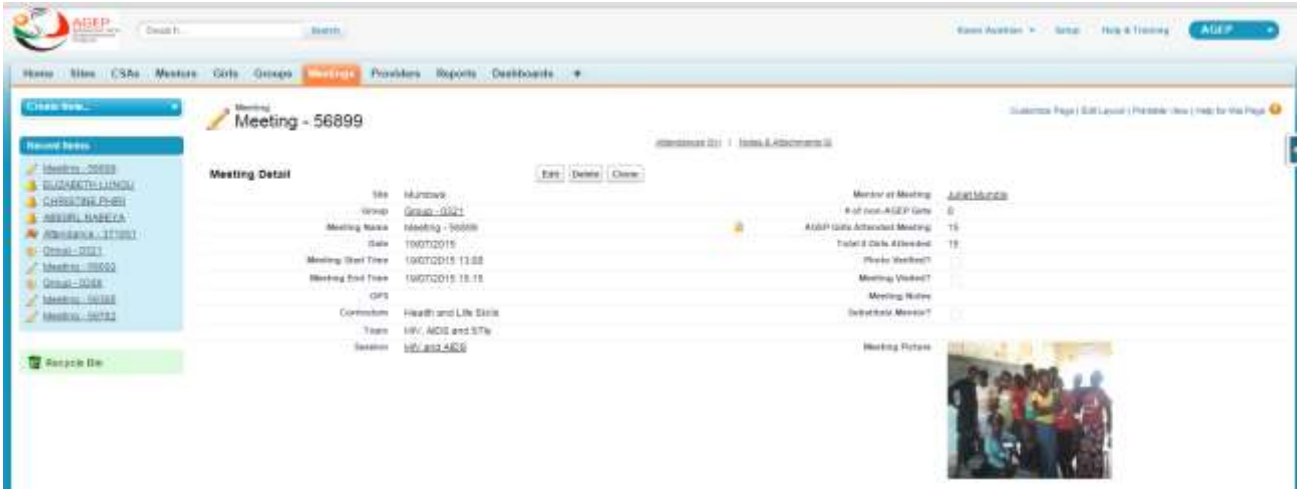
> *When we talk about abstinence it's not working…so the only thing to do is tell them abstain but if you can't, use a protective measure.* (Mentor, Ndola)

Programme monitoring data: Between August 2013 and March 2016, 405 Safe Space groups met in a total of 40,884 meetings. High priority was given to timely and high-quality programme monitoring data for two key reasons: 1) to maintain high programme implementation standards with strict fidelity to the research design and 2) to make use of detailed attendance data to better understand programme impact. Mentors were given basic Android phones with an Open Data Kit (ODK) app on them. ODK is an open source, free app that allows for online and offline data collection on phones and tablets. The app had preloaded forms specific to each particular site. A mentor would select her group from the app, input the date and meeting session completed, and then select which girls from her group's member list attended that day (see Figures 4 and 5). She would also take a picture of the girls in attendance for verification purposes. Mentors were also given a small amount of mobile data each month so that they could send the completed forms to an aggregate server.

Attendance data collected from ODK was then uploaded into a cloud-based database hosted on the Salesforce platform (see Figure 6). This allowed programme staff to see on a weekly basis which groups held meetings (or did not) and which girls attended (or did not). As a result,

there is a complete list of meeting sessions attended by each girl. Programme staff used these data to work with mentors who were underperforming.

Figures 4–6. ODK Forms and Salesforce Meeting Record



Key lessons learned from safe spaces: After making adaptations from the pilot phase, several key initial challenges had been solved by the time of the Safe Spaces implementation. Overall, the programme implementation was quite successful. Key elements that led to successful implementation were:

- monthly meetings with mentors to provide regular supervision and refresher training;
- regular community sensitization so that the community consistently felt aware of programme aims and activities; and
- access to the electronic monitoring data described above critical for the ability of the programme staff to successfully monitor and correct course when certain groups, mentors, or sites seemed to be going off track.

However, despite the overall smooth and successful activities, there were several challenges—the main one being maintenance of consistent attendance among AGEP participants over the course of two years. In order to incentivize attendance (in response to pilot lessons learned), a passport was developed in which a girl received a stamp for each meeting she attended. For every 10th meeting attended, the girl received a prize. The prizes were pre-determined (i.e., all girls in all sites got a school bag for their 30th meeting, an AGEP t-shirt for their 60th meeting,

and a chitenge for their 80th meeting) and girls knew ahead of time which prize they would get for each milestone. However, the logistics involved in delivering 10 different types of prizes to 10 different sites was overly burdensome and prizes were often late. In addition, girls and parents complained regularly about the quality of the prizes, and discrepancies between the number of meetings a girl thought she attended and what the cloud-based database confirmed left girls feeling frustrated. In the end, it is not clear that the prizes motivated girls to attend meetings, and the amount of money and human resources put into procuring and distributing the prizes was likely disproportionate to the benefits.

To compensate, sites held periodic fun days to motivate the girls, and mentors were each given a small budget to enhance their meetings with additional creative activities. However, as seen not only from programme uptake, but overall attendance, it was more challenging to keep the older girls engaged than the younger girls. The older girls had greater competing priorities outside the home: more domestic responsibilities, financial responsibilities, child care responsibilities, etc., and therefore the opportunity costs for them to come to the Safe Space meetings were higher. In addition, in terms of the financial curriculum, it is possible that they needed an extra set of activities that would extend beyond financial education to include income generation, as older adolescents have greater financial demands (either demands for contributions to their parents' household or demands to support their own personal needs and/or households).

Another lesson learned was that more could have been made of the community sensitization that was done. While it was handled well and consistently, and paid off with very high levels of community support for the programme, more activities could have been integrated that would have addressed community norms around adolescent girls—including the value of girls' education—so that the environment for girls could have been more enabling and increased the likelihood that asset acquisition would lead to broader and longer-term measures of health and well-being.

Health voucher implementation

While the experimental effect of the health voucher on usage of different types of health services, as well as on overall health, will be explored in the results section below, an analysis of who used the voucher and for which services is important in determining the programmatic lessons learned.

In keeping with the percentage of invited girls who actually joined a group (74%), three-quarters of girls in Arms 2 and 3 received a health voucher. However, when comparing girls who attended less than half of the sessions to those who attended more than half, only 60% of girls who attended less than half and over 95% of those who attended more than half, received a voucher. This is likely a function of the delay in voucher rollout. Due to the delays in receiving a formal agreement from the Government of Zambia to operate the voucher, the first round of vouchers was distributed approximately one year after the Safe Space groups started. Therefore, those who had dropped out of the programme prior to completing a year were unlikely to have ever accessed a voucher. As a voucher was valid for one year, girls were given a second voucher near the end of the Safe Space group meetings, valid for approximately one year after the Safe Space groups ended. Again reflecting overall participation trends, younger girls were more likely to have received a voucher than older girls, although with little to no difference between the urban and rural segments within an age cohort.

While distribution of the health vouchers was successful vis-à-vis the active girls in the Safe Space groups, the uptake in use of the voucher was lower than anticipated. Overall, 20% of girls who received a voucher in either the first or second year of voucher distribution used it. Use of the voucher for more than one service, among those girls who used the voucher, was quite common, with girls most commonly using the voucher twice (except for younger, urban girls who used it once). While all girls most commonly used the voucher for general wellness services (i.e., not for sexual and reproductive health [SRH] services), the older girls made much more use of the SRH services. When comparing the girls with the voucher who attended less than half the sessions with those who attended more than half, there was little difference in the breakdown between general wellness and SRH services. However, among girls who used an SRH service, 81% also used a general wellness service, which suggests that helping girls access general health services can provide a bridge to more specialized sexual and reproductive health services (i.e., it is easier for a girl to go to a clinic for general health services, but then while there to be counseled on SRH services).

Table 6. Health voucher distribution

	Did the respondent ever receive a health voucher?		
	No	Yes	Total
10–14/rural	424 18.2%	1,909 81.8%	2,333
10–14/urban	466 21.6%	1,694 78.4%	2,160
15–19/rural	479 31.5%	1,040 68.5%	1,519
15–19/urban	521 31.3%	1,146 68.7%	1,667
	1890 24.6%	5,789 75.4%	7,679

Table 7. Health voucher usage

	10–14 Rural	10–14 Urban	15–19 Rural	15–19 Urban	Total
Usage % used voucher Min/max # service Mean # services Median # services	 24% 1/11 2.14 2	 15% 1/6 1.56 1	 22% 1/13 2.79 2	 24% 1/13 2.26 2	 21% 1/13 2.15 2
Services used					
Wellness checks and all other non-SRH services	779 79.2%	331 85.8%	406 62.9%	412 66.8%	1,928 73.3%
Family planning	18 1.8%	3 0.8%	58 9.0%	40 6.5%	119 4.5%
Pregnancy test	0 0%	3 0.8%	24 3.7%	27 4.4%	54 2.1%

STI services	101 10.3%	44 11.4%	63 9.8%	87 14.1%	295 11.2%
HIV testing and counseling	69 7.0%	2 0.5%	71 11.0%	21 3.4%	163 6.2%
ANC	3 0.3%	1 0.3%	22 3.4%	19 3.1%	45 1.7%
Other SRH services	14 1.4%	2 0.5%	1 0.2%	11 1.8%	28 1.1%
	984	386	645	617	2,632

Overall, the implementation of the health voucher was successful, with a sufficient number of health facilities contracted in each site; each health facility participating in two trainings on adolescent-friendly health services;[17] and strong support and engagement from government officials at the district, provincial, and national level. The main benefit expressed by girls who used the voucher was that it increased their confidence that the clinic waiting times that used to deter them from seeking health services in the past would be reduced and they would be treated with respect by the providers because of the voucher.

> *When we were given cards (vouchers), we never had to queue up, we just used to go straight to be attended to and we were very happy.* (Girl, 18–21, SS+HV, Chipata/Chazanga)
>
> *The other thing that made it easy was because of the vouchers that we were given so upon seeing those vouchers, they had to attend to me so quickly.* (Girl, 17, SS+HV, Solwezi)
>
> *I am very confident (to seek health services)… It was easy for me because I have a health voucher, so they attended to me and gave me the medicine…. (Without the voucher) it would have been a bit slow and they would have asked me to buy some of the medicines…. They attend better to us now that we have the voucher and we are given all the meds that we need unlike before when we didn't have the voucher, they would even shout at us.* (Girl, 18–21, SS+HV+SA, Ndola)

In addition, the electronic platform that was developed for the AGEP voucher was well implemented and allowed programme staff to monitor, in real time, voucher usage so that any suspicious activity was immediately flagged and could be addressed. In addition, if there were periods of total inactivity, or inactivity at a particular facility, programme staff used the electronic platform to note the issues and troubleshoot immediately. The platform remains and the open source product is now available and can hopefully serve as an asset in other future health service programmes, whether targeting adolescent girls, other sub-populations with particular health needs, or the general population.

However, despite the smooth implementation of the voucher, uptake remained relatively low and challenged the programme staff to question why that was the case. Several reasons could explain this finding. First, overall, adolescents are relatively healthy and do not have the same need for health services as children under five or older women. Especially given the very low

[17] http://www.popcouncil.org/uploads/pdfs/2015PGY_AGEP_AFHSCurriculum.pdf

levels of sexual activity among the younger cohort, there is also not as high a demand for SRH services among younger girls. Second, in urban areas, voucher holders reported that the public facilities were quite crowded, did not have the commodities that they needed, and the voucher did not always give them preferential treatment in terms of waiting in the queue. The private facilities in the urban areas were often far from the compounds where girls lived and girls had to use public transportation in order to access the private facilities. It was notable that the one private facility that was located in the compound itself was frequently accessed by voucher holders in that site. The third issue was a bit of a "catch-22" situation because the providers were incentivized to provide better services to the adolescent girls because of the voucher payments—however, without higher uptake, the value of those incentives was quite low, and therefore in turn the actual incentive to prioritize adolescents was low, and providers did not encourage adolescent girls to make more use of the health voucher.

In addition, despite the voucher addressing some of the supply issues as well as some of the demand issues in accessing health services, there still remains among adolescents a strong psychological barrier to accessing services—especially sexual and reproductive health services. The message in society is clearly that unmarried adolescent girls should not be having sex, and thus should have no need for contraceptives. In order for an adolescent girl to make the decision to accept a method of family planning, she must internalize that she is sexually active and is at risk of getting pregnant. For many adolescents, this is a psychological leap they are not yet ready to make, especially in the context of non-supportive household and community norms.

Key lessons learned from the health voucher: In addition to reflections on the results of the health voucher in AGEP, if a health voucher for adolescents were to be replicated, there are several key lessons that should be taken into account:

- Use of an e-platform that allows for real-time monitoring of voucher usage is a key tool for successful implementation;
- Ongoing (bi-annual, if not quarterly) meetings with district, provincial, and national-level government staff are useful for maintaining overall support for voucher implementation;
- Staff turnover at clinics is quite high, therefore provisions for repeated training (annual at a minimum) should be included;
- For broad-based health service provision for adolescents, a voucher, even in the context of safe spaces, may not be enough to overcome all supply and demand barriers; and
- When Safe Space groups stopped meeting, voucher usage dropped even lower, therefore the ongoing platform in which girls met and interacted with one another, as well as their mentors, did encourage voucher usage.

Savings account implementation

The Girls Dream savings account was made available to girls who had joined Safe Space groups in clusters that were randomised to Arm 3 (which also included safe spaces and the health voucher). As the account-opening process was initiated in the safe space setting, it is very unlikely that girls not participating in an Arm 3 Safe Space group would open an account since the accounts were not available on the open market to girls who simply entered a NatSave branch.[18] At the group level, girls were sensitized about the account and invited to bring a female parent/guardian, or other female above the age of 18 as a co-signer, to an account-opening session that took place during a regular Safe Space group time. During that session all of the account-opening paperwork was completed and submitted to the bank. A few weeks later a trip to the nearest bank branch was organized so that girls could pick up their bank cards and make their first deposit. The account-opening process began about 8–10 weeks after the group was formed and started meeting.

Reflecting this process, 51% of girls registered to an Arm 3 Safe Spaces group opened an account, however when looking at girls who attended more than half of the Safe Spaces sessions, that number increased to 72%. Similar to programme participation overall, younger rural girls were most likely to have opened an account, followed by younger urban girls, older rural girls, and older urban girls in that order. This is perhaps counterintuitive since the urban girls lived in much closer physical proximity to the branches, and older girls had more access to cash than younger girls, but could be explained by reasons such as overall programme participation. In addition, the account-opening process was done at the Safe Space groups and the initial visit to the bank was facilitated by the programme—eliminating distance/transport barriers in the account-opening process.

Table 8. Percentage who opened Girls Dream accounts, by attendance, age, and geography

	10–14 Rural	10–14 Urban	15–19 Rural	15–19 Urban	Total
Attended 1–51 meetings	37%	33%	33%	31%	34%
Attended 52+ meetings	76%	72%	70%	67%	72%

The main barrier to account opening for girls who were actively participating in the Safe Space groups was an inability for some to produce a National Registration Card (NRC), in particular for the co-signer and girls who were ages 17 and above. The programme provided all of the information for girls who were eligible for NRCs to access them, yet in some cases there was a lack of will to complete the process, or competing priorities for time and financial resources. In addition, for girls who may have missed the two rounds of account opening that were completed at the group level, the process was likely too daunting to complete on one's own, preventing girls from opening accounts if they had been away or unavailable during the main and follow-up rounds of group-level account opening.

[18] Two percent of girls in Arm 3 who never attended a safe spaces group meeting, and 0.5% of girls in Arm 1 and 2, have a Girls Dream savings account. This is likely due to a) a small number of girls in Arm 3 who heard about the account and came only for that purpose, but never officially joined the group and b) contamination from girls moving in between baseline and account opening.

Overall, account usage, although low, was on par with other similar adolescent or child savings accounts developed through various financial inclusion programmes.[19] About a third of girls made an additional deposit into their account after the initial deposit made as part of the account-opening process. Here, however, there was little difference in age and geography. In addition, few girls (3%) withdrew from their accounts. It should be noted that this is one of the few indicators where urban girls were more likely to participate—perhaps reflecting more frequent use of cash in the urban economy. The market research conducted on account usage suggested that low withdrawals are because that money is being saved for long-term future use, not because girls are not able to access the money.

Table 9. Account usage summary

	10–14 Rural	10–14 Urban	15–19 Rural	15–19 Urban
# of deposits 1 deposit 2 deposits 3+ deposits	66% 20% 14%	67% 16% 17%	75% 19% 6%	70% 13% 17%
Deposit amount (ZMW) Median Mean	5 16	2.5 29	5 11	2.5 89
% withdrew	1%	4%	2%	9%
Withdrawal amount Median Mean	0 2	0 6	0 10	0 60
Account Balance (ZMW) Median Mean	5 15	2.5 18	5 12	2.5 30

Overall, saving was seen as a key step in the theory of change to better long-term health outcomes. This was confirmed in the qualitative analysis which asked girls, mentors, parents, and key community stakeholders to list and rank the key drivers of teenage pregnancy in their communities. Overwhelmingly, the key driver was transactional sex—that is, the money or material goods that girls received in exchange for sex. Sex could be transacted to make up for poverty and lack of funds in the household for food, shelter, school fees, and other basic needs, or for girls to satisfy their own personal desire for material things—whether it be new clothes, a mobile phone, or entertainment.

> *When there is no food in the home and she wants to find some money to buy relish only to be told that she has to sleep with a man in order for her to get that money. Upon doing that, and without using any protection then she becomes pregnant.* (Girl, 15–21, Control group, Ndola)

> *When someone is too poor and life is too hard it can make someone want to have a boyfriend who can at least help her with money. I have seen situations where a girl ends up going to the club to sleep with any man that she meets so that she can have something to eat, this may make one get pregnant from these men that she may be sleeping with.* (Girl, 15–21, SS Only, Ndola)

> *If you are lacking and you are getting pressure from the way your friends are looking (dressing), you will definitely find someone to sponsor you (sugar daddy) and in the end you get pregnant because you have no say you will give in and get what you want.* (Girl, 18–21, Control group, Chipata/Chazanga)

> *If girls are not being helped or given the things that they need (by parents)… girls who lack things like soap, lotion, and other things usually resort to getting those things from boys and men who ask for sex for*

[19] YouthSave Consortium. 2015. *YouthSave 2010–2015: Findings from a Global Financial Inclusion Partnership*. https://na-production.s3.amazonaws.com/documents/YouthSave_2010_2015.pdf

> *the things that they give them, and in the end they get pregnant…. The girls that get pregnant in this way are many because many families fail to help their daughters with things like soap, lotion, and other things…. When these girls lack basic needs they engage in sexual activities… these men trade material things for sex and girls can end up getting pregnant.* (Parent, Solwezi)

If lack of financial resources is a key driver of teenage pregnancy, and pregnancy is linked to many other longer-term outcomes such as timing of sexual debut, condom use, marriage, and educational attainment, then working with girls to save, as well as be aware of the risks of depending on others for money, has the potential to have a longer-term effect on sexual behaviour and health outcomes. Girls in the programme seemed to express an understanding of this link in the qualitative data:

> *If you have your own money, you cannot find yourself in (bad situations), if you make your own money you are your own boss and you can do whatever you want to do with your money, no implications in the end.* (Mentor, Ndola)

> *We did not know about how to save money; whenever we would have money we would just end up misusing it, but now we know and we save money such that we even buy things which are important and we really want.* (Girl, 18–21, SS Only, Solwezi)

As the financial education curriculum was a part of the Safe Space groups, all girls were exposed to lessons on the importance of saving, budgeting, and prioritizing different kinds of spending, as well as the risks of certain kinds of income. However, only girls in Arm 3 had savings accounts in addition to the financial education. Girls in Arm 3 showed a stronger positive effect in terms of financial literacy scores and savings behaviour (see results section for full details). Therefore, having a savings account reinforces what girls are learning and gives them an opportunity to put the skills they learn in theory into actual practice. This confirms the findings of two recent systematic reviews on financial education, which both found that financial education is more effective when given the chance for immediate application.[20] Market research conducted periodically throughout the intervention to assess the implementation of the savings account component confirmed that girls with a savings account felt motivated to save informally, even if they had not yet made a deposit, so they could save up enough for it to be worth a trip to the bank. Further spatial analysis is needed to assess if girls who live beyond a certain threshold distance from a bank branch experienced more, less or the same levels of added effect from the a formal savings account as compared to girls who live closer to the branches. For girls who live far from the branch, perhaps future interventions would explore the use of informal savings groups.

Key lessons learned from the savings account:

- Vulnerable adolescent girls can open savings accounts and save regularly (both formally and informally).
- Access to formal financial services can be a catalyst for both informal and formal savings activity.

---

[20] Miller, M., J. Reichelstein, C. Salas, and B. Zia. 2014. *Can You Help Someone Become Financially Capable? A Meta-Analysis of the Literature*. Policy Research Working Paper 6745. Washington, DC: World Bank.

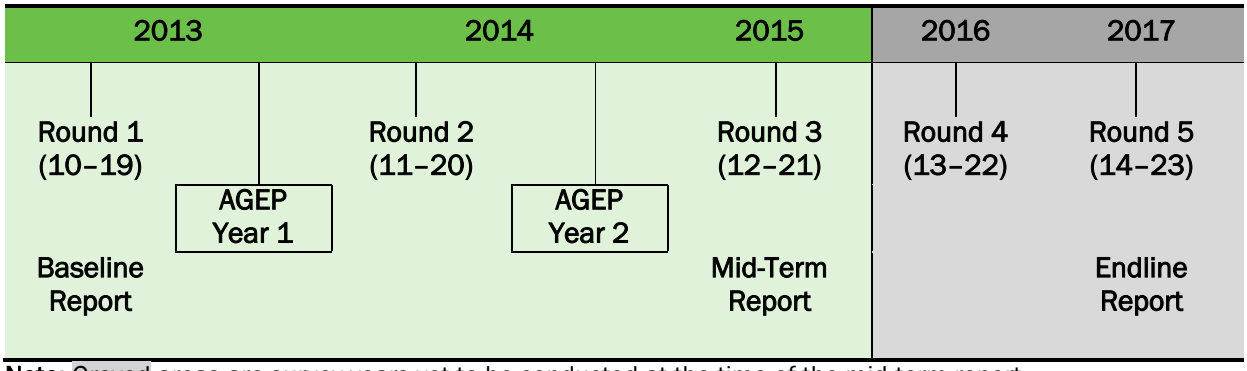
Fernandes, D., J. Lynch, and R. Netemeyer. 2014. "Financial literacy, financial education and downstream financial behaviours," *Management Science* 60(8): 1861–1883.

- Financial education is a critical component to the successful rollout of financial services for vulnerable adolescent girls.
- Support from the executive level of management at the financial institution is critical, as well as training of all branch-level staff (from the branch manager, to the tellers, to the security guards standing at the entrance).
- There may be a certain distance from the branch where alternative savings schemes should be explored.

