



Adolescent Girls Empowerment Programme, Zambia

End Term Evaluation Report

Volume I: Main Report

Mott MacDonald Evaluation Team

19 January 2017

M

**MOTT
MACDONALD**

M

Acknowledgements

This Mott MacDonald End Term Evaluation was conducted by Terri Collins (Team Leader), Braimoh Bello (Epidemiologist) and Catriona Waddington (Health Economist). The local research team was led by Chishimba Mulambia, and the fieldwork team included Mwangala Manyando and Michael Kambole. All team members participated actively in the process of data collection, verification and analysis.

The Evaluation Team received timely assistance from a number of individuals in completing this evaluation. From DFID Zambia, we would like to thank Valerie Roberts and Clare McCrum (Social Development Advisers), Cormac Quinn (Statistics Adviser and Results Lead) and Annelies Raue (DFID Economist) for their useful technical guidance. Ed Barney from DFID's Research and Evidence Division (RED) team in the UK also provided invaluable advice. In addition, we would like to thank the DFID Programme Officer, Florence Mabandhla, for her coordination and logistics support.

At the Population Council, Karen Austrian (Programme Director and Principal Investigator) and Natalie Jackson Hachonda (Senior Programme Manager) were open and helpful in providing us with information and data. We are also grateful to Chipso Zulu (Health Vouchers Coordinator), Paul Hewett (Lead Investigator), Erica Soler-Hampejsek (Research Associate), Fiammetta Bozzani (Economics Research Fellow) and the teams from YWCA and NatSave for the information they shared. We acknowledge the insights and experience of the mentors, parents and community leaders we interviewed in four AGEP Master Sites. Above all, we would like to thank the many AGEP girls we spoke to. They remain shining ambassadors for the AGEP programme, and testimony to the potential that can be unlocked by investing in adolescent girls.

Photographs: The photographs used in this report have been sourced from the Population Council's documents and website.

Contents

Abbreviations	i
Executive Summary	1
1. Introduction	6
2. AGEP overview	8
2.1 The AGEP Theory of Change	8
2.2 Programme components	8
2.3 Research overview	9
3. The independent evaluation design	13
3.1 A pragmatic approach	13
3.2 Applying a theory-based evaluation approach	13
3.3 Setting up the theory-based evaluation –conceptual analysis	15
4. Methodology	18
4.1 Overview	18
4.2 Data collection methods	18
4.3 Sampling strategy	18
4.4 Ethical compliance	20
4.5 Data analysis	20
4.6 Note on the economics evaluation	21
4.7 Limitations of the independent evaluation methodology	22
5. Findings Part 1: Theory-Based Evaluation	24
5.1 Findings from the empirical analysis of Hypothesis 1	24
5.2 Findings from the empirical analysis of Hypothesis 2	34
5.3 Findings from the empirical analysis of Hypothesis 3	42
5.4 Final reflections from the theory-based evaluation	49
6. Findings Part 2: Economic Evaluation	50
6.1 Understanding the different types of economic evaluation	50
6.2 Progress overview	52
6.3 Findings from verification/validation work	53
6.4 Commentary of the findings in the ‘Economic Evaluation End-of- Programme Report’	53
6.5 Review of Value for Money	60
7. Findings Part 3: Comparative Evaluation	62
7.1 Aim of the comparative evaluation	62
7.2 Comparison of epidemiological effectiveness	62
7.3 Economic comparison	70
8. Discussion	75
8.1 Evaluation Question 1	75
8.2 Evaluation Question 2	76

8.3	Evaluation Question 3	77
8.4	Evaluation Question 4	79
8.5	Evaluation Question 5	79
8.6	Summary of lessons	80
9.	Conclusion and recommendations	82
9.1	Conclusion	82
9.2	Recommendations	82

Table of tables

Table 1: Summary of theory-based findings by hypothesis	2
Table 2: Assumptions underpinning the hypotheses for the theory-based evaluation	16
Table 3: Approach to data analysis for each task area.....	20
Table 4: Main limitations to this evaluation	22
Table 5: Summary of key features of urban and rural contexts most relevant to the social assets of adolescent girls	30
Table 6: Review of logframe assumptions relevant to building of social assets.....	31
Table 7: Review of logframe assumptions relevant to building of health assets	40
Table 8: Review of logframe assumptions relevant to building of economic assets.....	47
Table 9: Types of economic evaluation	50
Table 10: Progress with activities in the economic evaluation of AGEP	52
Table 11: Unit cost figures for some DFID supported programmes in Zambia	54
Table 12: Incremental effectiveness and average cost-effectiveness, 2016 US\$	56
Table 13: Unit costs of achieving empowerment indicators and outcomes.....	58
Table 14: Cost of AGEP delivery, US\$ 2016	61
Table 15: Programmes contacted for inclusion in the comparison.....	63
Table 16: Comparison of AGEP short-term impact to MCP and ZCTP longer-term impact on educational and sexual and reproductive health outcomes.....	69
Table 17: Comparison of costs and effects relating to adolescent girls	71
Table 18: Estimated multiplier effects of the Child Grant Programme and Multiple Categorical Programme.....	73

Table of Figures

Figure 1: AGEP Theory of Change	8
Figure 2: Timelines for the AGEP programme and research, and the independent evaluation.....	10
Figure 3: Current empowerment outcome and long-term impact Indicators for the AGEP research... ..	12
Figure 4: Overview of the role of the independent evaluation team	13
Figure 5: Review of pathways to social assets for urban girls at Round 3	25
Figure 6: Review of pathways to social assets for rural girls at Round 3	27
Figure 7: Review of pathways to health assets for urban girls at Round 3.....	35
Figure 8: Review of pathways to health assets for rural girls at Round 3.....	36
Figure 9: Review of pathways to economic assets for urban girls at Round 3.....	43
Figure 10: Review of pathways to economic assets for rural girls at Round 3.....	44
Figure 11: Value for money criteria.....	60

Table of Boxes

Box 1: Five questions for the evaluation team	6
Box 2: Logic of the AGEP results chain as described in the 2014 analysis plan.....	11
Box 3: Key features of a theory based evaluation approach	14
Box 4: Step-wise approach to developing a theory-based evaluation of AGEP	14
Box 5: Proposed hypotheses to be tested through the theory-based evaluation	15
Box 6: Evidence cited in the AGEP Business Case to support Safe Spaces.....	17
Box 7: Secondary benefits of health vouchers.....	38
Box 8: Case study on saving from Kapiri Mposhi	45
Box 9: Costs, empowerment outcomes and impact of AGEP by month 24	59
Box 10: Overview of the selected comparison studies	64
Box 11: Cost-effectiveness ratios 2010-12, Zambia Child Grant Programme.....	72
Box 12: Spending \$50,000 on AGEP and on the MCP	74
Box 13: The Spill over dilemma	76
Box 14: The focus on vulnerability – some challenges	78
Box 15: Summary of lessons emerging from implementation of the AGEP model	80
Box 16: Specific Questions for the Round 5 Research.....	83

Abbreviations

ACASI	Audio Computer-Assisted Self-Interviewing
AGEP	Adolescent Girls Empowerment Programme
AIR	American Institutes for Research
ART	Anti- Retroviral Therapy
CAPI	Computer-Assisted Personal Interviewing
CBA	Cost-Benefit Analysis
CCA	Cost-Consequence Analysis
CEA	Cost-Effectiveness Analysis
CGP	Child Grant Programme
CHOICE	Choosing Interventions that are Cost-Effective (WHO)
CSA	Census Supervisory Area
CT	Cash Transfer
CUA	Cost-Utility Analysis
DALY	Disability Adjusted Life Year
DCMO	District Community Medical Officer
DDCC	District Development Coordinating Committee
DFID-Z	Department for International Development (UK) -Zambia
DiD	Difference in Difference [analysis]
DMO	District Medical Officer
ESC	Evidence Scrutiny Committee
FGD	Focus Group Discussion
FTE	Full Time Equivalent
GBV	Gender-Based Violence
GDP	Gross Domestic Product
HIV	Human Immunodeficiency Virus
HSV-2	Herpes Simplex Virus 2
I\$	International dollars
ICER	Incremental Cost-Effectiveness Ratios
InSTEDD	Innovation Support to Emergencies, Diseases, and Disasters
ITT	Intent to Treat
LSHTM	London School of Hygiene and Tropical Medicine
MCDSW	Ministry of Community Development and Social Welfare
MCP	Multiple Categorical [Cash Transfer] Programme
MoH	Ministry of Health
MTE	Mid-Term Evaluation
NatSave	National Savings and Credit Bank of Zambia
NGO	Non-Governmental Organisation
NRC	National Registration Card
ODI	Overseas Development Institute
ODK	Open Data Kit
PDA	Personal Digital Assistants
PI	Principal Investigator
PPAZ	Planned Parenthood Association of Zambia
RED	Research and Evidence Division
SRH	Sexual and Reproductive Health

TOT	Treatment on the Treated
UNFPA	United Nations Population Fund
UNICEF	United Nations Children’s Emergency Fund
UNZA-BREC	University of Zambia Biomedical Research Ethics Committee
US	United States
USAID	United States Agency for International Development
VFM	Value for Money
WHO	World Health Organisation
YWCA	Young Women’s Christian Association
ZCTP	Zomba Cash Transfer Programme

Executive Summary

The DFID-funded Adolescent Girls Empowerment Programme (AGEP) aims to empower vulnerable adolescent girls in Zambia. The AGEP model is based on a package of interventions that includes: 'Safe Spaces' for life skills and financial education, vouchers for accessing age-appropriate sexual and reproductive health (SRH) services and 'Girl's Dream' savings accounts. The Population Council has recently completed implementation of the interventions in four provinces. The interventions have been delivered over a 24-month period and have targeted 10 000 vulnerable girls in two age cohorts: 10-14 years and 15-19 years. The Population Council is also conducting a four-year impact evaluation of the model to assess intermediate empowerment effects and longer-term impacts. The research is based on a randomised cluster controlled study among a sample of 4 000 girls in 10 urban and rural 'Master Sites'. The study includes a qualitative research component and an economic evaluation. The research will conclude in September 2018.

Mott MacDonald has been contracted to conduct an independent End-Term Evaluation of the AGEP programme and research. Notably, this evaluation comes at the end of the programme of interventions, but mid-way through the Population Council's impact evaluation research. This report presents the findings of the independent End-Term Evaluation.

Evaluation design

The terms of reference for the independent evaluation defined the five evaluation questions to be addressed (see Annex 1). These related to review of: the quality of the programme implementation and research; cost-effectiveness and comparison against alternative approaches; potential sustainability and scale-up; recommended changes and efforts to support research uptake. Our primary evaluation design has been based on a pragmatic approach to address these questions.

DFID also asked us to frame the evaluation within a theory-based evaluation design to examine the AGEP theory of change. In order to do this, we have identified three core hypotheses within the theory of change (see Box 5, p.16). These hypotheses describe how the programme components are expected to build social, health and economic assets and contribute to longer-term impacts on education, sexual, reproductive and maternal health and experiences of gender-based violence. We set up our analysis to test each hypothesis based on a series of steps that included review of: disaggregated results to date; implementation 'mechanisms'; contextual factors; and validity of the design assumptions. These steps allowed us to assess the current strength of evidence for each hypothesis, determine the implications for the theory of change, identify lessons for future programming and assess the implications for sustainability and scale-up.

Methodology

In designing the evaluation methodology, our aim was to avoid duplicating the research of the Population Council; rather, we aimed to independently verify the research findings of the Population Council and add value to their analysis through peer review.

The fieldwork for the End Term Evaluation was completed in June 2016. We used mixed methods of sampling and data collection to capture the experiences of a range of programme stakeholders. We sought to maximise opportunities for triangulation and, as far as possible, avoid selection and confirmation bias. We visited two urban Master Sites in Lusaka and two rural Master Sites in Masaiti A and Kapiri Mposhi. We also conducted extensive data quality assurance and research verification work, along with reviews of programme and financial records and secondary literature. We observed all ethical, professional and data security protocols. Our data analysis procedures were tailored to the mixed types of data collected, and were internally quality assured to ensure technical rigour and evidence-based findings. We took active measures to respond to feedback on the first draft of this report and to correct any factual errors.

We acknowledge that there were some limitations to our methodology, such as potential biases in the selection of sites and individuals for interview; where possible we have taken mitigating action to limit any effects (see Section 4.7). There have also been some analytical constraints: for example, we have not yet had access to data analysis from the external control groups -this means we have not been able to fully assess the effects of spillover.

Findings Part 1: Theory-based evaluation

The main findings for the three hypotheses underpinning the AGEP theory of change are summarised in the table below. All results mentioned are drawn from the verified findings of the Population Council's Research and Evaluation Mid-Term Technical Report. For this analysis we have used the Intent To Treat (ITT), Difference in Difference (DiD) estimators for survey rounds 1-3 (R1-3).

Table 1: Summary of theory-based findings by hypothesis

Hypothesis	Summary of main findings
Hypothesis 1 on building social assets	<ul style="list-style-type: none"> □ From the ITT, DiD analysis for R1-3, there is evidence that Safe Spaces produced a significant positive ($p < 0.05$) effect for younger rural girls in having a safe space to meet friends, and for older rural girls on condom use at first sex. However, for the majority of indicators on social assets, there was no evidence of effects over Rounds 1-3. Of some concern were the significant negative effects on impact level indicators relating to sexual debut and school attendance for older urban girls. □ Evidence from the evaluation review of mechanisms suggested that, although the Population Council has done a commendable job in implementing the Safe Spaces model <i>as designed</i>, there is scope for strengthening several aspects of the approach –especially through customising the curriculum to different contexts and age cohorts; improving strategies for motivating attendance and increasing inclusivity; supporting the role of mentors; and improving community engagement. We note that several of the design assumptions could not be validated (e.g. the presence of other complementary initiatives working on structural and normative issues, and male involvement). □ We conclude that the evidence for Hypothesis 1 on building social assets is inconclusive at this stage. Considerations for sustainability and scale-up

Hypothesis	Summary of main findings
	<p>include the need to establish the scalable model and demonstrate that durable social assets results can be achieved in different implementation settings within acceptable timeframes.</p>
<p>Hypothesis 2 on building health assets</p>	<ul style="list-style-type: none"> □ From the ITT, DiD analysis for R1-3, there is evidence of significant positive effects ($p < 0.05$) on the contraceptive knowledge of younger, rural girls. However, for the majority of indicators on health assets, there remains no strong evidence of effects over Rounds 1-3 for either urban or rural girls across the two age sets. □ Evidence from our evaluation review of mechanisms confirms that there has been limited use of health vouchers –or just repeated use by a minority of girls. This, combined with the delayed roll-out of this component, could explain why the health voucher scheme has not yet translated into significant gains in perceived health status by Round 3. Important lessons for future programming relate to: the need to better calibrate voucher reimbursements to facilities in different sectors; the important leadership role of Heads of Facilities; and the importance of community and parental sensitisation work. Again, some design assumptions cannot be validated (e.g. the assumption that health vouchers alone are sufficient to leverage uptake of services). □ So, at this stage, the evidence for Hypothesis 2 on building health assets remains inconclusive. Implications for scale-up include the need to document lessons from the health vouchers scheme and the need to engage with national and district MoH strategies on promoting adolescent health.
<p>Hypothesis 3 on building economic assets</p>	<ul style="list-style-type: none"> □ From the ITT, DiD analysis for R1-3, there is strong evidence that the intervention led to significant positive effects ($p < 0.05$) on the saving activity of older and younger urban girls, and on the percentage of older urban girls working for cash or in-kind payments. There is also strong evidence of significant positive effects on the income earned by older rural girls working for cash, and on the financial literacy scores of younger rural girls. For most other indicators on economic assets, the evidence remains inconclusive; however, the significant negative effect on the percentage of older rural girls working for cash or in-kind payments remains difficult to explain. □ Evidence from our evaluation review of mechanisms suggests it is difficult to link positive gains in economic assets to use of the Girls’ Dream savings accounts –other than through reinforcement of learning. An important lesson is that distance to banks is a key barrier, especially for rural girls, and the design assumption on this theme needs to be reviewed. □ So, at this stage, the evidence for Hypothesis 3 on building economic assets appears somewhat mixed. Implications for sustainability and scale-up are that the Girls’ Dream savings account may not be sustainable in the longer-term without further collaborative working with the NatSave Bank and further investment from partners. The cost-effectiveness of this component thus remains uncertain.

In our reflections on the theory-based evaluation (Section 5.4), we note that results relating to the empowerment outcome indicators warrant particular attention at this time because they specifically relate to asset building. The AGEP theory of change posits that assets built during the two-year intervention phase will be protective for girls undergoing the transition to adulthood, and will thereby lead to longer-term impacts results (measurable in the 2017 Round 5 survey). Consequently, an emerging question is, how strong do the effects on empowerment outcomes need to be at this point (for younger and older adolescent girls respectively) to produce intended impacts in the longer term?

Findings Part 2: Economics evaluation

The economic evaluation of AGEP should be regarded as ‘cost-consequence analysis’ since it provides information on overall costs against a variety of types of outcome. Our review of the Population Council’s progress in conducting the economics evaluation study confirms that, despite the limited budget allocated for this work (<£20k or <1% of the research budget), reasonable progress has been made and a good deal of information has been assembled. Nevertheless, some work needs to be done on the most recent version of the economic evaluation report (August 2016) to align it with the indicators and analysis in the Population Council’s Mid-Term Technical Report; there also needs to be consideration of costs per negative effect and clarity on levels of significance.

One notable finding from the economic evaluation report is that the average total cost per participant is \$771, or \$394 for the Safe Spaces component alone. Based on the figures provided, and using a crude assumption of a 20% decrease in unit costs for an expanded AGEP, we have estimated that the unit costs for scale-up is £470 per girl, or £240 (August 2016) for the Safe Spaces-only version. We suggest that, once Population Council has re-worked the costs per outcome summaries, it will be possible to generate a cost-consequence analysis for the 24-month outcomes. The existing analysis already tells us that there were no significant additional benefits for the additional costs of Arms 2 and 3 of AGEP. However, the best time to look at an economic evaluation of AGEP is when the information on longer-term impacts is available.

With regards value for money (VFM), little has changed since the October 2015 Annual Review. AGEP has been a labour-intensive intervention, with salaries accounting for more than half of all spending.

Findings Part 3: Comparative evaluation

Our evaluation has included comparison work to assess how well the AGEP model compares with alternative approaches. In this report, we have demonstrated that setting up credible comparisons is detailed work that requires access to primary data from other studies, as well as careful alignment of objectives, indicators, timeframes and targeting strategies. We have established that two cash transfer programmes, the Zambia Multiple Categorical Programme (MCP) and the Malawi Zomba Cash Transfer Programme (ZCTP) offer useful comparisons, both in terms of comparison criteria *and* conceptually as an alternative approach. The comparative analysis shows that, after 24 months of interventions, AGEP compares reasonably well with the other programmes, since it has had significant positive effects on transactional sex and condom use at first sex. However, the MCP initiative

appears to have delivered better on educational outcomes, while the ZCTP delivered better on other sexual outcomes over similar timeframes. It should be noted that this analysis is based on comparable indicators only, and each approach offers distinctive additional benefits. Importantly, more recent results from the ZCTP study show that early effects have not been sustained over the longer-term so, once again, AGEP's 2017 survey will be critical for demonstrating epidemiological effectiveness over time.

Our literature review shows that costing data on alternative approaches is limited. We have compiled the costing data that is available for the comparison studies, but suggest that this data needs to be interpreted with caution. Based on the economics analysis at 24 months, AGEP appears expensive relative to other approaches. However, if AGEP shows significant positive impacts at 36 months, then the cost-consequence analysis could shift in favour of the AGEP model (based on the selected indicators of empowerment).

Discussion

In the discussion section of this report, we have reflected on the material covered by the evaluation findings to assess whether the five evaluation questions have been adequately addressed. We are confident that, together with the comprehensive annexes to this report (see Volume II), each of the questions has been fully addressed. We also provide additional evidence to demonstrate that the Population Council has been diligent in supporting research uptake through high quality publications and communication activities.

We suggest that there have been a number of important operational lessons from implementation of the Safe Spaces model and review of results to date -for example, lessons have surfaced on the importance of maintaining high-quality mentorship, and the need to address the challenges of sustained participation and community engagement. Questions have also emerged on how best to build economic and health assets within acceptable timeframes, and the feasibility of targeting the most vulnerable girls in urban and rural areas. These lessons and questions are likely to have implications for the final version of the AGEP model once impact-level results are known (see Annex 10, pp 55-57). Once the final version of the AGEP model has been established, it will be possible to make more definitive assessments of the cost of scale-up and the potential for institutional sustainability.

Conclusion and recommendations

We conclude this report with a reminder that, given the timing of the independent evaluation, our findings should only be seen as interim and as an assessment of the 'direction of travel' of the AGEP programme and research.

Our **recommendations** (Section 9.2) are based on the need to address a number of specific questions through the Round 5 research. We emphasise that these questions can only be satisfactorily addressed through stronger iterative links between the quantitative, qualitative and economics research. Additional recommendations relate to the need to: a) fully align indicators across key conceptual frameworks; b) document and disseminate lessons from programme implementation; and c) promote opportunities for collaborative learning partnerships on adolescent girls in Zambia.

1. Introduction

The Adolescent Girl's Empowerment Programme (AGEP), Zambia aims to empower adolescent girls aged 10-19 years by increasing school completion and improving sexual, reproductive health outcomes. To this end, the Population Council has implemented a programme of interventions based on three components: Safe Spaces;¹ Girl's Dream savings accounts and a health voucher scheme. Following an inception and pilot phase, the AGEP interventions were rolled out to almost 10,000 vulnerable adolescent girls over the period August 2013 –March 2016. In order to provide robust evidence on the impact of the interventions over the shorter and longer term, the Population Council is conducting an impact evaluation research study. This study is using a randomised cluster controlled design based on a sample of 4,000 unmarried girls in two age cohorts: 10-14 years and 15-19 years. The girls have been assigned to four treatment arms across 10 urban and rural Master Sites and surveys are being conducted annually over a five-year period (2013-2017) (see Section 2). The impact evaluation study is being accompanied by a qualitative study and an economic evaluation.

Mott MacDonald has been contracted to provide independent evaluation services for the AGEP initiative. The main aims of independent evaluation services are: to provide a peer review of the AGEP programme of interventions; to provide independent verification of the research findings; to compare the results from AGEP with alternative approaches; and to consider issues of cost-effectiveness, sustainability and research uptake. The main deliverables of the independent evaluation include: a technical briefing paper, an inception report, and comprehensive independent mid-term and end-term evaluations of the AGEP programme and research (see Annex 1, p.6). For the End Term Evaluation DFID has asked the Mott MacDonald evaluation team to focus on five main questions (Box 1).

Box 1: Five questions for the evaluation team

1. Has the Safe Spaces programme and research been well-implemented? What lessons have been learned from implementing the Safe Spaces programme and research in Zambia?
2. Is Safe Spaces the most cost-effective and efficient approach to empower adolescent girls? How does it compare with alternative empowerment programmes?
3. Is the Safe Spaces programme sustainable (i.e. are programme impacts likely to last) and can the model be scaled up (would processes allow and is it affordable)? Is there government commitment and will donors engage and coordinate to support this programme?
4. What changes should be made to the programme implementation in future?
5. Have the research findings from the Safe Spaces programme been sufficiently used to influence policy debate and the development and implementation of new programmes?

¹ The Safe Spaces model includes mentor-led life skills, SRH and financial education in a safe out-of-school environment.

This report forms the principal deliverable for the AGEP End Term Evaluation. Data collection for the independent evaluation was completed over the period June-July 2016. Notably, the End Term Evaluation took place at the end of AGEP's two-year programme of interventions, just after the Population Council's Round 3 research survey. Given the timing of the independent End Term Evaluation and the important opportunities for learning that have emerged from the AGEP initiative so far, DFID has asked the Mott MacDonald evaluation team to address the evaluation questions within a theory-based evaluation framework that allows for reflections on the AGEP theory of change.

This End Term Evaluation Report begins by presenting an overview of the AGEP programme and research (Section 2). Section 3 describes the independent evaluation design and the approach used for the theory-based evaluation, comparison work and economics evaluation. In Section 4, we present a more detailed account of the evaluation methodology. The findings of the independent evaluation are presented in three parts. Part 1 focuses on findings from the theory-based evaluation work and reviews the status of the evidence for three main hypotheses underpinning the AGEP theory of change. Part 2 focuses on review of the economics evaluation and a value for money assessment. Part 3 focuses on the comparison work to assess how the progress of AGEP compares with alternative approaches to adolescent girls' empowerment. The final sections of this report include a discussion in which we reflect on the five evaluation questions. We conclude with five recommendations for next steps.

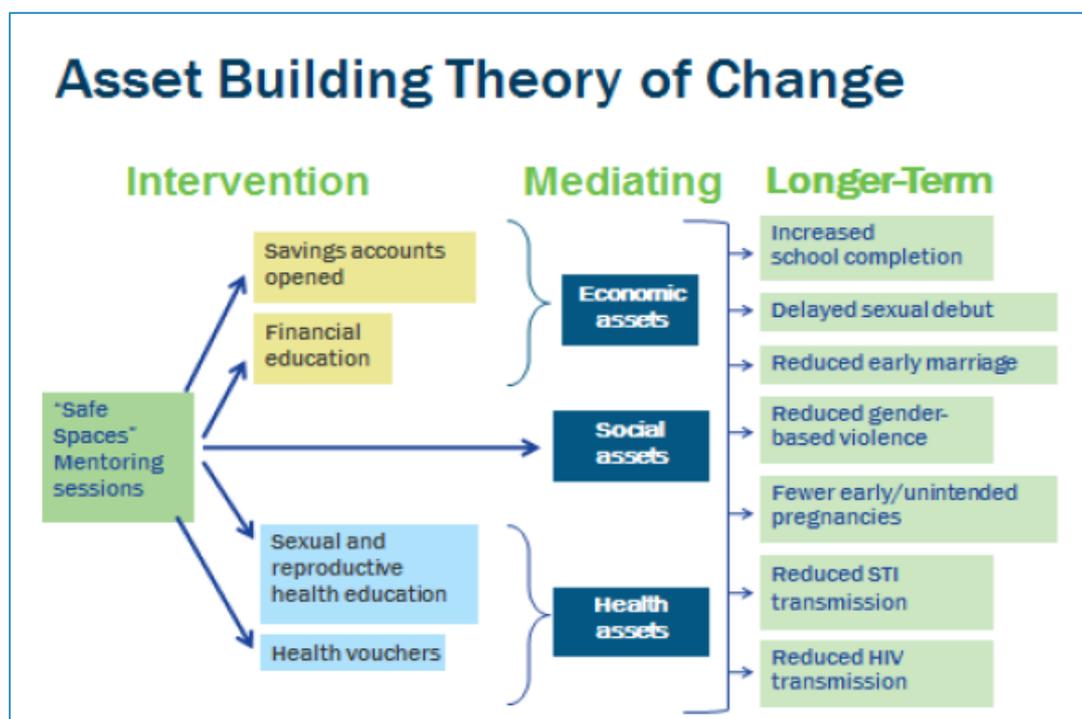
Since there are a large number of annexes to this report, we have submitted them separately as an additional volume to this report (Volume II). The annexes to this report include: the evaluation terms of reference; a map of AGEP Master Sites; the AGEP logical framework; a summary of AGEP survey findings; additional notes on our methodology and data analysis; a review of whether the AGEP research has been well implemented; verification work on survey results; a review of progress towards logframe milestones; additional notes on findings for the theory-based evaluation; considerations for sustainability and scale-up; a literature review of approaches to programming for adolescent girls; annotations to the theory of change; an overview of AGEP communication products and dissemination plan; and a list of people we have interviewed for this evaluation.

