“Urwaruka Rushasha”: A Randomized Impact Evaluation of Village Savings and Loans Associations and Family-Based Interventions in Burundi

---

Findings from the Mid-Term Survey

Tom Bundervoet, PhD, The International Rescue Committee

Jeannie Annan, PhD, The International Rescue Committee

Miranda Armstrong, CYPD Coordinator, The International Rescue Committee

Table of Contents

Executive Summary

Recovering from decades of conflict that claimed 300,000 lives and forced over a million people to flee their homes, Burundi is one of the poorest countries in the world. Since 2003, over 450,000 refugees have returned to the country and the potential for political instability is high. In this context, families, particularly those that have been most affected by conflict and displacement, lack the capacity to adequately respond to children's needs, and children face significant risks that impact their physical, cognitive, and social-emotional development. To address the risks facing children while also building evidence around effective approaches to improving outcomes for children affected by both poverty and armed conflict, the IRC is exploring two programmatic pathways: 1) increasing household wealth and 2) improving caregiver-child relationships.

The Urwaruka Rushasha project is a three-year project, funded by USAID's Displaced Children and Orphans Fund (DCOF). The project aims at improving the protection, development and well-being of highly vulnerable boys and girls in Burundi's Makamba and Bujumbura Rural provinces where rates of refugee return, population density, and potential for political instability are the highest. The project consists of two interventions that benefit 1,600 families and their children: 1) A Village Savings and Loans Association (VSLA) intervention to strengthen the participants' economic situation and 2) an added family-based discussion group called "Healing Families and Communities" (VSLA Plus). This report presents the results of the mid-term evaluation of the New Generation project, which marks the end of the first phase of both the VSLA and family-based interventions.

To measure the impact of the VSLA intervention and determine the extent to which the Healing Families and Communities discussion groups improved children's well-being, the IRC designed a randomized impact evaluation with a baseline survey and two follow-up surveys. Before the baseline survey, the IRC identified 77 self-selected VSLAs representing 1,600 households that met the project criteria. These VSLAs were randomized into either a waitlist control group (37 associations) or a treatment group (40 associations). Of the 40 VSLAs in the treatment group, half were selected to also participate in the family-based discussion groups (VSLA Plus) during the first project cycle (April 2010-March 2011). The baseline survey was conducted between January and March 2010. The mid-term survey, on which this report is based, was conducted between April and May 2011. Both surveys consisted of a household survey, completed by the VSLA member, and a separate child survey of a randomly selected child in the household between the ages of 10-14.

The randomized controlled trial was designed with enough statistical power that the impact of the VSLA intervention can be examined with the data from the baseline and the mid-term survey. To determine the full impact of the Healing Families and Communities discussion modules, data from the final survey, due to be undertaken in 2012, is needed.

Participation in the VSLAs increased consumption expenditures

Consumption expenditures are a key indicator of welfare in rural regions in Africa, where most people do not earn an income at all or have highly irregular incomes. Between the baseline and mid-term survey, average consumption expenditures increased from 30.8USD per person per month to 35USD per person per month. Increased consumption expenditures were seen in both the treatment group (households that received the VSLA intervention in 2010) and the control group (households that did not participate in the VSLA intervention in 2010). However, for the treatment group, the increase in average consumption expenditures was far greater. While per person monthly expenditures increased by

0.4USD for the control households (a 1.3% increase), the increase for the treatment households amounted to 7.4USD (a 24.4% increase). The net impact of VSLA-participation amounts to 7USD per month and is statistically significant.

VSLA participation led to poverty reduction

At baseline, 65.7% of households in our sample lived below the international poverty line of 1.25USD a day. At mid-term, and despite the overall increase in consumption expenditures, the percentage of households living in poverty had increased to 68.5%. The overall increase in poverty is the result of a sharp increase in poverty among the households in the control group. The percentage of households in the control group who were living below the poverty line increased by 10 percentage points, from 64.6% at baseline to 74.5% at mid-term. In contrast, the percentage of households in the treatment group who were below the poverty line modestly dropped from 67.7% to 63.5%. The results suggest that VSLA participation enabled the treatment households to escape a general downward economic trend in rural Burundi. According to the data, the net impact of VSLA participation amounts to a 14% reduction in the percentage of families living below the poverty line, an effect that is statistically significant.

Household assets increased as a result of participation in VSLAs

The impact of the VSLA intervention on asset holdings confirms the patterns found for expenditures and poverty. While the score on the asset index, a standardized indicator of the household's asset holdings, decreased for the control households, it increased for the treatment households. On average, VSLA participation led to an increase in the asset index of 0.22 standard deviations. This roughly corresponds to an extra head of cattle for those who participated in VSLAs.

Participation in the Healing Families and Communities discussion modules reduced harsh discipline

The results of the mid-term survey show that VSLA participation alone does not affect the ways in which VSLA members discipline their children. However, households that participated in the VSLA Plus intervention, show large reductions in harsh methods of disciplining their children, both physical and verbal (or psychological punishment, as termed by UNICEF). The impact is particularly remarkable for harsh verbal discipline: Among the households that participated in the discussion modules, there has been a 20 percent reduction in the number of parents who shouted or yelled at their children or called their children dumb or lazy or insulted them in another way. There are also notable reductions in corporal punishment: The percentage of respondents that hit their children on the hand, arm or leg halved, while the percentage that reported hitting their children with a stick or another hard object fell from 7% to 2.5%. Data provided by children themselves for the survey confirmed that the discussion modules were effective in reducing harsh child discipline practices.

Participation in VSLAs increased child well-being; data from children shows the effects are greater in households that also participated in the discussion modules

Both the household survey and the child survey found that participation in VSLAs had a positive impact on child well-being. However, while the discussion modules had no added value according to the caregivers' report, the children's report showed that the discussion modules had significant added value.

According to the household survey (responses from caregivers), overall child well-being increased substantially between the baseline and the mid-term survey for all groups. The score on the child well-

being scale increased by 43% for the control households (who did not receive any of the interventions), by 57% for the VSLA households and by 52% for the households who participated in both VSLAs and discussion modules. The finding that child well-being, as reported by the parents, increased drastically in control households may be explained by response bias: Even households that did not participate in the first cycle of the project (control households) are aware that the project targets child well-being. This may have led respondents to answer in ways that they thought would be desirable to the interviewer.

The children's answers on the child well-being scale present a slightly different picture, one that is less likely influenced by social desirability linked to the program. While the aggregate well-being score did not change for the children in the control households, it increased by 6% for children in VSLA households and 20.5% for children whose parents also benefited from the discussion modules.

Children whose caregivers participated in the discussion modules show improved mental health

Although data showed improvements in the aggregate well-being of children whose parents or caretakers participated in both the VSLA and VSLA Plus interventions, data show that increasing a family's economic means (through the VSLAs) does not by itself lead to better child mental health outcomes. Yet, as reported by both children and caregivers, adding a family-based discussion group component decreases children's distress and aggression.

VSLA plus family-based discussion groups reduced the incidence of family problems; however, participation in VSLAs has minimal impact on family well-being

Results from both caregiver and child surveys illustrate that the family-based discussion group reduced the incidence of family problems (intoxication of family members, violence among family members, adult in family who sells household property without consent). Relative to control and VSLA households, households in the VSLA Plus intervention experienced significantly lower levels of family problems. However, according to both the household and child surveys, family well-being did not change much between the baseline and the mid-term survey.

Conclusion and Future Directions

Overall, results from the mid-term survey are extremely encouraging and clearly highlight the positive impact the project is having on vulnerable families in post-conflict Burundi. The VSLAs have increased both financial and physical assets at the household level and there are clear indications that the Healing Families and Communities discussion modules are improving the protection, development and well-being of children. The positive impact of the discussion modules is particularly apparent in the reduction of harsh discipline in the home and improvements in child/parent communication. These results provide evidence that VSLAs can in fact improve economic outcomes for those living below the poverty line. They also provide evidence that together VSLAs and family-based interventions are an important approach for improving children's wellbeing.

The second cycle of the project began in June 2011 and will continue through August 2012. For the second cycle, the control group from cycle 1 (37 VSLAs) was randomized into two groups: one receiving VSLA support only and the other receiving both VSLA support and participating in family-based discussion groups. The original treatment groups (VSLA and VSLA Plus) are expected to continue functioning with minimum support from the IRC during the second cycle.

The final evaluation, scheduled for August 2012, will provide more robust evidence regarding the impact of the discussion modules, provide further insight into the process by which the VSLAs have led to improved outcomes, and offer evidence to guide future programming. However, a number of recommendations regarding future direction of the project can be made based on the clear, positive results of the midterm evaluation. First, planning should begin to scale up the project at the provincial and/or national level. This involves securing funding and exploring cost-effective approaches that maintain project quality. Other steps include exploring additional economic strengthening components that can be provided at low cost; documenting other interventions in the area in order to determine potential causes of unexplained results; expanding the final evaluation to include first-cycle VSLA participants and collect information on use of loans; and taking a comprehensive approach to addressing violence against children by also addressing violence in the schools.

1. Introduction

1.1 Background

Recovering from decades of conflict that claimed 300,000 lives and forced over a million people to flee their homes, Burundi is one of the poorest countries in the world. 68% of the population currently lives below the poverty line[1] and, with a population of almost 8.4 million people, it falls at the bottom of the Global Hunger Index. Since 2003, over 450,000 refugees have returned to the country and the potential for political instability is high. Economic opportunities are scarce and many families, particularly those that have been most affected by conflict and displacement, lack the capacity to adequately respond to children's needs.

The IRC works in two provinces that have been heavily hit by conflict: The southern province of Makamba, bordering Tanzania, and the western province of Bujumbura Rural, bordering the Democratic Republic of the Congo (see Figure 1). As the base of a rebel movement in the late 1990s and early 2000s, Makamba saw heavy fighting and mass displacement, with thousands fleeing to refugee camps in Tanzania. In recent years, thousands of refugees have made the journey home and Makamba is now the province with the highest percentage of returnees in Burundi. Bujumbura Rural also suffered greatly as a result of the civil war. As a stronghold of the last remaining rebel group *Forces Nationales pour la Liberation* who laid down their arms in April 2009, Bujumbura Rural experienced 16 years of fighting and has only recently become a more peaceful, stable environment. With a highly vulnerable population and large stocks of small arms, the province has a strong potential for renewed violence.

In this context, children face significant risks. Poverty and violence have been proven to have adverse effects on children's sensory-motor, cognitive, and social-emotional development. Deficits in these domains can lead to a lack of educational progress and negative later life outcomes.[2] Poverty and violence also heighten adult stress and increase the risk of child maltreatment and violence in the home. To address these risks and build evidence of effective approaches to improve outcomes for children affected by both poverty and armed violence, the IRC is exploring two programmatic pathways: 1) increasing household wealth through economic interventions and 2) improving parent-child relationships through family-based interventions.

Existing research shows that micro-economic interventions have had some success in alleviating poverty and improving child outcomes. Conditional cash transfers have been shown to increase health service

[1] The World Bank. (2011). *Burundi*. Retrieved from http://data.worldbank.org/country/burundi.

[2] Gardner, J.M, Lozoff, B., Pollitt, E., Wachs, T.D, Walker, S.P., & Wasserman, G.A., *et al.* (2007). Child development: risk factors for adverse outcomes in developing countries. *Lancet*, **369** 9556, 145–157; Grantham-McGregor, S., Cheung, Y.B., Cueto, S., Glewwe, P., Richter, L., & Strupp, B. *et al.* (2007). Developmental potential in the first 5 years for children in developing countries. *Lancet*, **369** 9555, 60–70.

utilization; improve children's health outcomes and nutritional status;[3] and reduce children's aggressive/oppositional behaviors (though with no effect on anxiety/depressive symptoms)[45]. Similarly, there is evidence that for those living in poverty, microcredit can lead to improved savings, expenditures, and accumulation of assets, as well as to increased expenditure on children's health, improved nutritional status and protective health behaviors for children.[6] Yet, there is no evidence of the effectiveness of conditional cash transfers or microcredit in conflict-affected settings or of their impact on families and children who live in extreme poverty. Regarding microfinance, some research suggests little impact on the poorest populations and on women. However, in targeting the rural poor who have little access to financial institutions, the Village Savings and Loans Association (VSLA) model has become enormously popular with an estimated 4.6 million people enrolled worldwide. Despite the popularity of VSLAs, however, little is known about their effectiveness: although there is extensive monitoring information pointing to positive results, a rigorous impact evaluation has not previously been published.

Over the last several decades, a great deal of attention has been placed in developed countries on improving the effectiveness of family-based interventions. Several programs have been shown to reduce child maltreatment and improve family wellness and parenting skills.[7] However, despite the fact that positive parenting and nurturing relationships with caregivers can protect children from the harmful consequences of violence, in humanitarian settings,[8] family-based interventions are rare. As such, there is a dearth of evidence around the impact of these interventions on child well-being in low resource countries and areas affected by conflict.

Recognizing the potential effectiveness of VSLAs and family-based interventions on the development and well-being of children in Burundi and other impoverished and conflict affected communities, but also recognizing the need to address gaps in evidence, the IRC is implementing and conducting a randomized impact evaluation of Urwaruka Rushasha (New Generation): a project that includes both VSLAs and a family-based intervention.

---

[3] Boccia, D., Hargreaves, J., Lönnroth, K., Jaramillo, E., Weiss, J., Uplekar, M., Porter, J. D. H., et al. (2011). Cash transfer and microfinance interventions for tuberculosis control: review of the impact evidence and policy implications. *The International Journal of Tuberculosis and Lung Disease*, *15*, S37-S49. doi:10.5588n/ijtld.10.0438; Lagarde, M., Haines, A., & Palmer, N. (2009). The impact of conditional cash transfers on health outcomes and use of health services in low and middle income countries (Review). *Cochrane Database of Systematic Reviews*, Vol. 7, (4). Retrieved from http://www2.cochrane.org/reviews/en/ab008137.html.

[4] Fernald, L. C., Gertler, P. J., & Neufeld, L. M. (2009). 10-year effect of Oportunidades, Mexico's conditional cash transfer programme, on child growth, cognition, language, and behaviour: a longitudinal follow-up study. *The Lancet*, *374*(9706), 1997-2005; Ozer, E. J., Fernald, L. C. H., Manley, J. G., & Gertler, P. J. (2009). Effects of a conditional cash transfer program on children's behavior problems. *Pediatrics. Vol 123(4)*, *123*, e630-e637.

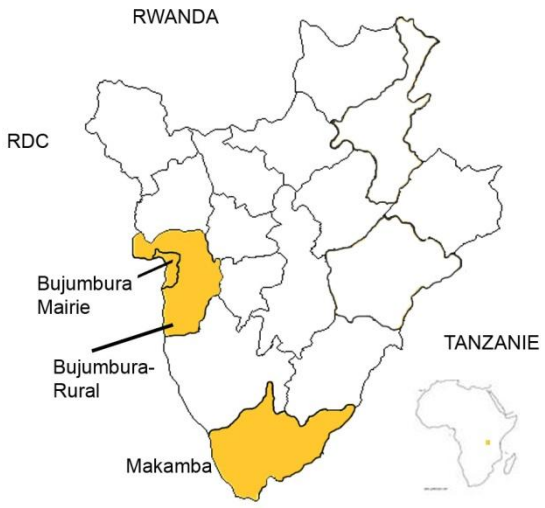
[5] Lund, C., De Silva, M., Plagerson, S., Cooper, S., Chisholm, D, Das, J., Knapp, M., & Patel, V. (2011). Poverty and mental disorders: breaking the cycle in low-income and middle-income countries. *Lancet*, 378, 1502-14.

[6] Stewart R, van Rooyen C, Dickson K, Majoro M, de Wet T (2010) What is the impact of microfinance on poor people? A systematic review of evidence from sub-Saharan Africa. Technical report. London: EPPI-Centre, Social Science Research Unit, University of London.

[7] World Health Organization (2009). Preventing violence through the development of safe, stable and nurturing relationships between children and their parents and caregivers. Series of briefings on violence prevention: the evidence. Geneva: WHO Press.

[8] Lustig SL, Kia-Keating M, Knight WG, et al. (2004). Review of child and adolescent refugee mental health. *J Am Acad Child Adolesc Psychology*, 43 (1), 24–36.

Figure 1: Provinces of Intervention, Bujumbura Rural (East) and Makamba (South)



1.2 The New Generation Project Model: Exploring two pathways to improving children's wellbeing

The New Generation project is comprised of two programmatic components: Village Savings and Loans Associations (VSLA) and an added family-based discussion group called "Healing Families and Communities" (VSLA Plus).

Village Savings and Loans Associations (VSLAs)

Through the New Generation project, the IRC is supporting the establishment and functioning of 84[9] Village Savings and Loans Associations (VSLAs). A Village Savings and Loans Association is a self-selected group of 15 to 25 members who save money by purchasing *shares* in the VSLA. The cost of a share is set by the group at a rate that allows the poorest in the group to save. The savings are pooled into a loan fund from which members can borrow, potentially enabling them to overcome entry-barriers to more lucrative and more reliable income-generating activities. Loans are repaid with a service charge (typically 10% of the loan) that is set and agreed upon by members. At the end of the VSLA cycle (12 months[10]), the accumulated savings and interest payments are distributed among the members. VSLAs typically have an average rate of return on savings of around 36% (see vsla.net).