AGEP Research and Evaluation

The ultimate objective of the AGEP evaluation has been to conduct a rigorous assessment of the impact of the programme on the trajectories of girls' lives in their adolescent years and beyond. This translates to assessing whether AGEP improves the knowledge, assets, and skills girls possess, as well as their longer-term demographic, reproductive, and health outcomes. The AGEP evaluation focuses on vulnerable adolescent girls, aged 10–19 in 2013 as they age to 14–23 in 2017. This mid-term report presents the progress in fulfilling the evaluation's ultimate objective, by reporting on the analysis of adolescent outcomes immediately after the programme ended, and recording the changes that have occurred since 2013 (see Table 10). Thus, we will focus on study participants across three rounds of data collection, with the latest interviews in 2015 when the girls were aged 12–21.

Table 10. AGEP evaluation and programme summary timeline



Note: Greyed areas are survey years yet to be conducted at the time of the mid-term report.

The baseline or Round 1 survey was begun in July of 2013 prior to implementation at the programme sites. The AGEP intervention was initiated in September 2013, following an incremental rollout process for each site that paralleled the site order of the baseline survey. The last site began to initiate AGEP in March 2014. Interviews with adolescents were conducted annually in 2014 and 2015, after one year and two years of AGEP, respectively. Additional rounds of interviews will be conducted in 2016 and 2017. Qualitative semi-structured interviews took place at Round 1 and Round 3 among a sub-sample of respondents. HIV and HSV-2 testing was conducted among adolescents aged 15 years and older at Rounds 1 and 3. Anaemia testing among adolescents aged 15 years and older was added in Round 3 and will be conducted yearly through 2017. HIV and HSV-2 status will again be tested at endline (2017).

Randomised cluster design

A randomised cluster design was implemented to rigorously assess the impact of AGEP, as well as to assess the impact of each of its core components. Clusters in each site were randomised at a public lottery to receive different combinations of AGEP components. All girls within the cluster who were selected for participation in AGEP were offered the intervention that had been randomly determined for that cluster; girls in clusters for the control arm received neither the AGEP intervention nor placebo exposures. As indicated in Table 1 above, study arm 1 consisted of girls participating only in the Safe Space girls groups, stratified by age and marital status as

described above. In study arm 2, in addition to participating in Safe Space groups, enrolled girls received the AGEP health voucher. In study arm 3, girls received an additional offer to open a girl-friendly bank account.

The number of clusters needed for the evaluation was determined by sample size calculations for cluster-randomised trials: 40 clusters per study arm, with a minimum of 20 girls per cluster by the end of the evaluation were needed. The 4-arm study design dictated that AGEP operate in 120 clusters (communities), with 40 additional control clusters. Each AGEP site, therefore, had 12 experimental and 4 control clusters that were randomly selected. Given the geographic proximity of experimental and control arms in urban areas and the potential for spillover from AGEP to control areas, an additional four clusters from nonadjacent but nearby areas were selected and designated as "external" controls for each of the five urban AGEP sites for a total of 20 external controls.

Clusters were delineated by Census Supervisory Areas (CSAs) as specified by the Zambia Central Statistical Office (CSO). A CSA contained a collection of adjacent standard enumeration areas (SEAs) that ranged in number from two to eight SEAs per CSA. SEAs were a convenient geographical demarcation, usually following geographic patterns (roads, natural terrain) that contained, on average, approximately 100 households in rural areas and 150 households in urban areas. CSAs contained approximately 750 households in urban areas and 300 households in rural areas. The number of households could vary considerably between individual CSAs. CSAs in urban areas were relatively small geographic areas, perhaps a few hundred metres long and wide, while in rural areas they could be much larger, encompassing numerous square kilometres.

Selection of girls: The adolescent girls who were to participate in AGEP were selected from the household listing. A vulnerability indicator was constructed and girls ranked by their vulnerability score. Those with the highest levels of vulnerability in each AGEP site were selected for the programme. Girls residing in boarding schools or mentally disabled were excluded, while all girls with physical disabilities were automatically invited. To reach the target of 10,000 participating girls, a total of 16,649 adolescent girls within the 10 AGEP sites were invited to participate. The following criteria were used for recruiting the selected vulnerable adolescent girls into the AGEP research component at baseline:

- Selected for AGEP (experimental areas only);
- Selected randomly if there was more than one eligible girl in the household;
- Between the ages of 10 and 19 years;
- Never married;
- Residence in selected CSAs in AGEP programme or control areas at baseline;
- Capable of meeting the obligations inherent in the research.

It is important to note that the **baseline research sample consisted only of never-married girls**. The decision to limit the research sample to never-married girls reflects the fact that many of the primary outcomes of the evaluation, including ever having sex, ever giving birth, and, of course, ever marrying would already have been observed and, therefore, could not be influenced by the programme. An implication of this decision is that the drawn sample could not represent all adolescent girls in the age group and, in particular, was less representative among the older adolescent girls who had a higher propensity for marriage. For instance, of the 19,978 girls ages 10–19 identified from the household listing to constitute the sampling

frame,[21] 5.9% were already married and therefore excluded from the evaluation, although they were still eligible for programme participation. The proportion married increased substantially by age: only 1.5% of 15-year-olds were married compared to 9.0% of 17-year-olds and 30.9% of 19-year-olds. These selection criteria have important implications for understanding outcomes and drawing inferences, particularly if the findings are compared to other data that include both married and unmarried participants.

Measuring vulnerability: One important consideration in identifying vulnerable girls for participation was not to over-represent girls who had already manifested their vulnerability in terms of the outcomes to be measured, e.g., by dropping out of school, becoming pregnant, or getting married. It was therefore preferable to capture girls who were vulnerable and at the cusp of manifesting adverse adolescent life-course outcomes. To identify vulnerable adolescent girls, behind-school-grade for age was used as a proxy indicator of vulnerability. Adolescent girls from the household listing were ordered by the estimated levels of their vulnerability and selected according to the number required for the sample at each AGEP site. Since the concentration of girls was higher in urban areas, the sample represented about 30% of girls in urban areas, and therefore the most vulnerable, while in rural areas the sample represented the majority of all adolescent girls, although still the most vulnerable of those girls.[22]

Adolescent survey

The adolescent survey instruments were intended to measure changes in attitudes, behaviours, transition status, social assets, and cognitive skills that might occur over time related to: 1) schooling attainment and transitions; 2) sexual activity, relationship status, and sexual partners; 3) marriage and marital dissolution; 4) sexual and physical coercion and violence; 5) gender attitudes, self-efficacy, and locus of control; 6) labour force participation and savings behaviour; 7) living arrangements and household resources; 8) mobility and migration; 9) literacy, numeracy, and cognitive skills and ability; and 10) financial literacy and knowledge. The information collected from adolescents is useful for assessing the potential differential impact of the AGEP programme among subgroups, as well as for increasing the power of the statistical assessment of the programme's impact by taking into account these measured covariates.

[21] Hewett et al. 2014. *AGEP Baseline Technical Report*. New York: Population Council.
[22] See the technical brief "Methodology: Reaching the Most Vulnerable Adolescent Girls" for more details on the vulnerability indicator (http://www.popcouncil.org/uploads/pdfs/2016PGY_AGEP-Vulnerability_brief.pdf)

Table 11. Summary of study instruments and measures

Instrument	Key elements	Ages (at survey)
Younger Adolescent Survey	Household sociodemographic characteristics Schooling history Social assets and networks Self-efficacy and locus-of-control Financial literacy, savings behaviour, and livelihood activities Self-reported health, reproductive health knowledge, and nutrition	10–14
	Experience of physical harassment and violence	13–14
Older Adolescent Survey	Household sociodemographic characteristics Schooling history Social assets and networks Self-efficacy and locus-of-control Financial literacy, savings behaviour, and livelihood activities Relationship history and marriage Sexual and reproductive behaviour Experience of physical harassment and violence Self-reported health, reproductive health knowledge, and nutrition HIV and AIDS risk perception Utilization of antenatal and postnatal care services	15–19
Adolescent Literacy, Math, Cognitive Skills	Reading ability in local language and English Excerpts from official mathematics assessments, multiple grades Ravens Progressive Matrices cognitive testing	10–19

The questionnaires were translated into the most common local languages spoken in the study sites. Surveys were implemented, where feasible, by electronic data capture using Samsung Galaxy tablets. Computer-Assisted Personal Interviewing (CAPI) was used for questions that were nonsensitive. CAPI is a process of data capture in which the interviewer reads the question from a computer screen and enters the participant's response directly into a handheld or tablet device. For sensitive questions, Audio Computer-Assisted Self-Interviewing (ACASI) was used. With ACASI the respondent listens on headphones to pre-recorded questions and response categories while simultaneously reading (if the participant is literate and desires to do so) the question on the tablet screen. The participant enters a response by touching a colour-coded number or option as specified in the audio script and on the tablet screen. ACASI maximizes confidentiality and privacy of response, since no one can hear or see the question being read, nor the response option selected.

Anthropometric data: To capture the shorter- and longer-term impact of nutrition on health outcomes, anthropometric data, specifically the participant's height and weight, were collected. The capture of height and weight allowed for the measurement of key indices to assess nutritional status, including body-mass index, height-for-age, weight-for-height, and weight-for-age. Anthropometric data were also collected from the living children aged five and younger of study participants to assess the impact of nutrition during pregnancy and postpartum on child growth. The equipment used and the procedures for implementing the anthropometric measurement were drawn from the Demographic and Health Surveys.

Biological markers for HIV and HSV-2 were collected from adolescents aged 15 and older in Round 1 (2013) and Round 3 (2015). Anaemia testing was added in Round 3. All biological

specimens were collected at the household or in a private space in the community in cases where confidential interviewing and testing could not be done at home. HIV rapid testing and counselling followed Zambia's national guidelines and were conducted by certified staff and results were provided directly to the adolescent. HSV-2 biological specimens were collected via finger stick capillary sampling. Approximately 250 µL of whole blood was collected, stored, and transported in microtainers and laboratory tested using the Kalon™ ELISA antibody test kit. The HSV-2 results and counselling were provided to the participant at the local health clinic. Anaemia testing was conducted among adolescents aged 15 and older and their children aged 6–59 months. A point-of-care Hemocue™ testing system was used, consisting of a battery operated spectrophotometer and a blood collection device called a microcuvette. A capillary blood sample of 10 µL was obtained and tested, and designations of severe, moderate, and mild anaemia based on haemoglobin levels were determined using age-appropriate WHO guidelines. All biohazardous waste bags—containing gloves, gauze, etc. —were taken to the nearest hospital with an incinerator. The sharp instruments (microcuvettes and lancets) were taken to the same location for disposal.

Assessed adolescent girl outcomes and impact indicators

AGEP was a community-based, multi-sectoral, girl-centred programme designed to address the multi-dimensional nature of girls' empowerment.[23] AGEP posited that adolescent girls are empowered when they are provided opportunities to acquire and maintain social, health, and economic assets and competencies, and that these assets and competencies can be drawn upon to reduce vulnerabilities and expand opportunities to improve girls' longer-term outcomes; outcomes that have significant consequences for their lives, their families, and their children. The evaluation of AGEP requires an impact assessment of how these assets and competencies were influenced or changed by the AGEP programme and, ultimately, what role they had in mediating the change observed in the longer-term outcomes.[24] Table 12 provides a list of the measures used in the AGEP mid-term report to capture girls' empowerment and assess impact. The number of indicators is large to, maximize, in a sense, the opportunity to comprehensively assess change as a result of the programme.

[23] The girl-centred approach stands in contrast to other types of interventions that address structural barriers faced by adolescent girls, such as community interventions addressing social norms or gender-based violence, adolescent-friendly health services, and cash-transfers for education, among others.

[24] The analysis presented in this report does not address the question of mediating influence because the report focuses on the impact assessment. Additionally, this question will be assessed using all five rounds of data.

Table 12. Empowerment outcome indicators, by domain and measure

Empowerment domain	Assessment outcome measures
Social assets	
Self-efficacy	1. Avg. score on self-efficacy scale 2. % confident regarding their ability to plan 3. % feel they make good decisions regarding money
Autonomy	4. % agree that permission to go to the health clinic is not a problem 5. % who jointly or solely make decisions with regard to money earned
Friends	6. Avg. number of friends 7. Avg. number of friends in school 8. Avg. number of friends who can be counted on if needed money 9. Avg. number of friends who can be counted on in an emergency
Safety nets	10. % have a safe space to meet with friends 11. % with adult female support in case of serious problem
Isolation	12. % attending any social groups/clubs within the past month 13. % who often/sometimes go to market[a] 14. % who often/sometimes go to community centre[a] 15. % who often/sometimes go to shops or restaurants[a]
Beliefs	16. Avg. score on gender equality scale 17. Avg. score on nonacceptability of intimate partner violence
Economic assets	
Financial literacy	18. Avg. score on financial literacy scale
Savings	19. % who have saved in the past year 20. Avg. amount saved in the past year among those who saved
Paid work and income	21. % working for cash or in-kind in the past year 22. Avg. reported income in the past year among those who worked for cash
Assets personally owned	23. % who own a bicycle 24. % who own a mobile phone
Health assets	
Health knowledge	25. % Understanding pregnancy risk during menstrual cycle 26. Avg. score on contraceptive knowledge scale 27. Avg. score on sexual reproductive health knowledge 28. Avg. score on HIV/AIDS knowledge scale
Self-assessed health	29. Avg. rating of health status in the past year 30. Avg. rating of health status in the past month
Experience of sickness	31. Avg. reported number of health problems in the past month
Skills and competencies[b]	
Performance on assessments	32. % who can read simple sentence in local language 33. % who can read simple sentence in English 34. Avg. score on numeracy assessment 35. Avg. score on cognitive assessment

[a] Indicator was added after the baseline survey.
[b] While these sets of asset indicators reflect important capabilities, they are not expected to be directly affected by the AGEP intervention as there are no programme components that build capacity in these areas. Hence, these indicators are not included in the AGEP impact assessment as outcomes. The baseline measures are, however, used as covariates in the adjusted estimation results.

It was hypothesized that exposure to AGEP would expand the assets that girls acquired over the two years of the programme. The following domains of empowerment were represented: 1) social assets, including self-efficacy, autonomy, social networks, support, isolation, and gender beliefs; 2) economic assets, specifically financial literacy and the material resources that could be productively used or provide a safety net; and 3) health assets: representing the knowledge required and the ability of the girl to stay active and productive. These domains and measures of empowerment were those specifically targeted by the AGEP intervention through its safe space curricula, weekly group meetings, interactive sessions, savings account, and health

voucher.[25] Additional measures of adolescent girls' skills and competencies are included, although these are not the target of the AGEP intervention as there are no programme components that build capacity in these areas. All empowerment measures were hypothesized to play a large role in mediating the impact of AGEP on the longer-term indicators of success.

Ultimately, the goal of AGEP was to transform the lives of adolescent girls in meaningful ways. It is the goal of the AGEP evaluation to measure that impact. The mid-term report will also assess the impact of AGEP on the longer-term adolescent outcomes, measured immediately after AGEP ended. These impact indicators are noted by domain in Table 13. While it would be a positive sign to see a significant difference between AGEP girls and control group girls immediately after the programme ended, there are two reasons such change may not yet be observed. The first is that many of the transitions (first sex, marriage, birth) occur later in adolescence and a significant proportion of girls will not yet have made those transitions. A second reason is that the study hypothesis tests were statistically powered to test differences over four years of observation. Unless there were even larger effects of AGEP than expected at this stage, the differences might not yet reach statistical significance.

Table 13. Longer-term impact indicators, by domain and measure

Outcome domain	Assessment impact measures
Educational	1. Avg. number of years completed 2. % completed primary school 3. % completed junior secondary school[a] 4. % currently attending school
Sexual risk behaviour[a]	5. % ever had sex 6. % agree that they have had unwanted sex 7. % agree that they have had transactional sex 8. % used condom at last sex with non-marital partner 9. % used condom at first sex
Marital[a]	10. % ever married 11. Avg. number of HIV risk-related topics discussed with partner 12. Avg. marital control score
Pregnancy & births[a]	13. % who have ever been pregnant 14. % who have ever had an unwanted pregnancy 15. % currently pregnant or who have given birth 16. % ever used modern contraception
Sexually transmitted infections[a]	17. % HIV positive 18. % HSV-2 positive
Experience of violence[b]	19. % ever experiencing physical violence in past year 20. % ever experiencing intimate partner violence in past year

[a] Indicators in this domain measured among those aged 15 and older.
[b] Indicators in this domain measured among those aged 13 and older.

[25] Other dimensions of empowerment exist but are were not specifically measured, for example: grit, motivation, negotiation and communication skills, longer-term planning, critical thinking, mental health, etc.

Statistical analysis and evaluation assessment

As a first approach for assessing the impact of AGEP on empowerment and longer-term outcomes, an intent-to-treat analysis (ITT) was implemented. An ITT analysis uses the original randomised assignment to study arm as the primary indicator of impact. The random or exogenous nature of this indicator assures that the estimate of impact is unbiased and therefore provides the highest degree of confidence in attributing any differences in outcomes specifically to AGEP. The ITT analysis estimates the average effect of the programme on the adolescents in clusters randomised to AGEP relative to girls in clusters randomised to the control. Mean and percentage differences across study arms can be assessed with relatively straightforward statistical tests, e.g., t-tests and chi-square tests, but the analysis can also be expressed within a regression framework, with and without adjustment for covariates.

The AGEP evaluation was powered to assess each of the AGEP study arms against the controls and against each other, allowing an assessment of the impact of each component (bank account, health voucher) added to AGEP. As the results did not vary significantly by each component of the intervention, they are not discussed in detail, except where appropriate. To control for any baseline differences between AGEP and control clusters that may have existed, a difference-in-differences (DID) calculation was made for each indicator. A DID approach captures the change (Δ) in an indicator over time (e.g., between Rounds 1 and 2, between Rounds 2 and 3, and between Rounds 1 and 3) separately for AGEP and control girls. From this, the objective is to determine whether the difference across AGEP and the control in the change over time is statistically significant. This can be expressed in the following general null and alternative hypothesis:

$$H_o: \quad \Delta \text{ AGEP} - \Delta \text{ Control} = 0$$
$$H_a: \quad \Delta \text{ AGEP} - \Delta \text{ Control} \neq 0$$

It should be noted that a number of hypotheses will be tested, i.e., for each of the asset and impact indicators, which raises the question whether statistical adjustments for multiple hypothesis tests should be conducted. As is well understood, as the number of comparisons increases, it becomes more likely that the groups being compared will appear to differ in terms of at least one outcome, if only by chance. The confidence that a result will generalize to independent data should generally be weaker if it is observed as part of an analysis that involves multiple comparisons, rather than an analysis that involves only a single comparison. The Bonferroni method, for example, often is used to correct statistical tests in such a case. The primary interest of the assessment, however, is not in a comparison such as whether for any of our output indicators, mediating indicators or long-run indicators differ between the AGEP and the control. Were this the comparison of primary interest, then adjustment for the multiple comparisons would be desired. Instead, the primary interest is in the specific hypotheses one by one, and thus separate questions such as: “Does AGEP increase SRH knowledge?” “Does AGEP increase performance on cognitive assessments?” “Does AGEP reduce the probability of having had sex?” For consideration of these individual hypotheses/questions one at a time independently, the standard statistical tests are considered appropriate.

Although ITT is the basis of the primary analysis, as was discussed above in the review of the implementation of AGEP, a proportion of girls invited to the programme did not enrol and, for those who did, the intensity of their participation varied. The ITT analysis does not account for the fact that many girls were never exposed to AGEP or had different levels of exposure. To account for actual programme participation in assessing the effect of AGEP on the evaluation outcomes, a secondary analysis was conducted using indicators that measured actual programme participation. It is very important, however, to recognize the statistical estimation problems that arise from self-selection into AGEP, and the degree of participation in it. Many of the same factors or characteristics that determined eligibility in AGEP were also directly related to the outcomes and behaviours of interest. Some of these individual, family, and community factors could be measured in the survey and controlled for in the analysis, but many either remained unmeasured or were hard to capture through survey interviews, such as innate capabilities, motivation, and ambition. If not properly accounted for, these unobservable factors could lead to inappropriate conclusions regarding programme impact.

We estimated two models of programme impact using AGEP participation rather than random assignment as the primary indicator of impact. The first model was naïve regarding potential self-selectivity issues, which rely on either 1) a strong assumption that the characteristics that determine participation in AGEP are not related or are exogenous to study outcomes, or 2) that the characteristics of self-selection are fully controlled by the included covariates. The second approach used instrumental variables estimation that attempts to relax these assumptions, yet obtain an unbiased estimate of impact.[26] The instrumental variables estimation was a two-step process in which the first step estimated participation in the AGEP programme using an exogenous variable not included in the second step to predict participation. If this exogenous variable was a good predictor of exposure, a second step used the predicted value of participation to assess the impact of the programme, given that the predicted participation was considered independent of any unobserved factors. This second-step estimation provided an unbiased estimate of the impact of AGEP participation. For the analysis, the original randomisation to study arm and whether participants received follow-up contact if they missed a meeting were used as instrumental variables to predict programme participation. A selection of statistical tests was used to assess the validity of the instrumental variable, including F-tests on excluded instruments, the Wald F-statistic, and the Hansen statistic for over-identification. The results of these tests indicated that the identification requirements of the instrumental variables approach were met.

Economic evaluation

The economic evaluation findings included in this report cover the inception and implementation phases of AGEP, between November 2011 and June 2016. The economic evaluation of AGEP consists of the following components:

- Collation of programme resource utilisation data from the Population Council.
- Collection of participant-specific out-of-pocket and indirect costs data.

[26] Two separate instrumental variables models were estimated and produced similar results. The first was a linear regression model with a random effects parameter at the individual level to adjust for the repeated observations on participants; it, however, could not control for within-cluster correlations. The second estimation was a linear regression model that provided robust standard errors for intra-cluster correlation; neither model could simultaneously adjust for both cluster correlations and repeated observations. The first model is presented in the results section.

- Micro-costing exercise at two health facilities, one urban and one rural, to estimate the costs of health services offered through the voucher scheme.
- Decision analytic modelling for combining programme costs and effect data to generate:
    - Average cost-effectiveness ratios (ACERs), comparing the costs and effects of Safe Spaces to no programme; and
    - Incremental cost-effectiveness ratios (ICERs), comparing the intervention arms to Safe Spaces only.

Programme costs: Direct costs were collected from AGEP budgets and expenditure reports compiled by the Population Council and its implementation partners, including the Young Women's Christian Association (YWCA), responsible for the Safe Spaces meetings; Making Cents International, who consulted on the development of the Girls Dreams savings account; and the National Savings and Credit Bank (NatSave) who rolled it out. This cost data collection technique is referred to in the literature as "gross-costing," as the total costs of service provision are collected at the institutional level and then broken down to the individual service units or end users, as opposed to "micro-costing," which collects data on resource use and unit prices at the level of the individual beneficiary of an intervention.[27] When gross-costing, resources are assumed to be equally distributed across end users, which was assumed to be appropriate for AGEP.