2. AGEP overview

2.1 The AGEP Theory of Change

For Zambian girls, social isolation, economic vulnerability, and lack of appropriate health information and services are critical problems that prevent a healthy transition from girlhood to womanhood. The design of AGEP is based on a theory of change that describes how Safe Spaces (including financial education) combined with health vouchers and savings accounts address these linked problems by building social, economic and health assets. These, in turn, are expected to support longer-term empowerment outcomes -increased school completion, delayed sexual debut, reduced early marriage, reduced gender-based violence, unintended pregnancies, human immunodeficiency virus (HIV) and sexually transmitted infections (STIs) (Figure 1).

Figure 1: AGEP Theory of Change



2.2 Programme components

2.2.1 Safe Spaces

Following the pilot period (August 2012-August 2013), the Safe Spaces component was implemented over a two-year period (August 2013-March 2016) using a phased roll-out strategy across ten sites in four provinces partnership. The Population Council implemented the Safe Spaces component in partnership with the Young Women's Christian Association (YWCA) Zambia. The intervention involved girls participating in a weekly meeting with a mentor (a young woman from their community). During the meetings, girls received training on a range of health, life skills and financial topics. The regular meetings were intended to serve two functions: 1) to provide a 'space' for girls to meet and receive a variety of interventions and educational topics, and 2) to build social assets –these included access to

a trusted adult, friendships and self-esteem that could have a positive influence on other aspects of their lives, such as livelihoods and health. Safe Spaces meetings were structured around two age cohorts, 10–14 and 15–19 years, with some additional groups for married girls. These cohorts were intended to provide girls with age-appropriate support over the course of adolescence, shape their goals and expectations and contribute to positive behaviour change.

2.2.2 Health vouchers

In partnership with the Ministry of Health (MoH) and the former Ministry of Community Development, Mother Child Health, the Population Council developed an e-health voucher that AGEP girls could redeem for a package of health services at contracted public, private and non-governmental health facilities. The services covered by the voucher included basic wellness examinations, as well as age-appropriate sexual and reproductive health services. Service providers used text messages to interact with a web-based system that issued real-time authorisation for the services. The Council trained providers at participating facilities in adolescent-friendly health services and conducted monitoring and quality-assurance visits. Providers were reimbursed per service based on pre-approved rates.

2.2.3 Savings accounts

The Population Council in partnership with the National Savings and Credit Bank (NatSave) and Making Cents International developed the 'Girls Dream' savings account for AGEP participants. This savings account required a very low minimum opening balance, and any amount could be deposited or withdrawn with no fee. Girls were able to make deposits on their own, but in order to comply with Zambian legal and regulatory requirements, girls under-18 years had to select a co-signatory to assist with account opening and withdrawals. This co-signatory could be a trusted adult female, such as the girl's mother or mentor.

2.3 Research overview

2.3.1 Impact evaluation design

The Population Council has been conducting a four-year impact evaluation of the AGEP interventions using a cluster randomised controlled design based on a sample of 4,000 girls across the two age cohorts. The study has been following girls for the two-year intervention (programme) period (survey rounds 1-3) and will continue to do so for two years after the interventions have concluded (survey rounds 4-5). Different versions of the programme were assigned to randomly selected communities (defined as Census Supervisory Areas (CSAs)) in ten Master Sites (urban and rural) (see Annex 2 for a map of AGEP Master Sites). Versions –or treatment arms- of the programme included: Safe Spaces only (T1); Safe Spaces with a health voucher (T2); and Safe Spaces with health vouchers and savings accounts (T3). The empowerment effects for AGEP participants are being compared with control group participants. Control sites include 'internal' controls at the Master Sites; they also include five urban 'external' controls for use in the analysis of spillover. Study instruments include a survey that measures self-esteem, social networks, attitudes and behaviours relating to gender, work and savings activity, nutrition status, literacy and numeracy skills, cognitive function, sexual and reproductive health knowledge, and sexual behaviour. In addition, HIV and Herpes Simplex Virus 2 (HSV-2) and height and weight

measurements are being collected for study participants aged 15-19 years in survey rounds 1, 3 and 5. Other components of the research have included 144 in-depth interviews for the purposes of supplementary qualitative research.

2.3.2 Economics component

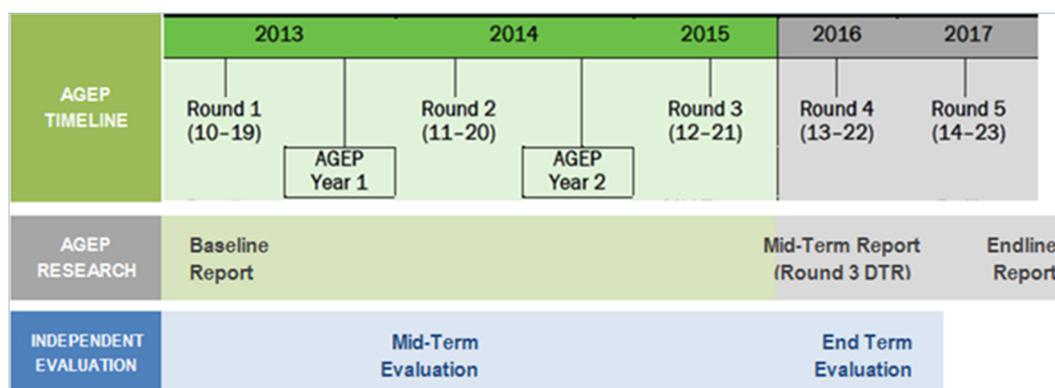
In addition to the impact evaluation research, the Population Council is conducting an economics evaluation of the AGEP programme. This study was designed to include the following components:

- Collation of programme resource utilisation data from the Population Council;
- Collection of participant-specific out-of-pocket and indirect costs data;
- A 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 incremental cost-effectiveness ratios (ICERs).

2.3.3 Progress to date

Figure 2 below provides an overview of the AGEP programme and research timelines.² We have added the independent evaluation timelines to help situate this report.

Figure 2: Timelines for the AGEP programme and research, and the independent evaluation



The Population Council conducted the Round 1 (R1) baseline survey from July of 2013, prior to programme implementation.³ The Round 2 (R2) survey was conducted between July and December 2014. Key findings from the Round 1 and Round 2 surveys are summarised in Annex 4. The Round 3 survey commenced in July 2015 and preliminary findings have been shared with the evaluation team in the Research and Evaluation Mid-Term Technical Report (July 2016). HIV and HSV-2 testing was conducted among adolescents aged 15 years and older at Rounds 1 and 3.

² Adapted from Population Council (2016). Adolescent Girls Empowerment Programme, Research and Evaluation Mid-Term Technical Report, July 2016: 29

³ See Hewett P et al. (2014). Adolescent Girls Empowerment Programme: Research and Evaluation Baseline Technical Report. Lusaka, Zambia. Population Council

Qualitative semi-structured interviews took place at Round 1 and Round 3 among a sub-sample of respondents. Findings from this qualitative work are described in the Population Council's 2016 qualitative report.⁴

In 2016, the Council also completed a preliminary analysis of the costing data for the economics evaluation. The preliminary economic evaluation findings covered the inception and implementation phases of AGEP between November 2011 and June 2016, and were described in the Population Council's Mid-Term Technical Report.⁵ However, this analysis did not include any form of cost-effectiveness analysis, so Mott MacDonald requested the Population Council to re-visit this section. The updated version was produced as the 'Economic Evaluation End-of-Programme Report' (August 2016). This is currently a stand-alone report and will need to be incorporated into future drafts of the main report using consistent information.

Additional survey rounds for the impact evaluation research will be conducted in 2016 and 2017. Anaemia testing among adolescents aged 15 years and older was added in Round 3 and will be conducted annually until 2017. HIV and HSV-2 status will be tested again at endline (2017).

2.3.4 Research indicators

DFID uses the AGEP logical framework (logframe) for tracking the progress of implementing the AGEP programme and research, as well as its contribution to results (see Annex 3 for the AGEP logframe). The AGEP logframe contains many indicators that overlap with those being used for the impact evaluation research.

In December 2014, the Population Council produced a statistical analysis plan that described the sequenced output, mediating outcome and longer-term impact indicators that would reflect the 'results chain' for the analysis of survey findings. The logic underpinning this results chain is summarised in Box 2.

Box 2: Logic of the AGEP results chain as described in the 2014 analysis plan

*"For AGEP girls, exposure to the programme is expected to result in an increase in a comprehensive set of social, human and financial assets that allow them to gain greater control of their health and economic decisions. In turn, these assets should serve to improve their life trajectories by increasing their educational attainment, delaying sexual debut, reducing unwanted pregnancy and STIs, increasing their ability to support themselves and their families financially, and increasing their control over health and financial decision making. These outcomes are hypothesized to ultimately reduce poverty for participants and their future families and communities."*⁶

In the Population Council's Mid-Term Technical Report,⁷ the indicators for measuring progress have been organised slightly differently and refer to "empowerment outcome

⁴ Duby Z et al. (2016). Adolescent Girls Empowerment Programme: Qualitative Evaluation Report. Lusaka: Population Council.

⁵ Population Council (2016). Op. cit. pp. 37-42 and 66-69

⁶ Population Council (2014). Statistical Analysis Plan for the Adolescent Girls Empowerment Program, 23 December 2014, pp. 10-11.

⁷ Population Council (2016). Op. cit.

indicators” and “longer term impact indicators”. This recent listing of indicators is shown in Figure 3 below.

We have agreed with DFID that we will reference the indicators listed in Figure 3 below for the End Term Evaluation analysis. However, the independent evaluation team would like to emphasise that changes in the conceptualisation of the results chain and lack of alignment of key conceptual frameworks has led to difficulties in assessing progress towards results and to internal inconsistencies in the analysis (e.g. the economics work is currently based on a different set of indicators from the impact evaluation analysis).

Figure 3: Current empowerment outcome and long-term impact Indicators for the AGEP research

Empowerment Outcome Indicators		Long-term Impact Indicators	
Empowerment domain	Assessment outcome measures	Outcome domain	Assessment impact measures
Social assets	1. Avg. score on self-efficacy scale	Educational	1. Avg. number of years completed
Self-efficacy	2. % confident regarding their ability to plan		2. % completed primary school
Autonomy	3. % feel they make good decisions regarding money		3. % completed junior secondary school ^a
	4. % agree that permission to go to the health clinic is not a problem		4. % currently attending school
Friends	5. % who jointly or solely make decisions with regard to money earned	Sexual risk behaviour ^a	5. % ever had sex
	6. Avg. number of friends		6. % agree that they have had unwanted sex
Safety nets	7. Avg. number of friends in school		7. % agree that they have had transactional sex
	8. Avg. number of friends who can be counted on if needed money		8. % used condom at last sex with non-marital partner
Isolation	9. Avg. number of friends who can be counted on in an emergency	9. % used condom at first sex	
	10. % have a safe space to meet with friends	Marital ^a	10. % ever married
11. % with adult female support in case of serious problem	11. Avg. number of HIV risk-related topics discussed with partner		
12. % attending any social groups/clubs within the past month	12. Avg. marital control score		
13. % who often/sometimes go to market ^a	Pregnancy & births ^a		13. % who have ever been pregnant
14. % who often/sometimes go to community centre ^a		14. % who have ever had an unwanted pregnancy	
15. % who often/sometimes go to shops or restaurants ^a		15. % currently pregnant or who have given birth	
16. Avg. score on gender equality scale		16. % ever used modern contraception	
Beliefs	17. Avg. score on nonacceptability of intimate partner violence	Sexually transmitted infections ^a	17. % HIV positive
Economic assets			18. % HSV-2 positive
Financial literacy	18. Avg. score on financial literacy scale	Experience of violence ^a	19. % ever experiencing physical violence in past year
Savings	19. % who have saved in the past year		20. % ever experiencing intimate partner violence in past year
Paid work and income	20. Avg. amount saved in the past year among those who saved		
Assets personally owned	21. % working for cash or in-kind in the past year		
Health assets	22. Avg. reported income in the past year among those who worked for cash		
	23. % who own a bicycle		
Health knowledge	24. % who own a mobile phone		
	25. % Understanding pregnancy risk during menstrual cycle		
Self-assessed health	26. Avg. score on contraceptive knowledge scale		
	27. Avg. score on HIV/AIDS knowledge scale		
Experience of sickness	28. Avg. rating of health status in the past year		
	29. Avg. rating of health status in the past month		
Skills and competencies ^b	30. Avg. reported number of health problems in the past month		
	31. % who can read simple sentence in local language		
Performance on assessments	32. % who can read simple sentence in English		
	33. Avg. score on numeracy assessment		
	34. Avg. score on cognitive assessment		

^a Indicator was added after in 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.

^a Indicators in this domain measured among those aged 15 and older

^b Indicators in this domain measured among those aged 13 and older

3. The independent evaluation design

3.1 A pragmatic approach

The original design for the independent evaluation of the AGEP initiative was based on a pragmatic approach that aimed to systematically address the five evaluation questions (Figure 4). This evaluation design was described

in our Inception Report (December 2012) and was highly focused on the verification of process, quality assurance of the impact evaluation research, validation of the economics evaluation, comparison work with other relevant initiatives, and strategic reflection. Importantly, in keeping with our terms of reference (see Annex 1), we have not sought to conduct an alternative study or duplicate the Population Council's research. Notably, the role of the independent evaluation team has been complemented by the role of the AGEP

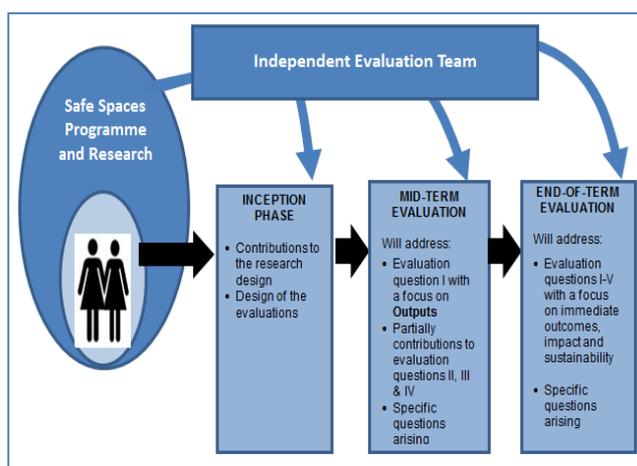
Evidence Scrutiny Committee (ESC). The ESC is responsible for ensuring the AGEP research is conducted in a rigorous and transparent manner, and that evidence generated is robust.⁸

Given that, firstly, the End Term Evaluation comes mid-way through the Population Council's impact evaluation research (although at the end of the AGEP intervention phase) and, secondly, the findings from the Round 3 survey have given some cause for reflection, DFID has asked the Mott MacDonald evaluation team to address the evaluation questions within the framework of a theory-based evaluation. This theory-based approach is intended to guide systematic reflection on the AGEP theory of change based on the current weight of evidence, and to help elicit lessons for future programme design.

3.2 Applying a theory-based evaluation approach

There are multiple perspectives on what constitutes a theory-based evaluation but some common features are summarised in Box 3.

Figure 4: Overview of the role of the independent evaluation team



⁸ See terms of reference for the AGEP Evidence Scrutiny Committee, December 2011. The Committee includes representation from DFID Zambia, DFID's Research and Evidence Division, Population Council, Ministry of Health and the University of Zambia.

Box 3: Key features of a theory based evaluation approach

Theory based evaluations are characterised by:

- **Policy relevance:** focused on addressing not only the question of what has worked, but also why and how it worked in order to produce a policy relevant evaluation.
- **Understanding the transformational relations** between treatment and outcomes, as well as contextual factors and aiming to identify the ‘mechanisms’ that make things happen. This goes from asking whether a programme works to understanding what it is about the programme that makes it work.
- **Having two key parts:** conceptual (developing the causal model or theory of change that underlies a programme, and using this model to guide the evaluation); and empirical (testing this theory of change to investigate how a programme causes intended or observed outcomes).
- **Being issues led**, and therefore, methods neutral.

In order to apply these principles while remaining sighted on the evaluation questions we have followed the steps described in Box 4.

Box 4: Step-wise approach to developing a theory-based evaluation of AGEP

STEP 1: Preliminary conceptual analysis

- Unpack the theory of change:
 - Isolate the underlying hypotheses (or main causal pathways) within the theory of change –ensure they specify process (mechanisms), outcomes (results) and context (for whom, where) constructs
 - Define the assumptions behind each hypothesis
 - Assess the strength of evidence for each hypothesis at baseline

STEP 2: Empirical analysis

- Test *each hypothesis* against implementation experience and impact evaluation findings to date:
 - Review the results: review the quality assured results reported by the Population Council at Round 3, taking into account differential results for younger and older adolescent girls and context (urban/rural).
 - Review the mechanism: review the process evidence to determine whether the interventions have been well implemented and factors that could influence survey findings.
 - Review contextual factors: Review the independent evaluation findings on urban/rural settings to identify factors that could influence survey findings
 - Review assumptions: review the assumptions for the theory of change/logframe to see whether they have been upheld.

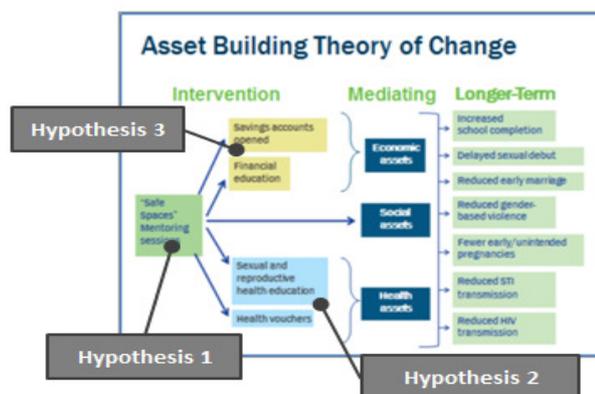
STEP 3: Strength of evidence assessment

- Assess the strength of evidence for the hypothesis at this point (Round 3). Consider:
 - Implications for the theory of change
 - Lessons for future programming
 - Implications for sustainability and scale-up

3.3 Setting up the theory-based evaluation –conceptual analysis

3.3.1 Defining the hypotheses

Box 5: Proposed hypotheses to be tested through the theory-based evaluation



- **Hypothesis 1 (social assets):** Vulnerable adolescent girls who participate in a programme of Safe Space meetings with curriculum-based training from a mentor will acquire social assets (measured by indicators on self-efficacy, autonomy, friends and isolation, safety nets, beliefs on gender and skills and competencies). These social assets will contribute to longer term empowerment impacts (measured by indicators on education, marital relations, pregnancy and births, sexually transmitted diseases and experience of gender based violence).
- **Hypothesis 2 (health assets):** Vulnerable adolescent girls who participate in a programme of Safe Space meetings with a mentor that includes SRH education *and* provision of age-appropriate health vouchers will acquire health assets (measured by indicators on health knowledge, self-assessed health, experience of sickness). These health assets will contribute to longer term empowerment impacts (measured by indicators on education, marital relations, pregnancy and births, sexually transmitted diseases and experience of gender based violence).
- **Hypothesis 3 (economic assets):** Vulnerable adolescent girls who participate in a programme of Safe Space meetings with a mentor that include financial educations *and* opening of bank savings accounts will acquire economic assets (measured by indicators on financial literacy, savings, paid work and income, and personally owned assets). These economic assets will contribute to longer term empowerment impacts (measured by indicators on education, marital relations, pregnancy and births, sexually transmitted diseases and experience of gender based violence).

In keeping with the step-wise approach described in Box X, we have reviewed the AGEP theory of change to determine the key hypotheses that underpin it. Since the theory-based approach is being applied somewhat retrospectively, we have defined the hypotheses through a process of deduction, rather than through a consultative process. However, we are confident that the hypotheses presented below are consistent with the original business case

and Population Council's descriptions of the programme approach. The three hypotheses identified for the theory based evaluation are defined in Box 5 below.⁹

3.3.2 Defining the assumptions

Since our three hypotheses are consistent with the conceptual framework underpinning the AGEP logframe (see Annex 3), we have sourced the assumptions for each hypothesis from the logframe. These assumptions will be reviewed in the course of our analysis and are listed in the table below.

Table 2: Assumptions underpinning the hypotheses for the theory-based evaluation

Hypothesis 1	Hypothesis 2	Hypothesis 3
<ul style="list-style-type: none"> □ Suitable female mentors can be identified in the community □ The most poor and vulnerable girls can be identified and will have sufficient autonomy to be able to attend weekly meetings □ Mentors will be able to successfully deliver the curriculum in such a way that it influences SRH knowledge and decision making □ Programme will be able to have an effect on self-esteem □ Girls willing to report unwanted sex during programme surveys; there is a link between assets and unwanted and/or transactional sex 	<ul style="list-style-type: none"> □ Health workers in post, especially at peripheral rural health facilities □ Support from MOH/MCDMCH and DMOs for voucher scheme □ A range of providers exists and has the potential to provide a package of quality services to adolescent girls □ Vouchers will encourage uptake of services among adolescent girls □ Vouchers will encourage providers to offer quality and adolescent-friendly services 	<ul style="list-style-type: none"> □ Financial institutions will extend services to rural areas and small savers in Zambia □ Financial goals translate into desire for savings accounts □ Poor adolescent girls will be able to save □ Girls are able to save money, especially in rural areas

3.3.3 Strength of evidence at the outset

In the design phase for the AGEP initiative, DFID developed a Business Case¹⁰ that described the rationale for the theory of change and the strength of evidence for each component at that time. It was noted that there was good evidence from the Population Council's experience in Ethiopia, Kenya and Uganda that Safe Spaces initiatives can have positive effects for vulnerable adolescent girls (see Box 6 below).

The Business Case was able to cite a number of studies that provided good or emerging evidence on the benefits of building social, health, economic assets for reducing the vulnerabilities and expanding the opportunities of targeted adolescent girls. However, the Business Case observed that the effect of Safe Spaces on *longer-term* measures of girls' empowerment or sexual and reproductive health had not been formally evaluated. It also

⁹ We note that the Population Council lists a number of statistical hypotheses in its 2014 Statistical Analysis Plan for AGEP; however, these are designed for highly specific statistical testing and serve a different purpose to the conceptual hypotheses behind the theory of change.

¹⁰ DFID Zambia (2011). Business Case for the Adolescent Girls Empowerment Programme, September 2011

observed that evidence on Safe Spaces to date was largely based on urban areas, so the applicability and cost-effectiveness of implementing Safe Spaces in rural areas of Zambia was uncertain.

Box 6: Evidence cited in the AGEP Business Case to support Safe Spaces

In Ethiopia, rural girls aged 10-14 years who participated in the *Berhane Hewan* programme that combined girls' groups, mentors and health training were three times more likely to be in school and less likely to be married. Girls who were sexually experienced were more likely to be using contraceptives. However, communities were also sensitised about the risks of early marriage and economic incentives were provided to families who did not marry off their daughters.¹¹ The *Biruh Tesfa* programme, which aimed to address the social isolation of girls in Addis Ababa, increased social participation and the likelihood of HIV counselling and testing.¹²

In Kenya and Uganda, positive changes were demonstrated in social networks and mobility, independence and self-confidence, gender norms, financial literacy, use of bank services and savings behaviour. Programme participants had greater understanding of family planning, longer term financial goals, greater understanding of money management and higher levels of savings. Also in Kenya, girls who had participated in the TRY programme, which involved group meetings, loans and savings, had greater ability to refuse sex and insist on condom use, compared to girls in a control group.¹³

The evaluation team has revisited this evidence and notes that, although the weight of evidence was moderately strong,¹⁴ the Safe Spaces initiatives mentioned were also accompanied by important additional components, such as community mobilisation, advocacy work, economic incentives and support of staying in school (Ethiopia) and micro-finance opportunities (Kenya). Moreover, the evidence for other AGEP mechanisms (namely health vouchers and opening of bank accounts for adolescent girls) was limited.

¹¹ Erulkar A and Muthengi E, 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

¹² Population Council, *Transitions to Adulthood Brief no. 21* January 2011

¹³ Erulkar A and Chong E, 2005. Evaluation of a savings and micro-credit program for vulnerable young women in Nairobi. Population Council

¹⁴ Our terminology here is based on the DFID *How to Note on Assessing Strength of Evidence* (2014).

4. Methodology

4.1 Overview

This section describes the methodology we have used for collecting data for the End Term Evaluation work. The section begins by describing the data collection methods we used. We then describe our sampling strategy and approach to data analysis for each part of the evaluation. We conclude this section with a summary of limitations to the methodology and the mitigating action taken.

4.2 Data collection methods

In keeping with the theory based evaluation approach and the requirements of the evaluation questions, we have used a mixed method approach to data collection that has allowed us use triangulation of data sources, data types and data analysis to build up reliable evaluation findings. Our data collection methods included:

- Key informant interviews with knowledgeable individuals
- Semi-structured interviews with programme and research stakeholders at national and sub-national levels
- Direct observation of operational context in urban and rural areas
- Desk reviews of programme records, including progress reports, monitoring databases and financial reports
- Desk reviews of AGEP research records, including research datasets and reports (covering survey data, qualitative data and economic research data)
- Desk reviews of comparison study records, including research datasets and reports
- Desk review of design documents, the evidence base on adolescent girls' empowerment and a selection of the Population Council's publications and communication products.