[9] The project as a whole is supporting 85 VSLAs; however, only 77 of these are included in the monitoring and evaluation component.

[10] This includes three months of training and nine months of actual savings and loans.

VSLAs plus Healing Families and Communities Discussion Modules (VSLA Plus)

Half of the VSLA members will benefit from "Healing Families and Communities" discussion modules. The discussion modules are designed to increase the caregivers' knowledge of actions to improve their children's protection, well-being and development. The curriculum was designed in country by the project team, drawing from other evidence-based family programs for guidance on sessions addressing communication and discipline. As part of the VSLA Plus intervention, following VSLA meetings, members participate in a series of ten discussion modules that aim to guide participants through a process of change. The topics include: 1) children's environment, 2) children's well-being and participation, 3) access to health and education, 4) positive discipline and communication, 5) child protection in the community, 6) family budgeting and 7) making a commitment to change. At the end of the discussion series, participants invite spouses and other family and community members to a public forum where the participants share what they have learned and describe the changes they will make.

The Target Population

The target population for the New Generation project was 1,600 poor families with children in zones with the highest percentage of returnees and a high rate of malnutrition[11]. Based on these criteria, the Makamba and Bujumbura Rural provinces were selected as areas of project implementation. Not only are these provinces heavily affected by the civil war and home to many who have only recently repatriated from neighboring countries, they are also home to a particularly vulnerable population. Over 40% of adults in the participating households never went to school, with the proportion being substantially higher for women. In beneficiary households, net enrollment in secondary school amounted to 12%, and was considerably lower for girls (9%, but in some communes dropping to only 4%). 65% of children under the age of five suffered from chronic malnutrition and almost 10% from acute malnutrition. At baseline, average daily per capita consumption expenditures amounted to 1.21 USD, which falls short of the international poverty line of 1.25 USD a day (in 2005 PPP prices). 65% of participating households fell below the poverty line (compared to 66.9% nationwide according to official Work Bank figures).[12] Setting the poverty line higher at 2 USD a day, over 85% of beneficiary households could be classified as living below the poverty line. Ownership of assets was low with only 17% of households owning a phone and only 13% having a mattress to sleep on.

[11] Malnutrition in zones was informally assessed by asking doctors at provincial hospitals to identify the zones with the highest rates of malnutrition.

[12] World Bank Data Catalog, consulted on December 19th 2011.

1.3 Methodology: Evaluation Design and Implementation

To rigorously examine the impact of the interventions on participants, the IRC implemented a randomized impact evaluation. Two questions drove the design of the evaluation:

1. Do Village Savings and Loan Associations improve economic outcomes of poor households?

2. Does the "Healing Families and Communities" discussion series offer additional benefits for child well-being beyond that which can be explained by increased economic outcomes? Or is money alone enough to improve child well-being in poor families?

The answers to these questions will help add to the knowledge base about: 1) whether humanitarian and development organizations should focus on VSLA interventions to reduce poverty and 2) whether VSLA interventions alone are sufficient to improve child well-being or whether there is also a benefit to implementing family-based interventions.

To address these questions the evaluation aimed to:

1. Assess the impact of VSLA programs on (a) household assets and consumption, including spending on children's education, health, nutrition and clothing; (b) children's education, labor, health and psychosocial well-being; (c) caregiver's use of harsh punishment and positive communication, and (d) family functioning;

2. Assess the incremental impact of a family-based intervention added to the VSLA program on (a) household assets and consumption, including spending on children's education, health, nutrition and clothing; (b) children's education, labor, health, and psychosocial well-being; (c) caregiver's use of harsh punishment and positive communication, and (d) family functioning.

The evaluation is a randomized controlled trial with a baseline and two follow up surveys. All households will be interviewed three times in three years. Figure 2 illustrates the evaluation strategy. For the baseline survey (January-March 2010), the IRC conducted household surveys with 96% of the members of 77 self-selected VSLA groups that fit the selection criteria for participation. After the interviews, the 77 VSLA groups were randomized into three categories:

1. Group 1 (20 VSLAs) receives VSLA support in 2010;
2. Group 2 (20 VSLAs)receives VSLA support and the "Healing Families and Communities" (HFAC) discussion series in 2010;
3. Group 3 (37 VSLAs) were the control group during the first cycle.

During the midterm survey, which was conducted in April-May 2011, all three groups of households were re-interviewed. Following completion of the mid-term survey, the control group from cycle 1 (37 VSLAs) was randomized into two groups: one that will receive VSLA support only and another that will receive both VSLA support and the family-based discussion sessions. The original treatment groups (VSLA and VSLA Plus) are expected to continue functioning with minimum support from the IRC during the second cycle.

The household questionnaire used in the baseline and midterm survey was a standard multi-topic household questionnaire adjusted for the specific purpose of this project. Certain sections of UNICEF's Multiple Indicators Cluster Surveys (MICS) and USAID's Food and Nutrition Technical Assistance (FANTA II) project were adopted. The major topics covered by the questionnaire are listed below:

1. **Demographics & Education:** included a standard household roster listing the demographic information of all household members as well as levels of education, literacy and school absenteeism (for those children who go to school).

2. **Economic Outcomes:** included a series of questions on economic outcomes, such as quality of housing, land endowments, household assets and livestock, and household private consumption.

3. **Health:** included several sections on adult and child health and use of health products and facilities. For adults, questions were asked about their current health status (injuries, chronic diseases) and their ability to perform daily routine tasks such as fetching water over relatively long distances, carrying produce to markets etc. This section also probed for alcohol use and abuse. For children under the age of five, standard questions were used (from UNICEF's Multiple Indicators Cluster Surveys -MICS) about mosquito net use and the incidence of fever and cough in the two weeks preceding the survey.

4. **Child Protection and Child Well-Being:** were assessed through several sections related to child protection and child well-being. These sections included: child labor (for children between 5 and 14-years old), child discipline (for children between 10 and 14 years old), child well-being and mental health (children between 10 and 14-years-old). The first two sections included items adapted from UNICEF's Multiple Indicator Cluster Survey, the Discipline Interview[13] used in a multi-country parenting study, and the Parental Acceptance-Rejection Questionnaire.[14] The last two sections were developed by the team of *Urwaruka Rushasha* during participatory exercises with children and adults in Bujumbura Rural and Makamba. These sections were administered with reference to one specific child in the household. If the household had several children in the relevant age range, one of them was randomly selected to be the "reference child" for the child modules.

5. **Family Well-Being:** included questions about the incidence of disputes and violence in the family, and the general atmosphere and functioning within the family. This was also developed

[13] Lansford, J. E., Chang, L., Dodge, K. A., Malone, P. S., Oburu, P., Palmérus, K., Bacchini, D., Pastorelli, C., Bombi, A. S., Zelli, A., Tapanya, S., Chaudhary, N., Deater-Deckard, K., Manke, B., & Quinn, N. (2005). Cultural normativeness as a moderator of the link between physical discipline and children's adjustment: A comparison of China, India, Italy, Kenya, Philippines, and Thailand. *Child Development*, 76, 1234-1246.

[14] Rohner, R. P., & Khaleque, A. (Eds.). (2005). *Handbook for the study of parental acceptance and rejection* (4th ed.). Storrs, CT: Rohner Research Publications.

during participatory activities and interviews with children and adults in Bujumbura Rural and Makamba.

After the questionnaire was pretested, it was programmed into Personal Digital Assistants (PDAs) for the purpose of Computer Assisted Personal Interviewing (CAPI). Research has shown that CAPI not only reduces interview time relative to Paper Assisted Personal Interviewing (PAPI) but is particularly efficient in reducing data inconsistencies and measurement errors.

In order to verify information reported by the caregivers, a questionnaire was also developed for children, which includes some of the same questionnaire sections as the household survey. To decide which children to interview, 400 households who participated in the household survey and had at least one child between 10 and 14-years-old were randomly sampled. For these households, the children's questionnaire was administered to the child who was selected as the "reference child" in the household survey. The children's questionnaire covers the following topics: Education; Child Well-Being; Child Discipline; Child Labor; Child Mental Health; and Family Well-Being. The child survey was carried out by IRC Child Protection Officers using a paper-based format for the baseline and midterm. 362 of the 400 selected children were interviewed during the same time period as the caregivers although the child survey took longer to carry out as the children were followed up separately and there were significant logistical challenges.

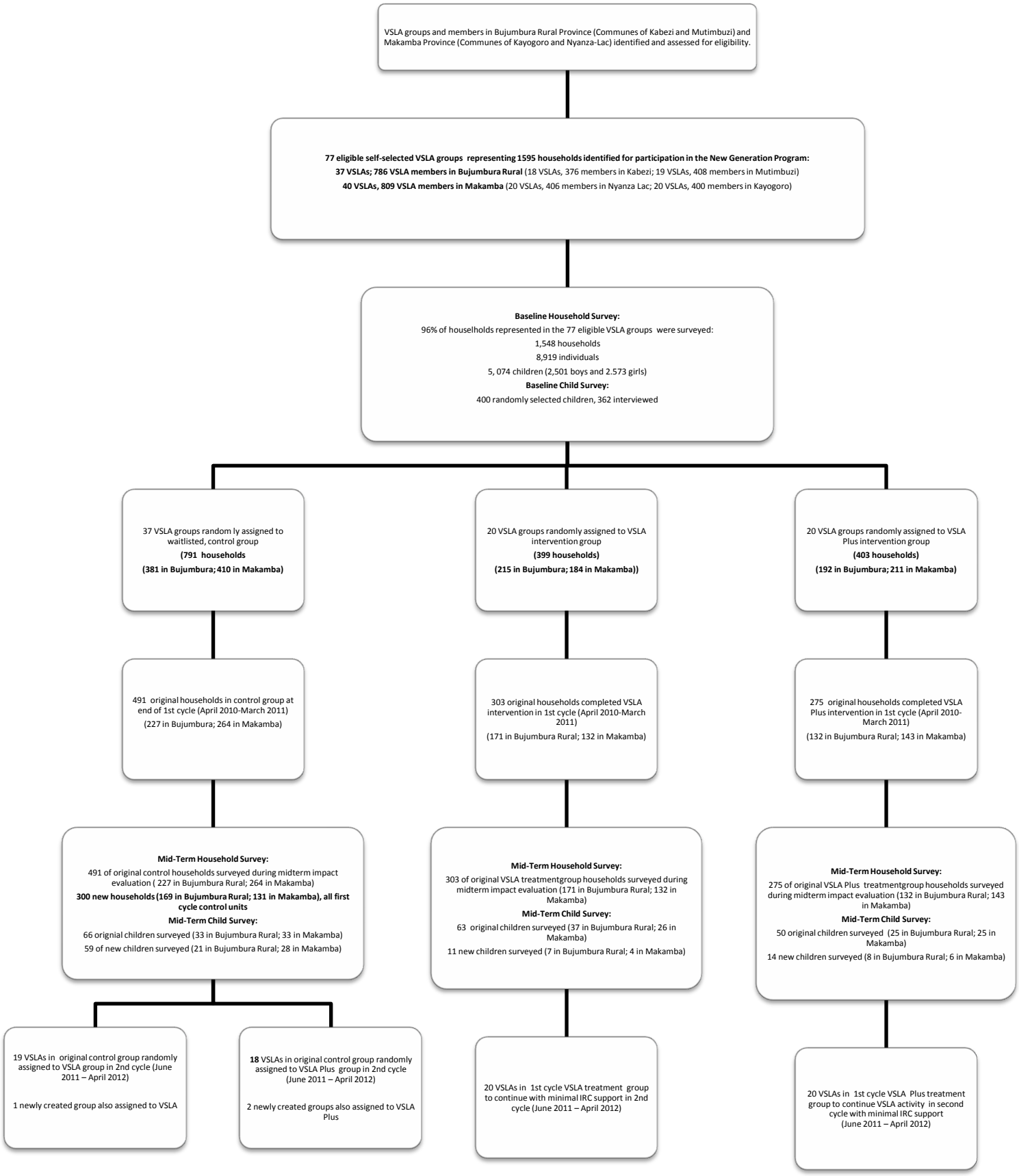
The randomized controlled trial was designed to provide sufficient statistical power to assess the impact of both the VSLAs and the family-based discussion modules (see appendix A1). Calculations of statistical power suggested that during the mid-term evaluation, we should focus on the impact of the VSLA program on economic outcomes. Nonetheless, despite limited power, the mid-term evaluation shows a range of significant outcomes from both the VSLA and family-based discussion interventions.

In order to feed into and complement the quantitative research, the IRC conducted participatory activities with children. The stories and experiences of children help to better understand what issues are most important to them and provide further knowledge about results. The following chart shows the topics, research questions and participatory activities that were conducted. See midterm qualitative report for full description of methods and findings.[15]

[15] The International Rescue Committee. (2011)."Urwaruka Rushasha (New Generation):Improving the Well-Being of Vulnerable Girls and Boys in Burundi: The results of participatory activities with children to evaluate child well-being, family relations, child discipline, children's problems and support environment, Mid-Term Evaluation: Design and Findings."

1. Children are free from violence, abuse, neglect and exploitation		
Item	*Research questions*	*Participatory exercise*
Discipline	Are discussions around discipline (linked to the body maps) changing? What are children saying about the feelings they have linked to different body parts? Are discipline methods changing?	Body Maps
Problems and social Environment	Are changes taking place in the children's social environment? What are the problems children are facing? Are these changing over time? Do children have people they can talk to about their problems? Who are these people? Is this changing over time?	Spider Diagrams
2. Children feel happy and safe (well-being)		
Item	*Research questions*	*Participatory exercise*
Significant changes	What changes are taking place in children's lives? What significant events have taken place in the last 6/12 months? Are any of these linked to the project?	Time Line
Communication	Are children able to talk to their parents/caregivers about their needs and feelings? How often can they do this? Are their needs met? How do they negotiate for what they need?	Time Line, Spider Diagrams, Wishes for 2010/2011
Relation / Quality time with parents	How much time do children get to spend with their parents? Do children play with their parents? Are children able to talk to their parents?	Time Line
3. Children have "agency" or control over choices in their lives		
Item	*Research questions*	*Participatory exercise*
Agency	Do children feel in control of their lives? Can children decide how they manage their time? Do children have time to play?	Time Line, Spider Diagrams, Wishes for 2011/2011

Figure 2: Flow of Participants throughout the course of the New Generation Project



2. Findings: The Impact of the New Generation Project on Participants

2.1 The Sample

The mid-term survey for New Generation was conducted in April and May 2011. Ideally, the mid-term survey would have re-interviewed all 1,548 households surveyed during the baseline. However, due to considerable drop-out during the first months of the project, only 1,069 (69%) of the 1,548 original households were re-interviewed at mid-term. Appendix A shows that – due to the nature of our randomization - drop-out only marginally lowers statistical power. Appendix B discusses the nature of the drop-out and its implications for the research design. We find that drop-out was selective, in the sense that less wealthy and less educated households were more likely to drop out of the treatment group, potentially causing an upwards bias in our treatment estimates. The econometric analysis will use sensitivity analysis to gauge the importance of this bias.

Overall, the mid-term survey represents 7,905 individuals (52.2% of whom were female)[16] in 1,369 households (1,069 households from the baseline and 300 new households). Average household size amounted to 5.8 persons and varied little by commune or province. Almost half of the sample (47.9%) was 14 years of age or younger (see Table 1).

2.2 The Impact of VSLA Participation on Economic Outcomes of Participants

In this section, the effects of VSLA participation on living standards will be examined. The first sub-section presents the process indicators that were gathered on a monthly basis during the VSLA cycle. In the second sub-section, the impact findings of VSLAs will be presented.

VSLA Process Data: Savings, Loan Disbursement, Loan Size and Rate of Return

Table 2 presents key indicators of VSLA performance: accumulated savings, the number of loans disbursed, loan size, and rate of return. Overall, the 928 members of the 40 first cycle VSLAs managed to save BIF 31,015,310 (USD 52,250 using the 2011 PPP exchange rate of *593.6*) during the first cycle. A total of 3,108 loans were disbursed, for a cumulative value of BIF 80,378,969 (USD 135,409). Average loan size amounted to BIF 25,862 (USD 43.6).[17]

[16] With original households we mean households that have been interviewed at baseline.

[17] According to the VSL methodology the value of a loan cannot exceed three times the cumulative savings of a person. Hence, loan value is rather small at the beginning of the cycle and grows bigger as participants accumulate more savings.