Start-up and implementation costs. The start-up period is defined as the time between the decision to implement an intervention and delivering to the first beneficiary.[28] Typically, start-up costs include (but are not limited to) costs incurred when recruiting personnel, procuring office space and equipment, and training field workers. Implementation costs are those incurred when the intervention is being rolled out and commonly include salaries, transportation costs, overheads, additional materials, and capital goods. It was noted in the literature that trying to separate start-up costs from implementation costs by labelling specific activity types can be time-consuming and impractical.[29] For this reason, in the present study start-up costs were defined as the costs of all activities undertaken from the beginning of AGEP in November 2011 to the end of the pilot in July 2013, irrespective of activity type. The inclusion of the first year of pilot activities in the start-up period was considered reasonable as the lessons learned during the first year were then incorporated into the final design of the AGEP intervention. While a detailed description of the start-up costs of AGEP is provided elsewhere,[30] this report will include a summary of the pilot costs, which were not fully available at the time the baseline report was prepared. Start-up and pilot costs are presented separately from other programme costs as they would not (or only partially) be incurred if the programme were scaled up or implemented in a new setting.

Programme delivery, support and joint costs. Direct costs were grouped by line items and then further categorised by applying the principle of functional classification, whereby expenditures

---

[27] Waters, H.R. and P. Hussey. 2004. "Pricing health services for purchasers: A review of methods and experiences," *Health Policy* 70: 175–184.
[28] Johns, B, R. Baltussen, and R. Hutubessy. 2003. "Programme costs in the economic evaluation of health interventions," *Cost Effectiveness and Resource Allocation* 1: 1–10.
[29] Batura, N., A. Pulkki-Brannstrom, P. Agrawal, A. Bagra, H. Haghparast-Bidgoli, F. Bozzani, T. Colbourn, G. Greco, T. Hossain, R. Sinha, *et al* 2013. "Collecting and analysing cost data for complex public health trials: Reflections on practice," *Global Health Action* 7: 23257.
[30] Bozzani, F.M. 2014. AGEP economic evaluation baseline report. Unpublished report.

are classified according to the activity they are used for.[31] The following line items were used: salaries; buildings; equipment; vehicles; utilities and insurance; supplies and services; and food, accommodation, and travel. The functional sub-categories used were: programme delivery costs (the costs of those resources employed directly in the delivery of programme components) and programme support costs (the costs of resources that are not directly consumed in programme delivery, such as training or international staff travel). All line items and their functional sub-categories used in this report are summarised in Table 14.

Table 14. Description of cost categories used in AGEP economic evaluation

Line items	Programme delivery costs	Programme support costs
Staff costs	Salaries and allowances of community sensitization fieldworkers, programme coordinator and officers for AGEP components, mentors	Salaries and allowances of other programme officers and coordinators, administrative and support staff (overhead), consultants
Buildings	Programme venue rental costs	Office space and training/workshop venues rental costs
Equipment	Programme laptops and mobile phones	Furniture, other IT equipment
Vehicles	Annualised costs, fieldwork fuel expenses, operating costs, registration and insurance of AGEP vehicles and motorbikes	Fuel, operating costs, and insurance of other vehicles
Utilities and insurance	Cell phone services for exploratory visits, sending invitations to participants, community sensitization and other programme fieldwork (excluding training)	Cell phone services for training and programme staff; bills for electricity, landline, waste collection, and internet; non-motor insurance
Supplies and services	Printing of curricula, mentor certificates, and programme material; branding and communications material for AGEP components	Stationery, office supplies, maintenance, security services, IT services, postage, storage and shipping, audit and legal fees, subscriptions, job advertisements and recruitment, banking fees
Food, accommodation, and travel	Meals, accommodation, taxis, and refunds for mentors and staff during field visits	Trainings, workshops, and international staff and consultant travel

All costs are presented in 2016 $US, using an average exchange rate for the year in which costs denominated in Zambian Kwacha were incurred[32] and inflating costs incurred in previous years using the consumer price index published by the International Monetary Fund.[33] As there is no consensus on the discount rates to be applied in the context of community participatory trials in low- and middle-income countries, the cost of all capital items, whose useful life is longer than one year, was annualised using the 3% discount rate recommended by WHO for

[31] Guinness, L. and V. Wiseman (eds.). 2011. *Introduction to Health Economics.* New York: McGraw-Hill Education.
[32] Currency historical rates [http://www.oanda.com/currency/historical-rates/]
[33] World economic outlook database. [https://www.imf.org/external/pubs/ft/weo/2016/01/weodata/weorept.aspx?sy=2014&ey=2021&scsm=1&ssd=1&sort=country&ds=.&br=1&pr1.x=43&pr1.y=14&c=754&s=PCPI%2CPCPIPCH%2CPCPIE%2CPCPIEPCH&grp=0&a=]

health programmes, which has yielded consistent results in Zambia in the past and is assumed to be appropriate given the 5-year time horizon of AGEP.[34,35]

Joint costs, arising when more than one activity, service, or output is produced from the same input, were allocated equally across intervention components (e.g., overhead costs of running the Kenya and Zambia Population Council offices). The costs of the research being conducted alongside the AGEP trial, which included this economic evaluation, were excluded from the analysis, as they were deemed to be a "once-off" cost that would not be incurred if AGEP were to be scaled up to other locations in Zambia or reproduced in another country in the future.

Direct costs of programme implementation per participant were calculated by dividing total programme costs by the number of invited girls in the intention-to-treat analysis. The rationale for including all invited girls rather than just those who completed the programme or those who participated in a given number of Safe Spaces meetings is that the programme was scaled to deliver the intervention to a target number of participants and a large proportion of the start-up and implementation costs could not be avoided even if this target could not be met. The costs per participant would thus be overestimated by using a denominator other than the number of invited girls.

Incremental costs. Incremental costs were defined as the additional costs of receiving the add-on interventions on top of the Safe Spaces meetings, which is the core AGEP intervention. These costs were calculated by adding up all costs incurred for both programme delivery and programme support for all three interventions.

Participant and mentor costs: The economic evaluation of AGEP takes a societal perspective, accounting for the costs of the programme incurred not only by the provider of the interventions (the Population Council and its implementing partners) but also for those incurred by other stakeholders such as the programme participants and mentors.

Two types of costs were captured from this perspective: the direct out-of-pocket costs of participating in the programme, such as the transport costs incurred to reach the Safe Spaces meeting venue or the health facility for redeeming the health voucher; and the indirect cost or opportunity cost of time, which includes any lost income from alternative sources as well as the cost of any substitute labour (e.g., for child care) incurred in order to participate in AGEP activities.

These costs were collected through surveys conducted with the girls as well as with the mentors. Survey data were analysed using standard cost data analysis methods and results were stratified by study site and intervention arm. The median and the arithmetic mean for out-of-pocket cost estimates were calculated and 95% confidence intervals were estimated around the mean.

Health provider costs: In an effort to provide a complete picture of the societal costs of AGEP, the economic evaluation also includes an estimate of the costs per client incurred by health providers to deliver the services covered by the health voucher.

Two study facilities were selected among the government providers enrolled in the health voucher scheme: urban Chawama clinic in Lusaka district and rural Mulungushi health centre

---

[34] Griffiths, U.K., F. Bozzani, L. Muleya, and M. Mumba. 2013. "Costs of eye care services: Prospective study from a faith-based hospital in Zambia," *Ophthalmic Epidemiology* ePub ahead of print.

[35] Chola, L. and B. Robberstad. 2009. "Estimating average inpatient and outpatient costs and childhood pneumonia and diarrhoea treatment costs in an urban health centre in Zambia," *Cost Effectiveness and Resource Allocation* 2009, 21: 7–16.

in Kapiri Mposhi district. Treatment costs incurred by healthcare providers were estimated using the so-called "ingredients" approach, whereby quantities and unit costs of all resource items were identified.

Expenditures were divided into capital and recurrent costs. Capital costs include durable items, such as buildings, vehicles, and equipment. Present values and life expectancies of capital items were approximated by consulting staff in charge. Items were annualized by 9% per year in line with the Bank of Zambia's policy rate. Recurrent costs were estimated from expense records and receipts from the accounting year 2015 and divided into major groupings, such as salaries, utilities, insurance, food and accommodation, and general supplies.

The incremental costs of providing the health voucher services to AGEP participants were then calculated by dividing the total annual costs by the number of AGEP girls who received the service at the health facility during the same year. As no health vouchers were redeemed at Chawama clinic during the evaluation year, only the results for Mulungushi health centre are presented in this report.

Cost-effectiveness modelling: Decision analytic modelling is a systematic quantitative approach for assessing the relative value of one or more different decision options, such as the different components of the AGEP trial.[36] Decision analysis is especially useful when the clinical or policy decisions are complex and information is uncertain. To combine the cost and effect data into a common analysis, a decision analytic model was constructed to generate estimates of the incremental costs per negative health outcome averted and positive progress achieved on non-health indicators from participating in AGEP.

Cost-effectiveness of safe spaces compared to no programme: To assist the independent evaluation carried out by AGEP's external contracted evaluator, which aims to compare the cost-effectiveness of AGEP to that of similar programmes in Zambia and in neighbouring countries, the incremental costs per unit of effect of Safe Spaces compared to no programme were estimated. Average cost-effectiveness ratios (ACERs) were calculated as follows for all indicators where any effect was detected between baseline and Round 3:

$$\text{ACER} = \frac{Total\ costs\ of\ Safe\ Spaces - Total\ costs\ of\ no\ programme}{Total\ impact\ among\ girls\ receiving\ Safe\ Spaces - Total\ impact\ in\ control\ group}$$

where the cost of no programme is assumed to be equal to zero.

[36] Drummond, M.F., M.J. Schulper, et al. 2005. *Methods for the Economic Evaluation of Health Care Programmes,* 3rd edition. Oxford: Oxford University Press.

AGEP Mid-Term Results

Of the 5,959 eligible adolescent girls who were targeted for interview at baseline, 5,235 completed the baseline or Round 1 survey, resulting in an overall interview rate of 88% from the household listing implemented to identify eligible participants. Approximately 90% of the adolescents interviewed in Round 1 were re-interviewed a year later in Round 2, and 89% of Round 1 participants were re-interviewed in Round 3. These overall rates mask some underlying differences. Table 15 provides further details of the survey interview completion rates for each of the study data collection rounds by location, age, and study arm; Pearson's chi-square p-values are also provided to show the statistical differences across the categories. Table 15 also provides completion rates for each of the additional components of the study, including anthropometric measurement, HIV, HSV-2, and anaemia specimen testing implemented in Rounds 1 and 3.

Table 15. Survey, anthropometry, and biomarker completion rates by study round

	Of HH listing who completed Round 1		Of R1 sample who completed Round 2		Of R1 sample who completed Round 3	
	N	%	N	%	N	%
Location						
Rural	2,273	89	2,056	91	2,008	88
Urban	2,962	87	2,637	89	2,626	89
Pearson's chi² p-value		.08		.09		.72
Age groups at baseline						
10–14	2,477 [a]	91	2,452	91	2,411	89
15–19	2,758 [a]	85	2,241	88	2,223	88
Pearson's chi² p-value				.01		.08
Age–Location groups						
10–14 rural	1,113 [a]	91	1,117	92	1,081	89
10–14 urban	1,364 [a]	91	1,335	90	1,330	89
15–19 rural	1,160 [a]	86	939	88	927	87
15–19 urban	1,598 [a]	85	1,302	88	1,296	88
Pearson's chi² p-value		.00		.01		.34
Study Arms						
SS	1,186	89	1,065	90	1,043	88
SS+HV	1,169	88	1,055	90	1,031	88
SS+HV+SA	1,160	89	1,023	88	1,030	89
Control Internal	1,146	86	1,042	91	1,020	89
Control External	574	87	508	89	510	89
Pearson's chi² p-value		.380		.202		.921

	15–19		15–21	
Completion rates	N	%	N	%
Anthropometry	5,232	99	4,634	99
HIV	2,424	96	3,054 [b]	97
HSV-2	2,360	93	2,960 [b]	97
Anaemia	--	--	2,681 [c]	97

[a] Due to misreporting of ages at the time of the household listing, the ages shifted somewhat between household listing and the baseline interview. Interviews starting in Round 2 use the baseline-reported age.
[b] Excludes those < 15 years old and those testing positive in Round 1.
[c] Half of the control sample was randomised not to receive anaemia testing, to assess the impact of testing on nutritional outcomes.

The completion rates at Round 1 for rural and urban areas differed marginally ($p < 0.10$), with a two percentage point higher completion rate for adolescents residing in rural areas. This difference continued in Round 2, reaching a similar level of statistical significance, but by Round 3, no meaningful or statistically significant differences in interview completion remained between urban and rural areas. Younger adolescents were also easier to interview at Round 1 and Round 2 ($p < .01$), but these differences largely disappeared by Round 3 as the completion rate for the older cohort remained steady, while the younger cohort's dropped. Comparing the combination of age and location characteristics, it can be observed that the younger age-cohort drove the differences in completion rates.

Data quality assessments

Response rates: A reasonable concern in programme evaluation was that individuals in areas receiving AGEP would be more willing to participate in the research than individuals in the control areas because they would have a greater vested interest in AGEP and/or because they were easier to track through the programme. If this were the case, the completion rates for AGEP girls would be statistically significantly different than for control girls. The completion rates in Table 15 by study arm, however, provide confidence that neither of these possibilities had a significant influence on the study follow-up, since the differences across study arms were not meaningfully or statistically significantly different. Thus, we do not expect differential attrition by study arm to be a factor for the analysis presented.

Baseline completion rates for the anthropometry component were very high: anthropometric data were obtained from 99.8% of the 5,235 girls who completed the Round 1 survey. Response rates for the biological specimen collection were also high. Of the 2,534 girls aged 15–19 who completed the Round 1 survey, 96% were tested for HIV and 93% provided a biological specimen for HSV-2. The difference between HIV and HSV-2 at baseline stemmed from the fact that the latter required a greater amount of blood and the fact that the specimen was not tested immediately, but taken away from the home to be tested at the laboratory. There were parents and adolescents who expressed concern about the taking of blood away from the household, leading to lower HSV-2 completion rates. By Round 3, the HIV and HSV-2 completion rates were higher (97%) among those eligible to be tested, likely due to the fact that the interviewed adolescents and guardians were more comfortable with the study and methods after two previous rounds.

Balance: Another important issue regarding the evaluation sample was whether the original randomisation was successful in distributing the characteristics of the population equally across study arms. The AGEP Baseline Report reviewed the balance of characteristics of the sample by urban and rural, by age group, and by AGEP and control arms, and concluded that the baseline results suggested no systematic differences across programme and controls at baseline that would indicate a lack of balance that might jeopardize programme assessment.[37] In Table 16, sociodemographic and behavioural characteristics are presented in summary for Rounds 1.

[37] Hewett, Paul C. et al. 2014. *Adolescent Girls Empowerment Programme: Research and Evaluation Baseline Technical Report.* Lusaka. Zambia. Population Council.

Table 16. Comparison of sociodemographic and behavioural characteristics, by study arm; % unless otherwise indicated (95% confidence interval)

Indicator	Round 1 AGEP (N=3,515)	Round 1 Internal control (N=1,146)	Round 1 p-value
Respondent level			
Age (mean)	14.3 (14.2-14.3)	14.3 (14.1-14.4)	0.765
Ever attended school	98 (98-99)	98 (97-99)	0.453
Highest grade completed (mean)[a]	5.5 (5.4-5.6)	5.6 (5.5-5.8)	0.135
Still attending school[a]	81 (80-83)	81 (78-83)	0.530
Ever had sex[b]	42 (40-45)	44 (40-48)	0.490
Ever been pregnant[b]	16 (14-18)	18 (14-21)	0.343
Ever given birth[b]	11 (9-12)	13 (10-16)	0.123
Ever used modern contraception[b]	10 (8-11)	10 (7-12)	0.949
HIV positive[b]	3 (2-4)	3 (1-4)	0.722
HSV-2 positive[b]	7 (6-9)	8 (5-10)	0.840
Parent level			
Mother is alive	88 (87-90)	88 (86-90)	0.751
Father is alive	78 (77-79)	78 (76-81)	0.856
Mother is coresident[c]	74 (73-76)	77 (75-80)	**0.038**
Father is coresident[c]	61 (60-63)	63 (59-66)	0.490
Mother completed grade 7	45 (43-46)	43 (40-46)	0.366
Father completed grade 7	50 (49-52)	51 (48-54)	0.600
Household level			
Girl is biological daughter of household head	61 (59-63)	62 (59-65)	0.637
Asset items (mean)[d]	5.9 (5.8-6.0)	5.8 (5.6-6.0)	0.540
Savings/assets			
Less than 100 kwacha	38 (37-40)	36 (33-39)	0.187
100 to less than 500 kwacha	35 (34-37)	34 (31-36)	0.282
500 to less than 1,000 kwacha	18 (16-19)	19 (17-21)	0.233
1,000 kwacha or more	9 (8-10)	11 (9-13)	**0.016**

[a] If ever attended school.
[b] Ages 15 and older at time of survey.
[c] If parent is alive.
[d] Out of the following 15 items: electricity/solar panels, radio, cassette player, television, mobile phone, fixed phone, refrigerator, table, sofa, bed, CD/digital music player, VCR/DVD player, car, motorcycle, bicycle.

Table 16 indicates that there are no statistically significant differences between the AGEP and control sample at baseline among the adolescent-level indicators. The mean age of the participants at Round 1 was 14.3 years in both AGEP and control arms, reflecting a slightly higher proportion of younger adolescents in the sample due to a marginally higher response rate at baseline for the younger cohort and a tendency for ages at the lower end of the older group to be shifted downward. There were also no meaningful differences in completed schooling or current attendance or in adolescent behavioural indicators, including the percentage having ever had sex, been pregnant, or given birth. Table 16 shows that among girls 15–19 at baseline, approximately 43% of the sample had initiated sexual activity at baseline. Among all girls 15–19, one in eight reported ever becoming pregnant, with more than one in three girls who had ever had sex reporting a pregnancy, reflecting the high rates of teenage pregnancy in Zambia.

Table 16 also indicates that the randomisation was effective in distributing girls in AGEP and control clusters equally by HIV and HSV-2 status. Among 15–19-year-olds tested at baseline, the existing HIV prevalence was 3% and the existing HSV-2 prevalence was 8%. The prevalence of HIV was lower than that observed in the 2013–2014 Zambia Demographic and Health Survey among women 15–19 years of age (5%). While more-vulnerable adolescent girls might have been expected to have even higher rates of HIV than a general population of adolescents, the baseline sample excluded ever-married girls who, in general, have higher rates of HIV.

Statistically significant differences ($p < .05$) were observed between AGEP and control clusters at baseline in two indicators at the parental and household level: the percentage of girls whose mothers were coresident was higher by 3 percentage points in the control sample, with a 2% percentage point difference in the same direction among those households who had savings of greater than 1,000 KW available. Given the overall number of adolescent and household socioeconomic baseline characteristics tested and the lack of any systematic differences between AGEP and control participants, there was no reason to believe that the models used to assess impact required adjusting by covariates, other than the potential benefits of increasing the precision of the estimated parameter of impact.

Attrition: Another important consideration in assessing the quality of the study sample for estimating the impact of the programme and for drawing inferences was the potential biases associated with attrition or loss-to-follow-up of study participants over time. A multivariable probit regression of attrition from the sample was conducted separately for each follow-up round and is presented in Table 17. The regression included the outcome measured as one if the participant could not be traced and zero if she was interviewed at Round 2 and Round 3. Each study arm and all baseline sociodemographic covariates were used as independent variables. Study Arm 3 (safe spaces, health voucher, and bank account) was marginally significantly different from the reference ($p <.10$) at Round 2, but in Round 3 there were no differences between the arms. These attrition findings by study arm were consistent whether only the internal controls were considered (Table 17) or whether both the internal and external controls were included in the analysis (data not shown). These results provide further evidence indicating that survey follow-up was independent from programme participation.

Table 17. Percentages and probit regression estimates of attrition across rounds, by study arm and baseline characteristics

Baseline indicator	Round 2				Round 3			
	Interviewed N=4,185 % (unless otherwise indicated)	Lost N=476 % (unless otherwise indicated)	Coeff. (95% CI)	p-value	Interviewed N=4,124 % (unless otherwise indicated)	Lost N=537 % (unless otherwise indicated)	Coeff. (95% CI)	p-value
Study arms								
SS	25	25	0.05	0.203	25	27	0.07	0.437
			(-0.10,0.20)				(-0.10,0.23)	
SS+HV	25	24	0.03	0.184	25	26	0.04	0.685
			(-0.13,0.18)				(-0.14,0.21)	
SS+HV+SA	24	29	0.15	0.317	25	24	0.02	0.808
			(-0.01,0.32)				(-0.16,0.20)	
Control internal	25	22	Ref		25	23	Ref	
Urban location	51	54	0.02	0.137	51	51	-0.04	0.545
			(-0.09,0.14)				(-0.16,0.09)	
Age (mean)	14.2	14.7	0.04	0.076	1423	1448	0.03	0.097
			(0.01,0.08)				(-0.01,0.06)	
Highest grade attended (mean)	5.4	5.3	-0.01	**0.027**	546	517	-0.02	0.248
			(-0.05,0.03)				(-0.06,0.01)	
Attending school	81	70	-0.15	**0.004**	81	72	-0.14	0.080
			(-0.30,0.00)				(-0.29,0.02)	
Literate (any language)	43	36	-0.03	0.104	43	37	0.02	0.788
			(-0.17,0.10)				(-0.11,0.14)	
Numeracy score at baseline (mean)	6.9	6.0	-0.02	**0.004**	686	608	-0.02	0.053
			(-0.04,0.00)				(-0.03,0.00)	
Cognitive score at baseline (mean)	6.5	6.1	-0.01	**0.014**	654	621	0.00	0.712
			(-0.02,0.01)				(-0.02,0.02)	

Tribe								
Bemba	35	38	-0.01	0.112	35	38	0.03	0.669
			(-0.13,0.11)				(-0.09,0.15)	
Nyanja	15	12	-0.18	**0.005**	15	11	-0.15	0.078
			(-0.35,-0.01)				(-0.33,0.02)	
Kaonde	8	8	-0.06	0.142	8	9	0.00	0.993
			(-0.25,0.14)				(-0.28,0.28)	
Other	41	43	Ref		41	42	Ref	
Mother is alive	88	88	0.16	0.327	88	88	0.14	0.056
			(0.00,0.33)				(0.00,0.29)	
Father is alive	78	76	0.03	0.170	78	76	-0.01	0.854
			(-0.10,0.17)				(-0.14,0.12)	
Mother completed grade 7	44	45	0.12	0.228	45	39	-0.06	0.288
			(0.01,0.23)				(-0.17,0.05)	
Father completed grade 7	51	46	-0.06	**0.038**	51	45	-0.05	0.418
			(-0.17,0.04)				(-0.16,0.07)	
Biological daughter of household head	62	50	-0.26	0.150	62	52	-0.21	**0.000**
			(-0.37,-0.15)				(-0.32,-0.10)	
Household assets[a]	6	6	-0.02	**0.005**	592	549	-0.02	0.052
			(-0.04,0.01)				(-0.04,0.00)	
Savings/assets								
Less than 100 kwacha	37	42	Ref		37	42	Ref	
100 to less than 500 kwacha	35	35	-0.02	0.102	35	35	-0.02	0.757
			(-0.15,0.10)				(-0.16,0.11)	
500 to less than 1,000 kwacha	18	15	-0.06	0.105	19	14	-0.10	0.228
			(-0.22,0.11)				(-0.27,0.06)	
1,000 kwacha or more	9	8	-0.04	0.185	9	9	0.05	0.616
			(-0.26,0.18)				(-0.14,0.24)	

Standardized vulnerability quintiles								
Quintile 1 (Least vulnerable)	24	17	-0.14	**0.046**	24	17	-0.17	0.071
			(-0.32,0.05)				(-0.35,0.01)	
Quintile 2	19	19	-0.02	0.138	19	15	-0.18	**0.035**
			(-0.18,0.14)				(-0.35,-0.01)	
Quintile 3	21	16	-0.20	**0.033**	21	19	-0.12	0.142
			(-0.36,-0.03)				(-0.27,0.04)	
Quintile 4	18	22	0.04	0.200	18	24	0.08	0.297
			(-0.11,0.20)				(-0.07,0.24)	
Quintile 5 (Most vulnerable)	18	25	Ref		18	25	Ref	
Among those who were 15 and older at baseline[b]								
Ever had sex	43	41		0.640	43	44		0.616
Ever been pregnant	16	17		0.661	16	18		0.337
Ever given birth	11	12		0.813	11	14		0.164
HIV positive	3	3		0.963	3	3		0.897
HSV-2 positive	8	6		0.314	7	9		0.436

[a] *Out of the following 15 items: electricity/solar panels, radio, cassette player, television, mobile phone, fixed phone, refrigerator, table, sofa, bed, CD/digital music player, VCR/DVD player, car, motorcycle, bicycle.*
[b] *p-values for indicator below from bivariate probits.*

Table 17 indicates that some sociodemographic characteristics were associated with attrition from the sample. For instance, completing an interview at Round 2 was associated with girls' schooling, those with a lower number of years completed and currently out of school being harder to reinterview. An 8% gap exists in the prevalence of loss-to-follow-up between the least (17%) and most vulnerable (25%). These differences persist and are jointly statistically significantly different in the probit regression ($p < .05$) when compared with the reference category of the most vulnerable. Other sociodemographic characteristics were significantly ($p < .05$) associated with loss-to-follow-up at Round 2, including age, numeracy skills, Nyanja ethnicity, mother being alive, mother's education, and whether the adolescent girl was the daughter of the head of household. However, these associations became less meaningful by Round 3, with smaller coefficients and four of the five indicators dropping in significance in the probit regression to $p < .10$, and mother's education no longer being significant. Overall, focusing on the signs of the probit coefficients rather than their significance seems to indicate that the very worst off and most vulnerable adolescent girls were the least likely to be interviewed at each round. With overall low loss-to-follow-up at around 10%, these factors were not expected to affect the impact assessments and were controlled for in the adjusted impact models.