We developed a customised tool-kit for data collection.¹⁵ We collected field data over the period June-July 2016 using a combined team of international evaluators and experienced Zambian fieldworkers who were fluent in relevant local dialects.

4.3 Sampling strategy

4.3.1 General approach

We used a mix of sampling approaches. These were mostly based on purposeful and convenience sampling that aimed to capture a range of stakeholder experiences and perspectives in urban and rural settings; however, where possible, we used random selection to reduce selection bias.

¹⁵ This tool-kit is available upon request.

The limitations of the sampling strategy are acknowledged and are discussed in Section 4.7. An important constraint was that programme implementation had concluded some months before the evaluation visit (phased close-out took place over the period December 2015-March 2016). This meant that most AGEPS staff had moved on or were no longer under contract, and many mentors and community-based stakeholders had dispersed. As a result, many of our interviews were, in practice, based on convenience sampling. Our sample size was limited, and there was some risk of selection bias. To mitigate we used triangulation techniques and, as far as possible, aimed for 'empirical saturation'.¹⁶

4.3.2 Sampling of Master Sites

To select Master Sites for the evaluation fieldwork, we used a combination of purposeful sampling and random selection. Purposeful sampling from the 10 Master Sites was based on the following criteria:

- A balanced mix of urban and rural sites;
- Sites that had most recently closed (to increase the likelihood of locating community stakeholders and accurate recall);
- A primary focus on sites that were not visited as part of the Mid-Term Evaluation, so that our findings could contribute to a more comprehensive assessment;
- Feasibility of reaching the sites within the time and resources available.

Master Sites were grouped around these criteria then four sites were randomly selected. The final sample of sites for visits included two urban sites (Misisi/Chawama and Chipata/Chazanga (Lusaka) and two rural sites (Masaiti A and Kapiril Mposhi).

In each of the selected Master Sites, we randomly selected one CSA from each implementation treatment arm (T1-T3 sites) to conduct interviews. For CSAs offering health care vouchers,¹⁷ we visited at least one participating health facility. For CSAs offering the savings account, we visited the nearest NatSave branch.

4.3.3 Sampling for interviews

Sampling of individuals and Safe Spaces groups for interview was based on a combination of random, purposeful and convenience sampling. For semi-structured interviews with individuals at national and district levels (such as government officials, AGEPS personnel, bank staff and health service providers), we generally used purposeful sampling. Sampling of mentors and girls from Safe Spaces groups was based on a combination of random and convenience sampling. For sampling of parents and community representatives, we mainly used convenience sampling (see Annex 5 (pp.28-29) for additional notes on the evaluation methodology).

¹⁶ See Fusch P and Ness R. 2015. Are We There Yet? Data Saturation in Qualitative Research in *The Qualitative Report* 2015 Volume 20, Number 9, How To Article 1, 1408-1416 <http://tqr.nova.edu/wp-content/uploads/2015/09/fusch1.pdf>

¹⁷ Since implementation of this component was delayed, the health voucher component was extended for a year after programme end, so vouchers were still valid in T2 and T3 sites.

4.4 Ethical compliance

With the assistance of DFID and the Population Council, we secured the required letters of introduction and official approvals before conducting interviews with stakeholders in the government and private sectors, and at the community level.

We ensured informed consent forms were signed for all community-level interviews. We also ensured that our interview teams were trained to observe the highest standards of ethical and professional research practice, and provided appropriate supervision to ensure these standards were observed in practice.

4.5 Data analysis

Data analysis for the End Term Evaluation fell into four main task areas. These task areas were: a) quality assurance of the AGEP research data and analysis; b) analysis of interviews and direct observation data obtained from fieldwork; c) technical review of the AGEP cost-effectiveness study and value for money performance; and d) analysis of the results and costing data from the comparison studies. Each task area required analysis and triangulation of quantitative and qualitative data generated using the data collection methods listed in Section 4.2 above. The analytical approach used for each of task area is summarised in Table 3. More detailed information on the steps we used for quality assurance of the AGEP research is included in Annex 6, and more detailed information on how we analysed data from field interviews and direct observation is included in Annex 5. Additional information on analysis steps used for the economic evaluation and comparison work is included in the respective parts of this report.

Table 3: Approach to data analysis for each task area

Task area	Data type	Analysis method
a. Quality assurance of AGEP research data and analysis	<ul style="list-style-type: none"> □ Quantitative data sets from the AGEP surveys and implementation data □ Population Council reports on the analysed survey data 	<ul style="list-style-type: none"> □ Data quality assurance review for completeness, consistency and credibility (see Annex 6) □ Re-run of analysis for key results indicators (see Annex 7) □ Technical review of the research analysis plan and survey reports
b. Analysis of fieldwork data from stakeholder interviews and direct observation (sampled sites)	<ul style="list-style-type: none"> □ Qualitative narrative data from stakeholder interviews and focus group discussions □ Direct observation notes 	<ul style="list-style-type: none"> □ Thematic 'closed coding' of interview transcripts □ Transfer of coded text to Excel matrix □ Synthesis and triangulation by theme/sub-theme, type of stakeholder and location (see Annex 5) □ Review of preliminary findings by data collectors

Task area	Data type	Analysis method
c. Validation and quality assurance of the AGEP cost-effectiveness study and an assessment of value for money (VFM) performance	<ul style="list-style-type: none"> □ Financial and costing data sets and records □ Population Council reports on cost-effectiveness analysis and survey findings □ DFID annual review reports □ Qualitative data from key informant interviews with AGEP finance staff and economist +direct observation 	<ul style="list-style-type: none"> □ Technical review of the cost-effectiveness study design and analysis plan □ Data quality assurance review of the costing data for completeness, consistency and credibility □ Re-run of analysis for key results indicators □ Thematic review of the qualitative data + financial management data files
d. Analysis of the data from the comparison studies	<ul style="list-style-type: none"> □ Quantitative data sets from the studies for comparison □ Study reports on the analysed survey data □ Secondary literature on approaches to adolescent girls empowerment 	<ul style="list-style-type: none"> □ Data mining: Alignment and systematic comparison for key results indicators to determine most successful approaches (disaggregated by indicator) □ Review of results against costing information □ Technical review of lessons from secondary literature

4.6 Note on the economics evaluation

As indicated above, the evaluation team has been asked to validate and quality assure the Population Council's costing study, and to use the findings as the basis for comparison work. The sections in this report on economic evaluation and value for money are largely based on the Economic Evaluation End-of-Programme Report from the Population Council.¹⁸ This document analysis was complemented by a number of interviews in Zambia with Population Council staff and the economist herself (Fiammetta Bozzani) to gain an understanding of both the resources used in AGEP and the methodology for conducting the costing. Interviews with the programme's accountant and economist clarified how financial data was re-classified into economic categories.

The second draft of the economic evaluation report addressed many of the detailed questions from Mott MacDonald about the first draft. However, some works still need to be done to: a) make sure the economic sections are fully aligned to the Round 3 survey findings (e.g. there is consistency in references to indicators and statistical significance) and b) there is additional explanation given on some numbers (such as the costs to mentors) -see Section 6.

¹⁸ This second draft (10th August 2016) was shared with us as a stand-alone document and updated the economics section of the Population Council's Research and Evaluation Mid-Term Technical Report (July 2016).

4.7 Limitations of the independent evaluation methodology

The main limitations to the methodology for the End Term Evaluation are summarised below, along with the implications for the evaluation work and the mitigating action taken.

Table 4: Main limitations to this evaluation

Limitation	Implications	Mitigating actions
1. Size of samples and use of non-probability sampling for interviews	Stakeholder interviews may not be fully representative and there may be some selection bias	<ul style="list-style-type: none"> □ Use of triangulation and data saturation techniques □ Balanced mix in purposeful sampling to ensure evaluation finding reflect urban and rural experience
2. Programme sites closed by time of evaluation visit	Findings from interviews based on retrospection; may be some memory effects and increased risk of self-selection biases	<ul style="list-style-type: none"> □ Use of triangulation and data saturation techniques □ Cross-referencing to the Mid-Term Evaluation
3. Research is ongoing (independent evaluation findings based on survey rounds 1-3 only)	End Term Evaluation findings do not capture impact level results or results delayed by adolescent transitions	<ul style="list-style-type: none"> □ Timeframes for the independent evaluation clearly stated and considered in the analysis and reflections on the theory of change
4. Draft report on the findings of the Round 3 survey received after End Term data collection; framework for the theory-based evaluation applied retrospectively	Small gaps in the evaluation data collected; data collection and enquiry based on preliminary findings	<ul style="list-style-type: none"> □ Evaluation data and analysis revisited to interpret data within context of the Population Council Mid-Term Technical Report □ Evaluation data revisited to apply the framework of a theory-based analysis □ Evaluation findings only based on the evidence available
5. Cost-effectiveness work ongoing (final outcome and impact level results not yet available)	Assessments of cost-effectiveness can only be preliminary	<ul style="list-style-type: none"> □ Focus on validating the technical quality of study design and preliminary analysis, and verifying programme costing data
6. Evaluation team did not visit control sites	Counterfactual analysis for the independent evaluation is based on Population Council's survey findings only	<ul style="list-style-type: none"> □ Data quality assurance extended to survey data and analysis from control sites; potential for spill-over reviewed
7. Comparison work based on published findings only	Published findings for the comparison studies may not	<ul style="list-style-type: none"> □ This limitation explicitly acknowledged and

Limitation	Implications	Mitigating actions
	contain the most up-to-date analysis or analysis on sustained effects or a robust cost-effectiveness analysis	considered in the comparison work

In addition, two particular **constraints** have been that:

- We have not yet had access to data analysis from the external control group -this means we have not been able to fully assess the effects of spill-over;
- We have not yet received data on the mediating effects of mentor quality¹⁹ –this could have important explanatory value in explaining differential results; it could also provide a basis for further operational learning.

¹⁹ This was not included in the account of the Population Council's research on mentor quality in the AGEP Research and Evaluation Mid-Term Technical Report (July 2016).

5. Findings Part 1: Theory-Based Evaluation

5.1 Findings from the empirical analysis of Hypothesis 1

Hypothesis 1 (social assets): Vulnerable adolescent girls who participate in a programme of Safe Space meetings that includes curriculum-based training from a mentor will acquire social assets (measured by indicators on self-efficacy, autonomy, friends and isolation, safety nets, beliefs on gender and skills and competencies). These social assets will contribute to longer term empowerment impacts (measured by indicators on education, marital relations, pregnancy and births, sexually transmitted diseases and experience of gender based violence).

5.1.1 Section overview

In this section, we review the empirical evidence for the AGEP hypothesis on social assets (Hypothesis 1) following completion of the two-year intervention phase. We begin with an in-depth review of the Round 3 survey results relating to social assets for urban and rural girls by age cohort.

Next, we review the evidence from the evaluation research to assess whether it is consistent with the survey findings and to identify areas where it adds explanatory value. Drawing on a 'realist evaluation' paradigm, we place particular emphasis on reviewing the *mechanism* (how the intervention was implemented) and the effects of *context* to trace the linkages to *outcomes* in order to contribute to an understanding of "what has worked for whom, in what circumstances and why?"²⁰ In keeping with the theory-based evaluation approach, we then review the assumptions that were posited for this component of the theory of change to assess whether they have been upheld.

Finally, in the concluding summary for this section, we assess the current strength of evidence for Hypothesis 1, and review the implications for the theory of change, lessons for future programming and any implications for sustainability and scale-up.

5.1.2 Review of Round 3 results on social assets

The evaluation team has reviewed the Population Council's findings for survey Rounds 1-3 (R1-3) for each of the outcome indicators relating to social assets as well as the evidence for contributions to impact.

In the Round 3 survey report,²¹ the Population Council reported on its ITT, DiD analysis comparing Round 1 (baseline) and Round 3 results against internal controls. For the purposes of this theory-based evaluation of Hypothesis 1, we conducted a quality assurance review of the

²⁰ Pawson, R. & Tilley, N. (1997). *Realistic Evaluation*. London: Sage. See also: http://betterevaluation.org/en/approach/realist_evaluation

²¹ Population Council (2016). *Adolescent Girls Empowerment Programme, Research and Evaluation Mid-Term Technical Report*, July 2016

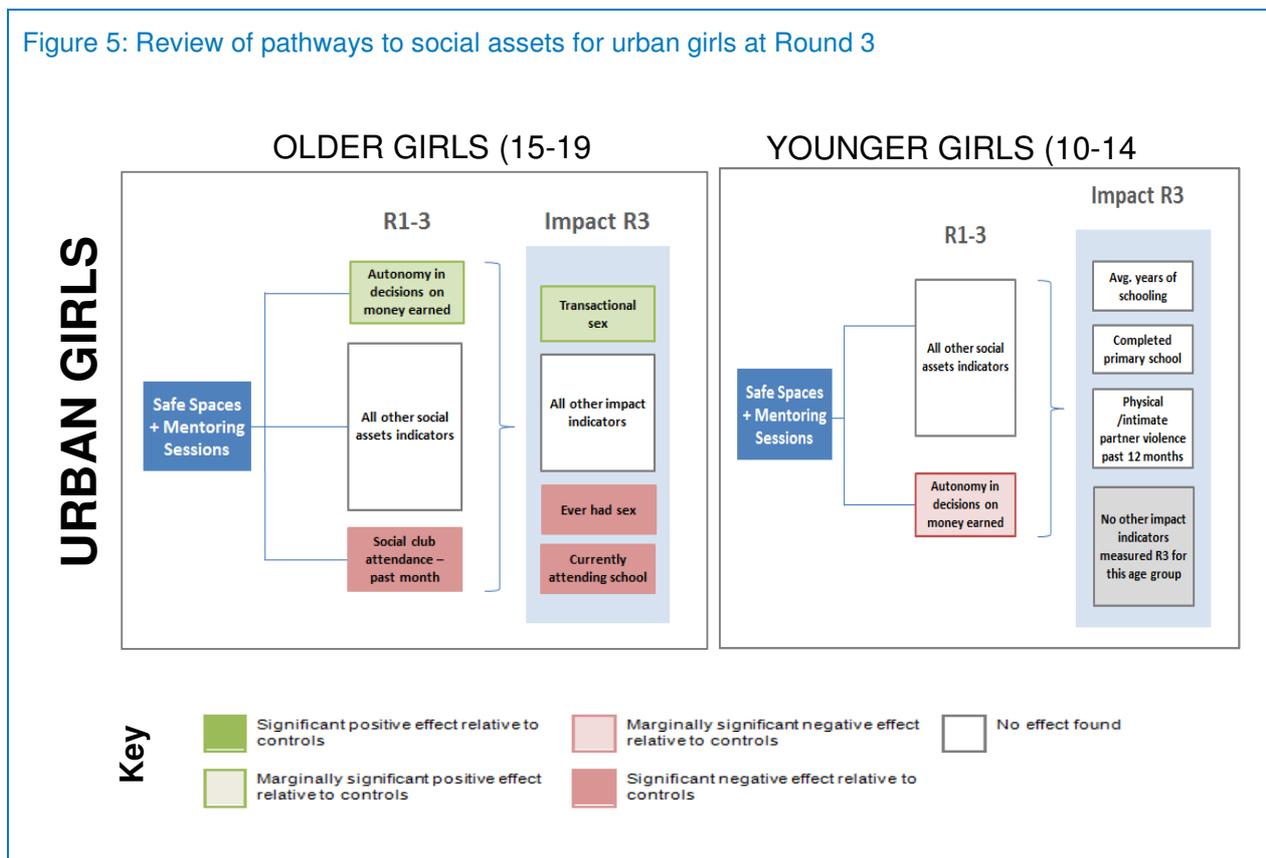
Population Council’s data and analysis for indicators relating to social assets. Next, using the ITT, DiD analysis, we isolated the effects for rural and urban girls by age cohort.²²

Our peer review indicates that the quality of the Population Council’s data and analysis relating to social assets is reliable (see Annex 7 for our quality assurance findings). We are, therefore, confident in using the data for the additional analysis relating to Hypothesis 1.

The analysis tables we have extracted from the Population Council’s work are contained in Annex 4. The main findings from our synthesis work on context and age cohort effects relating to social assets are summarised in the diagrams below. For the purposes of this review we have focused on effects found from Round 1 (the baseline) and Round 3 (at the end of the programme implementation phase). In the diagrams below we have captured marginally significant effects ($p < 0.10$), as well as significant effects ($p < 0.05$), to help us identify any transitional effects in the indicators underpinning the theory of change pathway for social assets.

Review of results for urban girls at Round 3

Figure 5: Review of pathways to social assets for urban girls at Round 3



²² Although the Population Council has also conducted a Treatment of the Treated (ToT) analysis on girls who have completed 52+ week of Safe Spaces meeting, we have agreed with DFID that, for the purposes of this theory-based evaluation, we will focus on the results from the ITT analysis.

Figure 5 shows that there have been differential effects from efforts to build social assets through Safe Spaces and mentoring for older and younger urban girls. For **older urban girls** (15-19 years), the ITT, DiD analysis comparing the baseline and 24-month evaluation data (R1-3) shows that there has been a *marginally* significant positive effect ($p < 0.10$) on 'the percentage of girls who can jointly or solely make decisions [autonomy] with regard to money earned'. For most other indicators of social assets there were no effects found in the ITT, DiD analysis for R1-3. However, for the indicator 'percentage attending any social groups/clubs within the past month' there was a significantly negative effect for older urban girls over Rounds 1 to 3 –perhaps reflecting declining participation by this population segment (see below).

Since the theory of change posits that building of social assets will contribute to positive gains in longer-term impact indicators, we also looked at these indicators for signs of change among older urban girls. At the impact level, the ITT, DiD analysis for R 1-3 did show a *marginally* significant positive effect on reports of transactional sex ($p < 0.10$) by older urban girls; however, there was also a significantly negative effect on the 'percentage who had ever had sex' and the 'percentage of girls currently attending school' relative to the internal controls. These negative effects are cause for concern and may require further investigation.

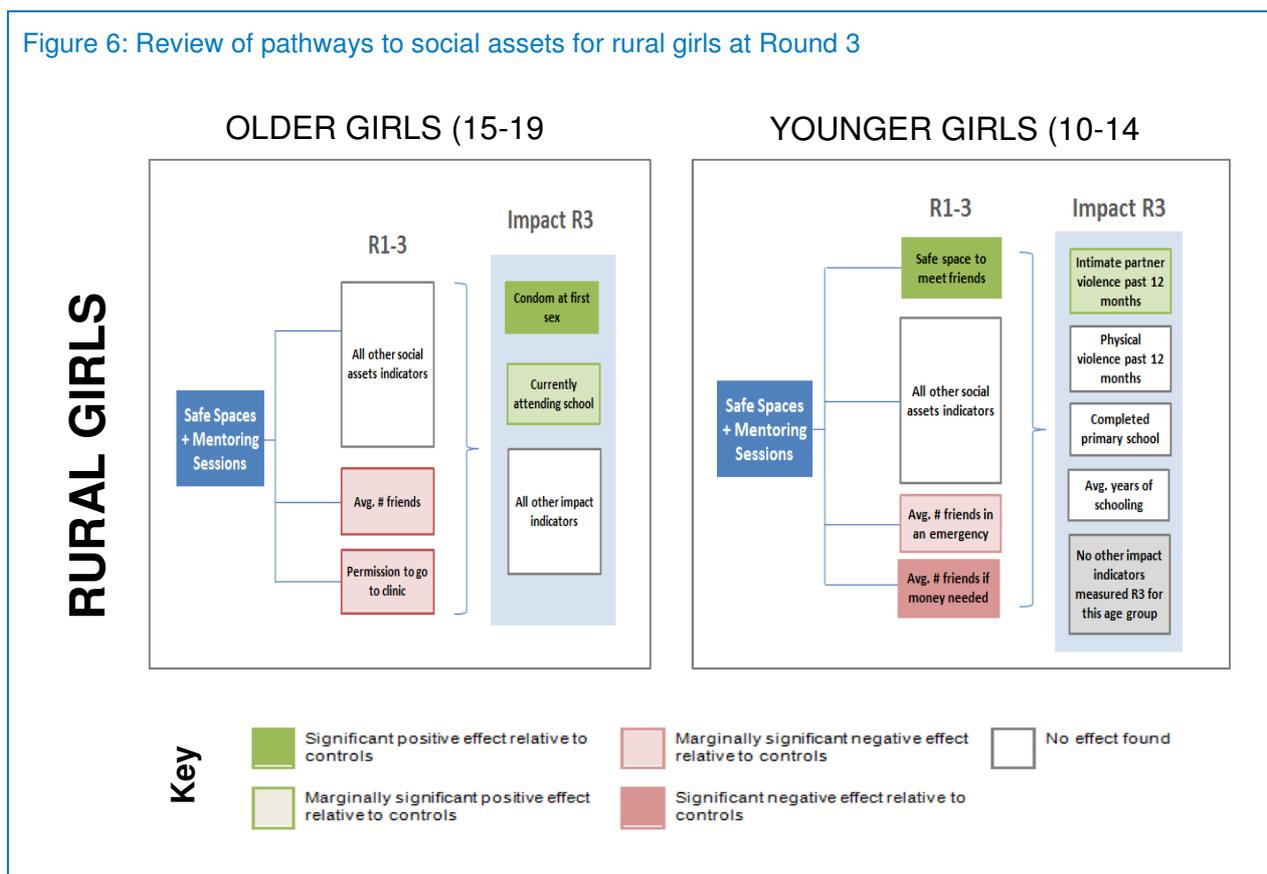
For **younger urban girls** (10-14 years), the ITT, DiD analysis for R1-3 found mostly no effects for the majority of social asset indicators. The exception was a *marginally* significant negative effect ($p < 0.10$) on autonomy in decisions on money earned. At the impact level, most indicators were not relevant to this younger age group over Rounds 1-3, so were not measured. However, for the four impact indicators that were measured (relating to school attendance and recent experiences of physical or intimate partner violence) no effects were found for younger urban girls over Rounds 1 to 3 relative to internal controls.

5.1.2.1 Review of results for rural girls at Round 3

Figure 6 below shows that there have also been differential results from efforts to build social assets among older and younger rural girls. For **older rural girls**, the ITT DiD analysis for R1-3 showed no effects for the majority of social asset indicators. However, for indicators on 'percentage who agree that permission to go to the clinic is not a problem' and 'average number of friends' there was a *marginally* significant negative effect ($p < 0.10$). Once again, these negative effects for older rural girls, while marginal, may warrant further investigation. At the impact level, the ITT DiD analysis for R1-3 showed a significantly positive effect on the indicator for 'percentage of girls who used a condom at first sex' ($p < 0.05$) and a *marginally* significant positive effect on the 'percentage of girls currently attending school' ($p < 0.10$). There were no effects found for other impact level indicators for older rural girls.

For **younger rural girls**, the ITT, DiD analysis for R 1-3 found a significant positive effect on the 'percentage of girls having a safe space to meet with friends'. For the majority of other social asset indicators, there were no effects found in this analysis. However, there were exceptions for measures relating to 'average number of friends who can be counted on if money is needed' and 'average number of friends who can be counted on in an emergency', for which significantly negative ($p < 0.05$) and *marginally* significant negative ($p < 0.10$) effects were found respectively. Once again, this negative effect on two 'safety net' indicators for younger rural girls requires investigation. At the impact level, there was no effect found for most of the indicators measured for this age group but a *marginally* significant positive effect was found on the 'percentage of girls who had experienced intimate partner violence in the past 12 months'.

Figure 6: Review of pathways to social assets for rural girls at Round 3



5.1.3 Review of the mechanism for building social assets

For this review, we have drawn on the End-Term Evaluation fieldwork, as well as findings from our desk review of programme records and reports, and findings from our Mid-Term Evaluation. We have examined both the safe spaces and the mentorship components of the intervention and considered both achievements and challenges.

5.1.3.1 Achievements

As indicated in Hypothesis 1, the principal strategy for building social assets was implementation of weekly Safe Space meetings among girls who scored high using a vulnerability index. Selected girls would attend weekly Safe Space meetings facilitated by trained local mentors. Meetings would include training based on standardised health, life skills and financial education curricula.²³ Selected girls were expected to attend 52+ meetings over the two year implementation period and received small prizes at every fifth meeting attended to encourage continuous participation. Girls' attendance was captured and reported using Open Data Kit (ODK) software on mobile phones provided to the mentors. Mentor performance was monitored and supported by Population Council/YWCA Local Coordinators.²⁴

The evaluation research confirms that, overall, the Population Council and its partners did a good job in implementing this component of the programme as designed. The evaluation team's

²³ In addition, 50% of the girls received nutrition education.

²⁴ Population Council. 2016. Research and Evaluation Mid-Term Technical Report. pp10-13

assessment of performance against 2016 logframe milestones is shown in Annex 8 and indicates that key milestones relating to Safe Spaces implementation were satisfactorily met. Key achievements in implementing the Safe Spaces intervention included:

- **Safe Spaces was fully implemented at scale:** The Population Council, in partnership with YWCA, delivered Safe Spaces at scale in urban and rural settings in 120 CSAs across 10 Master Sites in four provinces. Safe Spaces meetings were held weekly with girls in two age cohorts (10-14 and 15-10 years), mostly in groups of 20-30 girls. Almost 10,000 girls (9,536) joined Safe Spaces meetings; between August 2013 and March 2016, 405 groups met in a total of 40,884 meetings.
- **Well-designed curricula:** Our evaluation research confirmed that Safe Spaces meetings included short trainings for girls based on three technically well-designed and piloted curricula, covering health, nutrition, life skills and financial education. All educational sessions were designed to be interactive and inclusive. The evaluation research found that the **financial education curriculum** was especially popular across all sample segments. However, older, urban girls were able to describe more opportunities to apply this learning through access to resources and engagement in small-scale market activities. This qualitative finding could help to explain positive survey findings on autonomy in financial decision-making and reduced transactional sex among older urban girls.
- **Well-supported mentors:** The evaluation research confirmed that, in general, AGEP mentors were carefully selected, well trained and supported by Local Coordinators through regular meetings and monitoring visits. Overall retention rates were high, with 88% of Safe Spaces groups having the same mentor for two years. Interviews with mentors also pointed to **secondary benefits** for mentors themselves in the form of knowledge, skills and status gains, and sometimes career progression.
- **Compelling stories of positive change** emerged from evaluation interviews with the AGEP girls, mentors and parents regarding the positive effects of the Safe Spaces meetings. For community stakeholders, the benefits of Safe Spaces were mostly described in terms of improvements in girls' behaviour and improvements in school performance (due to improved attitudes of the girls and complementarities with the school curriculum). We note these 'participatory indicators' of success are not being currently being measured through the AGEP research.²⁵ In addition, we noted that a number of mentors (especially in rural areas) gave examples of girls **challenging abusive relationships**. Mentors suggested this was due to girls having a better understanding of acceptable behaviour, increased awareness of gender rights, and increased access to responsible adults. These accounts could go some way to explaining the positive survey findings on reduced intimate partner violence among younger rural girls.