Table 1: Demographic Distribution of the Sample

	Total	Kabezi	Mutimbuzi	Bujumbura Rural	Kayogoro	Nyanza-Lac	Makamba
Households Surveyed	**1369**	345	354	699	371	299	670
Individuals Surveyed	**7905**	2035	1932	3967	2184	1754	3938
Household Size	**5.8**	5.9	5.5	5.7	5.9	5.9	5.9
Male	**3780**	926	905	1831	1076	873	1949
Under 5	**632**	127	180	307	188	137	325
5 to 9	**680**	157	158	315	192	173	365
10 to 14	**516**	144	93	237	147	132	279
15 to 59	**1748**	464	404	868	491	389	880
60 and Older	**105**	26	26	52	26	27	53
Unkown	**99**	8	44	52	32	15	47
Female	**4125**	1109	1027	2136	1108	881	1989
Under 5	**666**	132	199	331	183	152	335
5 to 9	**702**	173	172	345	203	154	357
10 to 14	**587**	151	124	275	170	142	312
15 to 59	**1996**	602	486	1088	495	413	908
60 and Older	**103**	38	24	62	25	16	41
Unkown	**71**	13	22	35	32	4	36

Table 2: VSLA Process Indicators:
Accumulated Savings, # of Loans Disbursed, Loan Size, and Rate of Return

	Overall	Bujumbura Rural	Makamba
# of VSLAs	40	20	20
# of Members	928	509	419
Accumulated Savings (PPP USD)	52,250	21,014	31,236
# of Loans Taken Out	3,108	1,207	1,901
Accumulated Value of Loans (PPP USD)	135,409	40,896	94,513
Average Loan Size (PPP USD)	43.6	33.9	49.7
Average Return on Savings (%)	47.8	44.3	51.1

Notes: Based on Purchasing Power Parity (PPP) exchange rate of USD 1= BIF 593.6

On average, VSLA participants received an interest rate of 47.8%. In Makamba province, return on savings even topped 51%. The most successful VSLA achieved a return on savings of 78.2%. Only three of the 40 VSLAs achieved a return on savings of less than 30%, with the least successful group still managing to obtain a return of 26.6%.

VSLAs in Makamba handled substantially more money than the VSLAs in Bujumbura Rural. Accumulated savings and loans disbursed are 50% higher in Makamba than in Bujumbura Rural and the cumulative value of loans is more than double. This can mainly be explained by the fact that cost of shares in VSLAs in Makamba were higher than those in Bujumbura Rural : While the average value of a share was BIF 280 (USD 0.47) in Bujumbura Rural, it was BIF 391 in Makamba (USD 0.66). This means that the average participant in Bujumbura Rural could save a maximum of USD 2.35 per week,[18] while in Makamba this amounted to USD 3.3.

In general, the process indicators for the first VSLA cycle are impressive and justify the conclusion that the **VSLAs were highly effective in delivering basic financial services to project participants who do not normally have a means to save and access credit**. But did this possibility to save and access loans also have an impact on their economic outcomes and living standards? This is the subject of the next subsection.

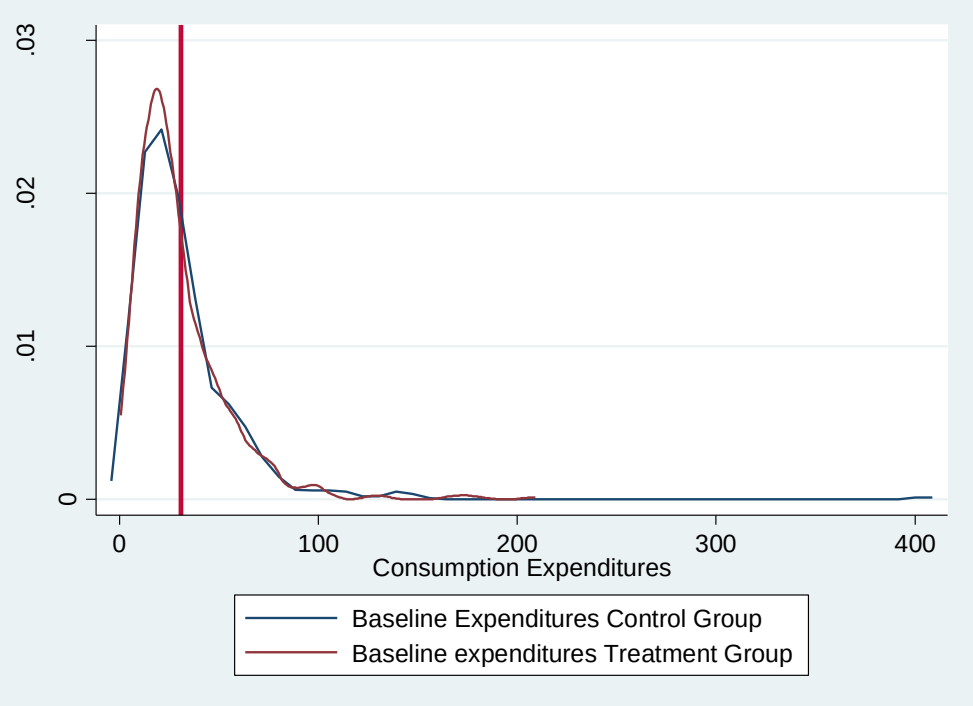
The Impact of the VSLA and VSLA Plus Interventions on consumption expenditures

To examine the impact of VSLA participation, the analysis focuses on the 1,069 households that were interviewed both at baseline and at mid-term. Average consumption expenditures at baseline were lower in the treatment group than in the control group: Food consumption at baseline amounted to

[18] According to the VSL methodology, members can only save a maximum of five shares per week.

USD 28.1 per capita per month for the treatment households vs. 28.5 for control households, and total expenditures amounted to USD 30.3 and USD 31.5, respectively. Despite average expenditures being somewhat lower in the treatment group at baseline, overall the distributions and means of the two groups are very similar (see Figure 3). Note that Figure 3 only takes into account the 1,069 households that were interviewed in both survey rounds (and does not include the households that were interviewed at baseline but dropped out).

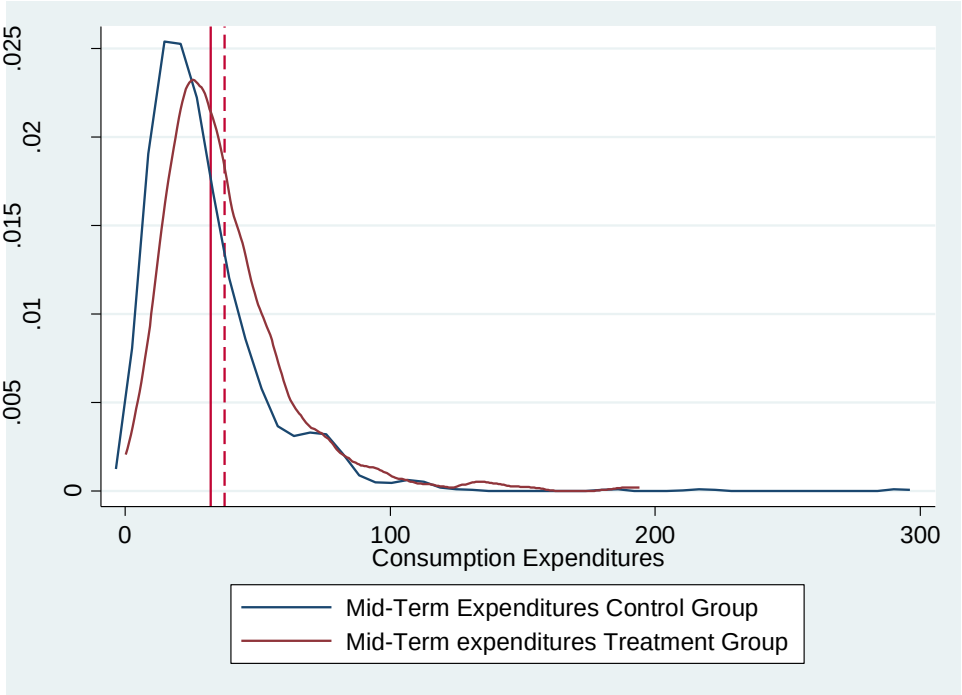
Figure 3: Mean and Distribution of Per Capita Consumption Expenditures at Baseline



Notes: Vertical lines represent baseline mean consumption expenditures for treatment (full line) and control (dashed line) households. Expenditures are in USD.

Figure 4 shows the distribution and mean of expenditures during the mid-term survey, following completion of one 12-month cycle of the VSLA intervention. The results illustrate that there has been an important shift in the distribution of expenditures for the treatment group: The whole distribution shifted to the right and is now clearly distinguishable from the control group distribution (in contrast to the situation at baseline, see Figure 3).

<u>Figure 4:</u> Mean and Distribution of Per Capita Consumption Expenditures at Mid-Term After VSLA Intervention, Treatment (full line) and Control (dashed line) group



Notes: Vertical lines represent baseline mean consumption expenditures for treatment (full line) and control (dashed line) households. Expenditures are in USD.

Following the intervention, the mean per capita consumption expenditures of treatment households was USD 37.7. This is USD 5.8 higher than the mean per capita consumption expenditures of the control households USD 31.9, a difference statistically significant at the 1% level. Figure 5 summarizes the change in food and total expenditures between baseline and mid-term. For the treatment households, food expenditures increased from USD 28.1 at baseline to USD 30.5 at mid-term, an increase of 8.4%. At the same time, food expenditures for control households *decreased* from USD 28.5 to USD 25.8. Total expenditures increased for both treatment and control households, though the increase is much higher in the treatment group (24.4%) than in the control group (1.3%). The net impact of VSLA participation amounts to USD 7 per capita per month (BIF 4,155). The impact is statistically significant at the 5% level. **For an average family of 5.8 members, this implies an increase in monthly expenditures of 40.6 USD thanks to VSLA participation[19].**

[19] Because the baseline (January to March 2010) and the midterm (April to May 2011) surveys were conducted during different seasons, with the period between January and April typically being a difficult period for agricultural households in Burundi, we would expect average consumption to be higher during the mid-term survey, regardless of whether the household was in the project (treatment household) or not. This, however, has no implication for the results of the impact evaluation: Since treatment and control households live in the same geographical area and are subject to the same agricultural seasons and climatic conditions, any seasonal and climatic effects are cancelled out by the randomized design.

Figure 5: Impact of VSLAs on the Monthly Expenditures of Treatment and Control Groups: A Comparison of Baseline and Mid-Term Results (in PPP USD)

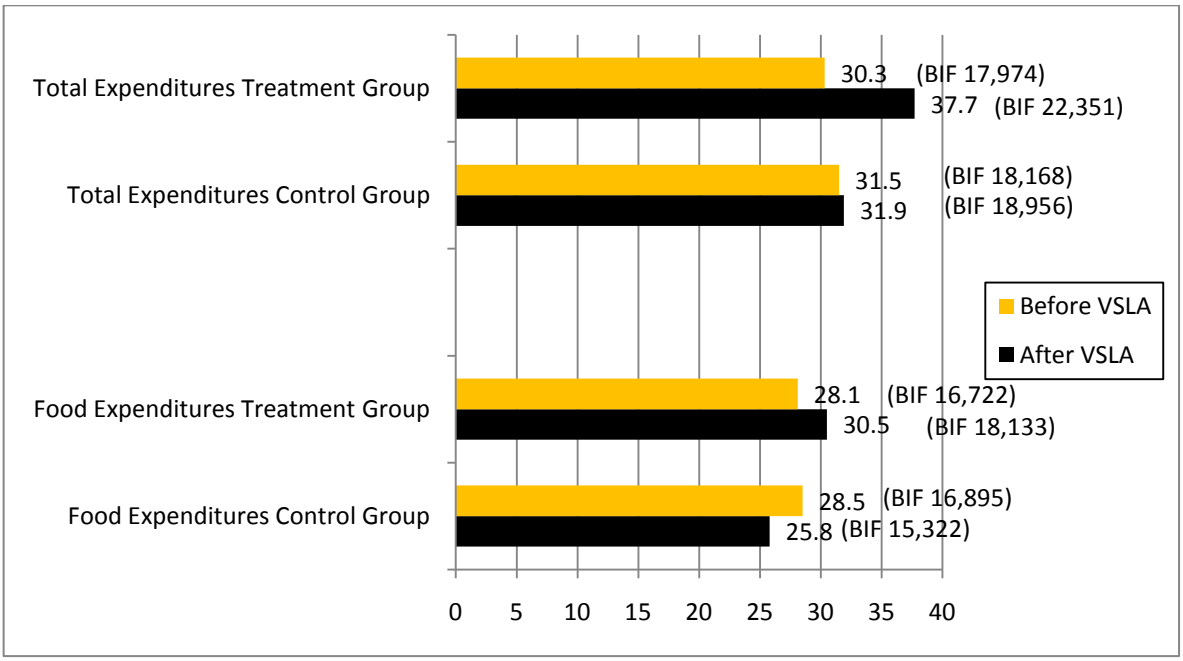


Table 3: Difference-in-differences Regression of Per Capita Consumption Expenditures (BIF) on VSLA Participation

	Control Households	Treatment Households	Difference
Total Expenditures Before VSLA	31.5	30.3	-1.2
	[1.83]	[1.53]	[2.39]
Total Expenditures After VSLA	31.9	37.7	5.8***
	[1.53]	[1.25]	[2.0]
Difference	0.4	7.4***	**7****
	[1.5]	[1.69]	**[2.25]**

Notes: Standard errors clustered at the VSLA level to account for clustered randomization; Dummies for strata included; ***: Statistically significant at 1%; **: Statistically significant at 5%. VSLA impact estimator in bold. Figures in PPP USD.

To appreciate the magnitude of the VSLA intervention's impact on consumption expenditures, it is informative to compare the size of the impact (7 USD per capita per month) with consumption expenditures in the sample at baseline (30.4 USD). This means that the magnitude of the VSLA impact equals 23% of pooled baseline expenditures. This is by all means a substantial impact.[20]

The Impact of the VSLA and VSLA Plus Interventions on Poverty Rates

Using the World Bank international poverty line of USD 1.25 (in 2005 PPP prices-see Box),[21] one can estimate that 65.7% of the 1,069 households surveyed were below the poverty line at baseline (that

[20] As mentioned before, the substantial rate of drop-out during the project likely introduces bias in the results. The direction and magnitude of the bias depends on the outcome evolution of the drop-outs had they not dropped out. To examine the sensitivity of the treatment effect estimated in Table 3, appendix Table D1 estimates the mean treatment effect under various missing data assumptions (see for instance Karlan and Valdivia (2010) for a similar approach). Scenarios (1) and (7) are the extremes: For scenario (1), we assume that consumption growth of the attrited treatment households equals the median growth rate in the bottom growth quintile of the non-attrited treatment group, while the growth of the attrited control households is assumed to equal the median growth in the top growth quintile of the non-attrited control group. Scenario (7) is the other extreme: We impute the median growth in the top growth quintile of the observed treatment group to the attrited treatment households and the median growth in the bottom growth quintile of the observed control group to the attrited control households. Scenarios (2) and (6) are similar but less extreme by using the second (instead of the bottom) and the fourth (instead of the top) quintile to impute the missing values for the attrited. Scenarios (3) and (5) are yet less extreme by using the fifth (between 40% and the median) and the sixth (between the median and 60%) growth deciles of the observed treatment and control group to do the imputations. Scenario (4) replicates the result without any imputation.

The extreme scenarios come up with largely implausible estimates: According to scenario (1) –the extreme lower bound- the VSLA intervention would have had a negative impact of *-11,096* (more than 59% of average baseline consumption expenditures), while scenario (7) –the extreme upper bound- results in a positive impact of *22,402*, meaning that the average treatment effect corresponds to 119% of baseline expenditures. The first important finding from the sensitivity analysis comes from scenario (2): Although this lower bound scenario still results in a negative treatment effect, the effect is not statistically discernable from zero. Under lower bound scenario (3), the treatment estimate is positive (*2,956*) and statistically significant at 5%. This is a reassuring finding: Even under a modest lower bound scenario we find a positive and statistically significant impact of the intervention.

Recently, researchers have argued that if autocorrelation in outcomes (in our case: consumption expenditures) is low, the ANCOVA estimator is more efficient than the difference-in-differences estimator (McKenzie, 2011). Since in our data, correlation between baseline and mid-term consumption expenditures is lower than 0.5 (*0.2695*) the ANCOVA estimator will be more efficient than the DiD presented in Table 3. This estimation strategy results in a treatment estimate of *USD 5.7,* which comes close to the difference-in-differences estimator presented in Table 3 (*USD 5.8*). The ANCOVA estimator is statistically significant at the 1% level (compared to 5% for the DiD estimator), which highlights the increase in power gained through ANCOVA estimation.

[21] Ravallion, M., Chen, S. and Sangraula, P. (2008) "Dollar a Day Revisited." *World Bank Policy Research Working Paper* 4620. Washington D.C.: The World Bank.

In 2005, the World Bank updated the "dollar a day" poverty line to "1.25 dollar a day" in 2005 PPP prices. To make the expenditure data from our survey comparable to the 1.25 dollar a day line, we deflated the 2011 expenditure data to 2005 prices (using data on inflation from IMF) and then applying the 2005 PPP exchange rate of USD 1=BIF 342.9.

is, they had per capita expenditures lower than USD 1.25 a day). At mid-term, 68.5% of the households were below the poverty line, an increase of 2.8 percentage points.

As shown in Table 4 and Figure 6, the net increase in poverty masks differences in trends for treatment and control households. While the percentage of control group households living below the poverty line *increased* from 65% to 75% (a 10% increase), the incidence of poverty *decreased from* 67% to 63% *(*a 4% decrease) among the households who participated in the VSLA intervention. Although the decrease in rates of poverty seen among the treatment households just misses statistical significance at conventional levels, the results indicate that the treatment households have been able to resist a general trend towards greater poverty in rural Burundi. **The net impact of the VSLA intervention amounted to a 14% reduction in poverty**. This is strongly significant at the 1%-level.