Impact of AGEP on girls' empowerment

The primary objective of the AGEP evaluation is to assess the impact of the programme over four years on the key longer-term outcome indicators. The study's statistical tests of impact were powered to assess the impact of AGEP on the longer-term indicators after four years, although that does not preclude the statistical assessment of indicators at other points in time. In this section of the report, study findings are presented that assess the impact of AGEP on the mediating empowerment outcomes after two years, or, in other words, immediately after the end of the programme. As discussed above, empowerment was defined as the condition of possessing the assets and capacities that would allow adolescent girls to exploit the opportunities they might encounter; it was the purpose of AGEP to build on and expand these assets during the programme. As empowerment was seen as a key ingredient to improving outcomes, these assets were seen as important in producing change in longer-term adolescent outcomes. The empowerment indicators used to assess the impact of AGEP are presented in Table 12 above.

Tables 18 (below) and C-1 present summary assessments of significant findings of programme impact on the development of girls' empowerment.[38] Table C-2 provides detailed statistics of the intent-to-treat assessment[39] of programme impact on the same indicators. Table 18 and C-1 indicate the significance ($p < .05$) and direction of programme impact for the full sample (All), then for each age group and location separately. The direction and significance of the coefficients are shown for after the first year of the programme (Round 1 to 2), after the second

[38] To keep the discussion manageable, a comparison is provided only for all the intervention study arms together, as this is where the greatest average effect for the full sample would be expected and given that few differences were observed by study arm; if differences between arms emerged in the results, they will be discussed separately. Further, the results by age and location groups are provided in the Summary table only, as there are too many estimates to convert to Excel at this time.
[39] As discussed in the statistical analysis and evaluation assessment of this report, the intent-to-treat analysis compared eligible girls who resided in clusters randomised to AGEP to girls who lived in clusters randomised to not receive AGEP (the controls). It did not account for the fact that participation rates varied for girls assigned to AGEP, a consideration that will be taken up in the treatment-on-the-treated analysis that follows.

year of the programme (Round 2 to 3), and for the full programme period (Round 1 to 3). If the coefficient was not significant, the cells of the table have been left blank. It should be noted that while blank cells do represent non-significant findings at the $p<.05$ level, they do not imply that there was necessarily no effect of the programme at all; complete information on the magnitude of impact is provided in Table C-2.

In Table C-2, descriptive statistics are presented for each annual round of data from 2013 (Round 1) to 2015 (Round 3). First, the mean scores by round and by AGEP versus control are provided, measured as a mean for a continuous variable or as a percentage for a binary indicator; a statistical test was conducted to assess differences in the estimates at each round. Second, a DID was used to assess the change between rounds. The DID estimates accounted for any potential baseline differences that might have existed in the indicator, whether or not the baseline differences reached levels of statistical significance. While the simple DID estimate provided was a straightforward calculation to reveal programme impact, it did not account for the correlations that might exist across individuals in the same clusters or the fact that observations were collected from the same person over time. The final set of results in the table, therefore, include a random effects parameter at the girl level and provide robust standard errors and are thus the best assessment of what conclusions can be drawn from the intent-to-treat analysis.

Table 18. Intent-to-Treat (ITT) Summary Results: Difference-in-Difference (DID) Estimators - Excludes external urban controls

Key	
✓	At p < .05 AGEP had a significant positive effect relative to controls
†	At p < .10 AGEP had a marginally significant positive effect relative to controls
x	At p < .05 AGEP had a significant negative effect relative to controls
†	At p < .10 AGEP had a marginally significant negative effect relative to controls
	Not measured at Round 1 or Round 2

		ITT - DID (xtreg models; two rounds at a time)[b]			ITT[c]
		R2-R1	R3-R2	R3-R1	ITT
	Empowerment indicators				
	Social assets				
1	Avg. score on self-efficacy scale (0-10)	✓			
2	% confident regarding their ability to plan				
3	% feel they make good decisions regarding money	†	†		
4	% agree that permission to go to the health clinic is not a problem				
5	% who jointly or solely make decisions with regard to money earned among those who earned money				
6	Avg. number of friends				
7	Avg. number of friends in school				
8	Avg. number of friends who can be counted on if needed money				
9	Avg. number of friends who can be counted on in an emergency				
10	% have place in community where feel safe to meet girl friends	✓	x	✓	
11	% with adult female support in case of serious problem				
12	% attending any social groups/clubs within the past month				
13	% who often/sometimes go to market[a]				
14	% who often/sometimes go to community centre[a]				✓
15	% who often/sometimes go to shops or restaurants[a]				
16	Avg. score on gender equality scale				
17	Avg. score on nonacceptability of intimate partner violence				
	Economic assets				
18	Avg. score on financial literacy scale	†		✓	
19	% who have saved in the past year			✓	
20	Avg. amount currently saved				
20a	% who currently have at least 20 kwacha saved			†	
20b	Avg. amount currently saved among those who saved	x			

21	% working for cash or in-kind in the past year			
22	Avg. reported work income in the past year			†
22a	Avg. reported income in the past year among those who worked for cash			
23	% who own a bicycle			
24	% who own a mobile phone		x	x
	Health assets			
25	% Understanding pregnancy risk during menstrual cycle			
26	Avg. score on contraceptive knowledge scale (0-9)	†		✓
26a	Avg. score on SRH knowledge scale (0-11)			✓
27	Avg. score on HIV/AIDS knowledge scale (0-11)			
28	Avg. rating of health status in the past year (0-10)			
29	Avg. rating of health status in the past month (0-10)			
30	Avg. reported number of health problems in the past month (0-7)			

Notes

It should be noted that while blank cells do represent non-significant findings, it does not imply that there was no effect.

All models include site fixed effects and other covariates.
Covariates: Age, school attendance, grade attainment, literate (any language), numeracy score at baseline, cognitive score at baseline, mother alive, father alive, biological daughter of HH head, HH items, HH assets/savings, tribes (Bemba, Nyanja, Kaonde, other), mother completed primary, father completed primary, standardized vulnerability quintiles.

Models for grade attainment, completed grade 7, and completed grade 9 do not include school attendance, grade attainment, literacy, numeracy and cognitive skills.

Models for school attendance do not include school attendance.

[a] ITT coefficients are estimated from linear regression models clustered at the CSA level given that we do not have a baseline measure for this indicator to estimate the difference-in-difference coefficient (see ITT (if not measured at R1) results). The difference-in-difference coefficient is estimated in the R2 & R3 results.

[b] DID estimates, girl random effects, robust SE clustered at CSA (160 clusters)

[c] If not measured at R1: Girl random effects (if two rounds), robust SE clustered at CSA (160 clusters)

A review of baseline girls' empowerment

Reviewing Table C-2 for Round 1 data, focusing on the mean and percentage differences between those receiving invitations to participate in AGEP and those not, it is apparent that the randomisation did an effective job of distributing the sample equally across the baseline indicators of empowerment. None of the empowerment indicators measured at baseline in the table were significantly different across the study arms (AGEP versus control) at the $p < .05$ level. Indicators of literacy cognitive skills were consistently higher by two or three percentage points in the control sample at baseline; although their significance reach only the $p < .10$ level. Overall, these findings indicate that the sample was well balanced at baseline across the empowerment indicators presented. With no meaningful differences between AGEP and control clusters at baseline, the sample statistics (means, percentages) can effectively be characterized for all girls, as is done immediately below.

Social assets. Reviewing a selection of key indicators at baseline reveals the characteristics of the girls' sense of empowerment at the start of the programme. A measure of self-efficacy among girls was obtained from a 10-point scale reflecting agreement or disagreement with whether girls believed themselves able to accomplish their goals and manage problems, unexpected situations, and difficulties.[40] The average number of statements in the affirmative was six out of 10, implying that self-efficacy was already moderately high at baseline. In addition, approximately 45% of the participants agreed that it was their active planning rather than luck or fortune that determined what happened to them, while 75% believed that they made good decisions regarding the use of money. While it is clear there is room to improve on these indicators, the baseline results indicate that despite their difficult circumstances and vulnerable situations, a sizable proportion of adolescents felt at baseline that they had existing capacity to effectively handle the demands of everyday life.

Other areas of girls' social empowerment are reflected by the ability to have a say in or make decisions for themselves on important issues and the holding of positive perceptions of the roles and abilities of girls and women relative to boys and men. Two domains in which autonomy was assessed at baseline were access to health services and decisions on spending of earned income. At baseline, 95% of the adolescent girls reported that they would be easily able to obtain permission to go to a health facility if they wanted to, while 90% of those earning money indicated that they participated in decisions made on the use of money they directly earned, with 59% of these girls indicating they were the sole decision makers.

Vulnerable adolescent girls were not without other social assets at baseline. For instance, the sample girls were not completely isolated from peers and without social support in times of need. On average, girls reported having three female friends and one additional male friend; although less than half (40%) felt that there was a safe space in the community to meet their friends. Of course, not every one of these friends was perceived as being supportive in times of need, whether financially or in case of a problem or emergency; in such situations, girls reported on average only one friend who could be reliably counted on in such circumstances.

Despite this limited level of peer support in difficult times, six in 10 girls reported having a woman in the community whom they could turn to if either a serious problem arose or they needed someone who would take them in. At baseline, a sizeable proportion (21%) of

[40] The self-efficacy scale's reliability at baseline, measured by Chronbach's alpha, was .82, which is considered good for Likert-type scales.

adolescent girls were also members of social groups or clubs other than AGEP and had participated within the group in the last month. These numbers, however, mask the existence of some girls who were extremely socially vulnerable, as 4% of the sample reported not attending school, having two or fewer friends, no older women whom they could rely on, and having not participated in any social groups and clubs in the past 30 days.

Girls at baseline largely, but by no means universally, held positive perceptions and norms regarding the role of girls and women relative to boys and men, agreeing to, on average, five of seven items on a gender equality scale.[41] Averages, however, mask broad agreement with specific statements that placed a higher value on boys and men than women; for instance, 45% of girls at baseline believed that boys were more intelligent than girls, although a lower percentage (23%) suggested that boys should be favoured for schooling when resources were limited. Girls were also close to evenly split (42% and 48% agreeing, respectively) that it is a father's final decision that matters regarding money and the number of children a family should have. Violence against women was not universally rejected either, with adolescents agreeing to, on average, at least two situations among five where hitting or beating a wife by the husband was justified. The most prevalent of these was 36% who felt beating was justified if the wife neglects the children and 35% who felt it was justified if the wife argues with the husband.

Economic assets. Additional aspects of empowerment noted in Table C-2 were financial literacy, possession of economic resources and access to them, and the possession of sufficient health assets to remain productive. Possessing economic resources (savings, employment, and productive assets) is thought to reduce the constraints on girls and women and to increase their ability to avoid difficult if not harmful choices,[42] while having the knowledge and capacity to manage such economic resources is perceived as critical to effectively translating resources into positive outcomes.[43] Overall, for adolescents 10–19 at baseline, while understanding of budgets, financial goals, and resource planning were already present to some degree, with girls correctly answering, on average, five of nine financial literacy questions, the possession of economic resources was limited, with low rates of savings, work in the past year, earned income, and ownership of productive resources, such as a mobile phone or bicycle. However, as one would expect, economic empowerment increased with age. For instance, while only 11% of those 10–14 years old were able to save in the past year, this percentage was 19% among the older age group. The situation was similar for working in the past year, whether paid in kind or cash, with 23% of adolescent girls aged 15–19 working as compared to 14% of the younger cohort. For those who did save or earn money, the overall amounts were modest, averaging approximately $10 in savings and $45 in earnings in the previous year.

Health assets. It is clear that the adolescents aged 10–19 at baseline did not possess all the knowledge and skills to effectively address critical issues of adolescence and challenges in this

---

[41] The Chronbach's alpha of reliability for the gender equality construct was very low at baseline ($\alpha = .47$), suggesting that there was not consistent agreement or homogeneity in perceptions of equality as measured by the indicators; in short, many of the girls were agreeing to some of the items, but not others, and agreement with some items was diverse.
[42] Duflo, Ester. 2012. "Women empowerment and economic development," *Journal of Economic Literature* 50(4): 1051–1079.
[43] Murphy-Graham, E. and Cynthia B. Lloyd. 2015. "*Empowering adolescent girls in developing countries: The potential role of education," Policy Futures in Education* 14: 556–577.

period of their lives.[44] For instance, very few of the adolescent girls (approximately 8%) understood the basic fact that there are times during the menstrual period when pregnancy is more likely; understanding was only marginally better among older adolescents 15–19 years of age, at 13%. This was followed by knowledge of, on average, only one traditional or modern contraceptive method for preventing pregnancies among those 10–19; this statistic increased to only two contraceptive methods for older girls. While performance on an assessment of HIV/AIDS knowledge was higher, with six out of 11 items answered correctly, given the magnitude of resources and effort in transmitting HIV-related information, the score reflects a significant gap in comprehensive knowledge of HIV. It is clear this gap remains throughout adulthood, as the 2013–2014 Demographic and Health Survey indicates that only 42% of women aged 15–49 had comprehensive knowledge of HIV/AIDS.

While there is no question that adolescent girls have specific sexual and reproductive health needs that often go unmet, adolescence is, in general, considered a relatively healthy time of life. This fact was reflected in the baseline data, both in terms of subjective assessments of health status and by objective indicators of illnesses experienced in the past month. Use of a 10-point scale to capture perceptions of health that ranged from very poor to very good indicated that adolescents, on average, felt their health was good to very good, in both the past year and the past month (8.0/10.0 and 8.4/10.0, respectively). Similarly, very few illnesses were reported, with on average less than one of seven potential illnesses experienced in the last month. Overall, the baseline data indicated that adolescent girls' generally positive health was an existing asset that could be relied on. That said, given that 42% of girls were sexually active and few were using contraception for pregnancy or HIV prevention, an HSV-2 prevalence of 8%, an HIV prevalence of 3%, and 16% of girls already having been pregnant, it should not be forgotten that there was a sizable proportion of adolescents in this age group with very important sexual and reproductive health needs.

Skills and competencies: Performance on assessments of reading and numeracy at baseline also revealed significant capability gaps. For instance, only one of three adolescents was able to read a simple sentence in a local language, with a slightly higher percent (40%) able to read a simple English sentence. While these indicators did not directly measure the percentage of adolescents who were able to effectively comprehend written information, by definition, the percentage that were able to do so was less, suggesting a wide gap between existing ability and capacity needed. Numeracy scores on grade three exam questions were also very low, with correct responses provided, on average, on only 7 items out of 15 on the assessment, a rate of 47%. Since all adolescents in the study should have completed grade three by the time of the baseline survey, this finding is particularly telling.[45] Again, while not directly measured, the lack of numeracy skills as measured suggests an even greater percentage of adolescents unable to effectively use numeracy skills in dealing with activities and exchanges that are dependent upon them.[46]

---

[44] These findings provide a strong rationale for knowledge- building programmes such as AGEP, as it is apparent that the educational system is not focusing on important topics of relevance to adolescent girls and/or has not yet reached a state of sufficient quality to address the skills and needs of students.

[45] The average grade attended at baseline was grade 6 (Hewett et al., 2014. *AGEP Baseline Technical Report*). This average is lower than should be for this age group due to late entry, grade repetition, and dropout.

[46] While these sets of asset indicators reflect important capabilities and are therefore discussed here, they were not expected to be directly affected by the AGEP intervention as there were no programme components that would build capacity in these specific areas. These skills are typically expected to be developed through the formal educational system.

Assessing change over time in empowerment, difference-in-differences, intent-to-treat

As noted above, in addition to the snapshot of adolescent girls' empowerment at baseline, the change from Round 1 to Round 3 in the selected empowerment indicators was measured and the impact of AGEP on such outcomes assessed. These are presented in summary form in Table 18 (above) and C-1, as well as for all relevant statistical estimates in Table C-2.
It should be noted that there may be a range of factors driving change in girls' empowerment, aside from the impact of AGEP. For instance, as girls age they may acquire new capabilities, acquire new information or hone the skills they already possess; attending good-quality schools would certainly contribute to their development. Adolescents may also be exposed in their communities to other programmes, interventions and/or information similar in nature to that provided through AGEP. Further, as they become older, girls may be provided or take on new responsibilities and challenges that naturally enhance their sense of efficacy, skills, autonomy, and control. It should be recognized, however that not all change is necessarily positive. It is also possible that girls lose skills and capabilities over time due to increasing social isolation, lack of practice, monotony of activities, and lack of environmental stimulation.[47] Ultimately, it takes a comparison between two groups that are similar in all other respects (AGEP versus control) to isolate the impact of AGEP on change.

Change in social assets: The results in Table 18 (above) and C-1 indicate that for two social asset indicators the change was significantly greater for all girls in AGEP relative to the controls, specifically for the measure of self-efficacy and agreement that there is a safe space in the community to meet. As can be noted, however, the change is significant only after the first year of the programme for the self-efficacy scale, although significance is reached for Round 1 to 3 for the safe space indicator. The results broken down by age and residence indicate that the effect is statistically significant at $p < .05$ only for younger adolescents living in rural areas, although it is still positive in other age groups. This is suggestive that the impact of AGEP in these social assets was greatest for younger, rural adolescent girls, perhaps as they were also the group most likely to participate in AGEP. As indicated by the changes in impact that occur between Rounds 2 and 3, young and rural AGEP girls began to lose ground across an array of social assets relative to control girls. This observation may be driven by stagnation in impact of the programme, a reduction in the AGEP participation rates in year two, and/or by the possibility that control girls were gaining efficacy through other means that effectively served as a substitute for AGEP.[48]

Table C-2 shows that girls' self-efficacy increased on average by approximately one unit of the 10-point scale from Round 1 to Round 3, with a slightly higher increase for girls in AGEP. While in absolute terms, this does not appear to represent dramatic change, it does translate to approximately a 15% increase in measured self-efficacy from the baseline measure. The unadjusted and adjusted, intent-to-treat, difference-in-differences estimates by round show that this change was statistically significantly different for those invited to participate in AGEP compared with controls, but only after the first year of the programme. After the second year,

---

[47] Soler-Hampejsek, E., B.S. Mensch, S.R. Psaki, M.J. Grant, C.A. Kelly, and P.C. Hewett. 2016. "Reading and numeracy skills after school leaving in southern Malawi: A longitudinal analysis." Manuscript.

[48] This finding may also occur if there were significant amounts of contamination or crossover from AGEP to control in the second year of the program. Given the lower-than-expected rates of participation and the relatively low number of control girls who reported participating in AGEP (12%), this is not likely to be the case, but needs to be assessed further. Contamination may also have occurred from other programmes operating in control areas, however, a programme mapping of the areas revealed no girl-centred programme addressing similar outcomes.

the control girls closed the self-efficacy gap.[49] This pattern of change was also apparent in the dichotomous indicator of internalized perceptions of money management skills. While there was an increase by 9% after one year of AGEP girls who felt they made good decisions about money and only a +4% change among control girls in the same period, after two years the percentage change in this indicator equalized between AGEP and control girls at +17% by Round 3.

Surprisingly perhaps, the AGEP programme did not greatly alter the network of friendships among the adolescent girls as measured in the quantitative indicators.[50] Girls did start the programme with a sizable number of friends and the average number of friends did not significantly change through the AGEP programme. In fact, whether in the programme or not, the average number of friends (female and male) declined over time (albeit not statistically significantly), likely due to increasing migration and marriage among girls as they aged. The average number of friends that could be counted on in times of need did increase, whether in terms of providing economic support or in terms of helping when problems or emergencies arose; however, no statistically significant differences were observed between AGEP and control girls in this respect. Similarly, access to older women who could be counted on or who would take the girl in if need be also significantly increased over time, by approximately 20% and 15%, respectively; this change, however, also did not vary by the study arms to which girls were randomised.

The AGEP programme introduced an opportunity for some adolescent girls that had not been widely available prior to the programme, specifically, a safe community space where they could meet with their friends. The percentage of girls reporting the availability of such a space grew by about 12% in the AGEP arm, while it remained stagnant over time for the control sample. The differences between AGEP and control were statistically significant at the $p < .05$ level in the regression models. It should be noted that the space was not available for all AGEP girls, since many weekly meetings (particularly in rural areas) did not take place in their immediate communities. The availability of such safe spaces was also reflected in the differential reporting of girls who said at Round 3 that they went often/sometimes to such spaces, with AGEP significantly ($p < .01$) reflecting a higher percentage of such cases at around 12%.[51] While AGEP did lead to greater use of community centres and such meeting places by Round 3, it did not appear to have much translational impact on girls' mobility to other common spaces, such as markets or stores/restaurants. Partially, this finding was due to the fact that girls' access to these areas was very high (>80%) with or without the programme.