5.1.3.2 Challenges

The independent evaluation research also identified several implementation challenges that could be factors in explaining weaker Round 3 results.

²⁵ See <https://www.ids.ac.uk/files/dmfile/PB12.pdf>

- **Variable or reducing participation:** By October 2015, of girls who had attended at least one meeting, attendance had declined to 63%²⁶ and mentors reported that attendance was often inconsistent. The Population Council reports that, of girls in the research, only a third of those invited continued to attend meetings over the longer term (52+) –these girls tended to be the younger girls, living in rural areas, in school, with higher numeracy skills, and the biological daughter of the head of household.²⁷ Annex 9 (p. 50) contains a synthesis of reasons for girls dropping out. Our evaluation research confirmed high levels of drop out among urban girls and older girls –often because they had competing obligations or opportunities; there also appeared to be poor retention of out-of-school girls and girls with disabilities. These participation patterns may go some way to explaining the weak results around social club attendance and friendship networks in some segments of the sample.
- **The system of attendance passports and prizes:** This system introduced after piloting to encourage sustained attendance. It was clearly important to monitor attendance carefully for both management and research purposes. However, the evaluation research found that there was widespread discontent about the quality of the prizes and how they were distributed.²⁸ Urban girls with better access to commodities were especially critical about the quality of the prizes. Several girls, parents and mentors (in urban and rural sites) described occasions where girls ceased attending –temporarily or permanently- due to grievances relating to the prizes.
- **Ensuring age appropriateness:** In the course of our evaluation research, a number of mentors and parents raised concerns about the age appropriateness of the curriculum for the youngest girls (10-12 years); some gave examples of younger girls finding the information distressing or using the information inappropriately. Mentors and Local Coordinators also described the distinctive challenges faced by older married girls, or girls with children and suggested that there needed to be separate groups for these girls. We suggest these observations could have implications for results relating to knowledge uptake, Safe Space attendance and establishment of friendship networks.
- **Maintaining community buy-in:** From our evaluation research in both urban and rural areas it was clear that parents, guardians, husbands and other community stakeholders can be important gatekeepers in the lives of adolescent girls. In interviews with mentors and community focus group discussions we were given several examples of parents or husbands blocking the participation of girls in Safe Space meetings –often due to misconceptions about the motivations of the programme implementers.²⁹ Our evaluation research highlighted the critical importance of maintaining community support and the potential added value of engaging community stakeholders.³⁰

²⁶ This also includes girls who may have moved away, but had been participating through April 2015.

²⁷ Population Council (2016). Adolescent Girls Empowerment Programme, Research and Evaluation Mid-Term Technical Report, July 2016: 13.

²⁸ This issue was raised in the Mid-Term Evaluation. We heard complaints about the prizes in every Master Site we visited and there were particular disputes about the final prizes in Masaiti A (related to the accuracy of the ODK records on attendance).

²⁹ Our research confirmed by the Population Council's own qualitative research on this theme, with some community stakeholders being concerned about 'Satanism', 'grooming' and discussion of topics, such as termination of pregnancy.

³⁰ For example, some of the mothers we interviewed in rural Kapiri Mposhi were keen receive training like the girls so they could support them better and teach other family members; they were also eager to help their daughters save and to work with them to start small businesses.

5.1.4 Contextual factors in building social assets

In the course of the evaluation field visits we identified a number of differences between rural and urban contexts that could influence the effectiveness of the AGEP intervention. These differences are summarised in the table below.

Our interviews with girls and key informants suggested that it is common for girls to move between urban and rural households during holidays and in times of economic hardship. Rural-urban differences within processes of modernisation and globalisation in Zambia, as well as the migration patterns of young women, are also described in several social studies.³¹ An understanding of these factors could help to explain some of the differential results among girls in urban and rural areas.

We note that the AGEP curriculum has not been explicitly adapted to urban and rural contexts in Zambia (although mentors were encouraged to give examples from their local context). We suggest that more specific engagement with urban and rural contexts and movement between them could help increase the relevance of the curriculum and the link to the reality of girl's lives.

Table 5: Summary of key features of urban and rural contexts most relevant to the social assets of adolescent girls

Urban areas	Rural areas
Mixed communities –exposure to multiple belief and value systems	Closer community and family ties –less diversity in belief/value systems
Greater exposure to processes of modernisation	Greater exposure to traditional systems and structures
More income-generations opportunities	Multiple domestic, childcare and agricultural responsibilities; food security a priority
More and diverse recreational opportunities	Fewer recreational opportunities but generally strong church/choir participation
Higher exposure to other relevant NGO/FBO/government activities	Lower exposure to other relevant NGO/FBO/government activities
Higher exposure to alcohol and substance abuse	Girls may have access to ribbon development along the line of rail/road and semi-urban contexts with truckers, markets and migrant labour (especially in mining areas and around plantations)
Greater availability of commodities	
Better infrastructure and access to services	Limited reliable access to electricity, potable water; rural roads can be impassable in rainy season. Often long distances to attend primary school, health services, banks etc. Secondary school

³¹ See, for example, Evans, A. (2016) Urban Social Change and Rural Continuity in Gender Ideologies and Practices [in Zambia]. Paper presented at 'Gendering the City' Annual Meeting of the Association of American Geographers, March 2016. **Also:** Evans, A. (2014) 'Women Can Do What Men Can Do': the Causes and Consequences of Growing Flexibility in Gender Divisions of Labour in Kitwe, Zambia. *Journal of Southern African Studies*. 40:5, 981-998.

Urban areas	Rural areas
	attendance often requires hostel accommodation (with associated costs and risks)
Greater exposure to the media, internet, social networking and 'youth culture'	Some/growing access to mobile phones, radio and TV; little access to internet

5.1.5 Review of assumptions

Table 6 below provides a review of the assumptions listed in the logframe³² that are relevant to the safe spaces component. These assumptions relate to Outputs 1 and 3 and outcome level results in the logframe.

As shown, in Table 6, it remains uncertain whether all assumptions have been fully upheld and, as indicated in our comments, this could have implications for the effectiveness of the Safe Spaces component.

Table 6: Review of logframe assumptions relevant to building of social assets

Logframe assumptions	Holds	Comment
Assumptions for logframe Outputs 1 and 3		
<input type="checkbox"/> Suitable female mentors can be identified in the community	✓	<input type="checkbox"/> Although suitable mentors were successfully recruited in each CSA, our interviews/observations pointed to some variability in mentor knowledge, experience and skills.
<input type="checkbox"/> The most poor and vulnerable girls can be identified and will have sufficient autonomy to be able to attend weekly meetings	✓ (mostly)	<input type="checkbox"/> Although girls were systematically selected against using vulnerability index, the need to meet recruitment targets meant there was greater variance in levels of vulnerability among rural girls (with possible effects on group dynamics). <input type="checkbox"/> We noted that parents, relatives, husbands and other 'community gatekeepers' can affect the autonomy of the girls to participate in programme activities.
<input type="checkbox"/> Mentors will be able to successfully deliver the curriculum in such a way that it influences SRH knowledge and decision making	? (unclear)	<input type="checkbox"/> Although mentors received good training and support, we found some variability in mentor skills. <input type="checkbox"/> Results on SRH knowledge and decision-making were not conclusive at the end of the intervention.
Relevant assumptions at the Outcome level		
<input type="checkbox"/> Programme will be able to have an effect on self-esteem	?	<input type="checkbox"/> Results on self-esteem were not conclusive at the end of the intervention (Round 3) for any sample segment.

³² Based on the March 2016 version of the AGEP logframe.

<ul style="list-style-type: none"> ☐ Girls willing to report unwanted sex during programme surveys; there is a link between assets and unwanted and/or transactional sex 	<p>?</p>	<ul style="list-style-type: none"> ☐ Survey data <i>suggests</i> girls are willing to report unwanted sex in AGEP surveys. ☐ Results for older urban girls (ITT, DiD analysis, R1-3) point to a <i>possible</i> link between resources and transactional sex (see Figure 5)
<ul style="list-style-type: none"> ☐ Condoms are available to adolescent girls, especially in rural areas 	<p>?</p>	<ul style="list-style-type: none"> ☐ Uncertain –the evaluation research indicated this might not be the case, especially in rural areas (due to lack of resources, social constraints and distance to providers); however, we note the positive result for older rural girls on condom use.
<ul style="list-style-type: none"> ☐ Complementary programmes address wider gender discrimination and social and cultural norms that affect girls' ability to use the assets they acquire 	<p>?</p>	<ul style="list-style-type: none"> ☐ There appears to be no up to date or systematic mapping of other initiatives in each CSA. We observed that girls in urban areas are potentially exposed to the activities of multiple NGO/FBOs, youth groups and clubs, as well as government programmes. None of the AGEP girls we interviewed reported direct involvement in other initiatives, but there may have been an effect on internal controls. ☐ There are programmes working on other approaches to adolescent girls' empowerment in other areas (see Annex 11) but these are unlikely to affect the structural or normative environment for AGEP girls.
<ul style="list-style-type: none"> ☐ Complementary programmes work with men and boys 	<p>?</p>	<ul style="list-style-type: none"> ☐ As above

5.1.6 Hypothesis 1: Concluding summary

5.1.6.1 Strength of evidence for Hypothesis 1 at Round 3

From the ITT, DiD analysis for R1-3, there is evidence that Safe Spaces produced a significant positive ($p < 0.05$) effect for younger rural girls in having a safe space to meet friends, and for older rural girls on condom use at first sex. However, for the majority of indicators on social assets, after two years of programme implementation, there was no evidence of effects. Of some concern were the significant negative effects on impact level indicators relating to sexual debut and school attendance for older urban girls. Evidence from the evaluation review of process suggests that, although the Population Council has done a commendable job in implementing the Safe Spaces model *as designed*, there is scope for strengthening several aspects of the approach to increase its effectiveness in intermediate outcome level results on social assets.

Since, at this stage, the evidence for building social assets remains inconclusive at the outcome level, it remains difficult to trace a credible pathway to impact level results, and thus confirm Hypothesis 1 in the AGEP theory of change.

5.1.6.2 Implications for the theory of change

We have observed that several of the assumptions underpinning the theory of change remain questionable. It might, therefore, be necessary to modify the theory of change to accommodate any assumptions that do not hold true. Based on the differential effects from building social assets, there may be value in developing a more disaggregated theory of change for different segments of the adolescent girl population.

5.1.6.3 Lessons for future programming

Lessons arising from our review of this programme component are described in Section 8.6 of this report. We suggest that key lessons for future programming relate to the need to: customise the curriculum to different contexts and age cohorts; improve strategies for motivating attendance and increasing inclusivity; better support the role of mentors; and improve community engagement.

5.1.6.4 Implications for sustainability and scale-up

Since the longer-term or sustained effects of building social assets for AGEP girls will not be established until the Round 5 survey has been completed, it is perhaps premature to discuss issues of sustainability and scale-up. However, some considerations are discussed in Annex 10 (p.56). Key considerations for the Safe Spaces component include the need to establish the scalable model and demonstrate that durable social assets results can be achieved in different implementation settings within acceptable timeframes.

5.2 Findings from the empirical analysis of Hypothesis 2

Hypothesis 2 (health assets): Vulnerable adolescent girls who participate in a programme of Safe Space meetings with a mentor that includes SRH education *and* provision of age-appropriate health vouchers will acquire health assets (measured by indicators on health knowledge, self-assessed health, experience of sickness). These health assets will contribute to longer term empowerment impacts (measured by indicators on education, marital relations, pregnancy and births, sexually transmitted diseases and experience of gender based violence).

5.2.1 Section overview

In this section, we review the empirical evidence for the AGEP hypothesis on health assets (Hypothesis 2) following completion of the two-year intervention phase.

Our approach will be structured in the same way as the review of Hypothesis 1, beginning with an in-depth review of the Round 3 survey results relating to health assets for urban and rural girls by age cohort and then reviewing the evidence from the evaluation research to assess whether it is consistent with the survey findings and to identify areas where it adds explanatory value. We will assess whether the assumptions posited for this component in the AGEP logframe have been upheld. Finally, we will provide a concluding summary that reflects on the current strength of evidence for Hypothesis 2, the implications for the theory of change, lessons for future programming and any implications for sustainability and scale-up.

5.2.2 Review of Round 3 results on health assets

The evaluation team has reviewed the Population Council's findings for survey Rounds 1-3 (R1-3) for each of the outcome indicators relating to health assets, as well as the evidence for contributions to impact.

In the Round 3 survey report,³³ the Population Council reported on its intent-to-treat (ITT) difference-in-difference (DiD) analysis comparing Round 1 (baseline) and Round 3 results against internal controls. For the purposes of this theory-based evaluation of Hypothesis 2, we conducted a quality assurance review of the Population Council's data and analysis for indicators relating to health assets. Next, using the ITT, DiD analysis, we isolated the effects for rural and urban girls by age cohort.³⁴

Our peer review indicates that the quality of the Population Council's data and analysis relating to health assets is reliable (see Annex 7 for our quality assurance findings). We are, therefore, confident in using the data for an additional analysis relating to Hypothesis 2.

The analysis tables we have extracted from the Population Council's work are contained in Annex 4. The main findings from our synthesis work on context and age cohort effects relating to health assets are summarised in the diagrams below.

³³ Population Council (2016). Adolescent Girls Empowerment Programme, Research and Evaluation Mid-Term Technical Report, July 2016

³⁴ Although the Population Council has also conducted a Treatment of the Treated (ToT) analysis on girls who have completed 52+ week of Safe Spaces meeting, we have agreed with DFID that, for the purposes of this theory-based evaluation, we will focus on the results from the ITT analysis.

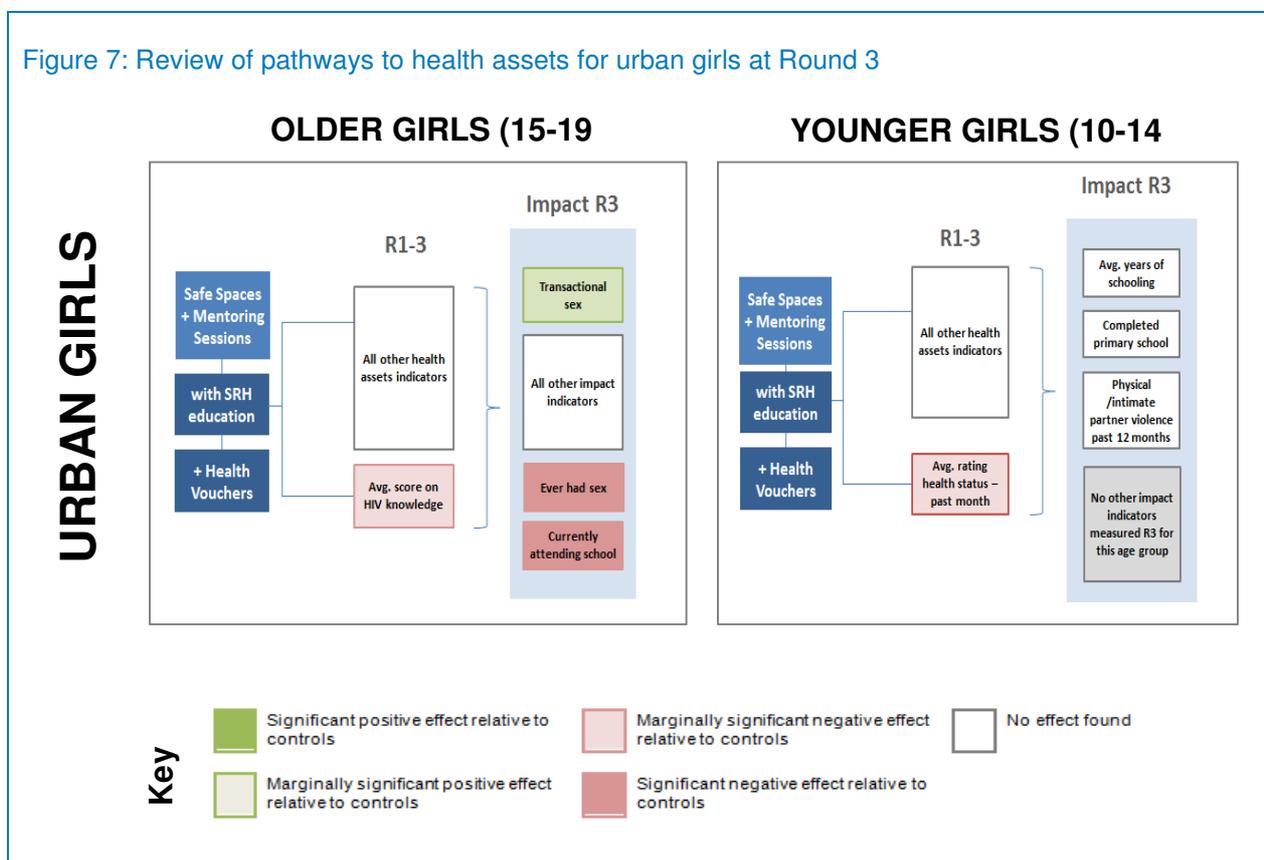
5.2.2.1 Review of results for urban girls at Round 3

Figure 7 shows that there have been differential effects from efforts to build health assets by including SRH education and health vouchers in the Safe Spaces model. For older urban girls (15-19 years), the ITT, DiD analysis for R1-3 shows that for most indicators of health assets there were no effects found. There was, however, a marginally negative effect ($p < 0.10$) on the indicator ‘average score on HIV/AIDS knowledge scale (0-11)’.

Since the theory of change posits that building of assets for adolescent girls will have a combined effect on longer-term impact indicators, the impact level results for older urban girls are the same as for Hypothesis 1. In summary, work on health assets may have contributed to a marginally significant positive effect on reports of transactional sex ($p < 0.10$) by older urban girls. However, for most impact level indicators there was no effect observed over Rounds 1-3 relative to internal controls, and significantly negative effects ($p < 0.05$) on the ‘percentage who had ever had sex’ and the ‘percentage of girls currently attending school’.

For younger urban girls (10-14 years), the ITT, DiD analysis for R1-3 found mostly no effects for the majority of health asset indicators. The exception was a marginally significant negative effect ($p < 0.10$) on the indicator ‘average rating of health status in the past month (0-10)’. Once again, impact level results are the same as for Hypothesis 1. In summary, for the four impact indicators that were measured (relating to school attendance and recent experiences of physical or intimate partner violence) no effects were found for younger urban girls over Rounds 1 to 3 relative to internal controls.

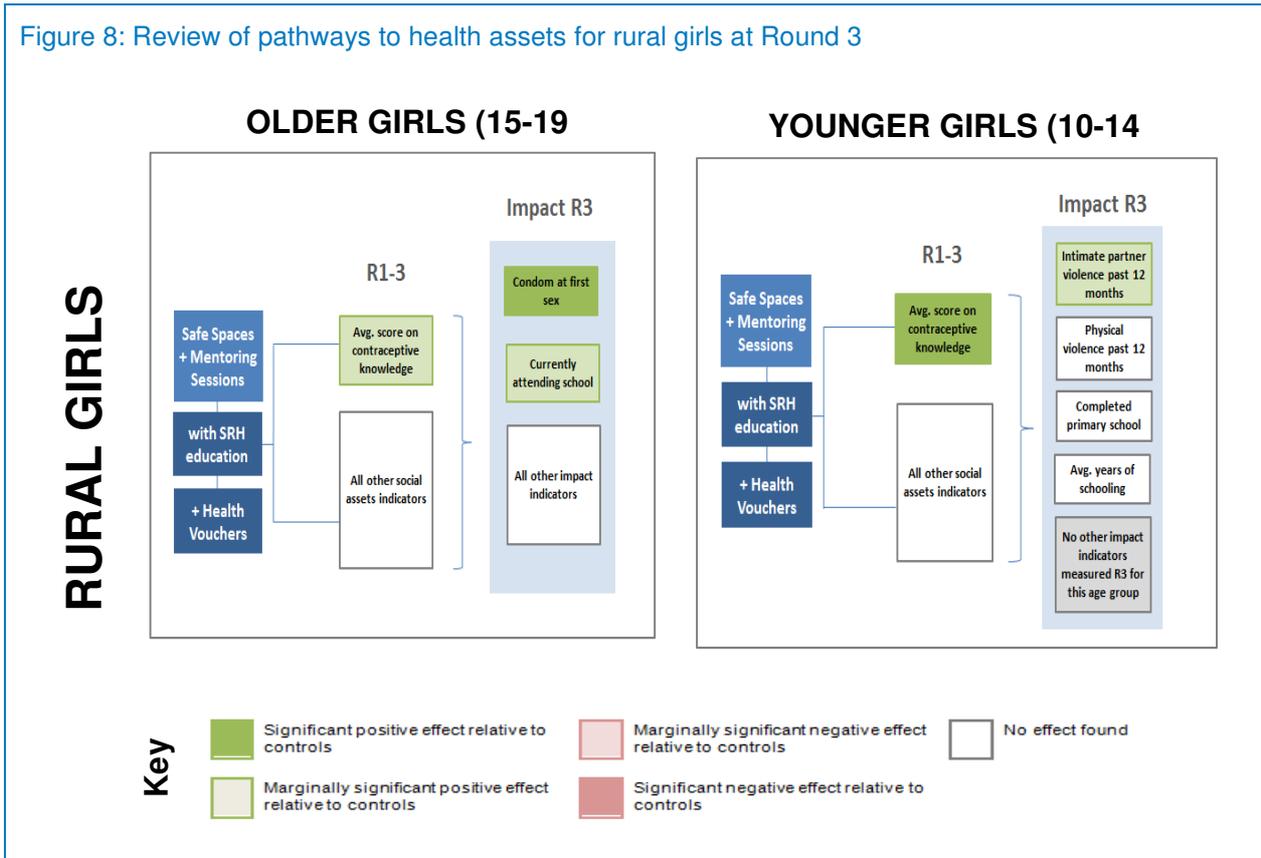
Figure 7: Review of pathways to health assets for urban girls at Round 3



5.2.2.2 Review of results for rural girls at Round 3

Figure 8 shows that there have also been differential results from efforts to build health assets among older and younger rural girls. For **older rural girls**, the ITT DiD analysis for R1-3 showed no effects for the majority of health asset indicators, although there was a *marginally* significant negative effect ($p < 0.10$) for the indicator, ‘average score on the contraceptive knowledge scale (0-9)’. Impact level results remain the same as shown for Hypothesis 1. In summary, the ITT DiD analysis for R1-3 showed a significantly positive impact level effect on the indicator for ‘percentage of girls who used a condom at first sex’ ($p < 0.05$) and a *marginally* significant positive effect on the ‘percentage of girls currently attending school’ ($p < 0.10$); there were no effects found for other impact level indicators for older rural girls.

For younger rural girls, the ITT, DiD analysis for R 1-3 found a significant positive effect on the indicator, ‘average score on the contraceptive knowledge scale (0-9)’; however, there were no effects found for other health asset indicators in this analysis. At the impact level (as shown for Hypothesis 1) there were no effects found for most of the indicators measured for this age group; however, a *marginally* significant positive effect was found on the ‘percentage of girls who had experienced intimate partner violence in the past 12 months’.



5.2.3 Review of the mechanism for building health assets

For this review, we have drawn on the End-Term Evaluation fieldwork, as well as findings from our desk review of programme records and reports, and findings from our Mid-Term Evaluation. We have discussed the health vouchers component with stakeholders at each level of the health system and considered both achievements and challenges.

5.2.3.1 Achievements

As indicated in Hypothesis 2, the principal strategy for building health assets was the addition of a health vouchers component to complement the SRH education provided by mentors as part of the Safe Spaces intervention.³⁵ The health vouchers were distributed to girls in Treatment Arms 2 and 3 of the research so that the marginal impact could be assessed. Girls could redeem the vouchers against a package of health services at contracted public, private and non-governmental health facilities. The services covered by the voucher included basic wellness examinations, as well as age-appropriate SRH services. Service providers used text messages to interact with a web-based system that issued real-time authorisation for the services. The Population Council trained health care providers at participating facilities in adolescent-friendly health services and conducted monitoring and quality-assurance visits. Providers were reimbursed per service based on pre-approved rates.³⁶

The evaluation research confirms that, overall, the Population Council and its partners did a good job in implementing this component of the programme as designed. The evaluation team's assessment of performance against 2016 logframe milestones is shown in Annex 8 and indicates that key milestones relating to the implementation of the health vouchers component were satisfactorily met or exceeded. Key achievements in implementing the health vouchers component included:

- **Successful establishment of the health voucher scheme:** The Population Council reports that a total of 5,789 vouchers were distributed to 75% of eligible girls (with the shortfall due to delayed start-up).³⁷ From our evaluation research we found that the voucher scheme was developed in consultation with the Ministry of Health (MoH) and the former Ministry of Community Development, Mother Child Health and rolled out to a range of public and private (including NGO) facilities in all Master Sites. District administrations had been well engaged and saw the health voucher schemes as consistent with their strategic priorities. At facilities where vouchers were used more successfully, Heads of Facilities and service providers suggested that the voucher scheme had helped them prioritise the needs of adolescent girls. Several also suggested that Safe Spaces meetings had influenced provider attitudes and provided a way to reach out to the girls. These effects on health care providers may be seen as secondary benefits of the health voucher scheme (Box 7).

³⁵ The SRH curriculum is discussed under Hypothesis 1.

³⁶ Population Council. 2016. AGEP Research and Evaluation Mid-Term Technical Report. pp10-13

³⁷ Ibid

Box 7: Secondary benefits of health vouchers

The words of a health care worker in Kapiri Mphoshi are testimony to some important secondary benefits of the health voucher scheme relating to provider attitudes.

“A colleague of mine attended to an AGEP girl who was being sexually abused by an uncle. The girl did not even tell her parents but decided to come to the clinic. For me this was real empowerment, it made me see that these are the things we ignore even as health workers in adolescents. It made me realise as a health worker that these are underserved populations and AGEP was a superb linkage... Lastly I can just say that AGEP has changed the mind-set of health workers like me. We have been discussing with my colleagues here, a lot goes on in these communities and we are not aware. It only comes to us when it turns out into a serious health problem but if we can address and prevent reproductive health problems in adolescents, then we are setting a stage for healthy future mothers”.