Table 4: Impact of the VSLA Intervention on the Percentage of Families Living Below the Poverty Line

	Control Households	Treatment Households	Difference
Poverty Headcount Before VSLA	0.65	0.67	0.02
	[0.022]	[0.020]	[0.045]
Poverty Headcount After VSLA	0.75	0.63	-0.12***
	[0.020]	[0.020]	[0.028]
Difference	0.1***	-0.04	**-0.14*** **
	[0.027]	[0.028]	**[0.041]**

Notes: Standard errors clustered at the VSLA level to account for clustered randomization; Dummies for strata included; ***: Statistically significant at 1%; **: Statistically significant at 5%. VSLA impact estimator in bold. Table shows marginal effects obtained from logit difference-in-differences regression.

Figure 6: VSLA Impact on the Percentage of Families Living Below the Poverty Line

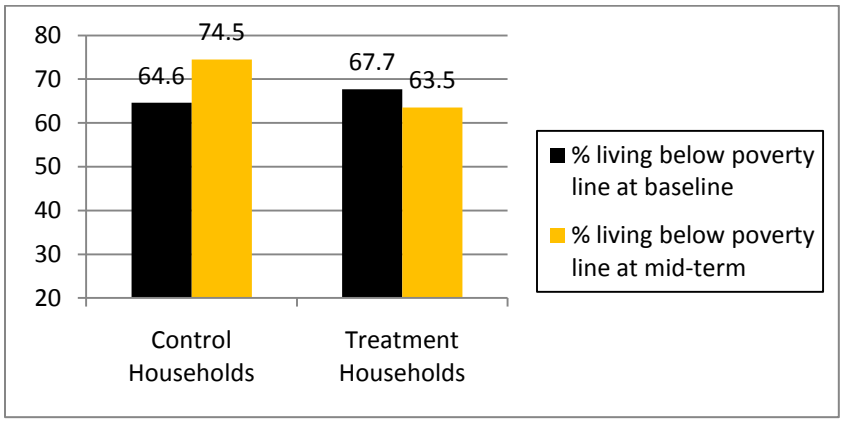


Table 5 shows that **poverty is more persistent in the control group than in the treatment group**: Of all control households that were below poverty at baseline, 85.7% were still below the poverty line at mid-term, showing little upward economic mobility in the control group. In the treatment group, 30% of households who were below the poverty line at baseline managed to cross the poverty threshold. This is twice as high as the proportion in the control group.

Table 5: Moving in and Out of Poverty, Treatment and Control Households

	Control Group	
	Below Poverty Line at Mid-Term	Above Poverty Line at Mid-Term
Below poverty line at Baseline	85.7	14.3
Above poverty line at Baseline	54.1	45.9
	Treatment Group	
	Below poverty line at Mid-Term	Above poverty line at Mid-Term
Below poverty line at Baseline	70.1	29.9
Above poverty line at Baseline	49.7	50.3

At the same time, more households that were not below the poverty line at baseline managed to stay above the poverty line at mid-term in the treatment group (50.3%) than in the control group (45.9%), although this difference is small.

The Impact of the VSLA on Household Assets

There are reasons to believe that a single VSLA cycle will not be enough to increase VSLA members' assets. Assets take a considerable lump sum investment to obtain and the first cycle in any VSLA is typically characterized by relatively low values of shares and small loans.[22] Table 6 presents the change in asset ownership for treatment and control households.

The difference between columns (3) and (1) shows the change in asset holdings of treatment households between baseline and mid-term. **These results show large increases in asset ownership between baseline and mid-term for the treatment households**:

- 16 % increase in ownership of radio
- 11 % increase in ownership of bicycle
- 13 % increase in ownership of mobile phone
- 14 % increase in ownership of bed
- 13 % increase in ownership of mattress
- 63% increase in number of Tropical Livestock Units[23]

The difference between columns (4) and (2) shows the change in asset holdings between baseline and mid-term for control households. Asset holdings of control households also increased:

- 4 % increase in ownership of radio
- 3 % increase in ownership of bicycle
- 10 % increase in ownership of mobile phone
- 9 % increase in ownership of bed
- 9 % increase in ownership of mattress
- 41% increase in number of Tropical Livestock Units

[22] According to experienced practitioners, the more important effects of VSLAs start showing after approximately three cycles (personal communication with Hugh Allen of VSL Associates in Bujumbura, 2010).

[23] Tropical livestock units (TLU) are used to make different species of livestock comparable to arrive at a single aggregate indicator of livestock holdings. One head of cattle equals 0.7 TLU, one goat and one sheep 0.1 TLU, one pig 0.2 TLU and one chicken and one rabbit 0.01 TLU.

Table 6: Change in Asset Holdings for Treatment and Control Households

	Baseline (Jan-Mar 2010)		Mid-Term (Apr-May 2011)	
	(1) Treatment	(2) Control	(3) Treatment	(4) Control
Owns Radio	0.35	0.31	0.51	0.35
Owns Bicycle	0.23	0.21	0.34	0.24
Owns Watch	0.13	0.12	0.2	0.17
Owns Mobile Phone	0.17	0.17	0.3	0.27
Owns Bed	0.56	0.61	0.7	0.7
Owns Mattress	0.13	0.13	0.26	0.22
Lives in House with Brick Walls	0.45	0.47	0.62	0.62
Lives in House with Iron Roof Sheeting	0.7	0.72	0.69	0.66
Lives in House with Concrete Floor	0.03	0.03	0.05	0.05
Number of Rooms in House	2.32	2.18	2.49	2.36
Uses Charcoal for Cooking	0.061	0.054	0.14	0.1
Owns Land	0.57	0.57	0.65	0.65
Number of Tropical Livestock Units	0.126	0.093	0.206	0.131
Asset Index	0.050	0.039	0.107	-0.125
N	578	491	578	491

Although both the treatment and the control group experienced a net increase in asset holdings over the course of the intervention, the increase for treatment households is consistently higher than the increase for control households. This translates into an improvement in the asset index for treatment relative to control households. While the average score on the asset index was higher for treatment households than for control households, at baseline, the standardized difference was only 0.011. At mid-term, this difference had increased to 0.234 highlighting the growth in asset holdings in the treatment group between baseline and mid-term (see Appendix C for the construction of the asset index). As reflected in Table 7, **the net impact of the VSLA-intervention amounts to 0.222, meaning that on average, a household that participated in a VSLA has an asset score that is 0.222 higher than a control household. This corresponds roughly to one extra head of cattle for each of the treatment households.**[24]

[24] Appendix Table D1 examines the sensitivity of the impact estimate to various scenarios about missing data. Again, scenarios (1) and (7) are extreme cases: Scenario (1) (scenario (7)) imputes the median asset change in the bottom (top) quintile of the observed treatment group to the treatment dropouts and the median asset change in the top (bottom) quintile of the observed control group to the control dropouts. Scenarios (2) and (6) repeat this, but use the second and

Table 7: Difference-in-Difference Estimation of the Impact of VSLA on Asset Holdings

	Control Households	Treatment Households	Difference
Asset Score Before VSLA	0.039	0.050	0.011
	[0.091]	[0.078]	[0.120]
Asset Score After VSLA	-0.125	0.107	0.234**
	[910.6]	[0.042]	[0.105]
Difference	-0.164***	0.057	**0.222*** **
	[0.053]	[0.057]	**[0.078]**

Notes: Standard errors clustered at the VSLA level to account for clustered randomization; Dummies for strata included; ***: Statistically significant at 1%; **: Statistically significant at 5%. VSLA impact estimator in bold.

Figures 7 and 8 illustrate the shifting asset position of treatment households. At baseline, the distribution of the asset index for treatment (red curve) and control (blue curve) households largely overlap. The dashed and full vertical lines, representing the average score on the Asset Index for treatment and control households, are similar. After the VSLA-intervention however (Figure 7), the distributions are clearly distinct, with the treatment distribution being located to the right of the control distribution. The dashed vertical line (treatment households) is now located well to the right of the full vertical line (control households). This points towards higher asset accumulation for treatment households during the intervention.

fourth quintile to do the imputations, while scenarios (3) and (5) use the fifth and sixth deciles. Scenario (4) is the non-imputed scenario presented in Table 7.

Under the extreme lower bound scenario (1), the VSLA intervention would have an estimated impact of *-0.491*. Under lower bound scenario (2), we find almost zero impact on assets (coefficient of *-0.032*, not statistically significant). Lower bound scenario (2) already results in a positive treatment effect, statistically significant at 1%. Needless to say, all higher bound estimates are big and statistically significant.

For assets we also estimated the ANCOVA estimator as we did for consumption. Using this estimator, the treatment effect amounts to *0.224* (compared to 0.222 for the difference-in-differences estimation), statistically significant at the 1% level.

Figure 7: Distribution of the Asset Index for Treatment and Control Households at Baseline

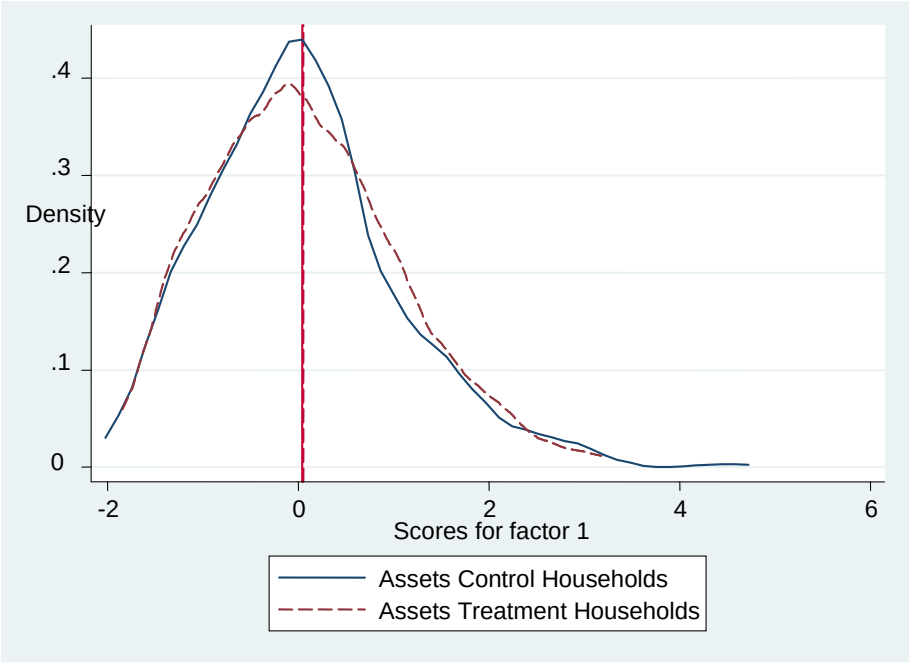
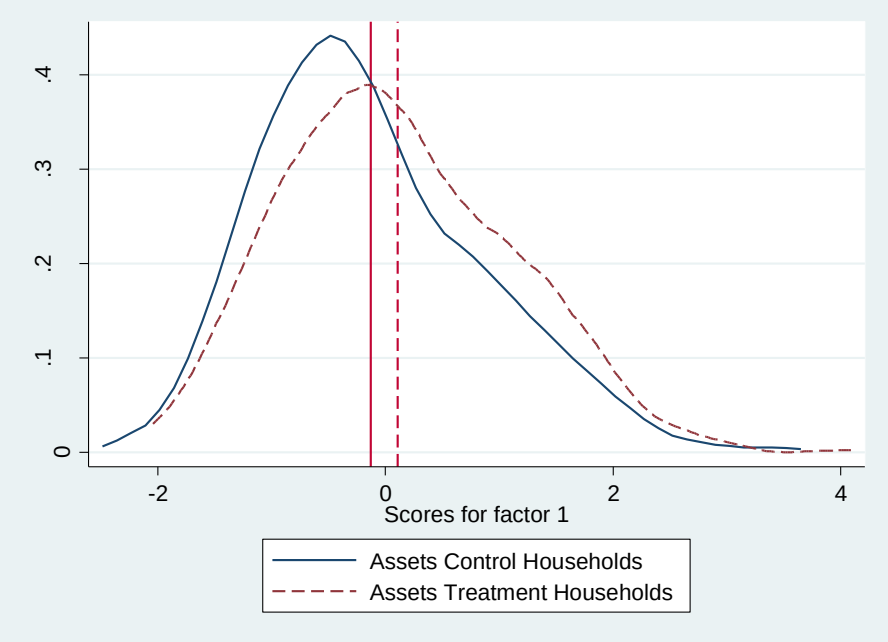


Figure 8: Distribution of the Asset Index for Treatment and Control Households at Mid-Term



2.3 The Impact of VSLA and VSLA Plus Participation on Outcomes for Children

The previous section suggests large and statistically significant impacts of the VSLA intervention on economic outcomes. The ultimate goal of the New Generation project is, however, to increase child protection, development and well-being. There are two pathways through which the project attempts to increase child well-being:

1. **Through increased financial assets of the participating households:** The assumption is that the households will use their increased financial means to invest in the development and well-being of their children;
2. **Through the family-based discussion groups**: The assumption is that engaging caregivers in discussion and skills building on parenting and issues of child protection, development and well-being will have a behavioral impact that will positively influence the way they interact with their children.

As explained in Section 2, the midterm evaluation was originally powered mainly to analyze the impact of the VSLA intervention on economic outcomes. However, in this section we analyze the impact on child outcomes and find significant impact in several key areas.

As described earlier the mid-term caregiver survey was complemented by a mid-term child survey. In order to triangulate the data provided by caregivers with the data provided by their children, the child questionnaire contained many of the same questions/modules as the caregiver questionnaire. This section will present findings from both caregivers and children

At baseline, 362 children (whose parents were part of one of the 77 original VSLAs) were administered a short questionnaire containing questions on topics similar to those seen on the household questionnaire (child well-being, mental health, discipline, family well-being, and child labor). Ideally, the mid-term child survey would have interviewed the same 362 children (aged between 10 and 14 years at baseline). However, 94 children lived in households that dropped out of the project, and hence did not qualify for re-interview in 2011. Of the remaining 268 children interviewed at baseline, 179 were re-interviewed at mid-term. Of the 89 original children that were not re-interviewed at mid-term, 32 did not live in their household anymore (36%) and 2 had died (2.2%). The remaining 55 original children could not be found (they were either not in school when the survey team visited their school or did not show up despite appointments being made).

Overall, 262 children were interviewed during the mid-term survey, 148 girls and 114 boys. Average age of the children was 12.4 years. Most children (63.4%) lived with their two biological parents. 28.8% of the children lived only with their mother. Of the 262 interviewed children, 179 are “original” (interviewed at baseline) and 83 are new children (to replace the ones who dropped out). For the new children, no baseline data exists.

This section will proceed as follows: For each topic, we will first present the data provided by the caregivers. Next, we will present the data provided by their children. For the children, the full mid-term data (all 262 children) will be presented, disaggregated by treatment status: Control, VSLA, VSLA Plus. This has the advantage of using all the data, but the disadvantage of not using the baseline data. Although the patterns uncovered by this method will be informative, using this method will make it impossible to fully attribute impact of the interventions due to the lack of baseline data. To address this challenge, data from the baseline and mid-term surveys conducted with the 179 original children will also be presented.

The Impact of the VSLA and VSLA Plus Interventions on Spending on Children

At the start of the project, three key spending areas that directly benefit children were identified: Spending on education, health and clothes. To track changes in expenditures in these categories, data were collected from the 1,069 households that were interviewed in both survey rounds (491 control households, 303 VSLA households and 275 VSLA Plus households).

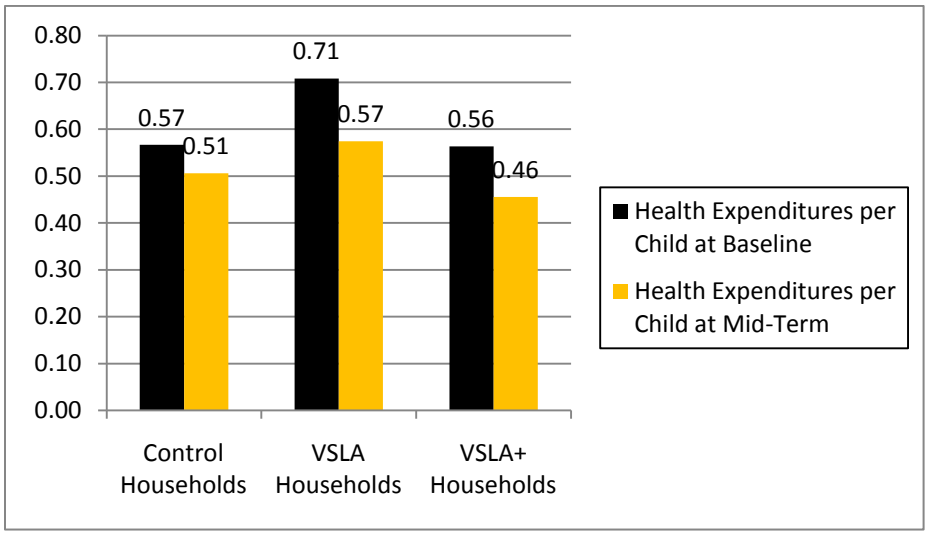
Spending on education

Caregivers were asked about expenditure on schooling for the preceding school year. The change in monthly education expenditures per child (of school-going age) is illustrated in Figure 9. **The results show spectacular increases across the board: Education expenditures in the control group increased by 82%, in the VSLA group by 115% and in the VSLA Plus group by 90%.** Just as with expenditures on food, expenditures on schooling show a more favorable trend in the treatment groups (VSLA and VSLA Plus) than in the control group, although the difference is not statistically significant (and the increase in the control group is impressive in its own right).