Whereas many of the indicators changed over time as girls matured and/or learned new things, perceptions of gender equality and norms regarding gender-based violence against women remained intractable. The notion that boys and girls, men and women are equal across innate characteristics, such as intelligence, or in norms concerning who should attend school or make decisions in the household, remained unchanged, whether among AGEP girls or among controls. Represented as a mean value of seven items, the average number reflecting positive gender attitudes remained fixed at five overall. Also, significant percentages of girls remained

---

[49] This difference-in-differences estimate in favour of change in this indicator for the control sample between Round 2 and Round 3 was statistically significant at $p < .05$.

[50] The quality of girls' friendships with fellow AGEP girls as compared to non-AGEP girls is discussed in the qualitative data collected at the same time. A theme that emerges from the qualitative data is that girls perceive their AGEP friends to be of higher quality – less likely to be involved in "bad behaviour," a more positive peer influence, etc. For more details see the full qualitative report available at www.popcouncil.org.

[51] Measures of travel to community centres, markets, and stores/restaurants were not included at baseline and, therefore, these indicators could not be assessed as changes since baseline.

supportive of a husband's right to hit or beat his wife for certain reasons, although as was the case at baseline, on average, such behaviour was more rejected than accepted. The lack of change in perceptions of gender, gender roles, and violence against women points to the deeply ingrained nature of these normative attitudes and beliefs.

Change in economic assets. The findings presented in Table 18 (above) and C-1 indicate that two important economic assets were positively and significantly changed by participation in AGEP for all girls over the full period (Round 1 to Round 3), specifically adolescent girls' financial literacy and the percentage of girls who saved in the past year. These results were only significant when assessing the full programme period, but this does not mean they were not positive for girls after the first and second year of the programme, even though the cells in the table are blank. For financial literacy, it is only in the younger cohort of adolescents living in rural areas that the estimates reach a level of statistical significance, suggesting stronger impact from AGEP for rural adolescent girls. As observed previously, the statistical significance likely emerges from the fact that it was the younger rural adolescent girls who were most likely to participate in AGEP and therefore to be influenced by the programme. Statistically significant impacts are observed for both older and younger urban girls in the percentage saving in the past year. Older urban girls in AGEP are also significantly more likely to have worked for cash or in-kind in the past year. The access for urban girls to the formal banking, markets, and employment provide a rationale for these findings. It is, however, the AGEP girls who were most likely to take advantage of these opportunities.

Focusing on Table C-2 indicates that as girls aged, they were more likely to work and earn money for labour and to save more of those funds. The percentage of girls working in the past 12 months increased by approximately 35% between Round 1 (28%) and Round 3 (38%), while the percentage earning money from work increased by approximately 33% over the same time period. Although not significant, the average amount earned for working AGEP girls increased by approximately 38% between Rounds 1 and Round 3, while earnings for working control girls remained stagnant. This work and earnings increased the proportion and amounts of girls' savings, and by Round 3 girls in the AGEP programme were significantly more likely to have saved in the past year (R1–R3, $p < .05$) relative to control girls; although for those who did save, their mean savings amount was largely the same across arms. This increased savings may have been the result of greater financial literacy among AGEP girls relative to controls observed over time (R1–R3, $p < .05$). As with self-efficacy, this result was largely driven by the changes in the indicator observed between Round 1 and Round 2.

Change in health assets. A primary focus of the AGEP Health and Life Skills curriculum was to address low levels of sexual, reproductive, and HIV knowledge among adolescent girls. Table 18 (above) and C1 indicate that the AGEP programme had a positive and statistically significant impact in the ITT analysis on only one of the health asset indicators, specifically contraceptive knowledge. Girls participating in AGEP, and particularly younger rural adolescents, were significantly more likely over time to acquire greater contraceptive knowledge relative to controls. This result was observed mainly after two years of the programme's implementation, but differences between AGEP and control girls were emerging after the first year as indicated in the detailed findings in Table C-2. Sexual and reproductive health knowledge did emerge as significant in the Safe Space group arm only (Table 22), although there were no significant impacts by arm on self-reported health assessments or on health issues experienced.

As indicated in Table C-2, AGEP did not appear to greatly increase the percentage of girls who came to understand the variation in pregnancy risk during the menstrual cycle; even by Round 3 when girls were aged 12–21, only 9% understood this basic information. Nor did the programme seem to differentially affect HIV knowledge over time. While the average score in HIV knowledge increased by approximately 40% over time, it increased equally for both the control and AGEP samples. In contrast, AGEP did have statistically significant ($p < .01$) and sustained effects on contraceptive knowledge. There was a 70% increase in the mean number of contraceptive methods known by AGEP girls, whereas the mean number of methods known by control girls increased by only 50%. To put this finding into perspective, however, the change translated into girls knowing one additional contraceptive method by Round 3, with an average total of only two methods known. The average number of methods known for girls 15–19 was only marginally higher, with just over two methods reportedly known on average by Round 3.

The AGEP Safe Spaces training curricula, including life skills, sexual reproductive health, and nutrition, addressed adolescent health in multifaceted ways: providing relatively low-resource intensive methods to reduce sexual reproductive health risks, avoid sexual violence, prevent anaemia, and improve nutrition. For girls in the health voucher arm, AGEP also provided access to adolescent-friendly health services for general wellness exams and, when desired, sexual and reproductive health care and treatment. While there was little indication of impact of these programme efforts in the main analysis (AGEP versus control), there was some indication that for two of the study arms (safe spaces only and safe spaces with the health voucher) there was improvement of self-reported health in the first year of the programme, with results significant at the $p < .10$ level (results not shown). As suggested previously, these early impact results with less impact overall may reflect the declining participation rates in AGEP safe spaces over time, as well as limited use of the health voucher by those receiving it.

Assessing change over time in empowerment, considering intensity of participation in AGEP

As discussed in the statistical analysis and evaluation assessment section of this report, the ITT analysis was considered the starting point of programme impact evaluation as it randomly distributed potential confounding participant characteristics (individual, family, community) evenly across arms. The ITT provided an estimate of the average effect of the programme on girls residing in clusters randomised to AGEP as compared to girls residing in clusters randomised to control. The ITT analysis, however, did not take into consideration that some of the girls randomised to receive AGEP chose not to participate and those who did participate, did so in varying degrees. The following analysis addresses this limitation by focusing on an indicator of programme impact that captures the intensity of AGEP participation. In the analysis to assess impact, intense participation is categorized as attending greater than 26 or half of all sessions in each year of the programme. An additional analysis also assessed change from Round 1 to Round 3 by using an indicator of attending more than 52 sessions over two years of the programme. Descriptive statistics of a summary measure of intense participation are provided in Tables 2–5 above in the review of the programme.

The results of the analysis using a measure of the intensity of programme participation are presented in Tables 19 (below), C-3 and C-4. Table C-3 indicates the significance ($p < .05$) and direction of programme impact for the full sample (All), then for each age group and location

separately. The direction and significance of the coefficients are shown for after the first year of the programme (Round 1 to 2), after the second year of the programme (Round 2 to 3), and for the full programme period (Round 1 to 3). If the coefficient was not significant, the cells of the table have been left blank. It should be noted that while blank cells do represent non-significant findings at the $p < .05$ level, they do not imply that there was necessarily no effect of the programme at all; complete information on the magnitude of impact is provided in Table C-4. In Table C-4, the columns in the table show the regression results for the indicator of intense programme participation separated by each AGEP year; between Rounds 1 and 2, between Rounds 2 and 3, and over the full programme from Round 1 to Round 3. All models have been adjusted for covariates noted at the bottom of the tables.

Table 19. Treatment-on-Treated (ToT) Summary Results: Difference-in-Difference (DID) Estimators - Excludes urban controls

Key	
✓	At p < .05 AGEP had a significant positive effect relative to controls
†	At p < .10 AGEP had a marginally significant positive effect relative to controls
x	At p < .05 AGEP had a significant negative effect relative to controls
†	At p < .10 AGEP had a marginally significant negative effect relative to controls

		TOT - DID[b]		
		TOT R2-R1[c]	TOT R3-R2[d]	TOT R3-R1[e]
	Empowerment indicators			
	Social assets			
1	Avg. score on self-efficacy scale (0-10)	✓		
2	% confident regarding their ability to plan			
3	% feel they make good decisions regarding money	✓	†	
4	% agree that permission to go to the health clinic is not a problem			
5	% who jointly or solely make decisions with regard to money earned among those who earned money			
6	Avg. number of friends			
7	Avg. number of friends in school			
8	Avg. number of friends who can be counted on if needed money			
9	Avg. number of friends who can be counted on in an emergency			
10	% have place in community where feel safe to meet girl friends	✓	x	†
11	% with adult female support in case of serious problem			
12	% attending any social groups/clubs within the past month			
13	% who often/sometimes go to market[a]			
14	% who often/sometimes go to community centre[a]	✓		✓
15	% who often/sometimes go to shops or restaurants[a]			
16	Avg. score on gender equality scale			
17	Avg. score on nonacceptability of intimate partner violence			
	Economic assets			
18	Avg. score on financial literacy scale	✓		✓
19	% who have saved in the past year			✓
20	Avg. amount currently saved		†	
20a	% who currently have at least 20 kwacha saved			
20b	Avg. amount currently saved among those who saved	x		
21	% working for cash or in-kind in the past year			
22	Avg. reported work income in the past year			
22a	Avg. reported income in the past year among those who worked for cash			

23	% who own a bicycle			
24	% who own a mobile phone		x	x
	Health assets			
25	% Understanding pregnancy risk during menstrual cycle			
26	Avg. score on contraceptive knowledge scale (0-9)	✓		✓
26a	Avg. score on SRH knowledge scale (0-11)	†		✓
27	Avg. score on HIV/AIDS knowledge scale (0-11)			
28	Avg. rating of health status in the past year (0-10)			
29	Avg. rating of health status in the past month (0-10)			
30	Avg. reported number of health problems in the past month (0-7)			

Notes

Indicator was denoted as significant at the p <.05 level if both the girl random effects and the robust SE clustered at CSA models were significant at p<.05. It was denoted as significant at the p<.10 level if both models were significant at p<.10 or if one was significant at p<.05 and the other at p<.10.

It should be noted that while blank cells do represent non-significant findings, it does not imply that there was no effect.

All models include site fixed effects and other covariates.

Covariates: Age, school attendance, grade attainment, literate (any language), numeracy score at baseline, cognitive score at baseline, mother alive, father alive, biological daughter of HH head, HH items, HH assets/savings, tribes (Bemba, Nyanja, Kaonde, other), mother completed primary, father completed primary, standardized vulnerability quintiles.

Models for grade attainment, completed grade 7, and completed grade 9 do not include school attendance, grade attainment, literacy, numeracy and cognitive skills as covariates.

Models for school attendance do not include school attendance as a covariate.

[a] TOT coefficients are estimated from linear regression models clustered at the CSA level (Model B) at R2 (in R1 & R2 results) and R3 (in R1 & R3 results) given that we do not have a baseline measure for this indicator to estimate the difference-in-difference coefficient. The difference-in-difference coefficient is estimated in the R2 & R3 results.

[b] DID estimates using instrumental variable analysis: girl random effects and robust SE clustered at CSA (160 clusters)

[c] TOT = Attended 26+ meetings between R1 & R2

[d] TOT = Attended 26+ meetings between R2 & R3

[e] TOT = Attended 52+ meetings between R1 & R3

Three separate estimation models were used to obtain the impact or programme participation on outcomes, details of which are shown in Table C-4. The first estimation model was considered "naïve," as it did not attempt to empirically account for the potential selectivity of those deciding to participate in the programme, whether at the individual, family, or community level. The second and third models of impact provided in the table used instrumental variables estimation, which first modelled the characteristics of intense programme participation (results shown in the programme report) before estimating the impact of intense AGEP participation on empowerment. The coefficient of impact in the second and third model, therefore, took into account the potential unmeasured selectivity of girls who intensely participated before estimating programme impact. The difference between models 2 and 3 has to do with the ways in which correlations were modelled, one using a random-effects parameter and one using robust standard errors at the cluster level; it was not possible to estimate both random effects and robust standard errors simultaneously. As there were a limited number of differences between them, the following discussion of results does not differentiate them further. Only if the coefficient was significant at $p < .05$ in both models was it displayed as significant in the summary Table 19 (above) and C-3.

The estimation results presented in Table 19 (above) and C-3 indicate that intense AGEP participation was statistically significantly associated ($p < .05$) with a selection of indicators of social assets, although overall it seem to have a limited impact. For instance, focusing on the results after year one of the programme (survey Round 1 to Round 2), such intense participation was positively associated with girls' self-efficacy and self-confidence in decision making about money. Girls who intensely participated in AGEP were significantly more likely to have higher levels of efficacy and confidence relative to the control girls and to girls who participated less intensely in AGEP or not at all. As with the ITT results presented earlier, it is the in higher-participating younger rural girls where the effect of AGEP is greatest. AGEP also clearly improved such girls' access to safe spaces to meet with their friends and their mobility within the community to visit such spaces and these results applies to the Round 1 to 3 assessment as well.

Similar to the ITT results, intense participation in AGEP was associated with significantly greater changes in three indicators of economic assets and one indicator of health assets. For financial literacy and savings in the past year the estimated parameters were statistically significant and suggest that if girls participated intensely in the programme, they improved their understanding of financial matters and translated that understanding into a greater propensity to save; although it should be noted that the amount saved is no different between intensely participating AGEP girls and others. Older urban girls intensely participating in AGEP were also more likely to work for cash or in-kind in the previous year. For health assets, contraceptive knowledge changes at a significantly higher rate for intensely participating girls. As with other results, this effect is particularly the case early in AGEP and for the younger, rural cohort. The strength of the impact between Rounds 1 and 2 is large enough to lead to positive overall results for the programme between Round 1 and 3.

In Table C-4, the second set of estimation results (models 2 and 3) for each round separately, as well as for the overall results for Rounds 1 through 3, attempted to address a question whether the observed positive association between intense programme participation and empowerment were the result of AGEP or whether they were rather capturing the selective characteristics of girls that drove both intense participation and greater levels of empowerment. If the coefficient of impact declined and became insignificant in these models, it would be suggestive that unobserved characteristics of girls were motivating the results of

the naïve model. In such a case, it would be improper to infer that AGEP had had an impact on empowerment. If the coefficient for intense participation remained meaningful and significant, it would suggest otherwise. The pattern of results in the naïve models of impact showing a decline in year two of AGEP are suggestive that it was not the fixed characteristics of the girls that were driving the results observed.

In Table C-4, the results of models 2 and 3 indicate that while some indicators may have been attributable to the selectivity of girls who participated intensely in AGEP, a selection of the empowerment outcomes were not. Focusing on the impact where it was observed to be the greatest (Rounds 1 to 2) indicates that, while the coefficient of impact for gender equality norms, savings and earned income in the past year, and HIV knowledge, moved from significant to insignificant from the naïve to the instrumental variables regression model, AGEP's impact on two measures of self-efficacy were positive and statistically significant ($p < .03$). At least in the first year of the programme, AGEP was observed to increase girls' self-efficacy, independent of any potential confounding from self-selection in programme participation. This result was also paralleled in indicators and the availability of a safe space in the community, financial literacy, and contraceptive knowledge, all statistically significant at the $p < .05$ level. Additionally, when assessing the full programme period Rounds 1 to 3, one indicator—saved money in the previous year—revealed a coefficient of impact that was positive and statistically significant at $p < .01$. This result suggests that, overall, AGEP increased the percentage of girls who were saving money either formally or informally.

There is a pattern, observed best in Table C-4, in many of the empowerment indicators if tracked over time; specifically, the impact of intense participation—as measured by coefficient magnitude and significance—was greatest in year one of the programme, declined in year two, and was sometimes statistically significant and sometimes not statistically significant when assessed across the total programme period (Rounds 1 to 3). For example, changes in contraceptive knowledge among those having intense participation were observed as large and statistically significant after year one of the programme, but the magnitude of the coefficient of impact dropped in year two, becoming statistically insignificant ($p =. 25$, model 2 and $p=.23$, model 3). When changes in contraceptive knowledge were assessed for the full programme period, the weight of impact in year one was strong enough to lead to statistically significant impacts over the period ($p = .001$ for both models). A similar pattern was observed for the availability of a safe space in the community, travel to the community centre, and financial literacy.

Impact of AGEP on longer-term outcomes

While a change in longer-term outcomes for adolescents was not expected to occur until after the full four years of observation were over, presented below are the results of an assessment of impact after two years on the indicators listed in Table 13. In many cases, such as marriage, school dropout, first birth, etc., few outcomes had yet been experienced for the younger cohort of adolescents, and therefore it would be premature to draw conclusions about the ultimate impact of AGEP at this stage on the longer-term indicators. Further, as discussed in the statistical analysis and evaluation assessment section of the report, the study sample size was powered based on estimates of what the means and prevalences of the indicators would be when the sample reached the ages 14–23. While the sample size was conservatively powered to assure a rigorous assessment, and therefore provide some leeway in conducting this early

analysis of impact on those currently aged 12–21, it would be mistaken to conclude that a null effect suggests strong evidence supporting the alternative hypothesis.

Assessing change over time in longer term-indicators, difference-in-differences, intent-to-treat

As Round 1 descriptive statistics for the longer-term indicators were presented and discussed in the *AGEP Baseline Report*,[52] the following discussion focuses on the assessment of change in those indicators over time and an assessment of impact of AGEP two years after baseline. As with the presentation of results on the indicators of empowerment, intent-to-treat results are presented first followed by an assessment that includes a measure of actual programme participation. The summary of statistically significant results is presented in Table 20 (below) and C-1 and the parameter estimates and level of statistical significance are provided in Table C-2.

[52] Hewett et al. 2014. *AGEP Baseline Technical Report.*

Table 20. Intent-to-Treat (ITT) Summary Results: Difference-in-Differences (DID) Estimators - Excludes external urban controls

Key	
✓	At p < .05 AGEP had a significant positive effect relative to controls
†	At p < .10 AGEP had a marginally significant positive effect relative to controls
X	At p < .05 AGEP had a significant negative effect relative to controls
†	At p < .10 AGEP had a marginally significant negative effect relative to controls
	Not measured at Round 1 or Round 2

		ITT - DID (xtreg models; two rounds at a time)[e]			ITT[f]
		R2-R1	R3-R2	R3-R1	ITT
Impact indicators					
	Educational				
1	Avg. number of years completed	†			
2	% completed primary school				
3	% completed junior secondary school[b]				
4	% currently attending school				
	Sexual risk behaviour[b]				
5	% ever had sex				
5a	Avg. age at first sex among those who ever had sex and report knowing age at first sex[c]				
6	% agree that they have had unwanted sex				
7	% agree that they have had transactional sex			✓	
8	% used condom at last sex with non-marital partner	†			
9	% used condom at first sex			✓	
	Marital[b]				
10	% ever married				
11	Avg. number of HIV risk-related topics discussed with partner (0-5)[a]				
12	Avg. marital control score (0-6)[a]				†
	Pregnancy & births[b]				
13	% who have ever been pregnant				
14	% who have ever had an unwanted pregnancy				
15	% currently pregnant or who have given birth				
16	% ever used modern contraception				
	Sexually transmitted infections[b]				
17	% HIV positive				
18	% HSV-2 positive				

	Experience of violence[d]
19	% have experienced physical violence in past 12 months
20	% have experienced intimate partner violence in past 12 months

Notes

It should be noted that while blank cells do represent non-significant findings, it does not imply that there was no effect.

All models include site fixed effects and other covariates.

Covariates: Age, school attendance, grade attainment, literate (any language), numeracy score at baseline, cognitive score at baseline, mother alive, father alive, biological daughter of HH head, HH items, HH assets/savings, tribes (Bemba, Nyanja, Kaonde, other), mother completed primary, father completed primary, standardized vulnerability quintiles.

Models for grade attainment, completed grade 7, and completed grade 9 do not include school attendance, grade attainment, literacy, numeracy and cognitive skills.

Models for school attendance do not include school attendance.

[a] ITT coefficients are estimated from linear regression models clustered at the CSA level given that we do not have a baseline measure for this indicator to estimate the difference-in-difference coefficient (see ITT (if not measured at R1) results). The difference-in-difference coefficient is estimated in the R2 & R3 results.

[b] Indicators in this domain measured among those aged 15 and older.

[c] Models will be estimated at endline using survival analysis.

[d] Indicators in this domain measured among those aged 13 and older.

[e] DID estimates, girl random effects, robust SE clustered at CSA (160 clusters)

[f] If not measured at R1: girl random effects (if two rounds), robust SE clustered at CSA (160 clusters)

Table 20 (above) and C-1 indicate that, to-date, statistically significant differences in longer-term outcomes were observed in two indicators after two years of AGEP. AGEP was positively associated with reductions in transactional sex and increases in the percentage of participants who reported having used a condom at first sex. For the latter indicator, this finding was strongest for older rural adolescent girls relative to older urban adolescent girls, but was nonetheless positive for both. While other indicators were not statistically significant after two years of programme impact, given the fact that these tests are not fully statistically powered and are not observed for the youngest adolescents in the sample, it should not be assumed that statistically significant impact will not be observed after all four years of observation are completed in the sample.

As would be expected, adolescence is a period of significant change and this was no less the case for the AGEP evaluation sample. For instance, Table C-2 (summary statistics section) indicates that an additional 20% of adolescent girls completed their primary education over the two-year period, with an additional 15% completing junior secondary. Thus, by Round 3, approximately 56% of the sample had completed primary school, and 25% had completed junior secondary. Despite these significant milestones reached by a number of girls, in total, 34% had left school by Round 3 and, therefore, had completed all the formal education they were likely to acquire in their lifetime; 13% had left school between Rounds 1 and 3 on top of the 21% who were not in school at the time of the baseline survey. Both the absolute differences and the intent-to-treat, difference-in-differences estimates between programme and control girls revealed no significant differences in schooling outcomes. There was a marginally higher ($p < .10$) average number of grades completed among AGEP girls relative to controls between Round 1 and Round 2, but that difference faded to insignificance by Round 3.

By Round 3, nearly two out of three girls aged 15 and older had initiated sex. This represents a 50% increase over the percentage of girls who were sexually active at baseline. While AGEP girls were slightly more likely to have ever had sex by Round 3 (66% versus 63%), there are some promising trends for AGEP girls relative to control girls in the sample. For instance, AGEP girls were significantly less likely (by a margin of approximately 13% [$p < .01$]) to have reported having engaged in transactional sex between Round 1 and Round 3 as shown in the intent-to-treat analysis. Also, there was a difference in condom use at first and last sex, with AGEP girls increasing their use over time (by 3–4%) relative to no change in condom use among controls; in the DID regression model, condom use at first sex among AGEP participants was significantly different compared with controls at $p < .01$. That said, where girls had less control over their sexual activity, specifically reporting unwanted sex, there was no difference between AGEP and control girls; more than one out of every three girls who had initiated sexual activity reported experiencing unwanted sex by Round 3.