- **A well-designed voucher system:** Our evaluation research confirms that the health voucher scheme covered 12 age-appropriate health services; it also included an electronic platform that allowed programme staff to monitor, in real time, voucher usage and irregular activity. The platform remains available as an open source product that could be used by future health programmes. The voucher system included strong mechanisms to prevent fraudulent activity and quality-assure services received.
- **A bridge to SRH services:** The Population Council reports that, among girls who used the voucher, usage for ‘general wellness’ services were the most popular (accounting for 73% of usage). However, among girls who used an SRH service, 81% had also used a general wellness service, suggesting that helping girls access general health services can provide a bridge to more specialised SRH care.³⁸
- **An influential model:** From our key informant interviews we learnt that the Planned Parenthood Association of Zambia recognised the potential of the health vouchers scheme and adapted it to its own work on SRH service provision for adolescents.³⁹

5.2.3.2 Challenges

The independent evaluation research also identified several implementation challenges that could be factors in explaining some of the weaker Round 3 results.

- **Delayed implementation:** The voucher scheme took almost two years to agree with government, including reaching agreement on the voucher package and getting contracts signed.

³⁸ Ibid: 22

³⁹ This was done by working with local businesses to sponsor vouchers to cover fee-based services, and distributing vouchers widely to female and male adolescents during community outreach activities.

- **Low usage:** The Population Council reports that, overall, only 21% of girls who received health vouchers made use of them during the intervention phase.⁴⁰ Although some key informants suggested there could be under-reporting of voucher use at busy urban facilities (because providers did not have enough time to process the vouchers), and the data point to more frequent use of SRH services by older girls, it does appear that the vouchers were not used as often as expected. Our interviews also indicated that there might need to be more sensitisation work with local stakeholders (the girls, mentors, service providers, and communities). Several service providers reported being confronted by angry parents who were concerned about family planning and abortions being offered to their daughters through the vouchers.
- **Supply-side challenges:** The voucher scheme required health care workers to be trained and orientated but high levels of staff turnover in some facilities meant this often had to be repeated. The reimbursement against the vouchers was not always seen as motivating: where usage was low it could take some time for a worthwhile sum to be accrued to benefit the facility; also some private facilities did not think their costs were adequately covered (and at least one pulled out for this reason).
- **Small potential for fraudulent use:** Although there were strong security measures built into the voucher system, there remained a small chance that girls could exchange or swap the voucher. Key informants suggest photo identification could have prevented this.
- **Perceptions of other service users:** Since the girls were accessing services that are provided free in the public sector, the main benefit was in being seen more quickly at facilities. However, our interviews suggested that some other service users saw this as unfair, while girls were sometimes concerned that this drew attention to them.

5.2.4 Contextual factors in building health assets

The Population Council's data does not point to major differences in voucher use among urban and rural girls, although it is noted that younger urban girls tended to just use the vouchers just once –the reasons collated in Annex 9 (p.52) may go some way to explaining this. As indicated above, there may also have been some under-reporting in urban areas.

From our evaluation research, we found that girls in rural areas may need to walk a considerable distance to reach a health facility and may need to take time from school to attend a facility during opening hours. In rural areas, adolescent girls (especially younger girls) generally have to be accompanied when walking any distance from home (reasons given related to safety and moral behaviour). Also, some younger girls interviewed in rural areas expressed concern about being recognised by other community members and arousing suspicion about their health.

It was also notable that there is little choice of service providers in rural areas and health services tend to be provided under the public sector. Since primary health care services are offered free of charge, the benefit of vouchers was that they allowed girls to be seen more quickly –however, as indicated in Annex 9 (p.53), this may not always be perceived as a benefit.

⁴⁰ Population Council. 2016. AGEP Research and Evaluation Mid-Term Technical Report pp.22-23. Of girls who used their vouchers, many used them more than once (except for younger, urban girls who tended to use them only once). In general girls most commonly used the voucher for general wellness services, but older girls made more use of the SRH services.

5.2.5 Review of assumptions

Table 7 below provides a review of the assumptions listed in the logframe⁴¹ that are relevant to the health voucher component. In the AGEF logframe, these assumptions relate to Output 4. As shown, in Table 7, it remains uncertain whether all assumptions have been fully upheld and, as indicated in our comments, this could have implications for the effectiveness of the health assets component.

Table 7: Review of logframe assumptions relevant to building of health assets

Logframe assumptions	Holds	Comment
Assumptions for logframe outputs 1 and 3		
□ Health workers in post, especially at peripheral rural health facilities	✓	□ In all facilities visited for the independent evaluation, there were health care workers in post. However, in many facilities there were staff shortages –this could have led to reduced processing of vouchers, especially in busy urban facilities. Staff turnover was also reported to be a constraint, as there was often a need for re-training in the health assets interventions.
□ Support from MOH/MCDMCH and DMOs for voucher scheme	✓	□ While there was support from government institutions, it took some time to secure buy-in at the national level –this led to delays in roll-out. Our interviews at MoH suggested that some senior decision makers remain skeptical about the sustainability of a health voucher scheme for adolescent girls.
□ A range of providers exists and has the potential to provide a package of quality services to adolescent girls	?	□ A range of providers is available in urban areas but this is less common in rural areas (see Section 5.2.4 on context above). Some private providers did not see the voucher scheme as cost-effective within their business model.
□ Vouchers will encourage uptake of services among adolescent girls	?	□ As discussed under the review of the mechanism (Section 5.2.3), the use of vouchers was not as high as anticipated. Interviews with stakeholders suggest that there may be a number of normative, access and supply side constraints that continue to deter girls from using the vouchers for preventative services.
□ Vouchers will encourage providers to offer quality and adolescent-friendly services	✓ (often)	□ This was largely confirmed by the independent evaluation research –and was especially the case when there was strong buy-in from the head of facility.

⁴¹ Based on the March 2016 version of the AGEF logframe.

5.2.6 Hypothesis 2: Concluding summary

5.2.6.1 Strength of evidence for Hypothesis 2 at Round 3

Based on the ITT, DiD analysis for R1-3, there is evidence of a significant positive effects ($p < 0.05$) on the contraceptive knowledge of younger, rural girls. However, for the majority of indicators on health assets, there remains no strong evidence of effects over Rounds 1-3 for either urban or rural girls across the two age sets. Evidence from our evaluation review of process confirms that there has been limited use of health vouchers –or just repeated use by a minority of girls. This, combined with the delayed roll-out of this component, could explain why the health voucher scheme has not yet translated into significant gains in perceived health status by Round 3. Given the limited use of vouchers, it is most likely that the positive gains in contraceptive knowledge for rural girls are linked to the SRH education component delivered through Safe Spaces.

So, at this stage, the evidence for building health assets remains inconclusive at the empowerment outcome level. It, therefore, remains difficult to trace a credible pathway to impact level results, and thus confirm Hypothesis 2 in the AGEP theory of change.

5.2.6.2 Implications for the theory of change

We have observed that several of the assumptions underpinning this component of the theory of change need qualification (e.g. the assumption that health vouchers alone are sufficient to leverage uptake of services). It might, therefore, be necessary to modify the theory of change to accommodate learning relating to these assumptions. Based on the

differential effects from building health assets, there may be value in developing a more disaggregated theory of change for different segments of the adolescent girl population.

5.2.6.3 Lessons for future programming

Lessons arising from our review of this programme component are described in Section 8.6 of this report. The evaluation team concurs with the lessons identified by the Population Council on government engagement and repeat trainings for high staff turnover.⁴² Useful lessons have also emerged on: the need to better calibrate voucher reimbursements to facilities in different sectors; the important leadership role of Heads of Facilities; and the importance of community and parental sensitisation work (see Annex 9 (pp.52-53) for interview extracts and quotes on this theme).

5.2.6.4 Implications for sustainability and scale-up

Implications for sustainability and scale-up include the need to document lessons from the health vouchers scheme and the need to engage more fully with national and district MoH strategies on promoting adolescent health (for a more complete discussion of this theme see Annex 10, pp. 55-56).

⁴² Population Council (2016). Ibid: 24.

5.3 Findings from the empirical analysis of Hypothesis 3

Hypothesis 3 (economic assets): Vulnerable adolescent girls who participate in a programme of Safe Space meetings with a mentor that includes financial education and opening of bank savings accounts will acquire economic assets (measured by indicators on financial literacy, savings, paid work and income, and personally owned assets). These economic assets will contribute to longer term empowerment impacts (measured by indicators on education, marital relations, pregnancy and births, sexually transmitted diseases and experience of gender based violence).

5.3.1 Section overview

In this section, we review the empirical evidence for the AGEP hypothesis on economic assets (Hypothesis 3) at the end of the two-year intervention phase.

Our approach is structured in the same way as the reviews of Hypotheses 1 and 2. We will begin with a review of the Round 3 survey results relating to economic assets for urban and rural girls by age cohort. We will then consider the evidence from the evaluation research to assess whether it is consistent with the survey findings and adds explanatory value. This will be followed by a review of the assumptions posited for this component in the AGEP logframe. We will conclude with a summary that reflects on the current strength of evidence for Hypothesis 3, the implications for the theory of change, lessons for future programming and any implications for sustainability and scale-up.

5.3.2 Review of Round 3 results on economic assets

The evaluation team has reviewed the Population Council's findings for survey Rounds 1-3 (R1-3) for each of the outcome indicators relating to economic assets, as well as the evidence for contributions to impact.

For this theory-based evaluation of Hypothesis 3, we first verified the quality of review of the Population Council's data and analysis for indicators relating to economic assets in its Round 3 survey report.⁴³ We were then able to use the ITT, DiD analysis for R1-3 to isolate the effects for rural and urban girls by age cohort.⁴⁴ The main findings from our synthesis work on context and age cohort effects relating to economic assets are summarised in the diagrams below. Once again, we have focused on effects found from Round 1 (baseline) to Round 3 surveys and have included both significant ($p < 0.05$) and marginally significant ($p < 0.10$) effects.

5.3.2.1 Review of results for urban girls at Round 3

Figure 9 shows that there have been differential effects from efforts to build economic assets by adding financial education and savings accounts to the Safe Spaces model. For **older urban girls** (15-19 years), the ITT, DiD analysis for R1-3 shows that there have been significant positive effects ($p < 0.05$) on the 'percentage of girls who have saved in the past year' and the 'percentage working for cash or in-kind in the past year'. However, for all other indicators of economic assets

⁴³ Population Council (2016). Adolescent Girls Empowerment Programme, Research and Evaluation Mid-Term Technical Report, July 2016

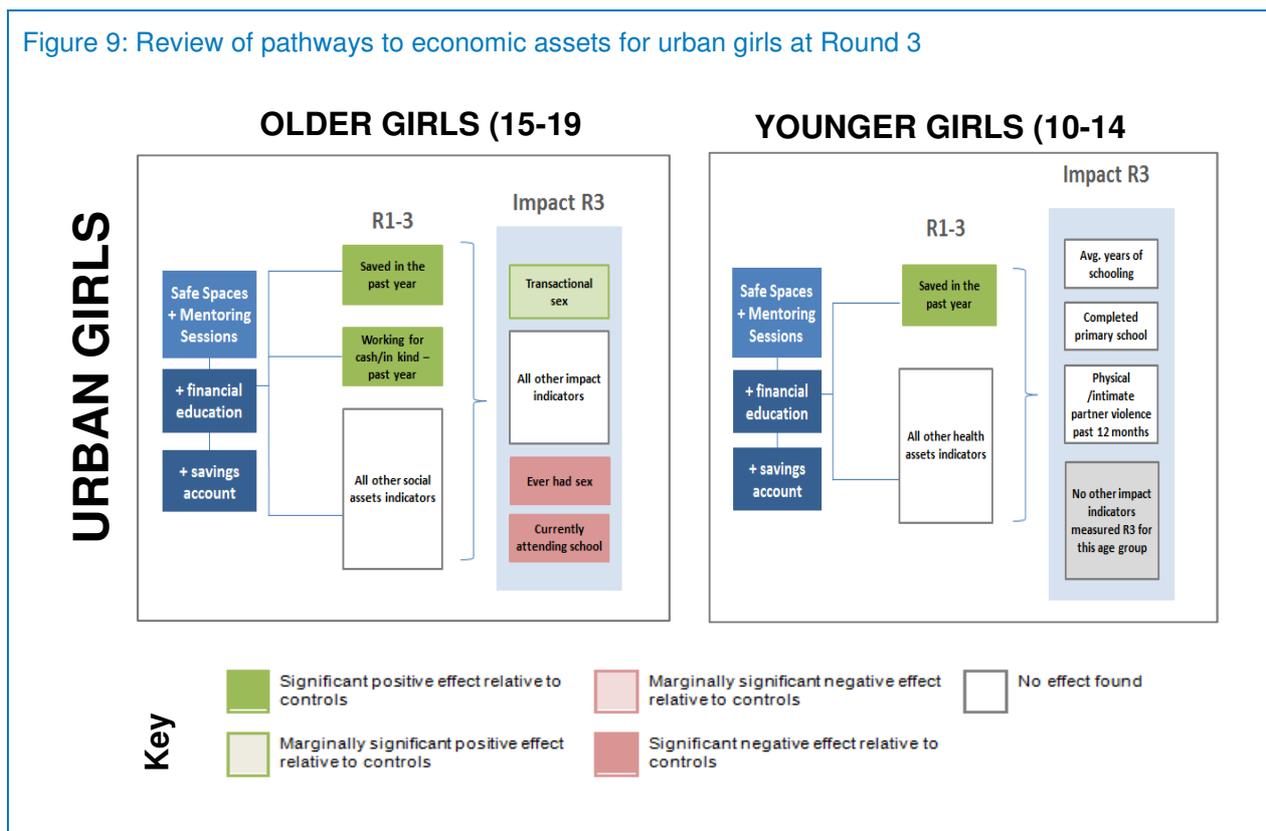
⁴⁴ Although the Population Council has also conducted a Treatment of the Treated (ToT) analysis on girls who have completed 52+ week of Safe Spaces meeting, we have agreed with DFID that, for the purposes of this theory-based evaluation, we will focus on the results from the ITT analysis.

there were no effects found. Once again, impact level indicators are cross-cutting, so are the same as those described for Hypotheses 1 and 2.

For **younger urban girls** (10-14 years), the ITT, DiD analysis for R1-3 found a significant positive effect ($p < 0.05$) on the ‘percentage of girls who have saved in the past year’; however, for all other indicators of economic assets, there were no effects found.

Again, impact level results are the same as for Hypotheses 1 and 2.

Figure 9: Review of pathways to economic assets for urban girls at Round 3

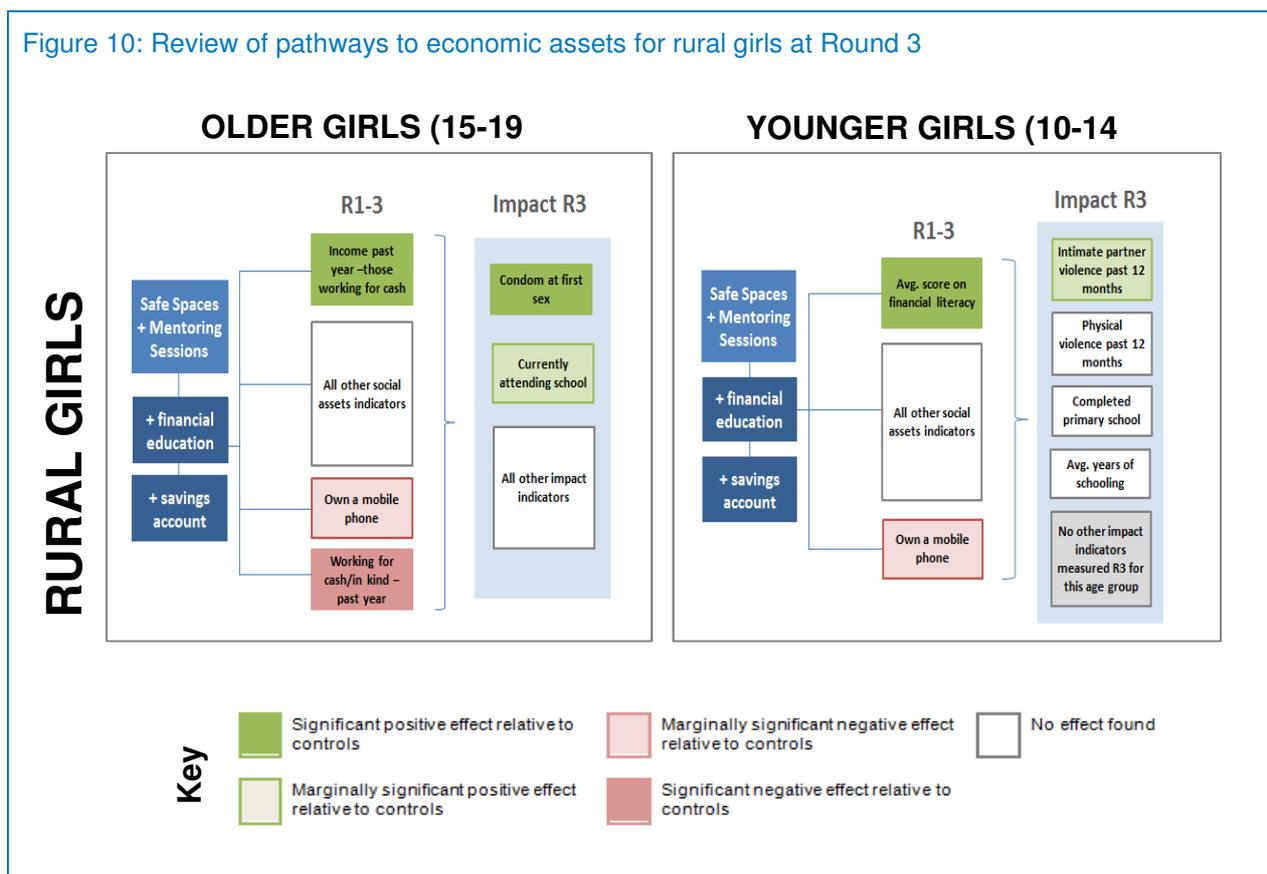


5.3.2.2 Review of results for rural girls at Round 3

Figure 10 shows that there have also been differential results from efforts to build economic assets among older and younger rural girls. For older rural girls, the ITT DiD analysis for R1-3 showed a significant positive effect ($p < 0.05$) on ‘average reported income in the past year among those who worked for cash’. There were no effects for most other economic asset indicators, although there was a *marginally* significant negative effect ($p < 0.10$) on ownership of a mobile phone and a significant negative effect on the ‘percentage working for cash or in-kind in the past year’. Impact level results remain the same as those described for Hypotheses 1 and 2.

For younger rural girls, the ITT, DiD analysis for R 1-3 found a significant positive effect ($p < 0.05$) on the indicator, ‘average score on the financial literacy scale’. There were no effects found for most other economic asset indicators in this analysis; however, there was a *marginally* significant negative effect ($p < 0.10$) found for ownership of a mobile phone.

Figure 10: Review of pathways to economic assets for rural girls at Round 3



5.3.3 Review of the mechanism for building economic assets

For this review, we have drawn on the End-Term Evaluation fieldwork, as well as findings from our desk review of programme records and reports, and findings from our Mid-Term Evaluation. We have discussed the ‘Girl’s Dream’ savings accounts with NatSave officials at national and branch levels and considered both achievements and challenges.

5.3.3.1 Achievements

As indicated in Hypothesis 3, the principal strategy for building economic assets was the opportunity to open a customised NatSave Girls Dream savings account to complement the financial education provided as part of Safe Spaces.⁴⁵ The savings account was available to girls in Treatment Arm 3 of the research so that the marginal impact could be assessed. The savings account required a very low minimum opening balance, and any amount could be deposited or withdrawn with no fee. Girls were able to make deposits on their own, but needed a registered adult co-signatory to assist with account opening and withdrawals. This co-signatory was usually a trusted adult female, such as the girl’s mother or mentor.⁴⁶

The evaluation research confirmed that, overall, the Population Council and its partners made concerted efforts to implement this component of the programme as designed. The evaluation

⁴⁵ The financial education curriculum is discussed under Hypothesis 1.

⁴⁶ Population Council. 2016. AGEP Research and Evaluation Mid-Term Technical Report, p. 25.

team's assessment of performance against 2016 logframe milestones is shown in Annex 8. Key achievements in implementing the savings accounts component included:

- **Opening of accounts for vulnerable girls:** In total, 2,435 accounts were successfully opened for AGEP girls across all 10 Master Sites. In total, 60% of eligible girls opened an account; however, for girls who attended more than half of the Safe Spaces sessions, this figure increased to 72%. At the group level, girls were assisted in completing the NatSave application process. After the initial deposit, 34% of girls made a further deposit and 4% of account holders made a withdrawal.
- **Institutional buy-in:** Our interviews with NatSave senior management and product staff at national and branch levels confirmed there was commitment and enthusiasm for the Girls Dream initiative. The product was seen as consistent with NatSave's social responsibility objectives and its mission to foster financial inclusion. It was part of a portfolio of products targeting young people, who are increasingly seen as a sound strategic investment.
- **Motivational effect:** The Population Council argues that although there were not high levels of account usage, having a savings account reinforced what girls were learning through financial education sessions in Safe Spaces meetings.⁴⁷ Market research conducted periodically over the course of the programme intervention, confirmed that girls with a savings account felt motivated to save (even if this saving was informal).
- **Creating savings:** The evaluation research identified a number of occasions where mentors encouraged girls to identify ways to start small money-making enterprises (see for example, Box 8). In keeping with the Population Council's qualitative research, several girls described how they had set up a rotating fund (or *Chilimba*) to save money.

Box 8: Case study on saving from Kapiri Mposhi

A 14 year old girl from the rural areas of Kapiri Mposhi proudly described how she had used the information she received through AGEP to save money:

"I joined my parents in the farming of groundnuts so I could sell them and earn money. It is from the knowledge I got from AGEP. I used the money from the groundnuts to buy sweets wholesale and sell them so I could save money. I later decided to use this money for cultivating a portion of land. I bought groundnut seeds from my mother to plant, and at the end of the season I harvested 10 bags. I sold half and made about 350 Kwacha...My parents are amazed at how I learnt to do that...Sometimes I buy things for myself –you know, I bought myself two pairs of school shoes that am using for school at the moment."



5.3.3.2 Challenges

The independent evaluation research also identified several implementation challenges that could be factors in explaining some weaker Round 3 results relating to economic assets.

⁴⁷ Ibid:27

- **Dormant accounts:** NatSave reported that by the end of the programme, around 50% of accounts had been dormant for over six months and the low levels of account usage made the product less viable (although NatSave was eager to continue the initiative and possibly extend it to boys). Some NatSave branch officers suggested that the accounts are only likely to be actively used if there is a greater investment in supporting the girls to build skills for income generation activities.
- **Opening and sustaining accounts and preventing misuse:** It took a huge investment of time and effort to process the initial account applications, especially as many applicants and co-signatories did not have a National Registration Card (NRC) number, so this needed to be processed first. Many AGEP and NatSave stakeholders suggested that there might be better ways of opening accounts on such a large scale. Some branch officers were concerned about the potential for misuse of accounts by co-signatories, and about the sustainability of the accounts at the end of the programme (our interviews in Kapiri Mposhi indicted that NatSave has asked Girls' Dream account-holders to convert their accounts to conventional 'minor's accounts' or ordinary adult savings accounts once the programme had ended).⁴⁸
- **Money and relationships:** Our interviews suggested that, in some cases, mentors had to give the girls extra support to deal with the effects of having some money –these effects included peer pressure and jealousy, as well as pressure from family members and boys. In addition, some parents reported that they felt pressure to give the girls extra money for their savings, and this created difficulties for them too.

5.3.4 Contextual factors in building economic assets

Several stakeholders interviewed observed that it was difficult for girls in rural areas to maintain the accounts because of the distance to the banks and the cost of transport. NatSave had hoped that agency banking operations in rural areas might address this issue, but this had not materialised during the intervention phase.

The findings from the Round 3 survey analysis above do suggest that urban girls may have more opportunities to save. However, in our interviews, several girls and mentors from urban sites observed that it can also be difficult for girls (especially younger girls) to get to a bank branch if it requires travel across town or the need to go with a co-signatory.

5.3.5 Review of assumptions

Table 8 below provides a review of the assumptions listed in the logframe⁴⁹ that are relevant to building of economic assets. In the AGEP logframe, these assumptions are found at the Outcome level and under Output 2. As shown, in Table 8, it remains uncertain whether all assumptions have been fully upheld and, as indicated in our comments, this could have implications for the effectiveness of the economic assets component.

⁴⁸ This was reported by parents in focus group discussions and corroborated by interviews with bank officials.

⁴⁹ Based on the March 2016 version of the AGEP logframe.

Table 8: Review of logframe assumptions relevant to building of economic assets

Logframe assumptions	Holds	Comment
Assumptions for logframe Output 2		
<ul style="list-style-type: none"> ☐ Financial institutions will extend services to rural areas and small savers in Zambia 	?	<ul style="list-style-type: none"> ☐ NatSave as a government-owned national bank has been a committed partner in the AGEF initiative. The Girls Dream savings account is consistent with NatSave’s mission to foster financial inclusion. However, NatSave branches in rural areas remain inaccessible to many rural girls. ☐ It is unclear whether private sector banks would consider this approach commercially viable. NatSave itself suggests the model would be more sustainable if it were extended to all young people.
<ul style="list-style-type: none"> ☐ Financial goals translate into desire for savings accounts 	?	<ul style="list-style-type: none"> ☐ From our interviews, most girls were happy to have the opportunity to learn about banks and were keen to save for their goals. However, it appeared that most preferred to save at home or to use traditional savings methods (such as Chilimba rotating funds).
<ul style="list-style-type: none"> ☐ Poor adolescent girls will be able to save 	✓	<ul style="list-style-type: none"> ☐ From the available data it appears girls were able to save, even if this was just small amounts. However, several stakeholders suggested that girls need more support and skills to start income generating activities in order to earn money safely.
Relevant assumptions at the Outcome level		
<ul style="list-style-type: none"> ☐ Girls are able to save money, especially in rural areas 	✓	<ul style="list-style-type: none"> ☐ As above. Some parents in rural areas complained that they felt under pressure to give their daughters money to save so they would not embark on risky activities.