Spending on health

Figure 9 shows the change in expenditures on children's health between the baseline and mid-term survey, disaggregated by type of household (control, VSLA or VSLA Plus). The change in expenditures on health care per child in the household is puzzling. **For all groups, child health expenditures dropped, and the drop is higher for VSLA Plus (-19%) and VSLA households (-19%) than for control households (-11%).** The difference between VSLA Plus and control is statistically significant. The reason for this decrease overall and in the VSLA Plus group is puzzling and we will further investigate this in the final evaluation.

Figure 9: Health Expenditures per Month per Child, Baseline and Mid-Term (PPP USD)



Spending on clothing

The monthly expenditures on clothing per child in the household increased for all groups, though the increase is substantially larger for households in the VSLA (+27%) and VSLA Plus (+42%) groups than for households in the control group (+16%). The finding that the increase is substantially larger for the VSLA Plus households than for the VSLA households might suggest that discussion group participation had additional impact, although this is not statistically significant at this stage.

The Impact of VSLA and VSLA Plus on Child Labor

Changes in Child Labor According to Caregivers

This subsection provides an overview of the change in the incidence of child labor for the 1,069 households interviewed at baseline and mid-term. The baseline survey provided information on child labor for 1,104 children between 5 and 14 years of age (in the 1,069 households). The mid-term survey provides information on 995 of these children. Of the 109 children who were not interviewed during baseline and mid-term, 62 had left the household, 9 had died. It was also found that 38 children had been invented during the baseline survey.[25]

[25] This is a well-known phenomenon in many poor countries with a heavy NGO presence. Respondents tend to exaggerate the number of children in their household hoping that this would attract material assistance towards their household.

To measure child labor, UNICEF's questionnaire and widely-accepted definition of child labor was used. For children between 5 and 11-years of age, UNICEF defines child labor as at least 1 hour of economic work or more than 28 hours of domestic work per week. For older children (12 to 14-years old), child labor is defined as at least 14 hours of economic work or more than 28 hours of domestic work per week. While the incidence of child labor decreased for the control households (- 4 percentage points) and VSLA households (-2.5 percentage points), it actually increased for the VSLA Plus households (+ 4.8 percentage points). The differences are not statistically significant.

For older children between 12 and 14 years of age, the pattern is opposite of that of the younger children. While the incidence of child labor increased for the control households (+ 2 percentage points) and the VSLA households (+ 5.5 percentage points), it remained stable at 6.8% for the VSLA Plus households. The differences are not statistically significant.

Following UNICEF's definition, domestic work consists of domestic chores such as cooking, cleaning, and caring for children, fetching water and firewood, etc. Economic labor consists of paid or unpaid work for someone who is not a household member and work on the family farm or in family-run businesses.

.

Overall, regardless of age category the incidence of child labor dropped from 30.8% at baseline to 28.3% at mid-term for control households, and increased from 29.2% at baseline to 30.6% at mid-term for the households with VSLA or VSLA Plus. **Despite these changes, it is not possible yet to draw a conclusion about the impact of VSLAs on child labor.** It is important to note, however, that there is a potential link between VSLAs and increased child labor: As economic activity at the household level picks up as a result of VSLA participation, the household may be in need of extra labor. The data from the final survey planned in 2012 should show whether these differences are statistically significant.

Changes in Child Labor According to Children

In line with the findings from the baseline survey, children report a higher workload than their parents do: Applying UNICEF's definition of child labor, 59.5% of children aged 10 or 11 in the sample were engaged in child labor during the week preceding the survey, compared to the 37.7% reported by parents. For the older children (12 to 14-years-old), 15.4% were involved in child labor.

Due to the smaller sample size of children, groups were aggregated (10 or 11-years old and 12 to 14-year old) to arrive at the overall incidence of child labor in the mid-term sample. The incidence of child labor among the 262 children interviewed at mid-term amounts to 31.3%, and is higher for the VSLA Plus children (35.9%) than for the VSLA (31.5%) and control group children (28.8%).

When focusing exclusively on the children interviewed both at baseline and mid-term, we find decreases in child labor in all groups. The improvement in the VSLA groups is particularly pronounced (-8 percentage points) and higher than the improvement in the VSLA Plus group (-3.8 percentage points). Due to the small sample size however, none of the impacts are statistically significant.

The household survey and the child survey are similar in that both surveys are inconclusive on the impact of the project on child labor. The trends discovered in both surveys seem, however, to be at odds with each other: While the household survey suggested that child labor actually increased in the treatment groups (VSLA and VSLA Plus) relative to the control group, the child survey shows a reduction in child labor that is higher for the treatment than for the control group. The final survey in 2012 should shed more light on the impact of the intervention on child labor.

The Impact of VSLA and VSLA Plus on Child Discipline

Changes in Child Discipline According to Caregivers

The surveys measured the incidence of physical and psychological punishment with the standard UNICEF MICS scale. The baseline survey documented use of harsh child discipline, both physical and (especially) psychological. Because of the strong negative relationship between harsh child discipline and the level of child well-being (as documented by the baseline survey), the "Healing Families and Communities" discussion modules were designed to include sessions on the issue of child discipline. The results show that the discussion modules have affected parents' use of harsh methods to discipline children.

Table 8 shows the percentage of parents/caretakers who reported that someone in their household has used a specific method for disciplining children in the month preceding the survey. Overall, in all groups (control, VSLA and VSLA Plus) we find reductions in the use of harsh child discipline methods. In the control group, all but two of the nine child discipline techniques decreased between the baseline and the mid-term survey. Two show a substantial and statistically significant decrease: "Shouting, yelling or screaming at the child", from 63.6% at baseline to 55.8% at mid-term, and "shaking the child", from 25.3% at baseline to 12.6% at mid-term. For the VSLA households, the improvements in child discipline seem slightly more modest than for the control households. Five of the nine child discipline indicators show an improvement (vs. seven for the control households). When considering the aggregate score on the child discipline scale (minimum of 0 and maximum of 9, where 0 means that none of the child discipline methods were used and 9 that all the methods were used), the results show a reduction of *0.3* for the control households and a reduction of *0.2* for the VSLA households.

However, **for households that participated in the discussion groups (VSLA Plus) the mid-term survey found substantial and across-the-board improvements in child discipline. Six improvements are statistically significant.** In terms of absolute values, the biggest improvements were found in verbal punishment: The percentage of caregivers that reported that they "shouted, yelled or screamed at the child" decreased by over 19 percentage points and the percentage that insulted the child ("called the child dumb, lazy or another name like that") decreased by almost 20 percentage points. Results also showed a large reduction in physical punishment for children in VSLA Plus households:

- The percentage of households that shook the child almost halved (from 25.8% at baseline to 13.6% at mid-term);

- The percentage of households that spanked, hit or slapped the child with the bare hand dropped by 41% (from 17.2% at baseline to 10.2% at mid-term);
- The percentage of households that hit the child on the bottom or elsewhere on the body with a belt, stick or other hard object dropped by 64% (from 7% at baseline to 2.5% at mid-term);
- The percentage of households that hit or slapped the child on the face, head or ears dropped by 47% (from 4.7% at baseline to 2.5% at mid-term);
- The percentage of households that beat the child up, hit the child over and over again dropped by 79%[26] (from 3.9% at baseline to 0.8% at mid-term);
- The percentage of parents/caretakers who think physical punishment is necessary in order to properly raise a child dropped by 65% (from 2.3% at baseline to 0.8% at mid-term).

These substantial improvements in child discipline translate to a 0.8 reduction in the aggregate score on the discipline scale for the VSLA Plus households, from 2/9 at baseline to 1.2/9 at mid-term. This reduction is strongly statistically significant at the 1% level. Here the results illustrate the first unambiguous positive effect of the Healing Families and Communities Discussion Modules.

[26] When there were reports of a child being beaten up and hit over and over again in the survey in any of the groups, they were followed up by a child protection IRC staff to assess the safety of the child and find an appropriate solution or referral to reduce the risk of harm to the child.

Table 8: Change in Child Discipline Methods between the Baseline and Mid-Term Survey

	Control			VSLA			VSLA Plus		
	Baseline	Mid-Term	*Diff*	Baseline	Mid-Term	*Diff*	Baseline	Mid-Term	*Diff*
Shouted, Yelled or Screamed at the Child	63.6	55.8	*-7.8**	66.9	60.9	*-6*	70.3	50.9	*-19.4****
Shook the Child	25.3	12.6	*-12.7****	26.1	18.8	*-7.3**	25.8	13.6	*-12.2****
Spanked, Hit or Slapped the Child with the Bare Hand	17.8	14.7	*-3.1*	23.9	18	*-5.9*	17.2	10.2	*-7**
Hit the Child on the Bottom or Elsewhere on the Body with a Belt, Stick or Other Hard Object	4.2	4.2	*0*	8.5	9	*0.5*	7	2.5	*-4.5**
Call the Child Dumb, Lazy or Another Name like That	39.7	39.5	*-0.2*	45.8	39.9	*-5.9*	47.7	27.9	*-19.8****
Hit or Slapped the Child on the Face, Head or Ears	3.3	4.2	*0.9*	4.2	6	*1.8*	4.7	2.5	*-2.2*
Hit or Slapped the Child on the Hand, Arm or Leg	16.4	12.6	*-3.8*	18.3	20.3	*2*	16.4	8.5	*-7.9***
Beat the Child Up, Hit the Child Over and Over Again	4.2	3.7	*-0.5*	3.5	4.5	*1*	3.9	0.8	*-3.1*
Believe that a Child Has to be Physically Punished for a Good Education	5.6	5.3	*-0.3*	4.2	3	*-1.2*	2.3	0.8	*-1.5*
Score on the Discipline Scale	**1.8**	**1.5**	**-0.3****	**2**	**1.8**	**-0.2**	**2**	**1.2**	**-0.8****

The mid-term survey asked a number of questions on positive ways of disciplining children that were not included on the baseline. Items were adapted from the Multiple Indicator Cluster Survey (UNICEF) and two other measures of parenting behaviors.[27] Six positive ways of disciplining children were considered:

1. Explaining to the child why his/her behavior was wrong
2. Telling the child to stop what s/he is doing and giving the child something else to do
3. Giving the child a "time-out" away from other people and fun things to do
4. Setting rules for the child's behavior at home
5. Complementing the child when s/he has done something good
6. Giving the child extra work (appropriate chores)

Table 9 shows the incidence of positive discipline for the control, VSLA and VSLA Plus households. Overall, it seems that households that received an intervention (VSLA or VSLA Plus) were more likely to use positive disciplining techniques than control households. This is especially the case for "Explaining to the child why his/her behavior was wrong" and "Giving the child a time-out". The discussion modules seem to have had a particularly strong impact on two positive discipline techniques: 76.3% of VSLA Plus households set rules for the child's behavior in the home, vs. 63.9% of VSLA households and 61.6% of control households (the differences are statistically significant). 79.7% of VSLA Plus households reported to have complimented the child when s/he had done something good compared to. 66.4% for VSLA households and 61.6% for control households (the differences are statistically significant).

[27] Rohner, R. P., & Khaleque, A. (Eds.). (2005). *Handbook for the study of parental acceptance and rejection* (4th ed.). Storrs, CT: Rohner Research Publications; Lansford, J. E., Chang, L., Dodge, K. A., Malone, P. S., Oburu, P., Palmérus, K., Bacchini, D., Pastorelli, C., Bombi, A. S., Zelli, A., Tapanya, S., Chaudhary, N., Deater-Deckard, K., Manke, B., & Quinn, N. (2005). Cultural normativeness as a moderator of the link between physical discipline and children's adjustment: A comparison of China, India, Italy, Kenya, Philippines, and Thailand. *Child Development*, 76, 1234-1246.

Table 9: The Incidence of Positive Discipline Techniques, Mid-Term Survey

	Control	VSLA	VSLA Plus
Explain the child why his/her behavior was wrong	57.9	64.7	65.3
Tell the child to stop what s/he is doing and give the child something else to do	53.2	60.9	58.5
Give the child a "time-out" away from other people and fun things to do	16.3	37.6	33.1
Set the rules for the child's behavior in the home	61.6	63.9	76.3
Compliment the child when s/he has done something good	61.6	66.4	79.7
Give the child extra work	12.1	13.5	11

Since there is no baseline data on positive discipline, it is not possible to examine whether there were changes in regard to positive discipline across the different groups (VSLA, VSLA Plus, control). However, randomization should in theory make sure that there are no observable or unobservable differences between groups, which means that the differences observed in Table 9 are due to the intervention.[28]

Changes in Child Discipline According to Children

The household survey revealed large improvements in child discipline practices since the baseline survey, in particular for the VSLA Plus households. The child data corroborates this finding. Table 10 shows the percentage of children who, at mid-term, reported having been subjected to the different discipline methods in the four weeks preceding the interview. For each discipline methods, the prevalence is lower among the VSLA Plus children than among the control and VSLA children. The differences are substantial:

- The percentage of children who have been shouted or yelled at is 13 percentage points lower in the VSLA Plus than in the control group;
- The percentage of children who have been hit with the bare hand is 10 percentage points lower in the VSLA Plus group than in the control group
- The percentage of children who have been hit in the face, head or ears is less than half in the VSLA Plus group than in the control group

[28] Keeping in mind the selective drop-out documented in Section 3.2, these are likely to be upper bounds of the effect.

Table 10: Incidence of Child Discipline Practices According to Children, Mid-Term (in %)

	Control	VSLA	VSLA Plus
Shouted, Yelled or Screamed at the Child	64.5	57.9	51.8
Shook the Child	35.2	35.3	26.4
Spanked, Hit or Slapped the Child with the Bare Hand	49.1	49.3	38.9
Hit the Child on the Bottom or Elsewhere on the Body with a Belt, Stick or Other Hard Object	46.4	40	32.1
Call the Child Dumb, Lazy or Another Name like That	48.6	47.8	30.9
Hit or Slapped the Child on the Face, Head or Ears	38.3	36.2	16.7
Hit or Slapped the Child on the Hand, Arm or Leg	43.7	40	20.7
Beat the Child Up, Hit the Child Over and Over Again	11.7	13	8.3
Score on the Discipline Scale	3	3	1.9

At mid-term, the aggregate score on the discipline scale amounts to 3 (out of 8) for both control and VSLA children.[29] For VSLA Plus children, the score is significantly lower (1.9), indicating that children in VSLA Plus households experienced less harsh discipline practices than other children (statistically significant at the 1% level).

The figures presented in Table 10 are indicative of a positive impact of the interventions. However, Table 10 only presents cross-sectional mid-term data. If the children in the three groups (control, VSLA, VSLA Plus) were different at baseline, then it would be incorrect to attribute the patterns in Table 10 to our interventions.