In addition to starting to have sex, a greater proportion of girls had transitioned to marriage, pregnancy, and first birth by Round 3. For each of these variables, the pace of change had doubled or more the percentage of girls who had transitioned along these milestones. It is interesting to note that while approximately 17% of the 14–21 sample had now ever married by Round 3, double that percent (34%) had ever been pregnant, highlighting the significant issue of early pregnancy in Zambia and representing an increase of 100% from the baseline percentage. Also, as was noted in the *AGEP Baseline Report*, the high pregnancy rate was associated with a high percentage of unwanted pregnancies; 23% of adolescent girls reported an unwanted pregnancy by Round 3, data that suggest there remains a large unmet need for contraception among adolescents. For instance, only 26% of girls who were sexually active had

ever used modern contraception, although it is interesting to note that this percent increased significantly to 70% among adolescents who had already given birth. These data point to lack of knowledge and access, as well as potentially stigma, as remaining barriers to girls' access to modern contraception; the provision of the health voucher had not as yet significantly changed the use of modern contraception.

Table C-2 also provides information on the HIV and HSV-2 infection rates in the sample of girls over time. As was noted in the Methods section, biological specimens for HIV and HSV-2 were collected and tested in Round 1 and Round 3 among all girls aged 15 and older. The baseline prevalence of HIV among the vulnerable adolescent girls aged 15–19 in the baseline sample was 3.0%, with a marginal, but insignificantly higher prevalence among AGEP participants relative to control cases. By Round 3, the HIV infection rate increased for those 15 and older by 1.6% in the AGEP sample of girls and 1.3% in the control sample. By Round 3, the prevalence of HIV was 4.7% among the AGEP sample of girls and 4.1% among the control sample; a nonsignificant difference. For HSV-2, baseline prevalence was more than double that of HIV, with approximately 7.5% of the adolescent girls having been infected with HSV-2; again, no meaningful differences existed in HSV-2 prevalence among AGEP and control girls at baseline. The two-year HSV-2 incidence rate was quite high for both the AGEP girls (at 7.5%) and for control girls (at 6.0%). While differences exist between AGEP and control, the difference-in-differences estimate remained insignificant.

Girls aged 13 and older were asked about their experience of physical violence in the previous 12 months. Among adolescents in the sample, the experience of physical violence had meaningfully dropped over the course of approximately two years, from an average of 38% to around 18% of adolescents reporting such incidents in the past 12 months. This trend may be a function of a lower experience of violence among the youngest cohort of adolescents, who were answering the question for the first time in Round 3, or actual declines in physical violence among all adolescent girls in the sample. The amount of change observed in the AGEP sample was virtually the same as that experienced by control girls over time, with no statistically significant difference in the intent-to-treat difference-in-differences estimate. Unlike general physical violence, the percentage of the sample reporting intimate partner violence remained virtually unchanged over time, despite the fact that more of the sample adolescents were entering relationships. That said, the percentage of adolescents reporting ever experiencing intimate-partner violence was very low, at approximately 2.6% of the sample at Round 3; no differences in this statistic existed between the intervention and control samples as well.

Table 21. Treatment-on-Treated (ToT) Summary Results: Difference-in-Difference (DID) Estimators - Excludes urban controls

Key	
✓	At p < .05 AGEP had a significant positive effect relative to controls
†	At p < .10 AGEP had a marginally significant positive effect relative to controls
x	At p < .05 AGEP had a significant negative effect relative to controls
†	At p < .10 AGEP had a marginally significant negative effect relative to controls
	Not measured at Round 1 or Round 2

		TOT - DID[e]		
		TOT R2-R1[f]	TOT R3-R2[g]	TOT R3-R1[h]
	Impact indicators			
	Educational			
1	Avg. number of years completed			
2	% completed primary school			
3	% completed junior secondary school[b]			
4	% currently attending school			
	Sexual risk behaviour[b]			
5	% ever had sex			
5a	Avg. age at first sex among those who ever had sex and report knowing age at first sex[c]			
6	% agree that they have had unwanted sex			
7	% agree that they have had transactional sex			✓
8	% used condom at last sex with non-marital partner			
9	% used condom at first sex			†
	Marital[b]			
10	% ever married			
11	Avg. number of HIV risk-related topics discussed with partner (0-5)[a]			
12	Avg. marital control score (0-6)[a]	x		†
	Pregnancy & births[b]			
13	% who have ever been pregnant			
14	% who have ever had an unwanted pregnancy			
15	% currently pregnant or who have given birth			
16	% ever used modern contraception			
	Sexually transmitted infections[b]			
17	% HIV positive			
18	% HSV-2 positive			

	Experience of violence[d]
19	% have experienced physical violence in past 12 months
20	% have experienced intimate partner violence in past 12 months

Notes

Indicator was denoted as significant at the p <.05 level if both the girl random effects and the robust SE clustered at CSA models were significant at p<.05. It was denoted as significant at the p<.10 level if both models were significant at p<.10 or if one was significant at p<.05 and the other at p<.10.

It should be noted that while blank cells do represent non-significant findings, it does not imply that there was no effect.

All models include site fixed effects and other covariates.

Covariates: Age, school attendance, grade attainment, literate (any language), numeracy score at baseline, cognitive score at baseline, mother alive, father alive, biological daughter of HH head, HH items, HH assets/savings, tribes (Bemba, Nyanja, Kaonde, other), mother completed primary, father completed primary, standardized vulnerability quintiles.

Models for grade attainment, completed grade 7, and completed grade 9 do not include school attendance, grade attainment, literacy, numeracy and cognitive skills as covariates.

Models for school attendance do not include school attendance as a covariate.

[a] TOT coefficients are estimated from linear regression models clustered at the CSA level (Model B) at R2 (in R1 & R2 results) and R3 (in R1 & R3 results) given that we do not have a baseline measure for this indicator to estimate the difference-in-difference coefficient. The difference-in-difference coefficient is estimated in the R2 & R3 results.

[b] Indicators in this domain measured among those aged 15 and older.

[c] Models will be estimated at endline using survival analysis.

[d] Indicators in this domain measured among those aged 13 and older.

[e] DID estimates using instrumental variable analysis: girl random effects and robust SE clustered at CSA (160 clusters)

[f] TOT = Attended 26+ meetings between R1 & R2

[g] TOT = Attended 26+ meetings between R2 & R3

[h] TOT = Attended 52+ meetings between R1 & R3

Assessing change over time in longer-term outcomes, considering intensity of participation in AGEP

As with the assessment of the array of indicators of changes in girls' assets or empowerment over time, estimates of the impact of AGEP on the longer-term adolescent outcomes based on measures of the intensity of programme participation were obtained and shown in Tables 21 (above) and C-3. The summary of results in Table 21 (above) and C-3 can be interpreted in the same manner as those discussed above for the discussion of impact of intense participation on empowerment. Table C-4 includes estimates from the naïve model of impact that do not account for potential selectivity of the girls (model 1), as well as estimates from models of impact that account for selectivity using an instrumental variables approach (models 2 and 3). The coefficient of impact was shown separately for Rounds 1 to 2, Round 2 to 3, and overall for Rounds 1 to 3. As noted previously, the changes in the magnitude and significance of the coefficients between the naïve and instrumental variable models would be indicators of the degree to which the self-selectivity of girls (whether at the individual, family, or community level) is a factor in how outcomes can be assessed and programme impact determined.

Similar to the results observed in the ITT analysis, the percent of girls reporting having experienced transactional sex and percent reporting condom use at first sex is significantly greater over time among those intensely participating in AGEP relative to other girls. The result of condom use at first sex is largely driven by the older rural girls, who have increased their condom use at greater rates than urban girls. Assessing the differences between the ITT and results of intense participation, Table C-3 reveals that there may be an emerging impact of intense AGEP participation on years attending school among younger rural adolescent girls, although it is only marginally statistically significant when assessing Rounds 1 to Round 3 and, therefore, too early to tell whether the impact will be sustained or built upon in the next two years.

Assessing the longer-term outcomes overall, comparing the results of the naïve model to the instrumental variable model shows a large degree of selection involved in girls' intense programme participation. For instance, girls who were more likely to intensely participate in AGEP were younger, less likely to be married, and less likely to be pregnant and have had a first birth, none of which are necessarily surprising given the requirements of participating in the weekly girls group meetings. The naïve models, not considering the selectivity of intense participators, would seem to suggest that AGEP has had a significant impact on early marriage and fertility to date. The instrumental variables estimates, however, indicate that AGEP has not had such an impact in the period immediately following the end of the programme. It should be noted, however, that this does not imply that AGEP will not have an impact on these indicators in the longer-term, as a greater number of younger adolescents transition to marriage, pregnancy, and first birth.

Furthermore, the results in Table C-3 reveal that, while there were no significant coefficients of impact from intense programme participation on sexual behaviour, experience of violence, or STI infection, there were observable impacts on girls' education. While two of the indicators of progression (completing grades 7 and 9) might suggest that selectivity is at play in the statistically significant coefficient of impact (and that, therefore, one should not yet infer a role played by AGEP in increasing transitions across schooling levels) there is some preliminary indication that AGEP may have added to girls' average grade attainment, at least early in the programme. Girls who intensely participated in AGEP, after one year of programme

participation, had a higher mean grade of school completed that was statistically significantly higher than did control girls, girls who participated less, or girls who did not participate at all. While this indicator of impact did not retain its impact in the second year of the programme or in the estimate of impact from Round 1 to 3, it is nonetheless a promising finding that may suggest potential after the girls are observed over the full four years.

Table 22. Intent-to-treat (ITT) by arm summary results: Difference-in-differences (DID) estimators

Key	
✓	At p < .05 AGEP had a significant positive effect relative to controls
†	At p < .10 AGEP had a marginally significant positive effect relative to controls
X	At p < .05 AGEP had a significant negative effect relative to controls
†	At p < .10 AGEP had a marginally significant negative effect relative to controls
	Not measured at Round 1 or Round 2

		ITT - DID (xtreg models)[e]			ITT[f]		
			R3-R1			ITT	
		SS	SS+HV	SS+HV+SA	SS	SS+HV	SS+HV+SA
Empowerment indicators							
	Social assets						
1	Avg. score on self-efficacy scale (0-10)						
2	% confident regarding their ability to plan						
3	% feel they make good decisions regarding money						
4	% agree that permission to go to the health clinic is not a problem						
5	% who jointly or solely make decisions with regard to money earned among those who earned money						
6	Avg. number of friends						
7	Avg. number of friends in school		†				
8	Avg. number of friends who can be counted on if needed money						
9	Avg. number of friends who can be counted on in an emergency						
10	% have place in community where feel safe to meet girl friends		✓				
11	% with adult female support in case of serious problem						
12	% attending any social groups/clubs within the past month						
13	% who often/sometimes go to market[a]						
14	% who often/sometimes go to community centre[a]				✓	✓	✓
15	% who often/sometimes go to shops or restaurants[a]						
16	Avg. score on gender equality scale						
17	Avg. score on nonacceptability of intimate partner violence	X					
	Economic assets						
18	Avg. score on financial literacy scale	✓	✓	✓			
19	% who have saved in the past year			✓			
20	Avg. amount currently saved						
20a	% who currently have at least 20 kwacha saved			✓			

20b	Avg. amount currently saved among those who saved					
21	% working for cash or in-kind in the past year					
22	Avg. reported work income in the past year	†				
22a	Avg. reported income in the past year among those who worked for cash					
23	% who own a bicycle					
24	% who own a mobile phone		X			
	Health assets					
25	% understanding pregnancy risk during menstrual cycle					
26	Avg. score on contraceptive knowledge scale (0-9)	✓	✓	†		
26a	Avg. score on SRH knowledge scale (0-11)	✓		†		
27	Avg. score on HIV/AIDS knowledge scale (0-11)					
28	Avg. rating of health status in the past year (0-10)					
29	Avg. rating of health status in the past month (0-10)					
30	Avg. reported number of health problems in the past month (0-7)					
Impact indicators						
	Educational					
1	Avg. number of years completed					
2	% completed primary school					
3	% completed junior secondary school[b]					
4	% currently attending school					
	Sexual risk behaviour[b]					
5	% ever had sex	†				
5a	Avg. age at first sex among those who ever had sex and report knowing age at first sex[c]					
6	% agree that they have had unwanted sex					
7	% agree that they have had transactional sex	✓	✓			
8	% used condom at last sex with non-marital partner					
9	% used condom at first sex	†		✓		
	Marital[b]					
10	% ever married	X	†			
11	Avg. number of HIV risk-related topics discussed with partner (0-5)[a]					
12	Avg. marital control score (0-6)[a]				†	X

	Pregnancy & births[b]	
13	% who have ever been pregnant	
14	% who have ever had an unwanted pregnancy	
15	% currently pregnant or who have given birth	
16	% ever used modern contraception	
	Sexually transmitted infections[b]	
17	% HIV positive	†
18	% HSV-2 positive	
	Experience of violence[d]	
19	% have experienced physical violence in past 12 months	†
20	% have experienced intimate partner violence in past 12 months	

Notes

It should be noted that while blank cells do represent non-significant findings, they do not imply that there was no effect.

All models include site fixed effects and other covariates.
Covariates: Age, school attendance, grade attainment, literate (any language), numeracy score at baseline, cognitive score at baseline, mother alive, father alive, biological daughter of HH head, HH items, HH assets/savings, tribes (Bemba, Nyanja, Kaonde, other), mother completed primary, father completed primary, standardized vulnerability quintiles.
Models for grade attainment, completed grade 7, and completed grade 9 do not include school attendance, grade attainment, literacy, numeracy, and cognitive skills.
Models for school attendance do not include school attendance.
[a] ITT coefficients are estimated from linear regression models clustered at the CSA level given that we do not have a baseline measure for this indicator to estimate the DID coefficient (see ITT [if not measured at R1] results).
[b] Indicators in this domain measured among those aged 15 and older.
[c] Models will be estimated at endline using survival analysis.
[d] Indicators in this domain measured among those aged 13 and older.
[e] DID estimates, girl random effects, robust SE clustered at CSA (160 clusters)
[f] If not measured at R1: Robust SE clustered at CSA (160 clusters)

Assessing the impact of the programme by study arm

Analysis of the impact of AGEP on girls' empowerment was also conducted for each intervention arm separately against the control arm. These results are presented in summary form in Table 22 (above) and in detail in Table C-5. The study was powered to assess differences between the intervention and control arms, as well as between each intervention arm separately. This latter fact offers an opportunity for the cost-effectiveness evaluation to provide an assessment regarding the cost-effectiveness of the AGEP components if scaling or expanding the programme to other settings is an objective. The study results by arm reflect the overall impact of the programme, in that most of the statistically significant findings are consistently significant across intervention arms relative to the control arm, suggesting that it is the impact is attributable to the Safe Space component. There are two indicators (% having saved in the past year, % having saved more than 20 kwacha) in which statistical significance is observed only in the arm that includes the additional bank account component. Hence, one may conclude that the addition of having a bank account, whether it is used very much, increases the propensity to save; this result is back up by the observation from the data that girls with a bank account are accumulating their savings informally as much as they are formally through the use of the account.

Analysis of the impact of AGEP on girls' empowerment and longer-term outcomes was also conducted for each intervention arm separately against the control arm. The study was powered to assess differences between the intervention and control arms, as well as between each intervention arm separately, although for brevity the results provided here are limited to comparisons of the intervention study arms against the control arm. For the most part, the study results by arm parallel the overall impact of the programme, in that most of the statistically significant findings are consistently significant across intervention arms. An assessment of the coefficients of these assessments in Table C-5 does not indicate meaningful differences across the intervention arms themselves. These results suggest that the impact is attributable to the Safe Space.

For instance, a key finding discussed elsewhere is that AGEP increased the likelihood that girls reported in Round 3 that they often or sometimes go to a community centre to meet with friends. The impact of each arm is also similar with regard the acquisition of knowledge from the Safe Space curricula, including the financial literacy, contraceptive, and sexual and reproductive health knowledge assessments. This is also largely, but not completely, the case with the longer-term outcomes that were found to be significant, including the percentage of girls reporting transactional sex and condom use at first sex. Two indicators, the percent having saved in the past year and the percent having saved more than 20 kwacha, are statistically significant only in the arm that includes the additional bank account component. These findings are suggestive that the addition of having a bank account increases the propensity to save over and above what is learned within the Safe Space groups. This finding is back up by the observation from the data that girls with a bank account are accumulating their savings informally as much as they are formally through the use of the account.

Assessing contamination from spillovers of the programme to controls

Spillovers may occur where girls in the internal control clusters benefitted from the programme either by direct exposure or through interactions with AGEP girls. Although the AGEP intervention was largely excludable, as girls were directly invited to the programme and were provided an invitation, it is possible that other girls were exposed either directly (by participating in the groups) or indirectly (by being friends with participants). For instance, approximately 12% of control girls report participating in the AGEP Safe Space groups. If there is significant programme exposure to the control girls, the impact of the programme may be underestimated in the intent-to-treat analysis. To explore whether there is evidence of spillovers undermining the programme effects, we replicated the ITT analysis for the urban sites excluding the rural sites, as external controls were not considered for these sites. In this exercise, comparisons were made from two sets of analysis:

1. Outcomes of girls in AGEP clusters compared to those of girls in internal control clusters.
2. Outcomes of girls in AGEP clusters compared to those of girls in external control clusters.

Results from this exercise (not presented in the mid-term report) indicate that among all the empowerment and longer-term outcomes only for the reproductive knowledge indicators is there a significant positive effect of AGEP when compared to the external controls at the same time where there is no significant effect when comparing to the internal controls. For other indicators, results are either similar or weaker against the external controls than against the internal controls. Thus, it is concluded that programme effects are not underestimated due to contamination.

Economic evaluation

Programme costs

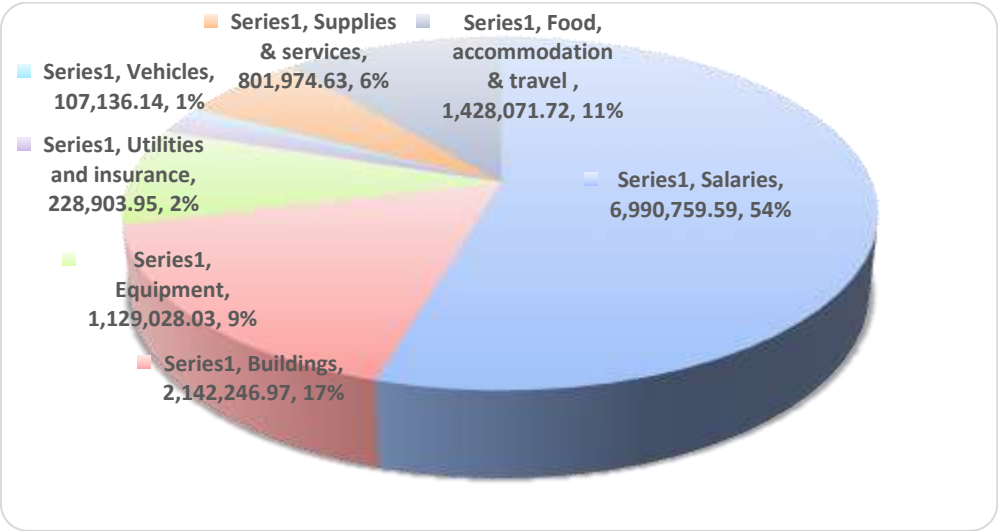
In total, 16,647 girls were invited to take part in AGEP (5,519 in Arm 1, 5,651 in Arm 2, and 5,477 in Arm 3, respectively). The inception phase of AGEP included a pilot of the programme which took place in one rural (Chibombo, n=829) and one urban district (Matero, n=762) between June 2012 and July 2013.

The total costs incurred in the start-up phase were $US 2,443,336, while the pilot delivery cost approximately $US 296,538.

Total costs of programme implementation

The total costs of delivering the programme, including the start-up costs but excluding the costs of pilot delivery, adjusted to 2016 $US values, were $US 12,828,121. Joint costs, that had to be split equally between the Safe Spaces, Health Voucher and Savings Account components, constituted approximately 11% of total costs at $US 1,429,000. The relative contributions of different line items to total start-up costs are shown in **Figure 7**, which highlights the importance of human resources as a cost driver.

Figure 7. Relative contribution of different line items to total programme costs



The total costs of the three experimental arms are presented in Table 23.

Table 23. Incremental costs of AGEP delivery, $US 2016

	Safe Spaces	Health Voucher	Savings Account	TOTAL
Salaries	3,110,268	2,006,135	1,874,357	6,990,760
Buildings	853,622	647,417	641,208	2,142,247
Equipment	963,734	82,906	82,387	1,129,028
Utilities and insurance	151,034	44,946	32,924	228,904
Vehicles	71,020	17,589	18,527	107,136
Supplies & services	460,824	186,570	154,581	801,975
Food, accommodation, & travel	940,758	271,539	215,774	1,428,072
TOTAL	6,551,260	3,257,102	3,019,758	12,828,121

The main drivers of Safe Spaces' higher costs were mentor salaries, as Safe Spaces is the only component that requires dedicated AGEP staff, and the costs associated with training. When total costs of programme delivery and support are considered, the main cost driver of all three interventions is staff costs, followed by buildings and by food, accommodation, and travel. Joint costs, which were equally allocated to the three components, represent 12% of Safe Spaces costs, 68% of health voucher costs, 23% of savings account costs, and 25% of total costs presented in Table 23. Since the largest share of joint costs is represented by support costs (82%), which are the most difficult to allocate to specific interventions, we conclude that the least costly component to deliver from the implementers' perspective is the savings account intervention, with roughly 33% of direct delivery costs, compared to approximately 75% direct delivery costs for the health voucher intervention and 73% for Safe Spaces, respectively.

Costs per participant

On average, AGEP implementation in Zambia cost approximately $US 770 per girl. This total takes into account the costs incurred by the Population Council and all its implementation partners.

The total costs per participant of the Safe Spaces programme as well as the incremental costs per participant of the add-on components (i.e., the additional costs of delivering the add-ons when Safe Spaces is in place, respectively) are shown in Table 24. The denominator for calculating the average cost of the Safe Spaces intervention is the total number of invited girls

in all three programme arms, since they all received the core Safe Spaces intervention. Similarly, for the health voucher average cost the denominator is the number of girls invited to receive the health voucher and savings account add-ons, while for the savings account average cost the denominator is the number of participants receiving the savings account add-on only.

Table 24. Incremental costs per participant, 2016 $US

	Safe Spaces	Health Voucher	Savings Account	Total
Salaries	$187	$180	$342	$420
Buildings	$51	$59	$117	$129
Equipment	$58	$7	$15	$68
Utilities and insurance	$9	$4	$6	$14
Vehicles	$4	$2	$3	$6
Supplies & services	$28	$17	$28	$48
Food, accommodation, & travel	$57	24	40	$86
TOTAL	$394	$293	$551	$771

Note: Incremental costs of actual expenditures (unadjusted for inflation) were $293 for the safe spaces arm, $208 for the health for the health voucher arm, and $409 for the savings account arm.

Scale-up costs

Several factors have contributed to making the implementation of AGEP more costly than it would have been under routine circumstances. First, the presence of a cluster randomised trial running in parallel to programme implementation made the presence of a project director in addition to the programme managers necessary for coordination purposes. Moreover, in this case the project director happened to be based outside of Zambia, which further increased the costs of programme support. Last, a nutrition add-on component that was not included in the evaluation study was introduced later during the implementation phase.

These factors all contributed to increasing programme costs but are unlikely to be incurred again should AGEP be scaled up in Zambia or established in a different setting. If these costs were not taken into account, total programme delivery costs would decrease by approximately 4% to $US 12.3 million ($US 740 per girl).