5.3.6 Hypothesis 3: Concluding summary

5.3.6.1 Strength of evidence for Hypothesis 3 at Round 3

From the ITT, DiD analysis for R1-3, there is strong evidence that the intervention led to significant positive effects ($p < 0.05$) on the saving activity of older and younger urban girls, and on the percentage of older urban girls working for cash or in-kind payments. There is also strong evidence of significant positive effects on the income earned by older rural girls working for cash, and on the financial literacy scores of younger rural girls. For most other indicators on economic assets, the evidence remains inconclusive; however, the significant negative effect on the percentage of older rural girls working for cash or in-kind payments remains difficult to explain.

Evidence from our evaluation review of process and the review of NatSave data suggests it is difficult to link positive gains in economic assets to use of the Girls Dream savings accounts – other than through reinforcement of learning. This suggests that the positive gains in economic assets are largely due to an effective financial education curriculum and guidance from mentors on income generating activities. This tentative conclusion would benefit from further investigation to strengthen the evidence base.

So, at this stage, the evidence for building economic assets is somewhat mixed at the outcome level. There are also no obvious links to the inconclusive impact level results. It, therefore, remains difficult to confirm Hypothesis 3 in the AGEP theory of change at this stage.

5.3.6.2 Implications for the theory of change

We have observed that, although some assumptions for this component have been broadly upheld, several need qualification. In particular, the assumption relating to the viability and accessibility of Girls Dream savings accounts needs to be revisited. This exercise could lead to revision of the theory of change on how best to build economic assets in cost-effective and sustainable ways. Based on findings showing the differential effects of building economic assets, a revised approach should also take account of the different constraints, risks and opportunities for younger and older girls in urban and rural settings.

5.3.6.3 Lessons for future programming

Lessons and reflections arising from our review of this programme component are summarised in Section 8.6 of this report. The evaluation team concurs with the lessons that the Population Council has documented for the savings account component – the lesson the distance to the bank can be a key barrier, especially for rural girls, is especially important.⁵⁰ We suggest that an additional lesson relates to the observation that traditional savings practices (such as *Chilimba* savings schemes) can offer a useful alternative to bank savings accounts.

5.3.6.4 Implications for sustainability and scale-up

Implications of sustainability and scale-up are discussed at more length in Annex 10 (p.56). Key considerations are that the Girls' Dream savings account may not be sustainable in the longer-term without further collaborative working with the NatSave Bank and further investment from partners. The cost-effectiveness of this component thus remains uncertain.

⁵⁰ Population Council (2016). Op. cit. p. 28

5.4 Final reflections from the theory-based evaluation

For the purposes of the theory-based evaluation, we have reviewed differential results relating to both empowerment outcome indicators and longer-term impact indicators. We have done this to help us assess the status of the results pathways in the AGEP theory of change. However, we suggest that, at this point in the AGEP initiative, particular attention needs to be given to the status of the *empowerment outcome indicators* because these relate specifically to building of assets. The AGEP theory of change posits that assets built during the two-year intervention phase will be protective for girls undergoing the transition to adulthood. This protective effect will lead to longer-term impacts that should be measurable in the 2017 Round 5 survey. So, an emerging question is, how strong (and complete) do the effects on empowerment outcomes need to be at this point to produce intended impacts in the longer term? Additional questions for consideration are: a) how long the effects of assets built over the two-year intervention phase likely to endure? and b) how strong are the longer-term impacts likely to be for older girls who have already made sexual transitions? We can only speculate at this stage; however, we suggest that these are critical questions for evaluating the AGEP theory of change going forward.

6. Findings Part 2: Economic Evaluation

6.1 Understanding the different types of economic evaluation

This section briefly describes different types of economic evaluation in order to explain the type of economic evaluation that is appropriate for AGEP and how it can be used. Different types of economic evaluation inform decisions in different ways. Table 9 describes four different types of economic evaluation: all measure costs in terms of money, but they differ in the way that the consequences are reflected. Note that “consequences” can also be thought of as benefits, effects or outcomes/impact.⁵¹

Table 9: Types of economic evaluation

Type of economic evaluation	Cost measure	Type of consequences identified for all alternatives	Methods for measuring and valuing consequences	Situation where this type of economic evaluation can be used
Cost-consequence analysis (CCA)	Money	Any types of impact that can be measured	Each individual consequence is measured and listed	All types of investment can be compared, but this is done through a list and there is no decision rule/ numerical “answer”.
Cost-effectiveness analysis (CEA)	Money	One single effect which is the main impact for all the alternatives being compared	One single indicator: e.g. number completing secondary school, number fully vaccinated children	To identify the cheapest way to achieve a given outcome, or how to maximise achievement of the outcome indicator with a given amount of money.
Cost-utility analysis (CUA)	Money	Effects that can be brought together into one index	An index: e.g. DALYs (disability adjusted life years)	Any choices where the consequences are captured by the index – e.g. health interventions for DALYs.
Cost-benefit analysis (CBA)	Money	Can be used for multiple types of impact, and these do not need to be common to all the alternatives	Money – all consequences measured and given a monetary value	All possible public sector investments can be compared.

⁵¹ This section is based on Chapter 1 of *Economic Evaluation* by J Fox-Rushby and J Cairns, 2005.

This assessment refers to adolescent girls' empowerment, which has multiple dimensions of impact: for example, AGEP measures impact in terms of effects on education, sexual and reproductive health and gender-based violence. In the next section, we will be exploring comparisons with some cash transfer programmes, which also have multiple dimensions of impact.

From the table above, we can see that cost-effectiveness analysis and cost-utility analysis are not relevant for this report, because they rely on either a single indicator of outcome (as in CEA) or an index that summarises the outcome (as in CUA).⁵² Neither of these applies to adolescent girls' empowerment.⁵³

That leaves cost-benefit analysis (CBA) and cost-consequence analysis (CCA). CBA is the only type of economic evaluation that describes both costs and consequences in monetary terms and is therefore able to compare interventions that have different types of consequences (e.g. comparing health, education and road-building interventions). CBA requires putting a monetary value on all consequences – in the context of adolescent girls' empowerment programmes, this would include putting a monetary value on effects such as delayed first pregnancy and reduction in gender-based violence. There has never been a plan to do a CBA for AGEP – rightly so, in our opinion, because giving monetary values to the consequences of AGEP would involve many value judgements which would be opaque in the conclusions of the CBA. Whilst CBA has the advantage of expressing findings in numbers, decision-makers can find a single figure that encapsulates costs and benefits difficult to understand and not credible.

This leaves cost-consequence analysis. CCA was developed because of the difficulties inherent in cost-benefit analysis. CCA ensures that all significant benefits are identified and quantified, even though they are not given a monetary value. The findings of a CCA are essentially a list, with each dimension of impact described separately.

The economic evaluation of AGEP is a cost-consequence analysis. It provides information on overall costs and a variety of types of outcome. It does not summarise the ratio of costs to benefits in one overall number, meaning that all comparisons with other programmes will require some form of value judgement about the relative importance of different outcomes.

Having explained the different types of evaluation, it is necessary to add a point about language. The term “cost-effective” has entered general non-technical usage and is commonly used to mean “better value”, rather than being a specific reference to Cost-Effectiveness Analysis. Cost-Effectiveness Analysis is a particular methodology, but the adjective “cost-effective” is a general term.

⁵² Some publications already refer to an “empowerment index” and there could one day be a widely agreed index that could be used for CUAs of empowerment programmes. For example see Bandiera et al (2012) *Empowering Adolescent Girls: Evidence from a Randomized Control Trial in Uganda*.

⁵³ Cost-effectiveness analysis (CEA) is frequently used, but it only offers numerical guidance to decisions if the alternatives being compared have a single shared outcome measured in the same way. It cannot be used to compare across programmes with different outcomes without missing out on some of the effects. When an intervention has a range of types of benefits (e.g. related to education, health, fertility and savings) and only one primary benefit is included in the CEA, this will under-value the overall benefits of the programme. This phenomenon has been documented for HIV, for example, by Remme et al, who demonstrate why HIV programmes with multi-sectoral benefits are less likely to be prioritised, financed and taken to scale when individual sectors are taking separate decisions about what interventions to fund. [Remme et al (2014) *Financing structural interventions: going beyond HIV-only value for money assessments*. AIDS 28:425–434]

6.2 Progress overview

The Population Council's Baseline Technical Report (April 2014) set out the work to be done within the economics component of AGEF. Table 10 lists the plans set out in this Baseline Report and comments on progress so far. In summary, Table 10 shows that reasonable progress has been made with most of the plans, although some work still needs to be done on the current draft of the Council's Mid-Term Technical Report (2016).

Table 10: Progress with activities in the economic evaluation of AGEF

Planned activity (from 2014 Baseline Report)	Progress (based on 2016 Economic Evaluation End-of-Programme Report)
<ul style="list-style-type: none"> □ Present programme costs by expenditure category from point of view of all service providers, including the National Savings and Credit Bank 	<ul style="list-style-type: none"> □ Almost completed. □ YWCA included as part of Population Council's costs. Not clear how any additional NatSave costs recorded, but possible that all the costs are captured through the Population Council's accounts. Requires a brief explanation.
<ul style="list-style-type: none"> □ Estimate average costs per participant, by programme arm 	<ul style="list-style-type: none"> □ Completed
<ul style="list-style-type: none"> □ Calculate differences in programme costs per participant, by programme arm 	<ul style="list-style-type: none"> □ Completed.
<ul style="list-style-type: none"> □ Analyse participants' out-of-pocket costs 	<ul style="list-style-type: none"> □ For girls: completed. Reasonable conclusion that these were too small to be worth including. □ For mentors: raw data presented, but needs explanatory narrative and linking this data to the costing as a whole. Where is the overall costing from a societal perspective, as mentioned in the report?
<ul style="list-style-type: none"> □ Estimate incremental costs per negative health outcome averted, by programme arm (using disability-adjusted life-years, DALYs). Develop incremental cost-effectiveness ratios (ICERs). 	<ul style="list-style-type: none"> □ The report gives an appropriate explanation of why this has not been done – no evidence about changes in DALYs.
<ul style="list-style-type: none"> □ Estimate incremental costs of positive progress achieved on selected output and impact indicators, by programme arm. Develop ICERs. 	<ul style="list-style-type: none"> □ The report gives an appropriate explanation of why this has not been done – no evidence about positive changes by arm.
<ul style="list-style-type: none"> □ Compare the incremental costs of implementation between urban and rural sites 	<ul style="list-style-type: none"> □ Very Limited – only about 2% of total costs could be assigned as "rural" or "urban". Not sufficient detail for discussing equity or costs of scaling up.
<ul style="list-style-type: none"> □ Micro-costing exercise at two health facilities 	<ul style="list-style-type: none"> □ Completed. Even though one facility had no voucher- users, information is fit for the purposes required. Report needs to relate findings to the number of vouchers used, and hence total estimated cost to health providers.
<ul style="list-style-type: none"> □ Create an index ranking the outcomes in terms of their impact on girls' lives (relative importance of different types of outcome) 	<ul style="list-style-type: none"> □ Agreed that focus group discussions will be written up at a later date.

It is worth pointing out the significance of the last row of the table above. Whilst a Cost-Consequence Analysis is realistically the only form of economic evaluation possible for AGEP, it would be useful to have some idea about which effects are regarded as the most important. For this reason, the economic protocol includes an action to create an index ranking the outcomes of AGEP in terms of their relative importance. This should be done largely through focus group discussions. It is our understanding that this work has been started by the Population Council, but has not yet been written up. Even with this “index” the evaluation will not move beyond CCA, but it will at least provide some insights into the relative values assigned by stakeholders to the different types of outcome produced by AGEP.

In addition to the original planned work – and at Mott MacDonald’s request - the second version of the economics report (the Economic Evaluation End-of- Programme Report)⁵⁴ contains an interesting table on average costs per measure of effectiveness. This is, in effect, the outline of a cost-consequence analysis. Whilst there are currently issues with the exact numbers used in this table, a future version of this table will be one of main findings of the economic work. The table is discussed in more detail in Section 6.4.3 below.

6.3 Findings from verification/validation work

The Population Council’s budget for the economics work was relatively small (<£20k or <1% of the research budget): this has imposed some limitations on the work. As an example, the Council’s economist could not investigate all spending lines herself in order to classify them as accurately as possible according to rural/urban, programme arm, specific location etc. The classification was done after the event by a member of Population Council staff. Given the resources available, a good deal of information has been assembled and – if properly presented with full explanations – will be useful. However, the information on rural/urban costs is too limited to be of practical value.

The revised economics report needs to be inserted into the main Population Council report and the figures about impact cross-referenced, so that it is clear where these figures come from and that only significant changes are being referred to. The economics work is still lacking in some detail – specifics are noted in Table 10 above.

6.4 Commentary of the findings in the ‘Economic Evaluation End-of-Programme Report’

6.4.1 Findings about costs

The Economic Evaluation End-of- Programme Report (August 2016) includes an analysis of the costs of AGEP. Since implementation has now ceased, these costs will not change for the final report (the research is not being costed). Information is presented on spending by arm, by type of expenditure and per participant. The report also describes the costs of health services available through the voucher scheme, as well as expenses incurred by AGEP mentors. This is all relevant information, but at times the report needs to give more explanation (e.g. explaining more about the mentors’ costs). The information on rural/urban costs is very limited, covering only 2% of total costs.

⁵⁴ A revised draft of the economic section of the Population Council’s Mid-Term Technical Report (2016) was presented as a separate 13-page report called the ‘Economic Evaluation End-of-Programme Report’, July 2016.

One notable finding in the report is that the average cost per participant is \$771, or \$394 for only the Safe Spaces component. It is important to be clear what these figures mean: they are not annual costs. AGEP is a one-off intervention which takes longer than one calendar year, but once finished, costs are zero. The cost of AGEP is more a one-off investment cost, rather than a recurrent cost. The investment is in the futures of individual adolescent girls. (Of course an AGEP-type programme would have recurrent costs over time, but these would be to work with changing cohorts of girls, not supporting the same girl year after year.)

Discussion of alternative uses of public money in Zambia can be a useful way of putting the average cost amounts into context (\$771 and \$394). Table 11 gives some costing information about DFID-supported programmes. It is very important that these numbers are not interpreted as value-for-money comparators. The programmes have different types of outcomes; moreover, some of the costs are recurrent annually, others are not. The reason for including the table is to place the AGEP figures in some sort of context – from this we can comment that AGEP is a relatively expensive programme, but no conclusions can be made about its value-for-money. (Expensive programmes can be a high priority: for example reducing maternal mortality to low levels is expensive, but good value if all the benefits for mother, baby and the wider family are considered.)

Table 11: Unit cost figures for some DFID supported programmes in Zambia⁵⁵

Unit	Cost (when original information in US\$, converted at August 2016 rate of \$1 = £0.76)	Comment
New sanitation user	£3.83	Excludes marketing costs.
Person reached with mass hygiene promotion	£0.12	
Drop hole in a school toilet	£578	Based on bill of quantities. Capital expenditure with multi- year benefits.
Nutrition programme beneficiary	£31 (converted from \$40)	
Adult HIV infection averted	£99	Based on number of infections averted in urban areas.
Person protected for one year against malaria by indoor spraying	£2.60	Recurrent cost
Impregnated net used by pregnant women and children under five years	£6.10	
One couple year protection with contraceptives	£20	Recurrent cost
Per girl to receive AGEP intervention	£588 (converted from \$771)	One-off intervention: hypothesis that this will have multi-year benefits

⁵⁵ See narrative on how to interpret this table.

Unit	Cost (when original information in US\$, converted at August 2016 rate of \$1 = £0.76)	Comment
Per girl to receive safe spaces part of AGEP intervention	£300 (converted from \$394)	
Social cash transfer per household	£123 (converted from \$161)	Average cost of current scale- up.

The bottom three rows of the table above are interesting because these are all expenditures that would fall under the remit of the Ministry of Community Development and Social Welfare (MCDSW). For every recipient of AGEP, the same money would fund almost two recipients of a Safe Spaces-only version of AGEP, or would provide between 4 and 5 households with social cash transfers for one year.

6.4.2 Scale-up costs

The cost information provided in the Economic Evaluation End-of-Programme Report is not very detailed, but it does allow for some points to be made about **scaling up**. As a programme expands, unit costs tend to decrease for two reasons: economies of scale and the fact that a proportion of costs are start-up costs which do not need to be repeated. The implementation cost given in the 2016 report is in 2016 US dollars, but the information on start-up costs is in 2014 US dollars. Nevertheless, we can estimate that start-up costs accounted for approximately 20% of total programme spending (\$2.4 million/\$12.8 million) and the pilot accounted for approximately 2% (\$297k/\$12.8 million). Some start-up costs are incurred whenever the programme moves to a completely new province (e.g. office space and local recruitment). Nevertheless, a crude rule of thumb would be that unit costs would decrease by 20% for an expanded AGEP, as long as this did not involve working in much more remote (and expensive) areas. A decrease of 20% would mean unit costs of £470 per girl, or £240 for the Safe Spaces-only version of AGEP.

The costs of scaling up depend on a number of factors, such as which arms are to be implemented and the rural/urban balance. Four cost categories account for 91% of AGEP costs: salaries (54%), buildings (17%), food/accommodation/travel (11%) and equipment (9%). When planning a scale-up, it is possible to predict what will be needed in terms of these four categories – this depends largely on the scale and geographic dispersion of the scale-up. An estimate based on these four cost categories can be used as the basis for a prediction of scale-up costs.

6.4.3 Findings about cost-effectiveness

Table 12: Incremental effectiveness and average cost-effectiveness, 2016 US\$

(NB These numbers should not be quoted as the text explains why we believe they are flawed.)

Indicators	Difference between R3 and baseline		Incremental effectiveness	ACERs (US\$)
	Intervention (n)	Control (n)		
<i>Impact indicators</i>				
Grade 7 completion	616	274	342	37,509.13
Grade 9 completion	459	227	232	55,293.63
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	56,263.69
HIV seroconversion	54	30	-24	-
HSV-2 seroconversion	211	86	-125	-
<i>Outcome/output indicators</i>				
Unwanted sex in last 12 months	162	41	-121	-
Condom use at last sex	46	31	15	855,208.07
Savings in the past 12 months	448	149	299	42,903.41
Good money management decision-making	108	53	55	233,238.56

Table 12 above has been extracted from the Economic Evaluation End-of-Programme Report⁵⁶ and presents the average cost of achieving various outcomes of AGEP.⁵⁷ Total cost is divided by effectiveness, indicator by indicator, using the control/intervention difference between baseline and Round 3. Each sum in the last column is simply the total cost divided by the number of units of outcome in that row: there is no attribution of particular costs to particular outcomes. This type of table is potentially useful, but at the moment we believe it uses the wrong indicators and data. (We suggest an alternative table below.) Specifically:

- Table 12 above focuses on impact indicators, but these are not the same indicators as given in 'Table 18' of the Population Council's Mid-Term Technical Report (2016) (Excel spreadsheet, Intent-to-Treat (ITT) Summary Results: Difference-in-Difference (DiD))

⁵⁶ This table is Table 9 in the Economic Evaluation End-of-Programme Report.

⁵⁷ The Economic Evaluation End-of-Programme Report states that this work was done "To assist the independent evaluation carried out by HLSP, which aims to compare the cost-effectiveness of AGEP to that of similar programmes in Zambia and in neighbouring countries". Even without the comparison work, this is basic, useful information and should be included in the Population Council's findings.

Estimators). The list and naming of indicators need to be consistent throughout the unified report.

- In any case, given that the focus of the overall End-of-programme Report is on changes in the empowerment indicators, it would seem reasonable for the economic analysis to take the same approach and look at the unit cost of achieving improvements in these indicators. This is not done.
- The table includes indicators for which there were no statistically significant differences between the intervention and control group.
- The table should use DiD data.
- The table only looks at cost-per-positive change. A full interpretation would also take into account the apparent finding that the intervention seemed to make at least one indicator worse.

This table needs to be re-done to align fully with the results discussed in the wider Population Council report.

We have discussed the cost-effectiveness table presented by Population Council at some length because it is the only information available. We have explained what we believe are the problems with Table 12 – Table 13 below demonstrates what we believe the table should look like.

Table 13 uses the same suite of empowerment and impact indicators that we have used throughout this report (see Figure 3 above). As discussed at the end of Chapter 5 (Section 5.4), it is the empowerment outcome indicators which are of particular relevance at this stage. The vast majority of the empowerment indicators (22/27) show no significant change – however, these no-change indicators need to be included in a costing table because money has, in effect, been spent on trying to improve them. Because DiD information is measured as proportions, the information is scaled up to 10,000 girls, to allow a cost per girl to be calculated. For the improvements measured in absolute numbers (e.g. % who have saved in the past year) total cost is divided by effectiveness, indicator by indicator based on difference-in-difference estimators. Each financial figure in the last column is simply the cost of achieving that unit: there is no attribution of particular costs to particular outcomes and the same pot of money buys all the improvements combined.

The costs-per-unit-of-benefit in the last column should all be taken together because the AGEP intervention simultaneously produces all these benefits. So Table 13 shows that for around \$50,000 AGEP will produce about four girls who have a safe space to meet with friends, an average increase of 0.3 in the financial literacy scale and 0.2 in the contraceptive knowledge scale for about 65 girls (\$50,000/\$771) and almost four girls who have saved in the past year. However, at the same time, there was also a significant decrease in mobile phone ownership.

Table 13 also includes impact indicators. Strictly speaking these should be left until the final round, but they are included because there is already considerable interest in the impact results. The table shows that there was no change in 16/18 impact indicators and a positive change in two. These positive changes mean that two more benefits can be added to the list of what \$50,000 spent on AGEP buys: seven girls prevented from having transactional sex and two instances of condom use at first sex.

Table 13: Unit costs of achieving empowerment indicators and outcomes

Indicators	Change due to AGEP (DiD)*	Change due to AGEP (absolute number)**	Cost per girl affected (10,000 girls @ \$771 per girl = total cost \$7,710,000)
Empowerment indicators			
Social assets			
Avg. score on self-efficacy scale (0-10) + 12 other indicators related to social assets	No significant change therefore no value gained for the money spent.		
% have a safe space to meet with friends	5.6	560	\$13,768 spent to have one more girl with a safe space to meet friends
Economic assets			
Avg. score on financial literacy scale	0.3	0.3	\$771 per girl: average girl improved by 0.3 score
% who have saved in the past year	5.2	520	\$14,827 spent to have one more girl save money
Avg. amount saved in the past year among those who saved + 3 other indicators related to economic assets	No significant change therefore no value gained for the money spent.		
% who own a mobile phone	-4.3	-430	Economic evaluation is not geared up to deal with negative results, but they need to be factored into the Cost-Consequence description.
Health assets			
% Understanding pregnancy risk during menstrual cycle + 4 other indicators related to health assets	No significant change therefore no value gained for the money spent.		
Average score on contraceptive knowledge scale (0-9)	0.2	0.2	\$771 per girl: average girl improved by 0.2 score
Impact indicators			
Educational			
Four indicators related to education	No significant change therefore no value gained for the money spent.		
Sexual risk behaviour			
% ever had sex + 2 other indicators related to sexual risk behaviour	No significant change therefore no value gained for the money spent.		
% agree that they have had transactional sex	-10.0	1000	\$7710 per additional girl who did not have transactional sex
% used condom at first sex	2.9	290	\$26,586 per use of condom at first sex
Marital			
% ever married	No significant change therefore no value gained for the money spent.		

Indicators	Change due to AGEP (DiD)*	Change due to AGEP (absolute number)**	Cost per girl affected (10,000 girls @ \$771 per girl = total cost \$7,710,000)
Pregnancy & births			
% who have ever been pregnant + 3 other indicators related to pregnancy and births	No significant change therefore no value gained for the money spent.		
Sexually transmitted infections			
% HIV positive and % HSV-2 positive	No significant change therefore no value gained for the money spent.		
Experience of violence			
% have experienced physical violence in past 12 months + % have experienced intimate partner violence in past 12 months	No significant change therefore no value gained for the money spent.		

*Based on ITT results – for numerical indicators, the DID is measured as a difference; for categorical indicators, DiD is percentage points difference

**Based on 10,000 girls

Box 9 below is, in effect, one way of presenting the Cost Consequence Analysis for AGEP at Round 3 (after 24 months of interventions)

Box 9: Costs, empowerment outcomes and impact of AGEP by month 24

For around \$50,000 AGEP will produce the following protective assets in girls transitioning to adulthood:

- About four girls who have a safe space to meet with friends
- An average increase of 0.3 in the financial literacy scale and 0.2 in the contraceptive knowledge scale for about 65 girls
- Almost four girls who have saved in the past year

However, at the same time there was also a significant decrease in mobile phone ownership.

If we think that it is valid to include positive impacts at this stage, the \$50,000 also produces:

- Seven girls prevented from having transactional sex
- Two instances of condom use at first sex

6.4.4 Marginal effects

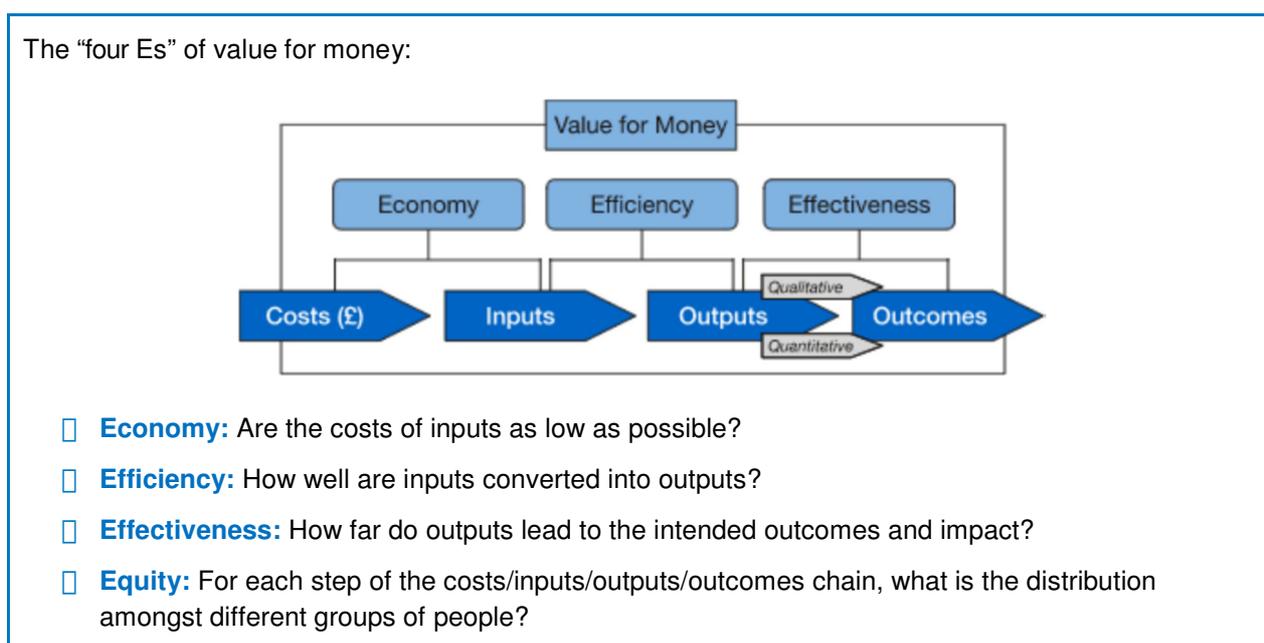
The Population Council’s Mid-Term Technical Report (2016) explains why one particular aspect of the planned economic evaluation has not been done. The plan was to look at the extent to which the more expensive arms of the programme produced more health benefits, with changes in health expressed in terms of averted DALYs. However, the end-of-programme results showed no evidence of a statistically significant effect from the ‘add-on’ interventions on the health-related indicators: i.e. no changes in DALYs were detected. Put another way, the \$6.277k spent on the health voucher arms (with and without the savings component) did not produce any additional health gains during the lifetime of the programme so far. Of course AGEP is not just about health – the reasons for mentioning the DALY analysis here are that there was an explicit

plan to include it, and the original business case was largely justified in terms of health improvements.