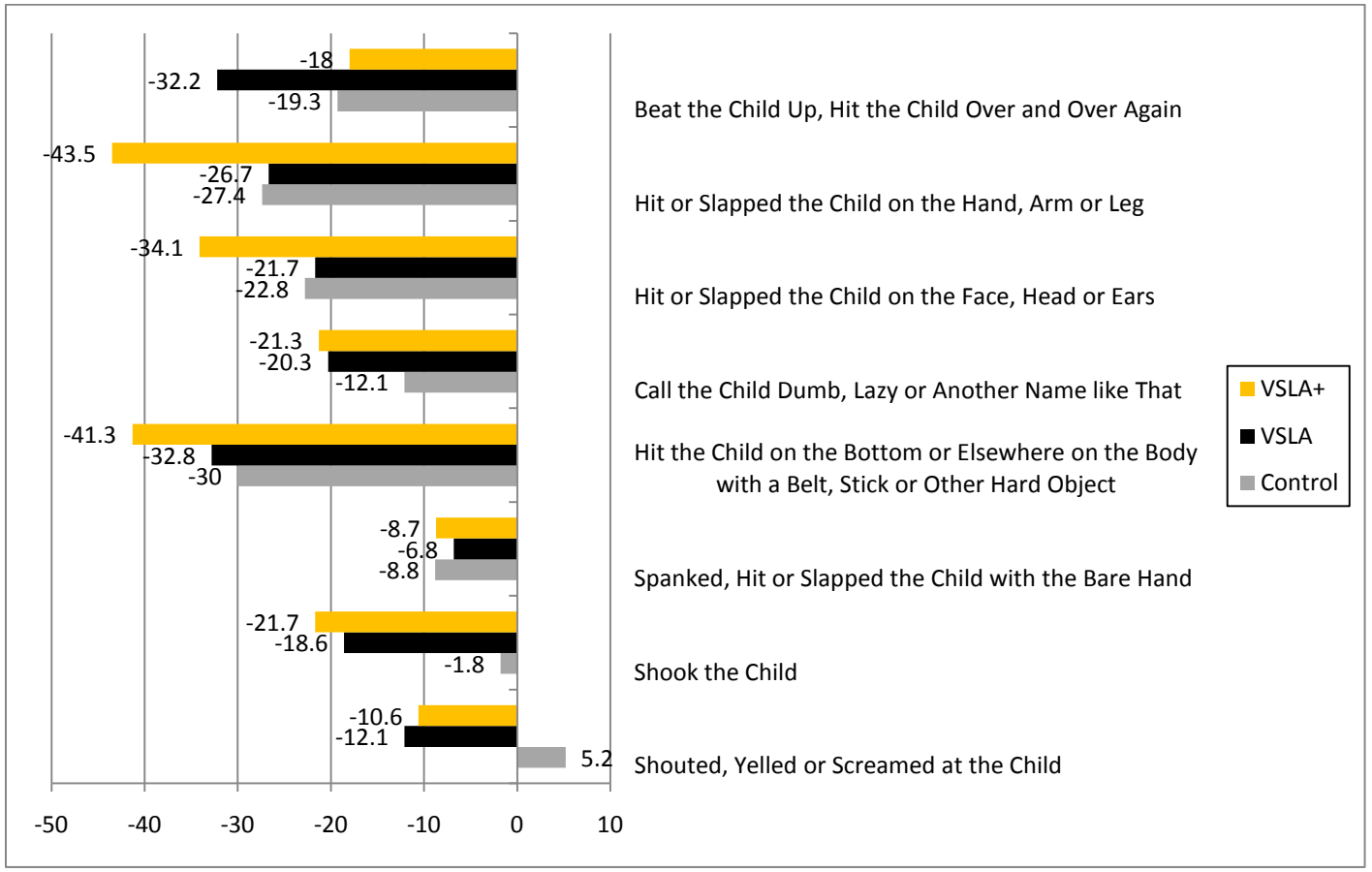
Figure 10 shows the percentage point change in the incidence of child discipline practices for the children who were interviewed at baseline and midterm. The picture that emerges is more complicated. We find across-the board improvements in child discipline, and while the improvements are generally bigger in treatment (VSLA and VSLA Plus) groups than in control groups, this is not always the case. Improvements for the VSLA Plus children are consistently larger than the improvements for the control children, except in regard to one practice (beating up the child). Improvements for the VSLA children (without discussion modules) are also generally better than those for control children (but again not for each child discipline practice). At the

[29] The aggregate discipline scale varies between 0 and 8 for the child survey, and between 0 and 9 for the household survey. This is because parents were asked one extra question ("Do you believe that in order to properly raise a child, s/he needs to be physically disciplined?").

aggregate level, the score on the discipline scale dropped (improved) from 4.4 (on 8) at baseline to 2.7 at mid-term. The improvement was higher for VSLA Plus (-48.7%) and VSLA children (-38.8%) than for control children (-32.6%).

Despite the differences in child discipline findings between the household (parent) and the child surveys, the key overlap and key message from both surveys is that to reduce harsh child discipline practices, VSLA participation does not suffice. **VSLA and discussion modules are significantly more effective in reducing harsh child discipline practices than VSLA participation alone.**

<u>Figure 10:</u> Percentage Point Change in the Incidence of Child Discipline Practices between Baseline and Mid-Term, According to Children



In line with the household survey, the mid-term child survey also included six questions on the incidence of positive methods of discipline that were not asked at baseline. Table 11 shows the percentage of children who reported that their parent/caretaker (who is in the VSLA) has used a specific method in the four weeks preceding the survey.

Table 11: The Incidence of Positive Techniques of Disciplining, by Treatment Status (Child Data)

	Control	VSLA	VSLA Plus
Explain to me why my behavior was wrong	77.4	79.5	74.2
Tell me to stop what I am doing and give me something else to do	63.7	68.5	71.4
Give me a "time-out" away from other people and fun things to do	20.8	32.9	33.3
Set the rules for my behavior in the home	58.9	63.01	57.1
Compliment me when I have done something good	84.7	83.6	90.5
Give me extra work	35.5	30.1	27

Overall, children reported a higher incidence of positive disciplining methods than their parents did (see Table 9). So, while children report higher incidences of harsh discipline, they also report higher incidences of positive discipline methods. Table 11 does not show a clear distinction by treatment status. While the incidence of certain positive discipline methods is higher among the children whose parents participated in the discussion modules than among other children (notably the second, third and fifth technique in the Table), other methods are more prevalent among control and VSLA children than among VSLA Plus children. No clear pattern emerges from the data. Since there is no baseline data, we have less certainty that the patterns in Table 11 are driven by the interventions. The final survey will shed more light on this.

Impact of the VSLA and VSLA Plus Interventions on Child Well-Being

Changes in Child Well-Being According to Caregivers

Caregivers were asked seven questions on the perceived well-being of their children at baseline and midterm with reference to the same child. These questions were developed based on qualitative interviews with caregivers and children. The seven items of child well-being were made into a scale measuring aggregate child well-being. The scale ranges from 0 to 14, with 0 indicating “no well-being at all” and 14 indicating maximum well-being. For all three groups of households, large improvements were found on the child well-being scale: From 4.9 to 7 for control households (+42.9%), from 4.7 to 7.4 for VSLA households (+57.4%), and from 5 to 7.6 for VSLA Plus households (+52%). Only the VSLA change is statistically significant. While participation in VSLA seems to have had an unambiguous positive effect

on child well-being (according to their parents), there is so far no evidence of an additional discussion-group effect.

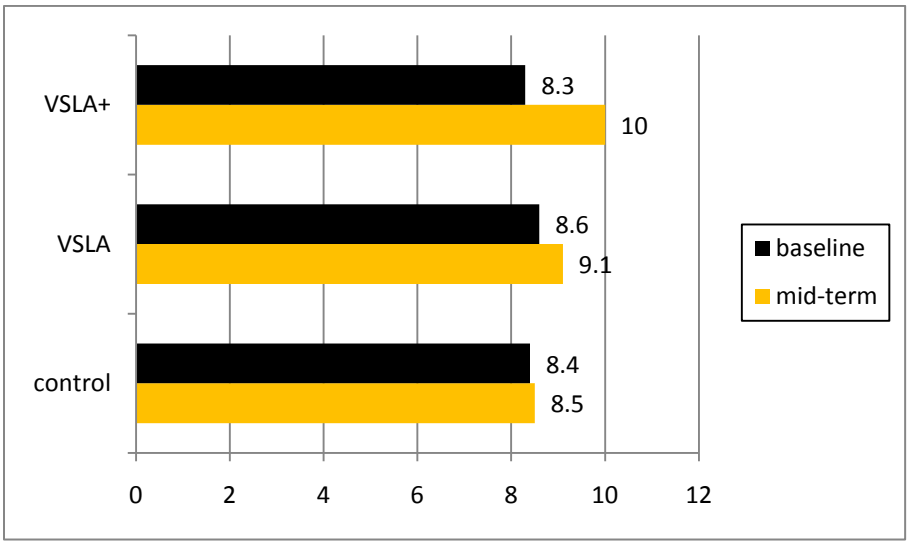
How can these improvements in child well-being among the control households who did not benefit from any of the IRC's interventions during the first cycle of the project be explained? There are a number of possibilities, all of which are hypothetical and currently un-testable:

- It is possible that for an unknown reason, overall child well-being in Burundi simply improved in 2010, regardless of the IRC intervention; this is the least likely explanation.
- It is possible that households in the control groups were exposed to projects of other NGOs or organizations working on issues of child well-being, which would explain the improvements;
- Although control households did not receive the interventions in 2010, it is likely they are aware of the fact that the intervention is essentially about children and their well-being. This might reflect in the data in two ways: First, they may have been inclined to give desirable answers to the questions on well-being (and discipline); second, knowing that the project is about child well-being may have brought about a real change in the way they interact with their children.

Changes in Child Well-Being According to Children

The children's report on well-being shows a different picture than the caregiver report.

Figure 11: The Impact of VSLA and VSLA Plus on Child Well-Being



While aggregate child well-being in the control group remained unchanged (from 8.4 to 8.5 out of a total of 14), the well-being score for VSLA children increased by 5.8% and by 20.5% for children whose caretaker was enrolled in a VSLA with discussion modules (from 8.3 at baseline to 10 at mid-term). While the effect of a VSLA alone just misses statistical significance at conventional levels, **the impact of VSLA combined with discussion groups is highly statistically significant.** Note the difference here between these findings and the findings from the household survey: According to the household data, being in a VSLA alone increased levels of child well-being more than being in a VSLA with discussion modules. According to the child data, however, it is the combination of VSLA participation and discussion modules that substantially improves child well-being.

The Impact of VSLA and VSLA Plus on Child Mental Health

Changes in Child Mental Health According to the Caregiver

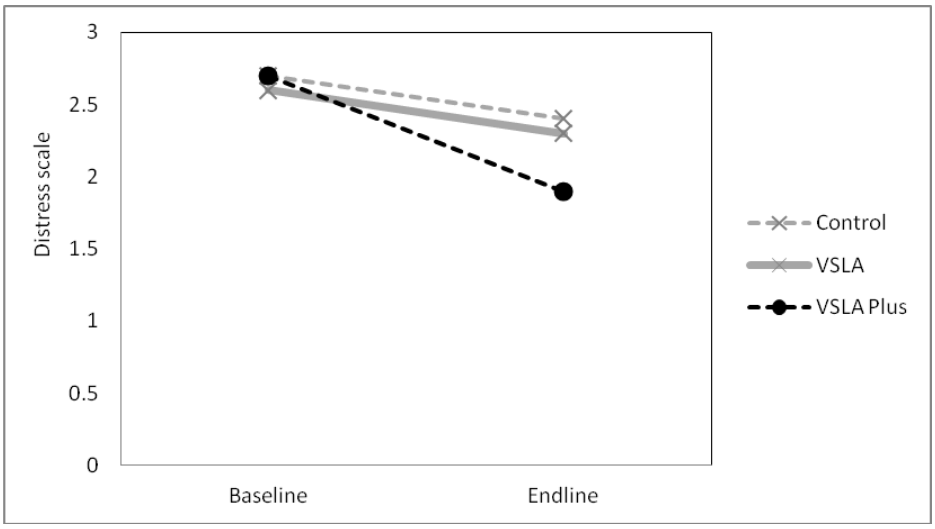
During the baseline survey caregivers were asked 10 questions on the mental health of their children (as perceived by the parent/caregiver). These items were derived from qualitative interviews with caregivers before the baseline who named important problems their children faced and common symptoms they saw in children with problems. The 10 items divided into two scales, one measuring the level of distress and the other the level of aggression (internalizing and externalizing behavior, respectively). The same questions were asked at mid-term, with reference to the same children.

Figure 12 shows the change in distress scores of children in the control, VSLA and VSLA Plus households. The distress scale consists of 7 items:

1. Feeling worried
2. Feeling anxious
3. Feeling dizzy (because of bad thoughts and worries)
4. Feeling sad
5. Being withdrawn
6. Crying
7. Isolating oneself from others

The distress scale ranges between 0 and 21, with 0 indicating that the child showed no distress symptoms at all and 21 being the maximum level of distress.

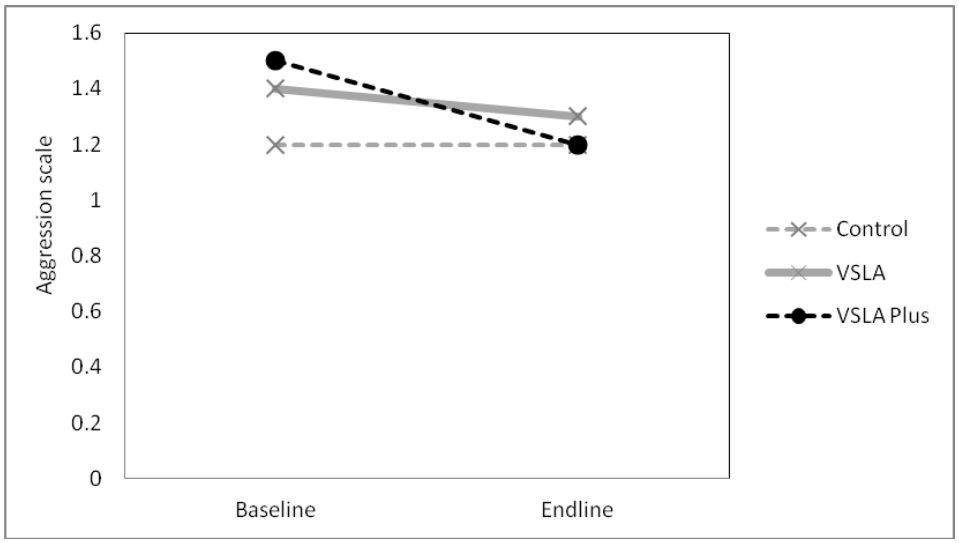
Figure 12: Changes in the Child Distress Score, Baseline and Mid-Term



At baseline, children in all three groups of households showed similar distress scores. At mid-term, all three groups showed a decrease in the child distress score, though the decrease is only statistically significant for the VSLA Plus group. While the distress score dropped by *0.3* for both control and VSLA households, it dropped by *0.8* for VSLA Plus households. This indicates that discussion groups reduced children's distress, as perceived by their caregivers.

The aggression scale consists of three items (insulting others, not respecting caregivers, being aggressive). The aggression scale ranges from 0 (no signs of aggression) and 9 (all signs of aggression). Figure 13 shows the change in the aggression scores between the baseline and the mid-term survey.

Figure 13: Changes in the Child Aggression Score, Baseline and Mid-Term



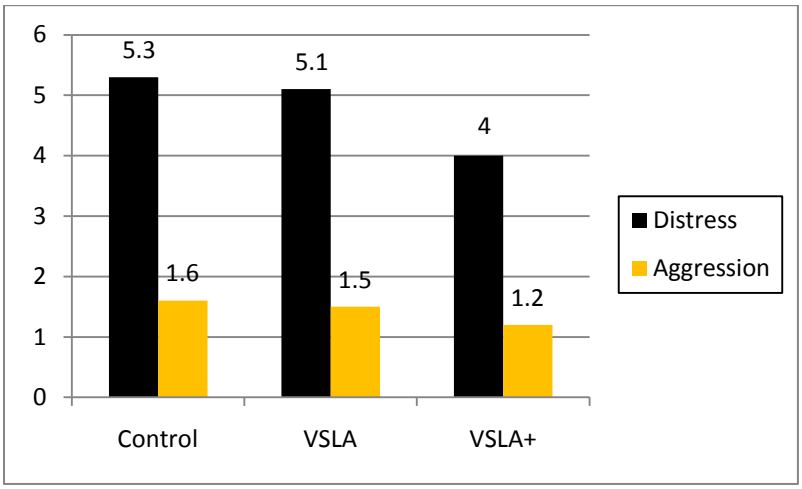
Overall, aggression scores were low at baseline (between 1.2 out of 9 for control households and 1.5 for children in VSLA Plus households) and changed little between baseline and mid-term. However, for children in the VSLA Plus households, a 20% drop in the aggression score was observed, which is statistically significant at the 5% level.

In general, the figures for the distress and aggression scores show that the discussion modules had a positive impact on indicators of child mental health (according to their caregivers).

Changes in Child Mental Health According to Children

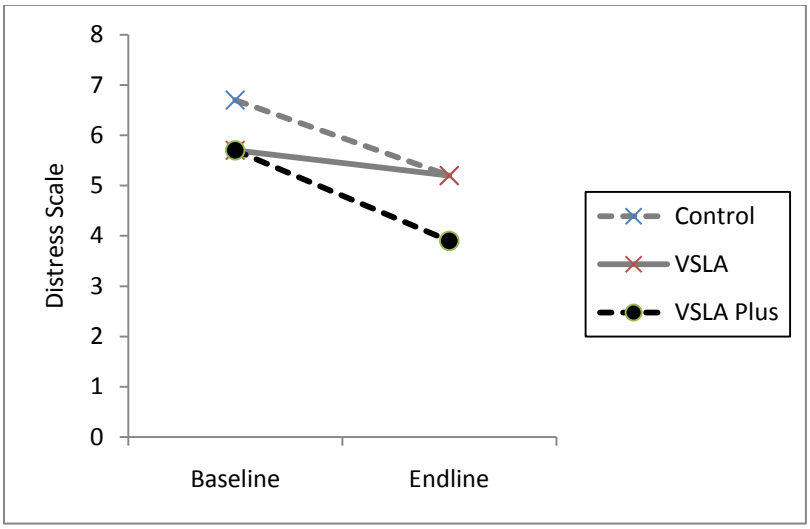
The mid-term child data confirms the findings from the caregiver survey. At mid-term, the level of distress among VSLA Plus children (4/21) is lower than that of VSLA children (5.1/21) and control group children (5.3/21). The difference in distress levels between VSLA Plus and VSLA children is statistically significant, as is the difference between VSLA Plus and control group children. For aggression, we also find that VSLA Plus children exhibited less externalizing behavior than VSLA and control group children (score on the aggression scale of 1.2/9 for the VSLA Plus children vs. 1.5 and 1.6 for VSLA and control children respectively). These differences are, however, not statistically significant.