Participant and mentor costs

Fewer than 1% of AGEP participants reported spending any money for or incurring any indirect costs from participating in programme activities. For this reason, cost data from participant surveys will not be presented in this report.

The mentor survey collected information on the costs of programme participation from 218 mentors. Approximately one-third of mentors (31%) reported spending their own money in order to participate in AGEP activities, in excess of what they received from the programme to cover transportation costs. The average out-of-pocket expenditure was $US 7 (CI: $US 4 - 9; range: $US 0 – 210). Eighty-seven percent of mentors declared that they would be performing an alternative money-generating activity if they were not volunteering as AGEP mentors. The average opportunity cost of mentors' time was $US 48 (CI: $US 39 – 58; range: $US 0 – 418).

Health provider costs

Mulungushi University health centre is a government-owned primary healthcare facility in rural Kapiri Mposhi district. The number of AGEP health voucher visits recorded during 2015 is summarized in Table 25.

Table 25. Health voucher services at Mulungushi University health centre, 2015

Health service	Visits in 2015
Wellness Check Plus	1059
Family planning (FP) consultation	602
Treatment of discharge (includes consultation)	123
RPR test (syphilis test)	146
HIV counselling, testing, and referral	61
ANC1 (includes laboratory tests)	176
TOTAL	2167

The incremental costs of each Health Voucher service are summarized in Table 26. HIV counselling and testing was the most expensive service to provide among those accessed through the AGEP voucher, followed by syphilis testing. In both cases, the high unit costs of diagnostics drive the costs of the procedure. The cost of "Wellness Check Plus" consultations, which include the management of illness that is not related to sexual and reproductive health, was calculated by allocating a proportion of the annual costs of medical supplies for diarrhea, malaria, and respiratory illness to these consultations.

Table 26. Total costs per visit of health voucher services at Mulungushi health centre, 2016 $US

	STI consultation	Wellness check plus	FP	Syphilis test	HCT
Staff	$3.57	$3.57	$3.57	$3.57	$3.57
Overheads	$0.05	$0.05	$0.05	$0.05	$0.05
Medical supplies	$1.50	$3.18	$9.78	$22.12	$329.71
TOTAL	**$5.12**	**$6.80**	**$13.40**	**$25.74**	**$333.33**

Table 27, below, shows the reimbursement rates agreed between the Ministry of Health and the Population Council for the services covered by the health voucher provided at government facilities. The reimbursement rates cover only a fraction of the total cost of the service: 30% for Wellness Check Plus, 35% for the average family planning consultation including pregnancy testing, 53% for STI treatment and consultation, 6% for syphilis testing, and 0.4% for HIV testing and counselling. However, the negotiated reimbursement rates were to act as an incentive to health providers rather than to cover the full costs of the services, since these are provided for free at public facilities and the aim of the health voucher was to stimulate demand.

Table 27. Health voucher reimbursement rates for GRZ clinics, 2016 $US

AGEP code and description of health service		Facilities Price
12	Wellness Check Plus	$2.00
23	FP consultation plus provision of pills (including condoms)	$2.00
31	Pregnancy test plus referral	$2.71
42	Treatment of discharge (includes consultation)	$2.71
44	RPR test	$1.40
51	HIV counselling, testing and referral	$1.40

Cost-effectiveness modelling

The study sample at the programme implementation sites included in the economic evaluation, from which effect estimates between Round 1 and Round 3 are calculated, are noted in Table 28 below.

Table 28: Total and by study arm sample of invited participants[53] of programme sites

	AGEP total sample	AGEP sample arm 1 (SS)	AGEP sample arm 2 (SS+HV)	AGEP sample arm 3 (SS+HV+SA)	Total controls (internal + external)
Urban sites					
Misisi/Chawama	355	122	113	120	119
Chipata/Chazanga	362	126	124	112	113
Kabwe	362	127	115	120	130
Ndola	349	120	124	105	112
Kitwe	379	135	118	126	107
Kanyama	-	-	-	-	117
Mandevu	-	-	-	-	98
Bwacha	-	-	-	-	132
Chifubu	-	-	-	-	106
Wusakile	-	-	-	-	121
Urban total	1,807	630	594	583	581
Rural sites					
Mumbwa	353	117	116	120	127
Kapiri Mposhi	322	103	104	115	97
Masaiti B	361	110	122	129	110
Masaiti A	349	118	119	112	113
Solwezi	323	108	114	101	118
Rural total	1,708	556	575	577	565
Total	3,515	1,186	1,169	1,160	1,146

[53] The denominator used in the cost-per-girl calculations was the total number of girls that the programme aimed to reach, not those who participated.

The ACERs comparing progress from baseline to Round 3 between the intervention arms receiving Safe Spaces and the controls are presented in Table 29. The numerator for the ratios is total programme costs (start-up and implementation, excluding pilot delivery), equal to $US 12,828,121. The denominator is the average difference in positive impact (or negative impact averted) from baseline to Round 3 between intervention arms and controls (or controls and intervention arms for negative impacts). The ACERs in Table 29 are the ratio of the difference in cost to the difference in effect comparing Safe Spaces to no programme (i.e., intervention vs control). The formula and calculations are shown in the above methodology section. The benefits have not been monetized – the cost-effectiveness analysis uses natural units as the outcome measure (i.e., additional girl reporting a unit of positive effect/one less unit of negative effect).

Table 29. Incremental impact of AGEP and average cost-effectiveness, 2016 $US

Indicators	Difference in outcome between baseline and R3		Incremental effectiveness	ACERs (US$)
	AGEP† (n)	Control (n)	AGEP (n) – Control (n)	
Impact indicators				
Grade 7 completion	616	274	342	$37,509
Grade 9 completion	459	227	232	$55,294
First marriage	408	144	-264	-
First birth	410	188	-222	-
First pregnancy	510	223	-287	-
First sex	821	332	-489	-
First use of modern contraception	424	196	228	$56,264
HIV seroconversion	54	30	-24	-
HSV-2 seroconversion	211	86	-125	-
Outcome/output indicators				
Unwanted sex in last 12 months	162	41	-121	-
Condom use at last sex	46	31	15	$855,208
Savings in the past 12 months	448	149	299	$42,903
Money management decision making	108	53	55	$233,239
Additional indicators				
Safe place to meet with girl friends	1	52	51	$251,532
Transactional sex	107	44	63	$203,621
Condom use at first sex	241	88	153	$83,844
Savings of at least 20 Kwacha (US$2)	416	169	247	$51,936

† Includes participants from all arms of the study

The indicators that were most efficiently improved by Safe Spaces were school attendance and financial security in the form of savings, followed by use of modern contraception. Sustained use of condoms during the last intercourse with a non-marital, non-cohabiting partner and good money management decision making were more expensive per unit of outcome, while no incremental effectiveness of Safe Spaces compared to no programme was achieved on the remaining indicators.

Limitations of economic evaluation

It was acknowledged in the economic analysis that the circumstances under which AGEP was implemented in Zambia are unlikely to present again should the programme be scaled up or implemented in a different setting. However, these particular circumstances might have had a positive effect not only on the costs, but also on the impact of AGEP. An example are the prizes given to girls in order to boost attendance and ensure that the sample size needed for the research was reached. Another example is the electronic attendance monitoring of the Safe Spaces meetings. While the costs of these extra activities can be easily calculated from programme budgets, their effect on the impact of the intervention is unknown and this increases the uncertainty around the potential cost-effectiveness of a scaled-up or expanded programme.

Conclusions

The Adolescent Girls Empowerment Programme (AGEP) was a social, health, and economic asset-building programme targeting vulnerable adolescent girls aged 10–19 that was successfully implemented in four of the 10 provinces of Zambia from late 2013 to early 2016. Over 240 mentors guided 11,390 adolescent girls aged 10–19 into the programme and through a planned two years of programming. In addition to the programme, a rigorous impact evaluation among 5,235 adolescent girls at baseline was embedded whose objective was to assess the impact of AGEP. A randomised cluster design with four study arms was implemented and the results of the study after two years of AGEP implementation are presented in this mid-term report. Observations on the cohort of adolescent girls will continue through 2017 to assess the impact of AGEP on longer-term adolescent outcomes.

The AGEP programme was shown to have an impact in each of the domains of empowerment measured – social, health, and economic. That said, the overall impact of AGEP on empowerment was modest compared to expectations at the mid-way point. AGEP did improve upon a number of social assets, including self-efficacy and confidence in decision making in the first year of the programme, and it lived up to its commitment to address social isolation by providing a safe space in the community for adolescent girls, as well as increasing their utilization of community spaces. AGEP's strongest results for the array of empowerment indicators were observed among the younger, rural adolescent girls and occurred primarily after the first year of the programme. The stronger impact on rural adolescents was likely a function of the fact that such girls had higher participation rates than other AGEP girls and, hence, more exposure to the programme.

Another area in which AGEP was successful was raising the financial literacy of adolescent girls and increasing the likelihood that the girls reported having saved in the previous year. This result was particularly strong for AGEP girls in study arm 3 who were provided access to bank accounts. For older urban girls, AGEP also improved the percentage of girls who reported having worked for cash or in-kind payments in the past year, potentially suggesting that urban AGEP girls were better situated to translate their skills into economic opportunity; for rural girls, these opportunities may have been fewer and farther between. In the realm of health assets, AGEP improved upon girls' contraceptive knowledge when compared to controls, although the overall impact was modest. Interesting to note is that in no group did AGEP lead to a positive change in gender norms. This speaks to the intractable nature of culturally accepted gender roles and suggests that future programmes may have to work not only with adolescent girls, but also address social and cultural norms at household and community levels so as to affect the enabling environment for girls.

The results from AGEP's impact assessment on adolescent girls' longer-term outcomes show that AGEP girls benefit from lower risky sexual activities, such as unprotected sex at sexual initiation and transactional sex. The results indicated that AGEP girls were statistically significantly less likely to engage in risky behaviours relative to controls. For condom use at first sex, the result was stronger for older rural adolescents relative to urban adolescents. In addition, while marginally statistically significant at this stage, the results indicated that younger rural adolescent girls who intensely participated in AGEP had a larger increase in the average number of grades of schooling completed, even after controlling for the selectivity of girls participating in the programme.

In assessing both the empowerment and longer-term outcomes, it is clear that girls who participated more intensely in AGEP experienced greater changes in empowerment over time relative to controls, those who never participated, and those who participated less frequently. The differences observed between the naïve model estimates of programme impact and the instrumental variable models of programme impact strongly suggest that unobservable factors existed that influenced both programme participation and adolescent outcomes. These unobservable factors may have included girls' personality, motivation, and ambitions, as well as parental involvement and household dynamics. While better outcomes were observed among girls who participated more intensively, once selectivity of participation was taken into account through an instrumental variable approach many of the programme's impact effects were not significant. This finding has important implications for programme design as it implies that even in very targeted programmes, it is difficult to reach those most in need.

Considerations on interpretation

One key consideration in interpreting the results presented in this mid-term report is the age of the sample, 12–21 years. The relatively young age of the research participants after the two years of the programme leaves many of the key adolescent transitions yet to be observed, e.g., completing grade 9 education, sexual debut, pregnancy, first birth, etc. The evaluation was powered based on four full years of observation at which time the sample will be 14–23 years old. As girls continue to age over the remaining two years of the evaluation, it is expected that a larger number of transitions will be observed to allow a more rigorous assessment of longer-term impact. For this reason, the mid-term report focuses on the mediating factors that are expected to translate into more positive longer-term impacts on adolescents' trajectory to adulthood, specifically empowerment and assets.

Another consideration when interpreting these results is that the vulnerable urban adolescent girls in the evaluation sample represented approximately 35% of the urban girls and, hence, could be considered the most vulnerable in their setting. Alternatively, the evaluation sample of vulnerable rural adolescent girls represents approximately 85% of the rural adolescent girl population. As such, the most vulnerable adolescent girls in the urban sample were worse-off in many respects than their rural counterparts; for instance, the sample of urban girls were less likely to be in school and performed worse on the literacy and numeracy assessments at baseline. As baseline competencies were a mediating factor in programme impact, affecting outcomes for the urban girls may pose a greater challenge.

Programme design implications

The mid-term results point to a number of programmatic lessons that can be used to inform conversations on programme design as the study continues to inform the longer-term impact of the interventions in the coming years. One key lesson is that safe spaces, while an effective platform to bring together some subsets of the vulnerable adolescent girls, may need to be coupled with additional interventions that take into consideration an ecological approach and address the other layers of influence in the lives of adolescent girls – including boys, young men, parents, teachers, and other key community leaders. There may be useful interventions, when working in tandem with activities to build girls' assets, which improve the enabling environment for girls.

Furthermore, AGEP was implemented in settings with entrenched, conservative social and cultural norms about the value of girls and adolescent sexuality—even held by some of the

mentors themselves. It seems likely that these norms will need to be addressed at individual, household, community—and perhaps even national—levels in order to create the kind of environment in which girls can truly make a safe, healthy, and productive transition into adulthood. Both engagement with the household and broader community, as well as work to address social norms, are likely missing pieces to the original AGEP theory of change.

An additional lesson is that the AGEP theory of change includes a very broad, and perhaps too broad, set of indicators on which we hypothesized the interventions would lead to positive impact. While there was a wide breadth of topics covered, perhaps that was to the detriment of depth on any one particular topic. Although the root causes of many of the long-term outcomes—education, pregnancy, marriage, HIV, and others—are intertwined, a focus on a narrower set of outcomes might have focused the components and content of the intervention to be more directly linked to the ultimate impact desired.

Following from that, additional targeted interventions at specific subsets of the adolescent girls population in Zambia may also yield more impact on the longer-term outcomes. This may also help to increase participation in light of the programme uptake analysis. For example, many of the older adolescent girls showed less participation, as well as less benefit. At that age, additional activities focused on linkages to livelihood strengthening may have kept girls participating in a wider range of activities, but also addressed an age-specific need. Furthermore, the livelihood activities would need to be adapted depending on geography, markets, and opportunities in any given site. Similarly, on the other end of the age spectrum, a more direct focus on an education support activity specifically focused on keeping girls in school and attending regularly may yield a more direct benefit on schooling completion.

Overall, the AGEP cohort data and the lessons it has generated from the Adolescent Girls Empowerment Programme are rich, nuanced, and important for informing the next generation of programmes for adolescents in Zambia and elsewhere. Even though the study is still underway, and the full longer-term effects of AGEP remain to be seen, the information presented in this report can be used to guide programme- and policy on programme areas of promise, gaps that need to be filled, and a range of questions that still need to be answered about how to best serve this population.

Annex A: List of Safe Spaces Training Session Topics

HEALTH AND LIFE SKILLS CURRICULUM

Introductory Sessions
Session 1: What to Expect – Part 1
Session 2: What to Expect – Part 2
Session 3: Teamwork
Session 4: His and Hers (Gender Roles)
Session 5: Communication
Session 6: Self-Esteem
Session 7: Goal Identification – The Road of Life
Session 8: Goal Setting and Achieving
Session 9: My Relationships

Reproductive Health
Session 1: Life Cycle
Session 2: My Body Is Changing – Am I Normal?
Session 3: How Does Pregnancy Happen, Anyway?
Session 4: Preventing Unintended Pregnancy (Contraceptives)
Session 5: Reproductive Myths
Session 6: Sexual Desire
Session 7: Unsafe Abortion
Session 8: Abortion and Stigma
Session 9: Maternal Mortality

Life Skills
Session 1: I Have Healthy Relationships
Session 2: Reasons to Delay Sex
Session 3: Strategies for Delaying Sex
Session 4: Passive, Assertive, Aggressive
Session 5: Drugs, Alcohol, and Other Mind-Altering Substances
Session 6: Peer Pressure
Session 7: Making Good Decisions
Session 8: How to Communicate with Adults
Session 9: How to Communicate with a Partner
Session 10: Managing Stress, Anger, and Conflict
Session 11: Conflict Resolution and Problem-Solving Skills

HIV, AIDS, and STIs
Session 1: HIV and AIDS
Session 2: Myth or Fact?
Session 3: HIV Testing and Counseling
Session 4: Risky Behaviour
Session 5: The Relationship of STIs and HIV and AIDS
Session 6: Stigma and Discrimination in HIV-Positive People

Gender and Gender-Based Violence
Session 1: Sexual Exploitation
Session 2: How to Report and Avoid Cases of Sexual Violence
Session 3: Rape and Gender Violence
Session 4: Preventing Unwanted Advances

Leadership
Session 1: Leadership
Session 2: Community Service – Putting Leadership into Action

Human Rights
Session 1: Human Rights and Children's Rights
Session 2: Sexual and Reproductive Health Rights
Session 3: HIV and AIDS and Human Rights

FINANCIAL EDUCATION CURRICULUM

Dream Big
Session 1: Dream Big!

The Steps for Saving and Earning Money
Session 2: Why Save?
Session 3: Choose a Savings Goal
Session 4: Make a Savings Plan
Session 5: Learning About Banks and Bank Accounts
Session 6: Exploring Options for Earning Money
Session 7: Girls' Money and the Risky Income Cycle
Session 8: Know the Difference Between Needs and Wants!
Session 9: Control Spending
Session 10: Think About the Future: Money In and Money Out
Session 11: Save Regularly

Talk About Money
Session 12: Save in a Safe Place
Session 13: Dealing with Setbacks in Saving
Session 14: Your Own Money vs. Someone Else's Money
Session 15: Talking About Money
Session 16: The DOs And DON'Ts of Talking About Money
Session 17: Resolving Conflicts About Money
Session 18: Role Play Resolving Conflicts

Conclusion and Reflection
Session 19: Our Journey to Good Money Management

NUTRITION CURRICULUM

Session 1: Nutrition Needs for Adolescent Girls
Session 2: The Role of Food in the Body
Session 3: Anaemia in Adolescent Girls
Session 4: Nutrition for Pregnant Adolescents (15–19 Only)
Session 5: Infant Feeding from Birth through Six Months (15–19 Only)
Session 6: Young Child Feeding and Growth Monitoring (15–19 Only)

ANNEX B: MENTOR EVALUATION RESULTS

Table B-1. Variables Used in the Construction of Dimension-Specific Mentorship Quality Scores*

Dimension-Specific Score	Questions
Knowledge	OBSERVATION QUESTIONNAIRE Grades for mentor, ranging from 1 to 3 for each item. ○ Shows comprehension of the material ○ Makes use of practical examples relevant to girls' lives ○ Asks questions to identify girls' knowledge and experiences on the session topic ○ Simplifies the language for girls' understanding • Level of girls' comprehension of material MENTOR SURVEY • Read the paragraph and answer the following questions: "Solid Waste Management is a major problem world-over and in Zambia. It offers several challenges from blocked drainage and sewers; waterborne diseases like typhoid, cholera, and diarrhoea; increased upper respiratory diseases from open burning of the garbage. There is no single solution to the challenge of solid waste management. Solid waste management is therefore the collection, transportation, treatment, and processing or disposal of waste." Do you agree or disagree with the following statements? ○ Open burning of garbage can increase the risk of upper respiratory diseases. ○ There is only one solution to solid waste management. ○ Solid waste management is not a major problem outside of Zambia. • Mary needs K2,000 to start a business. She has K1,000 in her account that she can use for the business. How much more money does Mary need to save to be able to start her business? • Grace needs K1,600 to buy a mobile phone. She earns an income of K300 per month. Of that amount, she is able to save K200 per month. How many months would it take Grace to save enough money to buy the phone? • After Grace has been saving for three months, she has an emergency that forces her to use K400 from her savings. How much more money does she need to save to reach her goal of K1,600? • Now we would like to ask you some questions about HIV/AIDS and other reproductive health issues. Do you agree or disagree with the following statements? ○ Pregnant women with HIV can take medicine to prevent their babies from getting HIV. ○ A healthy-looking person can be infected with HIV. ○ There is currently no medication for HIV -positive people to extend their lives and improve their quality of life. ○ If one has other sexually transmitted infections, it increases one's chance of getting HIV. • Please list all the ways you know that people can protect themselves from getting infected with HIV. † • Please list all of various methods that a couple can use to delay or avoid a pregnancy. † • Below are some statements about contraceptives. I would like you to tell me if you agree or disagree with the statement. ○ A condom can help prevent pregnancy but not HIV. ○ Using two condoms at the same time is better than using just one. ○ Condoms are effective in preventing STI transmission. ○ Condoms should not be reused. • Please list the signs and symptoms that might indicate that a woman has an STI. †

	• Please list the signs and symptoms that might indicate that a man has an STI. †
	• Which of the following is NOT considered a basic human right for adolescent girls?
	  o Food
	  o Sanitary Towels
	  o Education
	  o Shelter
	• Name three topics included in life-skills training for young people.
	• According to the recommended guidelines, what is the minimum amount of time that a mother should exclusively breastfeed a child?
	• Anaemia is a condition that can affect adolescent girls if they are lacking the following nutrient in their body:
	  o Vitamin C
	  o Calcium
	  o Iron
	  o Vitamin D
	  o Potassium
	• Which of the major food groups do the following foods belong to?
	  o Beans
	  o Eggs
	  o Maize
	  o Spinach
	  o Milk
	• What are two parts of a savings plan? †
	• What are two sections of a budget? †
	• When someone takes a loan, what are two of her responsibilities? †
	• What is one informal way to save money?
	• What is one formal place to save money?
	• What is the legal age of marriage for women in Zambia?
	
	GIRLS SURVEY
	• Now I'm going to read you some statements about your AGEP/Safe Space mentor. Please tell me if you agree or disagree with each statement.
	  o *My mentor is knowledgeable about the things that she teaches us.*
	  o *My mentor always makes time to answer all the girls' questions.*
	  o *My mentor can usually answer questions about the topics she's teaching.*

Safe Space creation	OBSERVATION QUESTIONNAIRE Grades for mentor, ranging from 1 to 3 for each item. • Patience (towards girls) • Has fun with girls, enthusiastic • Establishing a physical space that is supportive and conducive to learning • Time management and use of session guide • Mentor reaches general meeting session goals by doing all of the activities • Mentor comes to session prepared • Creates an atmosphere of well-being and acceptance among the girls • Dialogues about the problems shared by girls in the group • Use of body language to enhance the girls' engagement • Provides clear instructions during the session • Active listening • *Conflict resolution (how mentor managed conflict between girls)* • *Stimulates teamwork among the girls* • Allows girls to participate and take leadership MENTOR SURVEY • *Each week, how many minutes in total do you spend preparing for your Safe Space meeting?*

	GIRLS SURVEY
	• Now I'm going to read you some statements about your AGEP/Safe Space mentor. Please tell me if you agree or disagree with each statement.
	  o *My mentor always comes to AGEP/Safe Space meetings on time.*
	  o *My mentor makes it easy for all girls in the group to participate.*
	  o *My mentor always calls on the same girls to participate.*
	  o When there are problems between girls in the Safe Space group, my mentor helps them work it out.
	  o My mentor treats some girls in the AGEP/Safe Space group better than she treats other girls.