The same argument is true for all other types of indicators – Treatment Arms 2 and 3 did not show additional benefits over and above Treatment Arm 1, hence no incremental cost-effectiveness ratios by treatment arm could be calculated.

6.5 Review of Value for Money

Figure 11: Value for money criteria



Previous DFID Annual Reviews of AGEP have assessed value of money (VFM) in terms of economy and efficiency. This has involved discussion of issues such as procurement practices, sharing costs with other programmes and forward planning to minimise travel costs by having multi-purpose journeys. The October 2015 Annual Review of AGEP noted that VFM discussions could move on to discuss effectiveness in 2016 because “the end of programme cost-effectiveness analysis will be available in early 2016”. The discussion above about economic evaluation is therefore also highly relevant to considerations of VFM.

Box 9 and Table 13 above provide the information for a Cost-Consequence Analysis of AGEP at 24 months. We also know that there were no significant additional benefits for the additional costs of Arms 2 and 3 of AGEP. However, the most appropriate time to look at an economic evaluation of AGEP will be using the 36-month information on longer-term impacts, given that it may take time for some results to manifest themselves.

In terms of economy and efficiency, little has changed since the October 2015 Annual Review because the overwhelming majority of programme spending happened before that date. The Population Council’s technical report confirms previous findings about the cost drivers – as shown in the table below this is a labour-intensive intervention, with salaries accounting for more than half of all spending.

Table 14: Cost of AGEP delivery, US\$ 2016

	TOTAL
Salaries	6,990,759.59
Buildings	2,142,246.97
Equipment	1,129,028.03
Utilities and insurance	228,903.95
Vehicles	107,136.14
Supplies & services	801,974.63
Food, accommodation & travel	1,428,071.72
TOTAL	12,828,121.04

Source: Bozzani F (2016) *AGEP Economic Evaluation End-of-Programme Report*

In terms of equity, we know that the results are different for rural and urban girls, though we also know that there was less targeting of rural girls because of the smaller numbers in each location. We also know that the programme is more expensive to deliver in rural areas, but the information does not tell us how much more expensive. Unfortunately, only 2% of total costs were labelled as specifically “rural” or “urban”, so no meaningful analysis is possible around this.

The economic evaluation of AGEP is only about the intervention and does not include the costs of the research. However, it is worth thinking about the research in terms of value for money. About 22% of DFID’s total spending on AGEP has been on research.⁵⁸ By the end of AGEP the research should be able to guide DFID about whether it would be good VFM to adapt and scale up an AGEP-type intervention. In addition to this use, the research contributes to the literature on the economics of adolescent girls’ empowerment. The research findings thus have the potential to improve the way in which tens of millions of dollars are spent. Even if AGEP proves not to be effective, it can still be argued that the research has provided good value-for- money, because it can inform future spending decisions.

⁵⁸ Source: AGEP Annual Review 2015.

7. Findings Part 3: Comparative Evaluation

7.1 Aim of the comparative evaluation

The main aim of this part of the independent evaluation is to assess how well the AGEP model compares to alternative approaches to the empowerment of vulnerable adolescent girls, both in terms of performance (epidemiological effectiveness) and costs. In keeping with our terms of reference, we have prioritised comparison with cash transfer models (see Annex 1, p.5).

In developing our approach to the comparison work, we have found it useful to review the evidence on 'what works' in adolescent girls' empowerment and to situate the AGEP model and cash transfer approaches within a broader typology of approaches. Our findings from this review are summarised in Annex 11. Important observations from this review are that initiatives for adolescent girls often use combinations of approaches, and are rarely mutually exclusive. These approaches may operate at different structural levels and engage a range of stakeholders. They often focus on particular dimensions of empowerment, such as poverty, education or health; they may also be driven by locally-relevant issues, such as early marriage or female genital cutting. In this respect, this comparison work should be seen as a reflection on the evidence base; it is not intended to produce definitive conclusions.

7.2 Comparison of epidemiological effectiveness

7.2.1 Steps in the comparison

The comparison work on epidemiological effectiveness was conducted in three steps: firstly, we identified eligible studies for comparison; secondly, we confirmed the comparability of these studies; and thirdly we conducted the comparative analysis.

7.2.1.1 Step 1: Identification of eligible studies

For Step 1, we undertook an extensive search for potential comparison studies. We did this through online searches and 'snowballing' based on information received from key informants in DFID and country-based institutions. In conducting the search, we used the following inclusion criteria:

- The study should be set in Sub-Saharan Africa, preferably in the same region as AGEP, and should be conducted over roughly similar timeframes.
- The study should be based on interventions that could benefit adolescent girls directly or indirectly, and should potentially improve their life skills, education and reproductive health outcomes (although the interventions could be of any type e.g. skills training, microfinance programme, cash transfer etc.).
- The study should be based on a rigorous evaluation design to assess the impact of one or more interventions. The study design should include a programme intervention arm

and a counterfactual arm for measuring the impact of the intervention and establishing attribution.

- The study should have similar outcome measures as AGEP, especially relating to health and education.

Using the above criteria, six potential comparison programmes in/around the southern and east Africa region were shortlisted for potential comparison (Table 15).

With letters of introduction from the DFID, the evaluation team wrote to the Principal Investigators (PIs) of the six eligible studies. Despite considerable persistence and follow-up, responses were only received from three studies, one of which was still in the inception phase (Table 15). This meant that only two studies were available for follow-up (fortunately these were the two studies recommended by DFID in our terms of reference).

Table 15: Programmes contacted for inclusion in the comparison

Programme	Outcome
1. Zambia Multiple Cash Transfer Programme in Zambia	Included
2. Zomba Cash Transfer Programme in Malawi	Included
3. AGI-K programme in Kenya	Responded but programme still at inception
4. Kenya Cash Transfer Programme for Orphans and Vulnerable Children	Excluded – no response, despite multiple emails.
5. ISHAKA Programme in Burundi	Excluded – initial response, but then communication halted despite multiple emails.
6. BRAC ELA Project in Uganda	Excluded – no response, despite multiple emails.

The evaluation team had further communication with the PIs of the Zambia Multiple Categorical (Cash Transfer) Programme (MCP) and the Zomba Cash Transfer Programme (ZCTP). The exchanges were used to clarify the following aspects of the comparison work:

- The evaluation team will could work with the study focal person to confirm basic information on the study.
- The evaluation team would compare the impact evaluation methodology of both studies, including methods used for randomisation and other approaches for controlling measurement errors.
- The evaluation team could undertake a review of study design documents, protocols and reports to confirm the suitability of the study for valid comparison.

- Procedures and protocols for undertaking data mining for the comparative analysis if needed.

7.2.1.2 Step 2: Confirmation of comparability

Having established all of the above, the evaluation team completed a review of available study design documents, protocols, instruments and reports to confirm the similarity of expected results, indicators and survey questions. Where there were some differences in the way the questions were phrased, these were noted. After this analysis, a brief review document was sent to each of the PIs for confirmation or corrections where necessary.

A summary description of the two studies selected and confirmed for comparison is presented in the Box below.

Box 10: Overview of the selected comparison studies

The Multiple Categorical Cash Transfer Programme

In 2011, Zambia's Ministry of Community Development, Women and Child Health began implementing the Multiple Categorical Cash Transfer Programme (MCP) in two districts. The goal of the MCP is to reduce extreme poverty and the intergenerational transfer of poverty. An impact evaluation with experimental design accompanied the programme in order to determine its effects on recipients and provide evidence for making decisions about the future of the programme. The American Institutes for Research (AIR) was contracted by UNICEF Zambia to design and implement the evaluation. Although the MCP targeted households directly, there was interest in the impact the programme could have on adolescents. For each household, the evaluation included one female household head and up to two adolescents between 13 and 17 years old for interview. The MCP conducted 24-month and 36-month evaluations of the programme. Data on outcomes for adolescents were gathered through a dedicated youth interview, which was administered to the adolescents within a household at baseline and again during the 24-month follow-up data collection (when adolescents were aged approximately 15 to 19 years old).⁵⁹ To increase the sample size in the 36-month evaluation, the age range (16 to 20 years) was expanded to 16 to 23 years and three adolescents were interviewed per household.^{60 61}

The Zomba Cash Transfer Programme

The Zomba Cash Transfer Programme (ZCTP) was a two-year programme, in Zomba district in southern Malawi, that lasted from January 2008 to December 2009. The programme offered cash to households with schoolgirls aged 13-22 who had never been married. Some of the cash transfers were conditional on regular school attendance, while others were unconditional. The goal of the programme was to improve health and education outcomes of adolescent girls and young women through cash transfer social protection. The programme was run by an NGO, financed by the World Bank and its impact evaluated by a Research Support Team. The trial enrolled 3,796 girls and young women from Zomba. The beneficiaries were of two main groups – those who were enrolled in school at the start of the programme (baseline schoolgirls) and those who had dropped

⁵⁹ American Institutes for Research. (2014). *Zambia's Multiple Category Targeting Grant: 24-Month Impact Report*. Washington, DC: Author.

⁶⁰ American Institutes for Research. (2015). *Zambia's Multiple Category Targeting Grant: 24-Month Impact Report*. Washington, DC: Author.

⁶¹ Handa, S., L. Natali, D. Seidenfeld, G. Tembo and B. Davis (2016). Can Unconditional Cash Transfers Lead to Sustainable Poverty Reduction? Evidence from two government-led programmes in Zambia, Innocenti Working Paper 2016-21, UNICEF Office of Research, Florence.

out of school at the start of the programme (baseline dropouts). For the short-term evaluation of the ZCTP, behavioural assessments were done 12 months into the programme while serological assessments (HIV and HSV) were conducted at 18 months.⁶² The final – 60-month (5 year) - evaluation included behavioural and serological assessments.⁶³

The **key finding** from the confirmation work was that all three studies (AGEP and the selected comparison studies) targeted vulnerable girls directly or indirectly (MCP targeted households directly). Although there were small differences in the respective study designs, all three studies were based on randomised cluster controlled trials and used a DiD estimator, although this was only partially the case for the ZCTP study; all analysis adjusted for intra-cluster correlation and a number of covariates.⁶⁴ However an important *difference* is that the main aim of the MCP is household poverty reduction (girls' empowerment is almost an add-on extra benefit for the MCP), whereas it is central to the aims of the Zomba Programme and AGEP.

7.2.1.3 Step 3: Comparative analysis

Data sources and indicators

By mid-August 2016, the MCP 24-month and 36-month evaluation reports and the ZCTP 18-month and 60-month (five year) reports were available for review. So the comparison of results across AGEP and these two studies was conducted using these reports. Data mining for the more in-depth comparison analysis was not considered necessary for the following reasons.

- The results from AGEP showed that there had been minimal impact of the programme on the longer-term impact indicators at this stage.
- The evaluation reports for the comparison studies showed that these programmes had also observed minimal or no impact on most of the indicators being compared.
- The data and results contained in the respective study reports were sufficient for the comparison exercise.

Aligning the study analyses

Unlike AGEP and the MCP study which reported DiD as measure of impact, the ZCTP study mainly reported absolute risk reduction or risk difference (RD). It can be argued that in a randomised experiment, the RD is more or less the same impact estimator as the DiD and so can be interpreted in the same way for the purposes of this comparison.

⁶² Sarah J Baird, Richard S Garfein, Craig T McIntosh, Berk Özler. Effect of a cash transfer programme for schooling on prevalence of HIV and herpes simplex type 2 in Malawi: a cluster randomised trial. *Lancet* 2012; 379: 1320–29

⁶³ Suggested citation: Baird, S, Chirwa, E, McIntosh, C, and Özler, B, 2015. What happens once the intervention ends? The medium-term impacts of a cash transfer programme in Malawi, 3ie Impact Evaluation Report 27. New Delhi: International Initiative for Impact Evaluation (3ie).

⁶⁴ See the Mott MacDonald Preliminary Baseline Comparison Report submitted to DFID in January 2015.

The impact evaluation results for the ZCTP were presented separately for girls who were attending school at baseline and those who were school dropouts at baseline. We used the results for 'attending school' as these were the most relevant for comparison with AGEF.⁶⁵

The ZCTP study also reported impact separately by cash transfer conditionality – the first for conditional cash transfer and the second for unconditional cash transfer. We used both for this comparison exercise and their statistical significance was generally the same. So, in Table 16 below, ZCTP impacts are reported for conditional transfers and unconditional transfers, with the first figure (top of the cell) showing the impact for conditional cash transfers, and the second figure (bottom of the cell) showing the impacts for unconditional cash transfers in almost all cases,⁶⁶ the statistical significance was the same.

Due to the differences in the underlying theory of change between AGEF and the cash transfer studies, the cash transfer studies did not report on empowerment indicators in the same way as AGEF. However, the MCP did measure the effect of the intervention on adolescents' 'life outlook'. We, therefore, used this indicator as a comparison for the AGEF indicator on self-efficacy.

7.2.2 Results of the comparison

The findings of the comparative analysis are presented around the current version of the AGEF framework, namely against empowerment outcome indicators and longer-term impact indicators. Table 16 compares the impact of AGEF to shorter and longer-term impacts of the MCP and the ZCTP. Two expanded tables of results which contain the mean (frequency) of the outcomes are contained in Annex 12 (pp. 76-77). Although the means were not always available from the reports, in all cases, the size and significance of impact could be determined.

7.2.2.1 Comparison of effects on empowerment indicators

As explained above, it was difficult to find indicators that were common to all studies that could be compared to AGEF's empowerment outcomes. However, we considered that the MCP indicator on life outlook was similar to the AGEF indicator on self-efficacy, so we included a comparison of these indicators in Table 16 below. The results show that after 24 months, neither programme had a significant effect on these indicators. In AGEF, the impact on average score on self-efficacy scale was 0.14 while in the MCP, it was -0.02; neither was statistically significant (see Table 16). There were no similar empowerment indicators in the ZCTP.

7.2.2.2 Comparison of effects on longer-term impact indicators

Thirteen AGEF longer-term impact indicators were used for the comparative analysis (Table 16). Of these 13 longer term impact indicators, a positive effect was only seen on condom use at first sex and transactional sex at 24 months among AGEF girls.⁶⁷ While condom use at first sex increased by a mere 3%, transactional sex decreased by a considerable 10%

⁶⁵ At baseline and Round 3, most (81%) AGEF girls were attending school. At the 18-month follow-up, 90% of the girls in the ZCTP were currently attending school.

⁶⁶ See Notes beneath the table.

⁶⁷ This finding is based on the ITT, DiD analysis for all girls, and not the disaggregated sample (see Annex 4, pp.23-25)

points. Of the 13 indicators, eleven were reported in the MCP evaluation reports. Of these eleven, a positive effect was only seen for the education indicators at 24 months and 36 months.

Notably, the MCP reported school enrolment, while AGEP reported school completion; however, we suggest the results are comparable because the MCP analysis was restricted to narrow age bands: 11 to 14 years for primary school enrolment and 15 to 17 years for secondary school enrolment. Overall, the MCP had a positive impact on current school enrolment. It did not have an effect on primary school enrolment among girls aged 11 to 14 years; however, it did have a 19% points impact on junior secondary school enrolment among girls aged 15 to 17 years at 24 months. This impact was sustained at 36 months although it declined to 11 percentage points.

In the ZCTP study, primary and secondary school completion was measured in a similar way to AGEP. Although the ZCTP had a six percentage points impact on school enrolment at 12 months, it had no impact on girls' completion of primary or junior secondary school.

In the short-term, the ZCTP significantly decreased girls' likelihood of getting married by three percentage points and of acquiring HIV and HSV by two percentage points. However, these impacts did not remain in the long term (Table 16).

7.2.3 Concluding comments on the comparison of epidemiological effectiveness

For this comparative analysis, we compared AGEP to two cash transfer studies, one in Zambia and one in Malawi, using 14 indicators. The results showed that all three programmes had minimal impacts on the set of indicators used for comparison. These findings were based on the 24-month impact evaluation report for AGEP; we extended this comparison to the 24 and 36 month evaluation reports for the MCP, and the 18 and 60 month evaluation reports for ZCTP.

Several observations emerge from these results. First, each programme appeared to have impacts on different domains of outcomes. While AGEP mainly had an impact on sexual risk behaviours at 24 months, the MCP had an impact on educational outcomes only. The ZCTP impacts were mainly seen in sexually transmitted infections at 18 months -but, importantly, these effects were not sustained at 60 months. It should also be noted that, although the impact of the MCP on educational outcomes at 24 months persisted until 36 months, the size of the impact declined across the three indicators of interest.

It might be argued that AGEP has performed comparatively well because it has shown significantly positive effects on two sexual risk impact measures at 24 months, although the impact on condom use at first sex was rather minimal.⁶⁸ Moreover, the AGEP theory of change suggests that the interventions will produce additional positive effects on impact indicators after four years (once younger girls have made the transition to maturity). As we

⁶⁸ In addition, AGEP has produced significant positive effects across the study sample on empowerment indicators for safe space to meet, savings in the past year and financial literacy and contraceptive knowledge. However, these indicators could not be compared across the selected studies.

have seen, this would be a departure from trends found in the two cash transfer studies, and so would be an important finding.

We, therefore, suggest that a definitive assessment of how well AGEF compares to other approaches in terms of epidemiological effectiveness will only be possible once the AGEF Round 5 survey has been completed in 2017.

Table 16: Comparison of AGEP short-term impact to MCP and ZCTP longer-term impact on educational and sexual and reproductive health outcomes

	AGEP 24 Months		MCP 24 Months		MCP 36 months		Zomba 12/18 months		Zomba 60 months	
	DID (%)	p-value	DID (%)	p-value	DID (%)	p-value	RD ^a (%)	p-value	RD ^a (%)	p-value
Empowerment indicators										
Avg. score on self-efficacy scale (0-10) ^b	0.14	NS	-0.02	NS	1	NS	Not measured			
Longer term impact indicators: education										
% currently attending school ^c	0.1	NS	11	Sig	9	Sig	6	Sig	Not reported	
% completed primary school ^d	0	NS	5.9	NS	6.9	NS	3; 4	NS	-1.3; 1.6	NS
% completed junior secondary school ^e	-1	NS	19	Sig	11	Sig	-1.2; 0.3	NS	3.3; 1.4	NS
Longer-term impact indicators: sexual risk behavior										
% ever had sex ^f	3	NS	-3	NS	0.8	NS	0.7; -0.4	NS	0.4; 0.5	NS
Age at first sex	4	NS	-0.08	NS	-0.1	NS	0.19	NS	0.2; -0.2	NS
% transactional sex	-10	Sig	-3	NS	1.6	NS	Not reported			
Number of sexual partners past 12 months		NS	-0.09	NS	-0.33	NS	-2.1; -3.7	NS	-0.01; 0.1	NS
% condom use at first sex	2.9	Sig	-2	NS	1	NS	Not measured			
% condom use at last sex ^g	2.9	NS			3.2	NS	4	NS	1.0; 5.6	NS
Longer-term impact indicators: marital and pregnancy^c										
% ever married ^f	3	NS	0	NS	0.1f	NS	0.1; -3	Sig	-3.7; -1.2	NS
% currently pregnant or who have given birth	1	NS	1.8	NS	1.6	NS	0.8; -1	NS	-2.6; -0.4	NS
Longer-term impact indicators: sexually transmitted infections										
% HIV positive	0	NS	Not measured				-2.0	Sig	-0.1; -0.6	NS
% HSV-2 positive	1	NS	Not measured				-2.3	Sig		

Notes for Table 16:

- a=The Zomba evaluation mostly reported risk difference (the first figure (top of the cell) represents the impact for conditional transfer, while the second (bottom of the cell) is for unconditional transfers -in almost all cases, the statistical significance were the same);
- b=Self-efficacy was not measured in the MCP; life outlook was measured (% of adolescent girls who believed life would be better in five years);
- c=currently attending secondary school and it is not disaggregated by gender;
- d= The MCP reported primary school enrolment for children aged 11 to 14 most of whom would most likely complete grade 7 (primary);
- e=The MCP reported junior secondary school enrolment for children aged 15 to 17 most of whom would most likely complete grade 9 (junior secondary), not disaggregated by gender in 36-month report. Report says effect was the same in males and females;
- f= In the MCP, *ever had sex and ever married* included male adolescents.
- g=In the short-term Zomba report, condom use as defined as unprotected intercourse – inconsistent condom use. In the 36-month MCP report, this was defined as condom never used in the last three month;
- DiD=Difference-in-Difference;
- RD=Risk Difference;
- NS=not statistically significant at p=0.05;
- Sig=statistically significant at p=0.05.

7.3 Economic comparison

7.3.1 Adding costs to the comparison of AGEP, ZCTP and the MCP

For observations on the costing data found in our literature review, see Annex 11 (pp. 73-74).

Table 16 above shows the extent to which the three comparison studies – AGEP, the Malawi ZCTP and the Zambia MCP – had improved indicators relevant to adolescent girl empowerment at around the two-year mark. It is important to remember that the indicators in Table 16 have been selected because they are comparable between the studies and because they are relevant to adolescent girls. Table 16 does not capture all the benefits produced by each initiative – e.g. the improvements in financial literacy produced by AGEP do not feature in the comparison because this indicator was not measured by the other studies; similarly, Table 16 does not include the important improvements in household income associated with the MCP.

In terms of a cost comparison, another problem is that we do not know much about the cost structure of the Malawi ZCTP (the recent 60-month report did not discuss costs and a recent review of cost-effectiveness of interventions related to adolescents confirmed that the Zomba study did not look specifically at cost-effectiveness).⁶⁹ We know that the average

⁶⁹ Bolton L. HEART Helpdesk Report. *Cost-effectiveness of intervention for improving adolescent sexual and reproductive health*. June 2016.

total transfer per beneficiary household was \$10 per month for 10 months per year – i.e. \$100/year. We also know that secondary school fees were paid for girls, but we do not know how much this cost. It would also be appropriate to add a further 50% to the total to cover management costs.^{70 71} A figure of \$180 per girl is used here, but this is little more than an educated guess.

We have established that AGEP costs \$771 per girl. We know that the ZCTP cost approximately \$360 per girl over two years, and the MCP cost about \$322 *per household* over two years. To compare like with like, we take a notional sum of money (\$50,000 is used here) and see what each programme could “buy” in terms of girls’ empowerment. This is shown in Table 17.

Table 17 needs to be interpreted cautiously. The cost estimate of ZCTP is an educated guess. Most importantly, this table is only using selected indicators – what it does not capture is that the MCP was benefitting households as a whole, and concentrated primarily on improving household production and consumption.

Table 17: Comparison of costs and effects relating to adolescent girls

Programme	Units covered for \$50,000 spent over two years	Gains at 18/24 months relevant to girls’ empowerment
AGEP	65 girls	<ul style="list-style-type: none"> <input type="checkbox"/> 7 girls avoided transactional sex <input type="checkbox"/> 2 instances of condom use at first sex
Zambia MCP	155 <i>households</i>	Assuming each household included one adolescent girl: <ul style="list-style-type: none"> <input type="checkbox"/> 17 girls attending school <input type="checkbox"/> 29 girls completing junior school
Malawi ZCTP	139 girls	<ul style="list-style-type: none"> <input type="checkbox"/> 8 girls attending school <input type="checkbox"/> 4 delayed marriages <input type="checkbox"/> 3 cases of HIV prevented <input type="checkbox"/> 3 cases of HSV-2 prevented

⁷⁰ Baird, S J et al. *Effect of a cash transfer programme for schooling on prevalence of HIV and herpes simplex type 2 in Malawi: a cluster randomised trial*. Lancet 2012; 379: 1320–29

⁷¹ As acknowledged in the Lancet article, the management costs of the Zomba programme were particularly high.

7.3.2 Other relevant work on the economics of cash transfer programmes in Zambia

7.3.2.1 Review of relevant costing studies

For this section, we have included an additional comparison with the Zambia Child Grant Programme (CGP). Although this programme does not meet our criteria for comparing epidemiological effectiveness, it does provide useful information for contextualising the AGEP economic evaluation.

As far as we are aware, two relevant economic evaluations have been conducted on social cash transfer programmes in Zambia – a cost-consequence analysis of the Child Grant Programme (CGP) with calculations of multiplier effects for the CGP, and an economics assessment of the Zambia MCP.

The findings of the cost-consequence analysis for the CGP are shown in the Box below. This is a classic cost-consequence analysis (CCA), with a list of findings expressed in words and numbers. The example demonstrates the strengths and weaknesses of CCA. On the plus side, all the major outcomes of the programme can be seen without an undue emphasis on one particular type of outcome. More problematically, the findings do not allow for easy comparison with other programmes because the list of outcomes is tailor-made for that particular programme. The example of the CGP is included here for illustrative purposes: we note that there are no outcomes in common between this programme and AGEP.