Figure 14: Levels of Child Distress and Aggression According to the Children, Mid-Term



When focusing only on the children who were interviewed both at baseline and midterm, the impact of the interventions becomes less obvious. While there are still clear across-the-board improvements (i.e. declines) in distress levels, the improvement is actually substantially bigger for control group children (-20%) than for VSLA children (-8.8%), this is in contrast to the findings from the caregiver report. The improvement is still largest for the children in VSLA Plus households, but due to the large improvement among the control group children, the effect is not statistically significant. Regardless of this, the consistent improvement in child distress levels is of course good news.

Figure 15: Decrease in Distress Levels among Children Interviewed during both Baseline and Mid-Term Surveys



Parents tend to underestimate the level of distress amongst their children. While parents reported an average distress score of 2.2, children set the mark considerably higher at 4.9.

The Impact of VSLA and VSLA Plus on Family Well-Being

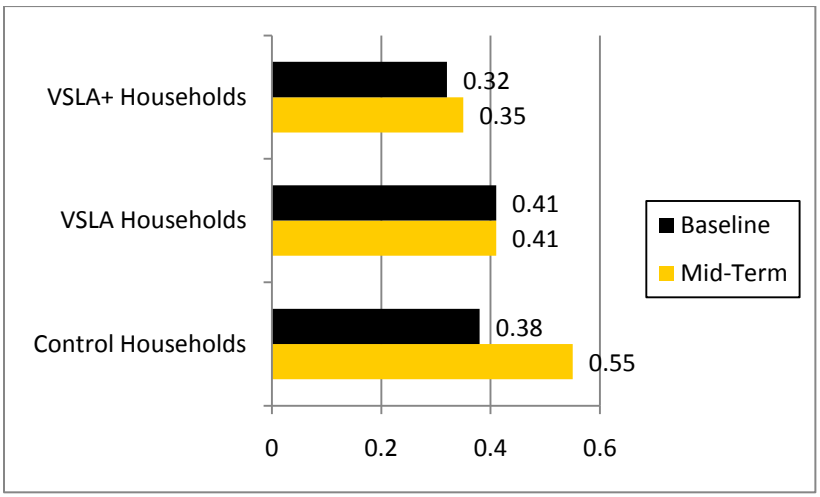
Changes in Family Wellbeing According to Caregivers

Family well-being was measured by two scales constructed from items identified by both children and caregivers during qualitative interviews. The family *well-being* scale consisted of the following items: good understanding among family members; dividing work together; and getting along well with neighbors. The family *problems* scale included the following items: violence among family members; intoxication; and a family member selling things in the home without consent. There was an overall decrease in family well-being during the first cycle of the project. In the full sample, the average score on the family well-being scale dropped from 4.4/6 to 4.0/6.

The decrease was, however, larger in the control households (decrease of 11.6%) than for the VSLA (-2.3%) and VSLA Plus households (-6.7%) although these differences are not significant. From this pattern, it does not seem that participation in discussion groups spurred overall family well-being (although participation in VSLA might have). This will be further investigated in the final survey.

The change in the occurrence of family problems is consistent with that of family well-being: While family well-being decreased between the baseline and the mid-term survey, family problems increased from 0.37/6 to 0.46/6. The increase in family problems is substantially higher among the control households (increase of 44.7%) than among the VSLA (no increase) and VSLA Plus households (increase of 9.4%).

Figure 16: Change in Family Problems Between the Baseline and Mid-Term Survey, Caregiver report



According to the caregivers' reports, it seems that participation in the Healing Families and Communities discussion groups insulated the treatment households from a general rising trend in family problems. While the incidence of family problems significantly increased among control households, it did not significantly change in either VSLA or VSLA Plus.

Changes in Family Wellbeing According to Children

Children report a slightly different picture from caregivers. Figure 17 shows the mid-term scores on the family well-being and family problems scales as reported by children. While the family well-being score is similar across the three groups (4.4/6), the level of family problems was significantly higher among control households (1.3/6) than among VSLA (0.86) or VSLA Plus households (0.55).

As shown in Figure 18, according to children, there have been large reductions in family problems for all three groups, but in particular for the VSLA Plus households. Thus, the discussion modules seem to have been effective at lowering the incidence of problems within the family. However, the results do not show evidence that the interventions were effective in increasing the level of family well-being. In contrast, the increase in family well-being was highest among control households (but not significantly so).

Figure 17: Levels of Family Well-Being and Family Problems According to the Mid-Term Child Survey

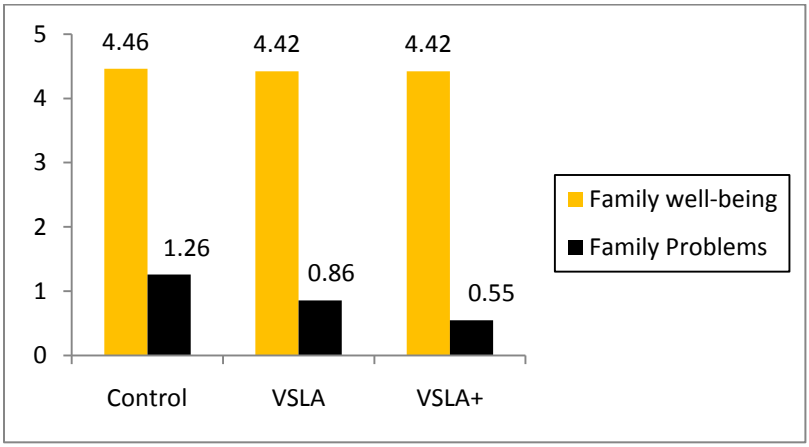
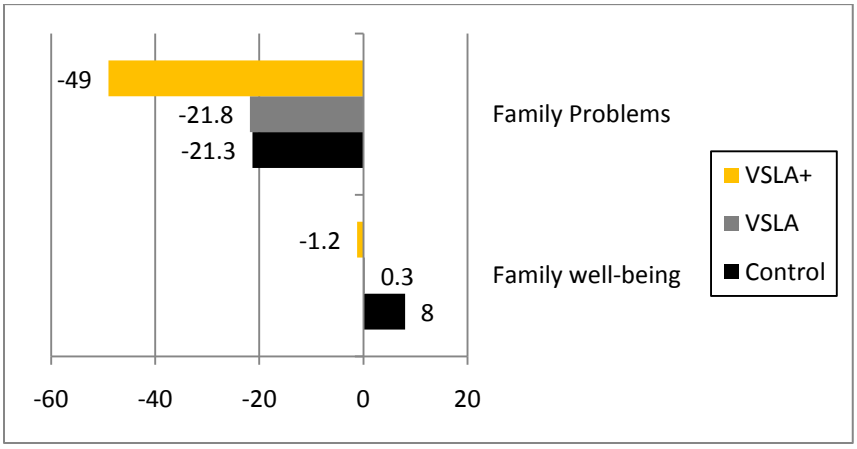


Figure 18: % Change on Family Well-Being and Family Problems Scales between Baseline and Mid-Term



The Impact of VSLA and VSLA Plus on Parent-Child Communication about Material Needs

Changes in Parent-Child Communication about Material Needs According to Caregivers

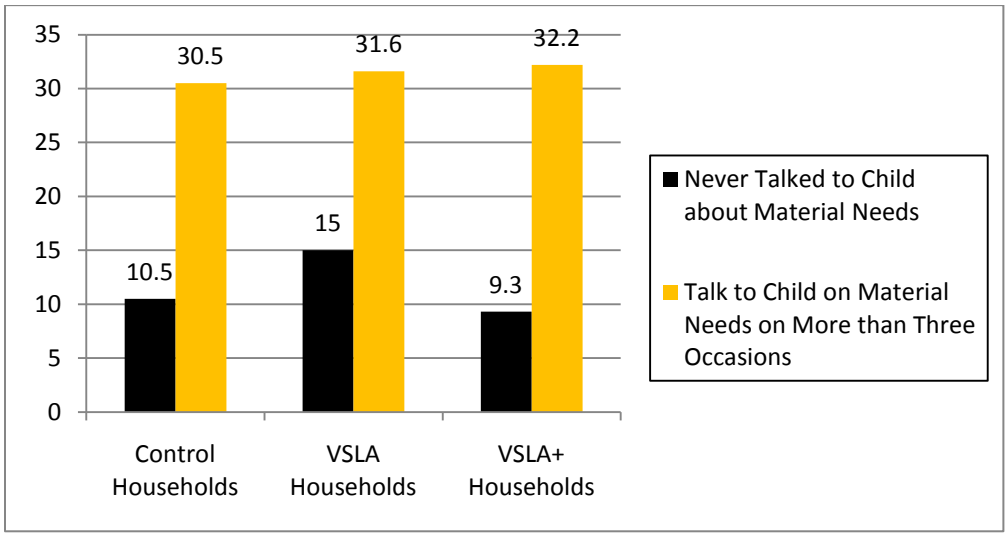
The mid-term survey included two new questions on parent-child communication. Both questions concerned the material needs of the child and were rated on a 4-point frequency scale:

1. In the past month, how many times have you discussed with your child his/her material needs?
2. In the past month, have you been able to respond to a material need expressed by your child?

The decision to only focus on material needs was inspired by the results of the baseline child participatory activities, which revealed that the bulk (>90%) of wishes expressed by children are material (having clothes, having school uniforms, having school material, having enough to eat, etc.).

Figure 19 shows the percentage of respondents who reported (a) never having talked with their child about his/her material needs and (b) having talked with the child about his/her material needs on more than three occasions during the past month.

Figure 19: Parent – Child Communication about Children's Material Needs

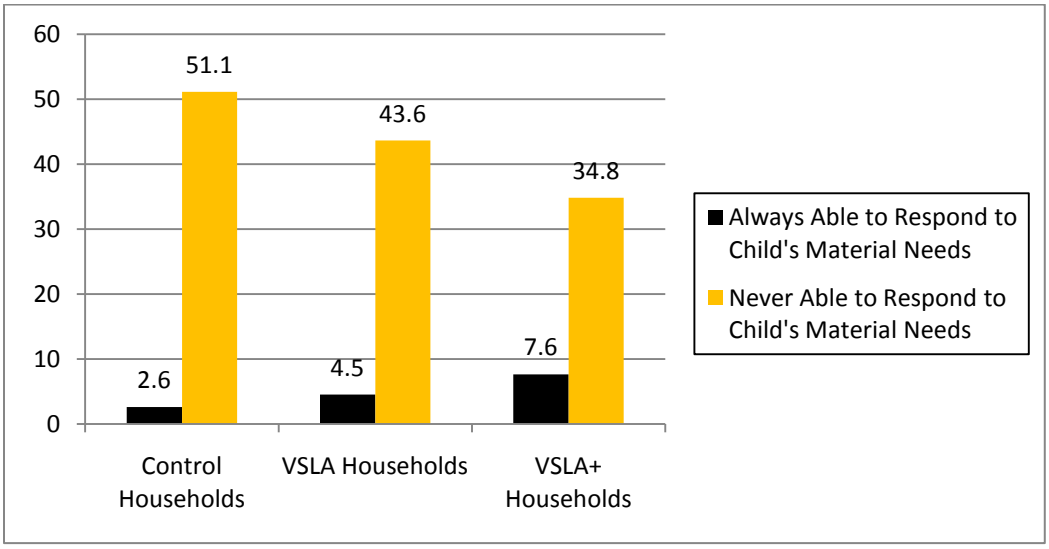


Overall, respondents who participated in the discussion groups talked a bit more frequently with their children about material needs than respondents in the VSLA and control groups. The differences are, however, small and are not statistically significant. 11.6% of respondents reported never having talked with their child about his/her material needs. This drops to 10.5% for control households and 9.3% for VSLA Plus households, but rises to 15% for VSLA households. 31.3% of respondents said they have talked with their child concerning his/her material needs on at least three occasions during the month preceding the survey. This was slightly higher for VSLA Plus households (32.2%) than for VSLA (31.6%) and control households (30.5%).

Figure 20 shows the percentage of parents who reported (a) never and (b) always being able to respond to a material need expressed by their child. Here we find an unambiguous effect, both of the VSLA and VSLA Plus interventions. More VSLA households (4.5%) relative to control households (2.6%) reported always being able to respond to children's material needs and less VSLA (43.6%) than control households (51.1%) reported never being able to respond to material needs expressed by their child.

This finding most likely reflects the VSLA's effectiveness in improving the economic welfare of the participants. The discussion modules seem to have had an additional impact on parents' ability, and possibly willingness, to respond to children's material needs: A higher proportion of VSLA Plus households (7.6%) than VSLA households (4.5%) reported always being able to respond to children's material needs and a smaller proportion of VSLA Plus (34.8%) than VSLA households (43.6%) reported never being able to respond to material needs expressed by their child.

Figure 20: Parents' Ability to Respond to Children's Material Needs



Changes in Parent-Child Communication About Material Needs According to Children

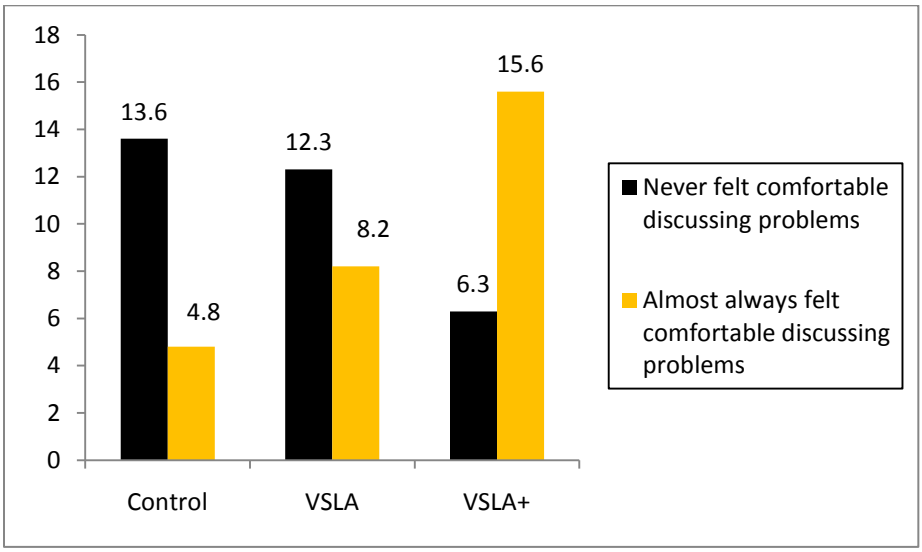
The mid-term child survey included two questions on parent child communication. These questions were not included at baseline and as such we can only show the mid-term distribution by treatment status. The two questions were:

1. In the past month, did you feel comfortable discussing your **problems** with an adult in your household?
2. In the past month, did you feel comfortable discussing your **emotions** with an adult in your household?

Both questions were rated on an ordinal four-point scale: Never, sometimes, quite frequently and a lot.

Figure 21 shows the responses to the first question. To simplify the figure the incidence of the two extreme answers ("never" and "a lot") is reported for the three groups. The patterns presented in Figure 21 indicate that the discussion modules have a positive effect on communication: The percentage of children who never felt comfortable discussing their problems with an adult in the household is approximately 50% lower for the VSLA Plus children (6.3%) than for the VSLA or control group children (12%-13%).

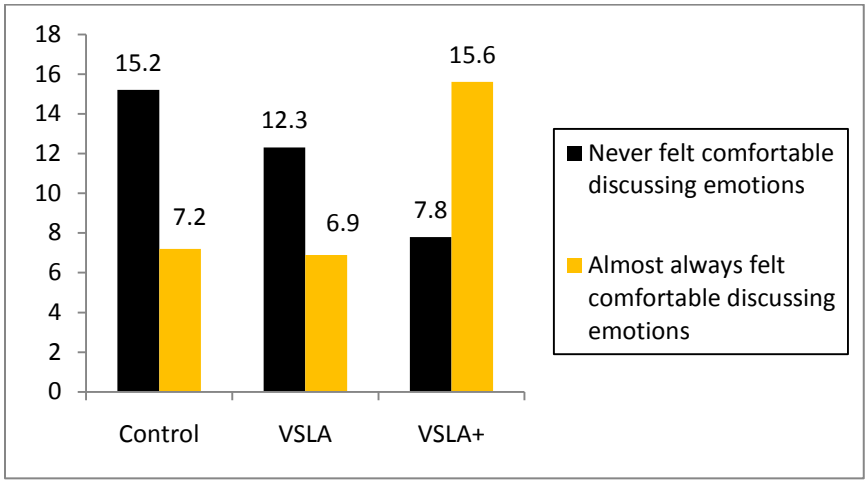
Figure 21: % of Children Who Reported Never (Always) Feeling Comfortable Discussing Their Problems with Their Parents



In line with this finding, the percentage of children who were always comfortable discussing their problems is substantially higher for children whose parents were in the discussion modules (15.6%) than the children whose parents were not (8.2% for VSLA children and 4.8% for control children). While the observed differences between the VSLA and the control group are not statistically significant, the difference between VSLA Plus and VSLA is.