Relationship with girls and community	OBSERVATION QUESTIONNAIRE
	Grades for mentor, ranging from 1 to 3 for each item.
	  o *Proves she knows the girls personally.*
	  o *Anger management (towards girls)*
	
	MENTOR SURVEY
	• As an AGEP mentor, in the last month, how many times did you do any of the following? For each activity, please indicate: every day, a few times a week, once a week, a few times a month, once a month, or never.
	  o How often visited a girl at her home due to missing meetings,
	  o How often counseled a girl regarding a personal problem.
	  o How often referred a girl to another organization.
	  o How often referred a girl to obtain contraceptives,
	  o How often mentor met with community leaders to discuss the program,
	
	GIRLS SURVEY
	• Now I'm going to read you some statements about your AGEP/Safe Space mentor. Please tell me if you agree or disagree with each statement.
	  o I do not feel comfortable discussing sensitive topics with my mentor.
	  o I think of my mentor as a role model.
	  o My mentor sometimes gets angry at girls in the group.
	  o I don't trust my mentor.
	  o *If I had a problem, my mentor is one of the first people I would tell.*
	  o *My mentor comes to find me after I miss a few Safe Space meetings.*
	  o If I told my mentor something private she would tell someone else.
	  o My mentor is not respected in the community.
	  o *When there are issues with parents, my mentor is usually able to help solve them.*

Self-efficacy	MENTOR SURVEY
	• Please tell me if you agree or disagree with each statement.
	  ○ I can always manage to solve difficult problems if I try hard enough.
	  ○ I have a budget that I usually stick to when spending my money.
	  ○ If someone is against me, it is difficult for me to find ways to get what I want.
	  ○ It is difficult for me to focus on my aims and accomplish my goals.
	  ○ I have a short-term and a long-term financial goal.
	  ○ I can be aggressive if I need to.
	  ○ I am confident that I could handle unexpected events very well.
	  ○ Because of the help I can get, I know how to manage unexpected situations.
	  ○ I can solve most problems if I make the necessary effort.
	  ○ When there is a conflict, I feel uncomfortable verbally disagreeing with people.
	  ○ I can remain calm when facing difficulties because I can rely on my own abilities.
	  ○ When I face a problem, I can usually find more than one solution.
	  ○ If I am in trouble, it is usually difficult for me to think of a solution.
	  ○ I can usually handle any situation that comes my way.
	  ○ During a conflict I can usually reach a mutually agreeable solution.

Attitudes: gender norms	MENTOR SURVEY
	• Please tell me if you agree or disagree with each statement.
	  ○ Girls are not as good as boys in school.
	  ○ Girls who do not actively participate in class should receive less attention from the teacher/mentor.
	  ○ When a family cannot afford to send all children to school, it is better to send boys than girls.
	  ○ Parents should decide whether or not their children go to school.
	  ○ Boys should do as much domestic work as girls.
	  ○ Girls can make as good leaders as boys.
	  ○ *Boys should not be asked to help their mothers prepare food.*
	  ○ If a rich man offers to marry a 16-year-old girl, she should leave school to get married.
	  ○ A 17-year-old girl should get married when she finds an appropriate partner, even if she is still in school.
	  ○ Girls from poor families deserve to have all of their basic needs met.

Attitudes: contraception	MENTOR SURVEY
	• Please tell me if you agree or disagree with each statement.
	  ○ Most girls who use contraceptives have many sexual partners.
	  ○ Condoms decrease intimacy or emotional closeness.
	  ○ *I would refuse to have sex with someone who is not prepared to use a condom.*
	  ○ Only married girls should have access to contraceptives.
	  ○ Moral people can use condoms.
	  ○ My religion forbids use of contraceptives.
	  ○ If a husband and wife disagree on family planning, the husband's opinion should come first.
	
	GIRLS SURVEY
	• *My mentor says that family planning is only for married girls.*

Attitudes: gender-based violence	**MENTOR SURVEY**
	• Please tell me if you agree or disagree with each statement.
	  ○ Some girls who are raped deserve it because of the way they dress or talk to boys.
	  ○ Men who force girls to have sex should be sent to jail.
	  ○ Men rape girls because they can't control themselves.
	  ○ If a man doesn't hit his wife, it means he doesn't love her.
	  ○ It is a man's right to have sex with his wife whenever he wants.
	• It is OK for a man to hit or beat his wife if…?
	  ○ She burns food.
	  ○ She argues with him.
	  ○ She goes to the neighbours without telling him.
	  ○ She refuses to have sex with him.
	  ○ She neglects the children.

* Items in italics were excluded from score computation.
† For these items, we used only responses collected in wave one because of lack of comparability across waves.

Table B-2. Correlations between overall and dimension-specific scores of mentorship quality

Overall	0.741	0.418	0.186	0.509	0.745	0.512	0.667	0.544
0.741	Knowledge	0.319	0.070	0.347	0.543	0.494	0.486	0.421
0.418	0.319	Safe Space Creation	0.026	0.097	0.202	0.211	0.187	0.115
0.186	0.070	0.026	Relationship with Girls and Community	-0.006	-0.007	-0.013	0.070	-0.057
0.509	0.347	0.097	-0.006	Self-Efficacy	0.375	0.379	0.352	0.335
0.745	0.543	0.202	-0.007	0.375	Gender Attitudes	0.644	0.618	0.773
0.512	0.494	0.211	-0.013	0.379	0.644	Attitudes: Gender Norms	0.371	0.405
0.667	0.486	0.187	0.070	0.352	0.618	0.371	Attitudes: Contraception	0.413
0.544	0.421	0.115	-0.057	0.335	0.773	0.405	0.413	Attitudes: GBV

Table B-3. Distribution of dimension-specific mentorship quality scores by mentor's characteristics:*

Scores for Knowledge, Safe Space Creation, Relation to Girls and Community, and Self-Efficacy

	Knowledge				Safe Space Creation				Relationship with Girls and Community				Self-Efficacy			
	Poor	Reg.	Good	Total	Poor	Reg.	Good	Total	Poor	Reg.	Good	Total	Poor	Reg.	Good	Total
Age																
≤25	43.3	32.7	24.0	100	41.7	38.9	19.4	100	35.3	30.2	34.5	100	37.2	25.7	37.2	100
26–30	25.8	36.0	38.2	100	28.0	31.2	40.9	100	37.6	36.6	25.7	100	30.8	39.4	29.8	100
31+	29.5	30.8	39.7	100	29.1	27.8	43.0	100	25.9	34.1	40.0	100	31.5	35.9	32.6	100
Secondary school																
No	57.3	32.0	10.7	100	47.4	30.3	22.4	100	37.0	30.9	32.1	100	46.5	33.7	19.8	100
Yes	24.5	33.7	41.8	100	28.4	34.3	37.3	100	32.1	34.4	33.5	100	28.3	33.2	38.6	100
SES level																
Low	41.0	39.0	20.0	100	38.1	40.2	21.6	100	34.9	31.1	34.0	100	40.7	26.9	32.4	100
Medium	42.2	26.5	31.3	100	30.8	36.3	33.0	100	33.0	29.9	37.1	100	36.0	33.0	31.0	100
High	17.0	33.0	50.0	100	31.5	22.8	45.7	100	32.3	39.4	28.3	100	22.8	40.6	36.6	100
Age at 1st marriage																
<18	38.5	26.9	34.6	100	40.0	24.0	36.0	100	17.4	47.8	34.8	100	42.3	34.6	23.1	100
18 or +	33.5	34.8	31.6	100	34.1	31.1	34.8	100	37.4	31.3	31.3	100	36.3	35.7	28.0	100
Never	32.2	32.2	35.6	100	30.8	39.6	29.7	100	30.0	34.0	36.0	100	25.7	28.7	45.5	100
Age at 1st pregnancy																
<18	23.4	42.6	34.0	100	27.9	30.2	41.9	100	35.6	40.0	24.4	100	32.0	48.0	20.0	100
18 or +	33.7	31.9	34.4	100	34.1	31.2	34.7	100	32.0	32.5	35.6	100	36.2	29.6	34.2	100
Never	41.0	29.5	29.5	100	36.7	40.0	23.3	100	35.5	32.3	32.3	100	25.8	33.9	40.3	100
Total	33.6	33.2	33.2	100	33.6	33.2	33.2	100	33.4	33.4	33.1	100	33.3	33.3	33.3	100
N	91	90	90	271	94	93	93	280	101	101	100	302	103	103	103	309

* Sample sizes vary because of missing values among the variable set used for each mentorship quality score.

Table B-4. Distribution of dimension-specific mentorship quality scores by mentor's characteristics: Global Score for Gender Attitudes, and sub-Scores for Attitudes on Gender Norms, Contraception, and Gender-Based Violence

	Gender Attitudes				Attitudes: Gender Norms				Attitudes: Contraception				Attitudes: GBV			
	Poor	Reg.	Good	Total	Poor	Reg.	Good	Total	Poor	Reg.	Good	Total	Poor	Reg.	Good	Total
Age																
≤25	34.2	38.7	27.0	100	32.2	39.1	28.7	100	37.5	40.2	22.3	100	33.6	39.7	26.7	100
26–30	36.0	30.0	34.0	100	40.2	27.5	32.4	100	33.7	26.7	39.6	100	41.0	19.0	40.0	100
31+	29.1	30.2	40.7	100	37.0	35.9	27.2	100	27.6	33.3	39.1	100	35.9	30.4	33.7	100
Secondary school																
No	48.8	30.5	20.7	100	62.8	26.7	10.5	100	42.2	34.9	22.9	100	53.9	19.1	27.0	100
Yes	27.4	34.4	38.1	100	26.0	37.2	36.8	100	30.0	33.2	36.9	100	29.9	34.4	35.7	100
SES level																
Low	36.0	32.0	32.0	100	36.1	42.6	21.3	100	37.0	39.0	24.0	100	37.0	29.6	33.3	100
Medium	38.9	43.2	17.9	100	49.5	27.3	23.2	100	43.3	28.9	27.8	100	49.0	30.4	20.6	100
High	25.5	25.5	49.0	100	23.5	32.4	44.1	100	20.4	33.0	46.6	100	24.3	30.1	45.6	100
Age at 1st marriage																
<18	24.0	36.0	40.0	100	52.0	28.0	20.0	100	26.9	38.5	34.6	100	32.1	25.0	42.9	100
18 or +	36.6	33.7	29.7	100	42.9	30.2	26.9	100	31.2	35.2	33.5	100	40.8	33.2	26.1	100
Never	29.9	32.0	38.1	100	20.6	43.1	36.3	100	38.8	29.6	31.6	100	30.7	25.7	43.6	100
Age at 1st pregnancy																
<18	24.5	40.8	34.7	100	46.9	26.5	26.5	100	16.0	52.0	32.0	100	30.0	24.0	46.0	100
18 or +	34.4	33.3	32.3	100	37.8	32.7	29.6	100	36.3	27.9	35.8	100	41.0	30.5	28.5	100
Never	37.3	27.1	35.6	100	23.8	44.4	31.7	100	38.3	36.7	25.0	100	29.0	33.9	37.1	100
Total	33.3	33.3	33.3	100	36.2	34.3	29.4	100	33.3	33.7	33.0	100	36.7	30.0	33.2	100
N	99	99	99	297	112	106	91	309	100	101	99	300	115	94	104	313

* Sample sizes vary because of missing values among the variable set used for each mentorship quality score.

Table B-5. Distribution of overall mentorship quality score* by mentor's characteristics

	Poor	Reg.	Good	Total
Age				
≤25	40.3	32.8	26.9	100
26–30	33.3	32.4	34.3	100
31+	24.7	35.5	39.8	100
Secondary school				
No	50.6	37.1	12.4	100
Yes	26.8	32	41.2	100
SES level				
Low	42.6	30.6	26.9	100
Medium	38.5	36.5	25	100
High	19	33.3	47.6	100
Age at 1st marriage				
<18	30	36.7	33.3	100
18 or +	35.1	34.6	30.3	100
Never	31.4	30.4	38.2	100
Age at 1st pregnancy				
<18	23.1	42.3	34.6	100
18 or +	33.8	32.8	33.3	100
Never	41.3	27	31.7	100
Total	33.4	33.4	33.1	100
N	106	106	105	317

*Defined as the mean of the following standardized scores: Knowledge, Safe Space Creation, Relation to Girls and Community, Self- Efficacy, Attitudes on Gender Norms, Attitudes on Contraception, and Attitudes on Gender-Based Violence. Missing scores were ignored when calculating means.

Table B-6. Logistic regressions explaining the probability of being a good mentor according to overall mentorship quality score and dimension-specific scores (odds ratios)

	1	2	3	4	5	6	7	8	9
	Overall	Knowledge	Safe Space Creation	Relationship with Girls and Comm.	Self-Efficacy	Gender Attitudes	Attitudes: Gender Norms	Attitudes: Contracep.	Attitudes: GBV
Secondary school	6.43***	8.11***	2.77**	1.17	1.99*	2.33*	5.00***	2.06*	1.11
Age, Ref≤25									
26–30	1.25	1.55	2.57*	0.69	0.7	1.37	0.82	1.93+	2.53**
31+	3.22**	3.07*	3.47**	1.63	1.29	2.73*	1.14	2.46*	2.37*
SES Level, Ref: Low									
Medium	1.26	2.60*	1.79	1.23	1.41	0.55	1.87	1.26	0.58
High	3.06**	4.78***	2.21*	0.78	2.09*	2.27*	5.18***	2.42*	2.08*
Age 1st pregnancy, Ref: Never									
<18	2.3	1.55	2.26	0.94	0.98	1.4	1.45	1.77	2.73+
18 or +	1.47	1.13	1.17	1.65	1.86	1.01	1.3	1.65	0.96
Never married	2.91**	2.27+	1.56	1.65	4.01***	2.17*	2.63*	2.21*	3.94***
N	316	271	279	301	308	297	308	300	312

Exponentiated coefficients.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table B-7. Linear probability models estimating the impact of dimension-specific high mentorship quality on programme outcomes: Pregnancy and contraception

	1	2	3	4	5	6	7	8
	Ever been pregnant	Ever unwanted pregnancy	Ever given birth	Ever given birth or currently pregnant	HIV-positive	HSV-2-positive	Ever used modern contraception	Condom use at first sex
Mentorship Quality Dimensions, Ref: Medium/Low								
Attitudes: Contraception	**-0.120***	**-0.082***	-0.046	-0.062	0.006	-0.002	-0.043	0.159+
	[0.050]	[0.041]	[0.044]	[0.046]	[0.030]	[0.033]	[0.046]	[0.090]
Self-Efficacy	-0.029	-0.006	-0.021	-0.01	**-0.076***	0.013	0.03	0.026
	[0.042]	[0.037]	[0.042]	[0.041]	[0.030]	[0.041]	[0.042]	[0.111]
Safe Space Creation	-0.041	-0.055+	-0.057+	-0.053	-0.01	0.032	0.021	0.104
	[0.040]	[0.033]	[0.033]	[0.036]	[0.018]	[0.037]	[0.038]	[0.110]
Rel. with Girls and Comm.	-0.024	-0.012	-0.025	-0.031	-0.013	-0.048	-0.014	0.028
	[0.043]	[0.038]	[0.033]	[0.037]	[0.017]	[0.034]	[0.042]	[0.083]
N	348	348	348	348	340	340	348	140

Standard errors in brackets. + 0.10 * 0.05 ** 0.01
Covariates not shown: Age, school attendance, grade attainment, mother alive, father alive, biological daughter of HH head, HH items, HH assets/savings, tribes (Bemba, Nyanja, Kaonde, other), mother completed primary, father completed primary, standardized vulnerability quintiles.

Table B-8. Linear probability models estimating the impact of dimension-specific high mentorship quality on programme outcomes: Violence and unwanted sex

	1	2	3
	Agrees has had sex at times didn't want to	Physical violence anyone in past 12 months	IPV in past 12 months
Mentorship Quality Dimensions, Ref: Medium/Low			
Attitudes: GBV	0.035	0.029	0.002
	[0.097]	[0.040]	[0.008]
Self-Efficacy	**-0.224***	-0.001	-0.001
	[0.105]	[0.039]	[0.007]
Safe Space Creation	-0.038	-0.02	0.006
	[0.106]	[0.035]	[0.008]
Rel. with Girls and Comm.	-0.177+	**-0.077***	-0.005
	[0.094]	[0.033]	[0.009]
N	138	529	529

Standard errors in brackets. + 0.10 * 0.05 ** 0.01
Covariates not shown: Age, school attendance, grade attainment, mother alive, father alive, biological daughter of HH head, HH items, HH assets/savings, tribes (Bemba, Nyanja, Kaonde, other), mother completed primary, father completed primary, standardized vulnerability quintiles.

Table B-9. Linear probability models estimating the impact of dimension-specific high mentorship quality on programme outcomes: Marriage and sex

	1	2	3
	Ever married	Ever had sex	Agrees has had transactional sex
Mentorship Quality Dimensions, Ref: Medium/Low			
Gender Attitudes	0.02	-0.113+	-0.065
	[0.024]	[0.065]	[0.111]
Self-Efficacy	-0.007	-0.036	-0.088
	[0.023]	[0.067]	[0.128]
Safe Space Creation	-0.041+	-0.021	0.103
	[0.022]	[0.060]	[0.122]
Rel. with Girls and Comm.	-0.028	0.07	-0.033
	[0.025]	[0.055]	[0.083]
N	347	348	134

Standard errors in brackets. + 0.10 * 0.05 ** 0.01
Covariates not shown: Age, school attendance, grade attainment, mother alive, father alive, biological daughter of HH head, HH items, HH assets/savings, tribes (Bemba, Nyanja, Kaonde, other), mother completed primary, father completed primary, standardized vulnerability quintiles.

Table B-10. Linear probability models estimating the impact of dimension-specific high mentorship quality on programme outcomes: Educational attainment

	1	2	3	4
	Highest grade complete, mean	Completed grade 7	Completed grade 9	Attended current school year
Mentorship Quality Dimensions, Ref: Medium/Low				
Gender Attitudes	-0.017	0.042	0.017	-0.05
	[0.113]	[0.036]	[0.030]	[0.039]
Self-Efficacy	-0.06	-0.047	-0.035	-0.022
	[0.136]	[0.035]	[0.027]	[0.034]
Safe Space Creation	-0.053	-0.042	-0.02	0.034
	[0.116]	[0.032]	[0.032]	[0.040]
Rel. with Girls and Comm.	-0.001	0.014	-0.043	-0.016
	[0.133]	[0.034]	[0.027]	[0.037]
N	609	609	609	609

Standard errors in brackets. + 0.10 * 0.05 ** 0.01
Covariates not shown: Age, mother alive, father alive, biological daughter of HH head, HH items, HH assets/savings, tribes (Bemba, Nyanja, Kaonde, other), mother completed primary, father completed primary, standardized vulnerability quintiles. Models for school attendance include grade attainment.

Annex C: Intent-to-Treat (ITT) and Treatment-on-Treated (ToT) Summary and detailed results.

The following Excel tables are available at:
http://www.popcouncil.org/uploads/pdfs/2016PGY_AGEPMidTermReport_AnnexC.xlsx

Table C-1. Intent-to-Treat (ITT) Summary Results: Difference-in-Difference (DID) Estimators - Excludes external urban controls

Table C-2. Intent-to-Treat (ITT) Detailed Results: Difference-in-Difference (DID) Estimators - Excludes external urban controls

Table C-3. Treatment-on-Treated (ToT) Summary Results: Difference-in-Difference (DID) Estimators - Excludes urban controls

Table C-4. Treatment-on-Treated (ToT) Detailed Results - Difference-in-Difference (DID) Estimators - Excludes external urban controls

Table C-5. Intent-to-Treat (ITT) by Arm Detailed Results: Difference-in-Difference (DID) Estimators - Excludes external urban controls

Annex D: SUMMARY OF KEY LESSONS LEARNED

PROGRAMME IMPLEMENTATION LESSONS

Safe Spaces

- Monthly meetings with mentors to provide regular supervision and refresher training were critical to ongoing mentor monitoring and support.
- Regular community sensitization was necessary so that the community consistently felt aware of programme aims and activities. However, community sensitization that goes beyond obtaining buy-in for the programme, which uses the opportunity to address the community as the beneficiary and work to increase knowledge and improve norms, could be a better leveraging of the time with the community.
- Access for the programme staff to real-time electronic monitoring data was critical in the ability of the programme staff to successfully monitor and correct course when certain groups, mentors, or sites seemed to be going off track.
- A system of a prize for each 10 meetings attended, while potentially motivating, was not successful enough to be worth the effort and cost. A more modest set of activities such as fun days, birthday celebrations, and creative activities could work, especially with the younger participants. Incentivizing participation of older girls will likely need to address the economic needs that they develop as they grow older.

Health Voucher

- Use of an e-platform that allows for real-time monitoring of voucher usage is a key tool for successful implementation.
- Ongoing (bi-annual, if not quarterly) meetings with district, provincial, and national-level government staff is useful for maintaining overall support for the voucher implementation.
- Staff turnover at clinics is quite high, therefore provisions for repeated training (annual at a minimum) should be included.
- For broad-based health service provision for adolescents, a voucher, even in the context of safe spaces, may not be enough to overcome all supply and demand barriers.
- When Safe Space groups stopped meeting, voucher usage dropped even lower, therefore the ongoing platform in which girls met and interacted with one another, as well as their mentors, did encourage voucher usage.

Savings Account

- Vulnerable adolescent girls can open savings accounts and save regularly (both formally and informally).
- Access to formal financial services can be a catalyst for both informal and formal savings activity.
- Financial education is a critical component to the successful rollout of financial services for vulnerable adolescent girls.
- Support from the executive level of management at the financial institution is critical, as well as training of all branch level staff (from the branch manager, to the tellers, to the security guards standing at the entrance).
- There may be a certain distance from the branch for which alternative savings schemes should be explored.

PROGRAMME DESIGN LESSONS

In the Zambian context, a programme that seeks to improve a wide range of long-term outcomes related to education, timing of sexual debut, pregnancy, and marriage, experience of sexual and gender-based violence and HIV and STI transmission should note the following lessons:

- For programmes in which girls choose whether or not to participate, typically with the permission of a parent/guardian, those who do not participate are likely to be more vulnerable than those who do.
- Safe spaces and direct asset-building for girls should be coupled with an "ecological" approach to working with girls to involve parents, boys and young men, and key community members. On their own, safe spaces may have a limited impact in the Zambian context.
- Address social and cultural norms around gender roles and adolescent sexuality, including use of SRH services, at individual, household, community, and even national levels.
- Take into account the specific needs of subgroups of the adolescent girl population. For example:
    - Encourage more formal education support for girls to remain in school; and
    - Link knowledge and skills to work opportunities and access to resources to enhance the effect of interventions.
- Consider the trade-off between a theory of change that focuses on a wide range of topics, but does not provide depth on any one of them, such as AGEP, versus one that focuses on a narrower range of outcomes, but provides more depth.

POLICY LESSONS

- Outcomes that are under more direct control of girls (knowledge, saving behaviour, transactional sex, condom use), might be easier to change with an asset-building programme for girls based on the safe spaces
- A health voucher, even coupled with safe spaces, is not sufficient to address all the barriers that girls face in accessing health services—particularly sexual and reproductive health services.
- Access to formal savings accounts is a catalyst for both formal and informal savings behaviour.
- Conservative norms on gender roles, acceptability of violence, and adolescent use of contraceptives are entrenched even in early adolescence. It will likely take an intervention at individual, household, community, and national levels to address these norms in a meaningful and impactful way.