Box 11: Cost-effectiveness ratios 2010-12, Zambia Child Grant Programme

Impact indices	Cost-effectiveness ratio (USD)
Increase of ZMW 1 in the monthly per capita consumption expenditure	\$7.68
One percentage point reduction in poverty gap and squared poverty gap	\$10.47
One percentage point increase in households with 2+ meals per day	\$71.96
One percentage point increase in proportion of children aged 6-24 months receiving minimum feeding requirements	\$5.23
One percentage point increase in households owning any livestock	\$27.41
One percentage point increase in households owning any non-farm business assets	\$127.92

Source: Jesse C et al (2014) Cost analysis of Zambia's social cash transfer programme. AIR.

Multiplier effects have been calculated for the CGP and the MCP initiative in Zambia, as shown in the Table 18 below.⁷²

⁷² Source: Handa S et al (2016) Can unconditional cash transfers lead to sustainable poverty reduction? Evidence from two government-led programmes in Zambia. Innocenti Working Paper 2016-21, UNICEF Office of Research, Florence.

Table 18: Estimated multiplier effects of the Child Grant Programme and Multiple Categorical Programme

	Child Grant Programme	Multiple Categorical Programme
24-month impacts	1.64 [0.96-2.33]	1.38 [0.63-2.10]
36-month impacts	1.23 [0.65-1.81]	2.08 [1.09-3.07]
Pooled impacts	1.46 [0.88-2.05]	1.72 [0.94-2.50]

The multiplier effect for the CGP (1.46) is computed as the ratio of the sum total of annualized spending impacts over the annual value of the transfer; spending impacts include: consumption, savings, loan repayment, livestock purchases and productive tools.⁷³

The average multiplier for the MCP so far, is 1.72. This means that, for each one Kwacha transferred to a household, there is an additional 0.72 Kwacha more in terms of net benefit to the household. The multiplier works through increased productive activity, including livestock rearing, agricultural production and diversification of income sources.

The multiplier is, in effect, a cost-utility analysis calculation, but one that is particularly easy to understand because both the costs and the consequences use the same unit of measurement, namely money. The multiplier only captures one dimension of the benefits of the cash transfers (consumption/ food security): it does not monetize the benefits of changes such as increased schooling. If we want to compare all the benefits of the cash transfer programmes – or if we want to compare cash transfer programmes with programmes such as AGEP which do not have consumption/food security as their primary outcome - we must revert to Cost-Consequence Analysis, i.e. a structured comparison of lists of measured impacts.⁷⁴

7.3.3 Concluding comments on the economics comparison

Box 12 below brings together the various analyses above about AGEP and the MCP in the style of a Cost Consequence Analysis. This still does not capture all the indicators measured for the MCP, but it does now include the financial multiplier.

⁷³ Impacts are based on econometric results. Only statistically significant (at the $p < 0.05$ level) impact estimates are considered. Loan repayments were not measured in the CGP at 24 months.

⁷⁴ Cost-Benefit Analysis has been ruled out as a possibility (see Section 6.1), even though in theory this would produce a purely numerical comparison between multi-impact programmes.

Box 12: Spending \$50,000 on AGEP and on the MCP

Spending \$50,000 over two years.....

AGEP produces:⁷⁵

- About 4 girls who have a safe space to meet with friends
- An average increase of 0.3 in the financial literacy scale for 65 girls
- An average increase of 0.2 in the contraceptive knowledge scale for 65 girls
- Almost 4 girls who have saved money in the past year.
- 7 girls prevented from having transactional sex
- 2 instances of condom use at first sex

However, at the same time there was also a significant decrease in mobile phone ownership.

MCP produces:

- \$36,000 net benefit shared amongst 155 households (greater productive activity, based on 1.72 multiplier)
- 17 girls attending school (assuming each household included one adolescent girl)
- 29 girls completing junior school (assuming each household included one adolescent girl)

It is useful to note that, in Zambia, the thinking about cash transfers is moving towards the idea that cash transfers can be complemented by specific interventions (e.g. nutritional supplements or behavioural change training). A recent Overseas Development Institute (ODI) review of cash transfers⁷⁶ also confirmed that the balance between spending on cash transfers and complementary interventions is a relevant policy question, but it did not discuss any evidence about when and whether the complementary interventions are cost-effective in terms of achieving marginal added value.⁷⁷

⁷⁵ Data from Table 13. Note the discussion about whether impact indicators should be included at the 24-month stage.

⁷⁶ Bastagli F et al. (2016). Cash transfers: what does the evidence say? A rigorous review of programme impact and of the role of design and implementation features. London: ODI

⁷⁷ We followed this up in correspondence with the authors of the ODI paper, who confirmed that they were not aware of any literature on the cost-effectiveness of interventions which complemented a cash transfer programme.

8. Discussion

In the course of presenting the findings of the independent End Term Evaluation of AGEP, we have indirectly referred to most of the evaluation questions. In this section, we revisit the five evaluation questions posed by DFID to ensure they have been adequately addressed.

8.1 Evaluation Question 1

Has the Safe Spaces programme and research been well-implemented? What lessons have been learned from implementing the Safe Spaces programme and research in Zambia?

We have aimed to address Evaluation Question 1 within a theory-based evaluation framework. We are confident that the programme of interventions has been well managed and that it has been fully implemented *as designed* by the Population Council and its partners. However, both the impact evaluation research and the findings of our independent evaluation point to the need to ‘fine-tune’ the Safe Spaces model to better meet the distinct needs of older and younger adolescent girls, and to respond to differences in urban/rural context. We have summarised key lessons emerging from the experience of implementing the AGEP model in Zambia in Box 15 at the end of this section.

In the course of this report we have noted that not all of the ‘assumptions’ defining the conditions for successful implementation have been met. These combined factors, along with evidence from our literature review (see Annex 11), point to the fact that the AGEP model could be strengthened by additional strategies to address structural and normative barriers –and that these strategies should include stronger community engagement efforts.

We acknowledge that, at this stage, our evaluation findings are based on *intermediate* results from the AGEP research. We accept that the AGEP impact evaluation research is powered to assess results over four years of observation, once the girls have transitioned to sexual maturity. We have established that the AGEP research to date is robust (see Annexes 6 and 7), and is likely to yield credible findings once the Round 5 survey has been completed in 2017. It is, therefore, somewhat premature to make definitive statements about the effectiveness of the AGEP interventions at this time.

All this said, results from the research point to an important question about whether the effects of asset building have been sufficiently strong to see the girls through transitions to maturity. There are also emerging questions about the durability of intermediate results, the likely impacts for older girls who have already made sexual transitions and the extent to which the research has curtailed some potentially positive effects of ‘spillover’ (see Box 13).

Box 13: The Spill over dilemma

In keeping with the requirements of a robust impact evaluation design that would allow intervention outcomes to be attributed to AGEP, the Population Council has sought to minimise the contamination effects of spillover. However, it seemed there was still some potential for spillover due to the geographical proximity of the intervention and ‘internal control’ sites -especially in urban areas. Consequently, the Mott MacDonald evaluation team recommended that the Council establish more distant ‘external control’ sites so that the effects of spillover could be estimated (results from the five external control sites established are still awaited).

It must be acknowledged however, that under non-research conditions, spillover effects might be encouraged to enhance the effects of AGEP by building up peer networks and a momentum for social change.⁷⁸ This means that the impact evaluation research could be underestimating the effects of a ‘fully fledged’ AGEP programme.

It is clear that the Round 5 survey will need to focus on addressing some important questions about the effectiveness of AGEP and the validity of the underlying theory of change (see Section 9.2.1). We have observed from interim results that there are likely to be differential results for urban and rural girls. There is a need to explain these effects for both older and younger girls and the pathways to longer term results. There is thus an important role for more complementary qualitative research. This research should focus on helping to explain *why* impact level results are or are not achieved, as well as any statistically weak or negative effects. There may also be a case for additional analysis to identify the subset of girls for which the programme had the highest impact in order to identify key factors in success.

8.2 Evaluation Question 2

Is Safe Spaces the most cost-effective and efficient approach to empower adolescent girls? How does it compare with alternative empowerment programmes?

We have aimed to address Evaluation Question 2 in Parts 2 and 3 of our evaluation findings.

We have shown that setting up credible comparisons is detailed work that requires access to data from other studies that are sufficiently similar to AGEP in terms of objectives, indicators, targeting, timeframes and context. We have established that the Zambia MCP and the Malawi Zomba studies offer useful comparisons, both in terms of our comparison criteria *and* conceptually as an alternative approach (see Annex 12). Our comparison work shows that, after 24 months of interventions, AGEP compares reasonably well with the other programmes in terms of epidemiological effectiveness. Using comparable indicators, AGEP has shown significant positive effects on transactional sex and condom use at first sex. However, the MCP and ZCTP studies have shown positive impacts on educational outcomes and other sexual health outcomes after 24 months respectively. The ZCTP study

⁷⁸ See UNFPA 2007. Framework for Action on Adolescents & Youth: Opening Doors with Young People: 4 Keys (pages 37-40). <http://www.unfpa.org/publications/framework-action-adolescents-and-youth>

is also a salutary reminder that good initial results may not be sustained over the longer term.

Several development partners and MoH key informants made the important point that the AGEP model and cash transfer initiatives are not necessarily mutually exclusive; rather they are potentially complementary. It was suggested, too, that cash transfers may be especially useful for supporting the economic empowerment of the *most vulnerable* girls (see discussion points in Section 8.3 below).

In order to support cost comparisons, we have made several suggestions to help ensure the Population Council's economic evaluation remains reliable, technically strong and well aligned to the impact evaluation research. We have explained that "cost-effectiveness" in the context of a complex programme is a value judgement, and not purely a technical decision. We have shown what \$50,000 can "buy" in terms of results at the two-year stage, while noting that the best time to do economic comparisons is when the final round of results is available.

8.3 Evaluation Question 3

Is the Safe Spaces programme sustainable (i.e. are programme impacts likely to last) and can the model be scaled up (would processes allow and is it affordable)? Is there government commitment and will donors engage and coordinate to support this programme?

In Part 1 of our evaluation findings we have commented on the potential for sustained impacts and programme scale-up for each component of the intervention. Once again, it is too early to discuss longer-term impacts. We are aware, however, that DFID is considering a second phase of programming for adolescent girls, so we have identified a number of factors for consideration, as well as some design questions that need to be addressed (see Annex 10 for a detailed account). We suggest that the work on Cost Consequence Analysis is also relevant to sustainability because it informs the question of whether there will be an appetite amongst funders to pay for a programme such as AGEP.

We note that the Zambian Government and a number of international development partners recognise the importance of a strategic focus on vulnerable adolescent girls, and there is potential for rolling out Safe Spaces interventions within the institutional setting of the Ministry of Community Development and Social Welfare. This focus on the 'most vulnerable girls' does, however, throw up some design and cost-effectiveness challenges that need to be addressed if the AGEP model is to be scaled-up (Box 14).

Box 14: The focus on vulnerability – some challenges

The Population Council based its recruitment of girls on a ‘vulnerability index’ that took account of individual and household characteristics, including material assets, socio-economic and educational status, and number/type of dependents.⁷⁹ A number of observations on targeting vulnerability have emerged from the independent evaluation research:

- **Development of a standard definition of vulnerability.** Our interviews suggested that there may be multiple perspectives on vulnerability that could make it difficult to arrive at a standard definition for targeting girls in urban and rural areas. For example, some community members took account of concepts, such as collective or ‘community vulnerability’ (e.g. due to lack of security, local gang, drug or criminal culture, a fragile agricultural environment, and lack of political affiliations).
- **Vulnerability may not be a fixed state.** In-depth interviews with adolescent girls indicated that their vulnerability status often changed over time and was sometimes mitigated by household coping strategies and support from social networks (such as movement of girls to a relative’s house or seasonal employment among family members).
- **Targeting or labelling?** Some mentors expressed concern that selection of girls for inclusion in AGEP could be associated with ‘labelling’. This, in turn, may have discouraged sustained participation in AGEP.
- **Targeting vulnerability requires resources and administrative time.** Interviews with AGEP implementation staff suggested that recruiting girls based on vulnerability scores and maintaining the exclusivity of Safe Spaces groups required a significant allocation of time and resources.⁸⁰ Key informant interviews also indicated that targeting of the most vulnerable girls could reduce the viability of voucher and savings schemes.

Our interviews with the MoH suggested that there may be little appetite for scaling-up health vouchers as they are not an explicit part of current government plans and budget allocations. Nevertheless, there is considerable interest in the health vouchers *experience* and lessons learnt from AGEP; it is, therefore, important for the Population Council document these thoroughly.

We found, too, that the Girls’ Dream savings account is unlikely to be sustained by NatSave, unless there are further partner investments or adaptations made to the product. This said, our interviews with NatSave suggest that dialogue on the product has stimulated discussion within the institution on the benefits of investing in products for young people.

We, therefore, suggest that, along with the longer-term impacts, the scalable model of AGEP has yet to be finalised. We have observed that, especially in urban areas, there are a number of other initiatives that could complement or compete with AGEP (e.g. church and sports groups). There remains a need to explore the marginal effects of these initiatives quantitatively and qualitatively. We also suggest that there are important opportunities to work collaboratively in learning partnerships with other organisations in Zambia to rapidly determine what works at scale in empowering vulnerable adolescent girls across a range of different contexts (see Annex 10, p.57).

⁷⁹ See the Population Council’s *Research and Evaluation Baseline Technical Report* (2014: pp7-8 and 26-28) for a full account of how the vulnerability index was applied and the implications for the impact evaluation research.

⁸⁰ These challenges relating to ‘targeting’ are similar to those associated with conditional cash transfer programmes - see DFID (2011), *Cash Transfers Evidence Paper*, Policy Division, April 2011

8.4 Evaluation Question 4

What changes should be made to the programme implementation in future?

As indicated under Evaluation Question 1, there have been a number of lessons from implementation of the AGEP that should inform any future programme. Once again, we acknowledge that final results from the impact evaluation of the AGEP model are not yet available, so our conclusions are tentative. We note, however that, based on evidence cited in the Business Case, the health vouchers component and savings account component were the most innovative elements of the model. At this stage (the end of the intervention phase), the Population Council reports that these components have not (yet) produced significant marginal effects. Given that these components are resource intensive (in terms of human and financial resources), it is important to continue to reflect on whether these approaches are the most cost-effective ways to build the health and economic assets of vulnerable adolescent girls.

Box 15 at the end of this section lists a number of operational lessons that have implications for programme implementation. Any changes to the programme *design* would need to be reflected in the theory of change and take account of any assumptions that have not been upheld. Based on the findings of the independent evaluation and a review of the literature on 'what works' in adolescent girls' empowerment (Annex 11), we have developed an annotated version of the theory of change to identify areas where available evidence points to *potential* ways of strengthening the AGEP model in the future (see Annex 13).

8.5 Evaluation Question 5

Have the research findings from the Safe Spaces programme been sufficiently used to influence policy debate and the development and implementation of new programmes?

The Population Council has systematically documented the design and implementation of its impact evaluation research. It has also reported on preliminary findings in a number of technical reports, programme briefs, conference papers and posters. These are listed in Annex 14. At the end of July 2016 the Population Council shared its draft Mid-Term Technical Report (which incorporated findings from the Round 3 survey) and a report from the qualitative research. Our review of all this documentation indicates that they are all of a high standard in terms of data quality, analysis, presentation and transparency.

The Population Council has been active in disseminating its communication products and publications through a series of meetings and events and has reached out to government ministries at all levels, as well as the non-governmental sector and donors in-country (Annex 14). We suggest that DFID and the Population Council will need to agree a joint communication strategy for the final survey findings to ensure that they continue to contribute constructively to the evidence base, and the wider learning agenda on the empowerment of vulnerable adolescent girls.

8.6 Summary of lessons

For ease of reference we have summarised the main lessons arising from review of the impact evaluation research and the independent evaluation in Box 15 below.

Box 15: Summary of lessons emerging from implementation of the AGEP model

Key lessons on Safe Spaces and building social assets

- **Lesson 1: The quality of mentors is critical for the success of Safe Spaces.** Important considerations relate to the recruitment and training of mentors and to their management. With regards recruitment and training of mentors, key factors are: a) *mentors may need specific competencies* for running Safe Spaces urban and rural contexts and for working with girls in different age cohorts; b) *mentors in rural areas face particular challenges* as they often need to cover long distances and manage multiple groups (as suitable mentors are more scarce in rural areas); c) *community support is important* -there are high expectations that mentors will be positive role-models; however, mentors may need to challenge harmful beliefs and practices and need to be able to manage conflict with community members; d) *values clarification is important* -mentors need to be able to present information in a values-neutral way; however, they must also be able to engage in relevant and authoritative dialogue with girls and their communities. With regards management of mentors, key factors are: a) *appropriate supportive supervision and performance monitoring is critical* and requires appropriate infrastructure and resources -notably, distances and lack of connectivity can make this more challenging in rural areas; b) *mentors may need to work longer hours* to follow up on girls who drop-out or who need additional support; c) *the system of stipends needs careful management* to take account of additional hours and work with multiple groups -opportunities for career progression and professional development also need to be considered over the longer term; d) *mentors could benefit from access to a network of specialist support* -e.g. from police officers and social workers, to deal with challenging problems or sensitive issues; e) *the system of 'alternates' has proved useful* as it is sometimes necessary to replace weaker mentors or those who drop out -this requires appropriate allocation of resources.
- **Lesson 2: The challenge of maintaining participation needs to be addressed.** Key issues for consideration are: a) *the system of 'prizes' needs review*: although anecdotally the system helped to maintain some participation, it was logistically challenging to administer -moreover, it was a frequent source of controversy and may have contributed to some cases of drop-out; b) *the optimum size for Safe Spaces groups remains uncertain*: the size of Safe Spaces groups fluctuated over time (there were often large numbers at the beginning because girls brought friends, and much smaller numbers at the end due to drop-out) -there is, therefore, a need to establish what group size creates the most effective dynamic and what is practical in terms of mentor management; c) *continuous efforts are needed to engage the most vulnerable girls*: anecdotal evidence suggests that the most vulnerable girls (such as girls with disabilities) often drop-out quickly due to access issues, while vulnerable girls often have multiple domestic responsibilities -additional efforts may be required to address such barriers.

- **Lesson 3: Time, resources and skills are required for community engagement.** Key factors for consideration are: a) *full parental and community buy-in is essential* for girls to participate in Safe Space meetings; however, this buy-in requires ongoing dialogue and inputs from mentors and the programme team -the concept of a “safe space” is open to misinterpretation and it takes time to build up community trust; b) *active community engagement can lead to useful exchanges* on social norms and generate complementary community initiatives.
- **Lesson 4: There is scope for strengthening the Safe Spaces curriculum.** Although the Safe Spaces curriculum was technically well-designed and popular, evidence from the independent evaluation points to a need to increase its effectiveness in terms of knowledge retention. Feedback from stakeholders suggests that the curriculum needs to be: a) *customised so that it is more age-appropriate*; b) *practically-oriented* for “less academic” girls and place more emphasis on *skills-building* for income-generation and health/well-being promotion. The independent evaluation research also suggested that there is scope for: a) *improving context relevance* for urban and rural girls; b) *establishing the optimum duration* of the Safe Spaces intervention and c) determining whether there *are benefits to ‘post-graduation’ follow-up* and support.

Key lessons on health vouchers and building health assets

- **Lesson 5: The health voucher scheme has been an innovative demand-side initiative, but its effectiveness and sustainability have yet to be demonstrated.** Key considerations include: a) *in Zambia, there is a need to calibrate voucher reimbursements* to optimise the motivation of staff in government, non-governmental/private facilities respectively; b) *the leadership role of Heads of Facilities* can be pivotal in determining responses to the initiative; c) there is a need to take account of *broader health system supply and demand side issues* in urban and rural contexts and across sectors, as well as health seeking behaviour and barriers to service uptake for adolescent girls; and d) there is a need to include *more intensive community and parental sensitisation work*, so that parents and guardians understand the benefits of girls attending facilities for preventive health services (see Annex 9 (p. 53) for interview extracts on this theme).

Key lessons on savings accounts and building financial assets

- **Lesson 6: Customised bank accounts can be catalytic for informal and formal savings activity** (see the Population Council’s Mid-Term Technical Report, 2016. However, important considerations are: a) *considerable time and resources are needed* to support opening of accounts at scale; b) *distance from the bank is a key barrier* to maintaining the account - especially in rural areas; c) *other traditional savings practices* may offer a useful alternative to bank savings accounts (such as *Chilimba* savings schemes).
- **Lesson 7: There is some evidence of progress in building financial assets** (although at this stage the specific effects are different for urban and rural girls across the two age cohorts). Evidence from the independent evaluation research suggests that the financial education curriculum is likely to have made an important contribution to positive results. However, an important consideration is that *the intervention design needs to take account of the amount of savings needed* to bring about intended impact-level results -for example, rural girls require almost 3 000 Kwacha per year to cover the cost of secondary school fees, hostel and material expenses, yet most AGEP girls were only able to save a few hundred Kwacha (formally and informally) over the duration of the intervention.

9. Conclusion and recommendations

9.1 Conclusion

The independent End Term Evaluation of the AGEPE initiative has been conducted at the end of AGEPE's two-year intervention phase, but before the four-year impact evaluation research has been completed. The findings of this independent evaluation are thus based on a review of the AGEPE programme and the research results available to date. Our findings should, therefore, be seen as interim and based on an assessment of the 'direction of travel'.

In the course of this independent evaluation, we have conducted a theory-based assessment to review the emerging strength of evidence for the AGEPE theory of change. We have also assessed the status of the Population Council's economic evaluation and conducted a preliminary comparison with alternative (cash transfer) approaches to empowerment of vulnerable adolescent girls. Based on the aggregation of these evaluation assessments, we have sought to address the five evaluation questions posed by DFID. However, conclusive answers to these questions all hinge on the results from the final phase of the impact evaluation research.

9.2 Recommendations

9.2.1 Recommendations for the Round 5 research

In order to maximise the contribution of the AGEPE impact evaluation research to the evidence base on adolescent girl's empowerment, we suggest that *analysis* of the Round 5 data should include a focus on a number of key questions (see Box 16 below).

The questions listed in Box 16 relate, not only to the effectiveness of the AGEPE model for delivering intermediate and longer-term results, but also to the cost implications and the explanations for any differential results (positive or negative). The latter elements will require additional efforts to ensure that: a) the economic evaluation remains aligned to the impact evaluation research and includes more detailed contextual information on rural and urban costs;⁸¹ and b) more focused qualitative research explores how and why successive results have emerged for different segments of the adolescent girl population.

⁸¹ Especially in terms of numbers of girls reached in each context with the precise locations of the clusters.

Box 16: Specific Questions for the Round 5 Research

The following specific questions for the Round 5 research arise from the findings of this independent End-Term Evaluation. As indicated by the sequencing of the questions below, we suggest that the usefulness of the Round 5 research for programme learning could be increased by building stronger *iterative* links between the quantitative, qualitative and economics research.

Cross-cutting questions

- What is the evidence that the AGEP model has led to building of social, economic and health assets and that these, in turn, have led to longer-term empowerment effects that are attributable to the programme?
- Are there differential results for younger/older girls and girls in different contexts (urban/rural or specific locations)?
- What should be the intermediate milestones for assessing whether longer term impacts will be delivered? Are measures for key intermediate milestones (such as self-efficacy) sufficiently sensitive? Does the qualitative research point to better ways of measuring changes in empowerment indicators?
- What are the explanations for any differential results? What are the implications for the AGEP theory of change?
- What are the cost-effectiveness implications of these differential results? Will there be enough cost information available to answer this question?
- Are there any marginal effects that can be attributed to the health vouchers component and/or the saving account component?
- Does the qualitative data support the AGEP theory of change for these marginal effects?
- Can these marginal effects be considered cost-effective?
- What is the evidence that the AGEP interventions have benefited the most vulnerable girls (based on analysis of individualised data from urban and rural areas)?
- What are the differentiated cost estimates for targeting/recruiting vulnerable adolescent girls in urban and rural areas? What are the implications for the cost-effectiveness of the AGEP model?
- Does evidence from the external control sites suggest that there may have been significant spill-over to the internal controls? What are the implications for the interpretation of results?
- What are the implications of the overall findings from the Round 5 research (the survey, qualitative research and the economics evaluation) for the AGEP theory of change and for the design of the scalable model?

Specific questions on the Safe Spaced component

- What are the qualities and competencies of a good mentor? Are there particular competencies required for mentors a) working in urban/rural areas and b) working with younger/older adolescent girls or married girls/girls with children? Can mentor characteristics help explain any differential results? What are the implications for the scalable model?
- Does the data (quantitative and qualitative) show that the effectiveness of Safe Spaces groups (for urban/rural, older/younger girls) is associated with an optimum group size (for older/younger, urban/rural girls)? What are the implications for the scalable model?
- What strategies have been most effective for community engagement. What are the implications for the scalable model?

9.2.2 Additional recommendations

The Mott MacDonald evaluation team's more general recommendations are:

- **Align indicators and conceptual frameworks:** For the remainder of the initiative, ensure there is clearer alignment between the logframe, theory of change and the research analysis plan, so there is less ambiguity around how success will be measured and interpreted. Alignment of indicators applies to the economics work too. The Round 3 work needs to be re-done, using the appropriate empowerment indicators, and incorporated into the main Round 3 report. The economics work for Round 5 should be planned well in advance to ensure that it addresses all the relevant issues.
- **Document and disseminate lessons from programme implementation:** Although final results are uncertain, there are a number of operational lessons relating to the implementation of Safe Spaces, as well as the innovative health vouchers and savings accounts components. These lessons need to be disseminated widely to inform future programming. They should also form part of longer-term AGEP communication strategy for publishing the findings of successive survey rounds.
- **Work within collaborative learning partnerships:** There are currently a number of other government, non-governmental and UN initiatives that are targeting adolescent girls in Zambia (see Annex 11). We suggest that, with coordination, there is huge potential to rapidly expand the evidence base on adolescent girls in Zambia through partnerships that work in dynamic and complementary ways (see Annex 10). A 'community of learning' for adolescent girls should seek to overcome the current fragmentation of approaches. It should be multi-sectoral, and aim to drive innovation and rapid learning at scale by building on existing programme platforms, while incorporating operational research or action/adaptive learning. Costing studies should remain an integral part of the learning programme.

This document is issued for the party which commissioned it and for specific purposes connected with the above-captioned project only. It should not be relied upon by any other party or used for any other purpose.

We accept no responsibility for the consequences of this document being relied upon by any other party, or being used for any other purpose, or containing any error or omission which is due to an error or omission in data supplied to us by other parties.

This document contains confidential information and proprietary intellectual property. It should not be shown to other parties without consent from us and from the party which commissioned it.