As shown in Figure 22, similar patterns were found in the responses to the second question: A higher proportion of children in VSLA Plus households always felt comfortable discussing their emotions with an adult in their household (15.6%), while a lower proportion of those children never felt comfortable discussing their emotions (7.8%). The corresponding percentages for the other groups are consistently lower for "always comfortable" and higher for "never comfortable. The differences are statistically significant between the VSLA and VSLA Plus group, highlighting the positive effects of the discussion modules.

Figure 22: % of Children Who Reported Never and Always Feeling Comfortable Discussing Their Emotions with Their Parents



Although the patterns presented in Figures 21 and 22 are indicative of a positive impact of the discussion modules, the lack of baseline data means that we are not certain of the causality of the changes. Again, the final survey will shed more light on this.

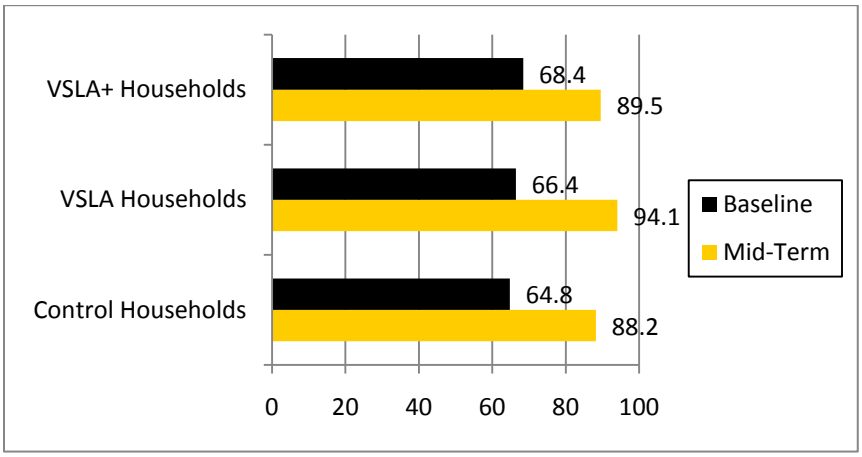
2.4 The Impact of VSLA and VSLA Plus on Children Under Five

Although the New Generation project does not specifically target the care and well-being of children under five years of age, it is nevertheless possible that the VSLA and VSLA Plus interventions have spillovers that positively affect the health and well-being of those young children. If, for instance, the lack of mosquito nets is predominantly due to a lack of means to buy them, then the VSLA intervention may have had a positive impact on mosquito net ownership and use. Or if the lack of health-care use is driven by a lack of money, then the interventions, through the effect on living standards, may have increased the use of formal health care.

Ownership and Use of Mosquito Nets

Only 66.2% of the 1,069 households interviewed at baseline and mid-term owned mosquito nets at the time of the baseline survey. By the mid-term survey, this proportion had increased spectacularly to 90.2%, an increase of 24 percentage points. Figure 23 shows, however, that this increase had very little to do with the VSLA or VSLA Plus interventions: Ownership of mosquito nets rose with 24 percentage points for control households, 28 percentage points for VSLA households and 21 percentage points for VSLA Plus households.

Figure 23: % of Households Owning Mosquito Nets, Baseline and Mid-Term Survey



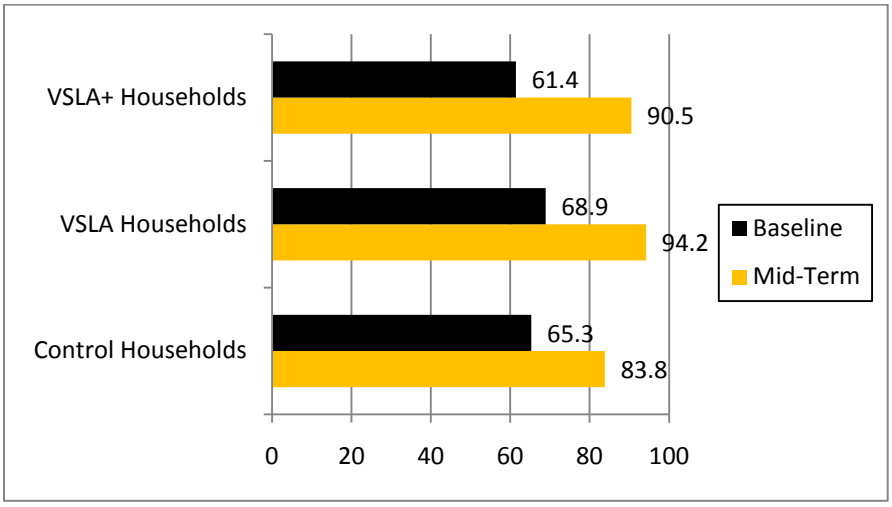
Combining the VSLA and VSLA Plus households, mosquito net ownership increased by 24 %, which is exactly the same as the increase in the control group. This increase is mosquito net ownership, although not brought about by the project, is of course a very positive change.[30]

In line with the increase in mosquito net ownership, the mid-term survey found a marked increase in the proportion of children under the age of five who slept under a mosquito net the night preceding the interview. During the baseline survey, 65.2% of children under five slept under a mosquito net. At mid-term, this had risen to 87.9%.

In contrast to mosquito net ownership, the interventions seem to have had an impact on mosquito net use for under-fives: While the use of mosquito nets for under-fives increased by 18.5 percentage points in the control group, it increased by 25.3 percentage points in the VSLA group and 29.1 percentage points in the VSLA Plus group (see Figure 24). **The increase in the VSLA Plus group is more than 10 percentage points higher than the increase in the control group and is statistically significant.**

[30] Each year a Belgian radio station collects money for a humanitarian purpose through its one-month action "Music for Life". In 2010 the money collected went to procuring mosquito nets for distribution in rural Burundi. This may explain the findings in Figure 21.

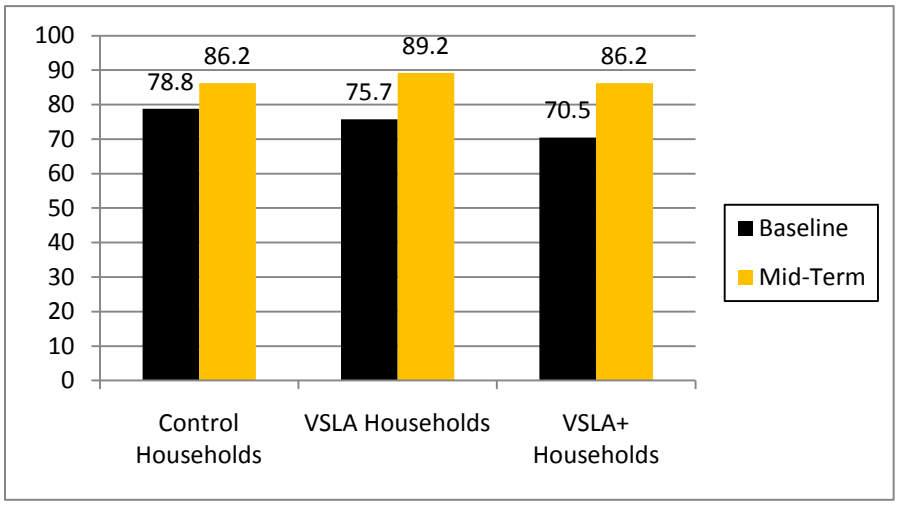
Figure 24: % of Children Under the Age of 5 Who Slept under a Mosquito Net the Night Preceding the Interview, Baseline and Mid-Term



Incidence of Fever and Health Care Consultations

Despite the substantial increase in the proportion of children under five who sleep under a mosquito net, there is only a small drop in the proportion of children who experienced a fever in the two weeks preceding the interview. While at baseline 50.6% of young children fell ill with a fever in the two weeks preceding the baseline interview, this amounted to 46.2% during the mid-term survey. The decrease is similar in magnitude for the control, VSLA and VSLA Plus households. However, the results do illustrate a differential trend in health care seeking behavior for the households in the control, VSLA and VSLA Plus groups (see Figure 25): **While the percentage of respondents that consulted a qualified medical care provider (in case of fever in a young child) increased by 7.5 % in the control group, it increased by 13.5% in the VSLA group and 15.7% in the VSLA Plus group.**

Figure 25: % of Respondents who Consulted a Qualified Medical Care Provider in Case of Fever of their Child, Baseline and Mid-Term



Overall, the increase in health care seeking behavior is twice as large for treatment (VSLA and VSLA Plus combined) than for control households.

3. Cost Effectiveness

In addition to the actual impact of a project, to rate a project's success and sustainability, we need to consider the costs of delivering the project. If the project delivery costs outweigh the benefits to the participants, the project is not cost-effective and should not be continued or scaled-up. Of course, analyzing the cost effectiveness for programs with social returns is a challenging exercise as it is particularly difficult to express certain outcomes in monetary terms.

New Generation is an expensive intervention, largely due to the costs associated with design of a new intervention, start up and the impact evaluation, all of which would not be as costly in future stages or scale up. Dividing the total cost of the project by the number of direct participants (the 1,600 VSLA members), we arrive at a cost-per-participant of 1,490 USD. The cost per household members who benefit from the program is 257 USD.

How does this relate to the project's benefits? Calculating the monetary benefits of the project is challenging and requires making several assumptions. The impact evaluation showed that VSLA participation increased the average household's consumption expenditures by 40.6 USD a month. A simple extrapolation is that over the course of one VSLA cycle, which usually lasts 10 months, the average participating household would have benefited from increased consumption worth 406 USD. Given that half of respondents will do two VSLA cycles during the project (and the other half only one cycle), this means that the expected increase in food and non-food consumption brought about by the intervention will amount to 609 USD per participating household. If

the VSLAs formed during the project keep on functioning for years to come, the expected consumption increase will of course be many times larger.

In addition to economic returns, the impact evaluation revealed a number of important social returns: an increase in child well-being and a decrease in harsh discipline and child mental health problems. It is difficult to express these payoffs in monetary terms. The monetary equivalent of these returns should reflect the importance attached to them. Without a monetary equivalent of these returns, the simple cost-benefit analysis above underestimates the benefit of the program, and as such, it is an incomplete metric for determining the cost-effectiveness of the program.

Based on the positive impact of this program and now that the design, set up and evaluation have been conducted, IRC will analyze the lowest estimated cost to effectively deliver the VSLA Plus program. IRC is not striving to implement the least expensive program but rather the most effective program, which maintains quality of delivery. This will be done in preparation for scaling up the program.

4. Limitations

This evaluation had several limitations which have been mentioned. First, the study uses self report which has the potential for social desirability bias and participants in the family program may be more likely to provide answers that they think the interviewers want to hear whether or not behavior had changed. However, including both caregiver and child report allows us to triangulate the data and one would assume that the children in the VSLA and VSLA plus households, who were not directly involved in the programs, would not be more biased than those who are in the control group. In most cases, caregiver and child report showed similar patterns. Second, as explained in the beginning of the report, the selective attrition and midterm may lead us to overestimate the impact of the program and sensitivity analysis was conducted to understand the potential for this bias. Finally, as also noted, the evaluation had limited power to detect the results of VSLA Plus and the final evaluation will provide further information of its impact.

5. Conclusions/ Programmatic Recommendations

Overall, results from the mid-term survey are extremely encouraging and clearly highlight the positive impact the project is having on vulnerable families in post-conflict Burundi. Participation in Village Savings and Loans Associations both increased assets –which are more indicative of longer-run income—and consumption, a measure of current income. Given that the mid-term survey was conducted not long after the cash-out, when the participating households received a big lump-sum of cash, it is possible that the effect on consumption is transitory, reflecting the sudden availability of money within the household. However, given that an increase in productive assets will in the longer-run increase consumption pay-offs, it can be hypothesized that VSLA participation has put the households onto a higher consumption path (this hypothesis will be tested with data from the final survey, due to be held in the Summer of 2012).

What does this tell us about the mechanism through which VSLAs have beneficial effects? Much like formal microfinance, the results suggest that VSLAs alleviate the credit constraints poor and isolated rural households face, enabling them to overcome entry barriers to engage in more profitable activities (e.g. raising livestock, buying a cassava mill, play on inter-regional arbitrage). This in turn allows them to reap the rewards of the higher-return activities and further accumulate assets. During the second cycle of the research, monitoring will focus on understanding how loans are used. This will shed more light on how and why small loans can have big effects.

In addition, the Healing Families and Communities discussion modules led to decreased harsh physical and verbal discipline in the home, improved communication between children and caregivers, and a decrease in family problems, including violence and intoxication of family members. Further, children in the families who participated in the discussion groups showed reduced distress and less aggression. This finding confirms evidence from other family-based programs that a relatively brief skills-focused program can improve caregivers' behaviors towards their children and improve children's emotional and behavioral health.[31] This evaluation provides some of the first evidence that this can be done in low resource and conflict affected settings. Because the need for parenting programs in low-resource settings is widely recognized, but evidence-based programming is scarce,[32]the evidence generated through this project will be important for the larger field of child protection and global health.

The final evaluation, scheduled for July/August 2012, will provide more robust results with regards to the impact of the Healing Families and Communities discussion series.

Moving forward, it is important to take the following programmatic recommendations into consideration:

⇒ **Scale-up**: Due to the clear, positive, results of the VSLA Plus intervention, within in an extremely challenging socio-economic environment, it is essential to begin planning scale-up at provincial and/or national level. The challenge now is to find new sources of funding and ensure that the intervention is cost-effective without losing quality. Further exploration also needs to be done to understand how best to deliver the family-based intervention in scale up.

⇒ **Document other major events**: In preparation for the final evaluation, it would be useful to develop a timeline of other major events that have occurred during the life of the project in order to understand some of the unexplained results i.e. the increase in mosquito nets, the reduction in health expenditure.

⇒ **Monitoring & Evaluation Design:** As the mid-term survey has clearly shown the positive impacts of the intervention, the project team proposes to cut the quarterly and bi-annual monitoring from the M&E design and focus more on: 1) a larger final evaluation that will include the first-cycle VSLA

[31] Kaminski, J. W., Valle, L. A., Filene, J. H., & Boyle, C. L. A (2008). Meta-analytic Review of Components Associated with Parent Training Program Effectiveness. *Journal of Abnormal Child Psychology*, 36, 4, 567-589.

[32] Engle, P., Black, M., Behrman, J. , Cabral de Mello, M., Gertler, P., Kapiriri, L., Martorell, R., Young, M.E., & the International Child Development Steering Group. (2007). Strategies to avoid the loss of developmental potential in more than 200 million children in the developing world . *The Lancet*, 369, 229-242; Patel, V., Flisher, A. J., Nikapota, A., Malhotra, S. (2008). Promoting child and adolescent mental health in low and middle income countries. *Journal of Child Psychology and Psychiatry*, 49, 313-334.

participants; and 2) on collecting monitoring information on the use of loans to better understand how this increases assets at the household level.

⇒ **Take a Comprehensive Approach to Addressing Violence against Children** – Results from the midterm survey show that violence against children in the home is reducing as a result of this intervention. However, what is clear is that much violence is occurring in school (see also results from the qualitative study). Therefore, to have a larger impact on violence against children, as they experience it, it would be necessary to add a schools component, perhaps using the IRC Healing Classrooms toolkit.

⇒ **Further explore the link between increased economic productivity and child labor:** This evaluation did not provide any conclusive evidence of the impact of VSLA or VSLA Plus on child labor. However, the direction of changes suggests that we should pay careful attention to the final results in this area and explore the potential link through monitoring and qualitative methods.

References

Altman, D. (1985). “The Comparability of Randomized Groups.” *The Statistician 34*: 125-136.

Boccia, D., Hargreaves, J., Lönnroth, K., Jaramillo, E., Weiss, J., Uplekar, M., Porter, J. D. H., et al. (2011). Cash transfer and microfinance interventions for tuberculosis control: review of the impact evidence and policy implications. *The International Journal of Tuberculosis and Lung Disease*, *15*, S37-S49. doi:10.5588/ijtld.10.0438

Bruhn, M. & McKenzie, D. (2008). “In Pursuit of Balance. Randomization in Practice in Development Field Experiments.” *World Bank Policy Research Working Paper 4752*. Washington D.C.: The World Bank.

Lagarde, M., Haines, A., & Palmer, N. (2009). The impact of conditional cash transfers on health outcomes and use of health services in low and middle income countries (Review). *Cochrane Database of Systematic Reviews*, Vol. 7, (4). Retrieved from http://www2.cochrane.org/reviews/en/ab008137.html

Engle, P., Black, M., Behrman, J. , Cabral de Mello, M., Gertler, P., Kapiriri, L., Martorell, R., Young, M.E., & the International Child Development Steering Group. (2007). Strategies to avoid the loss of developmental potential in more than 200 million children in the developing world . *The Lancet*, 369, 229-242.

Fernald, L. C., Gertler, P. J., & Neufeld, L. M. (2009). 10-year effect of Oportunidades, Mexico’s conditional cash transfer programme, on child growth, cognition, language, and behaviour: a longitudinal follow-up study. *The Lancet*, *374*(9706), 1997-2005.

Gardner, J.M, Lozoff, B., Pollitt, E., Wachs, T.D, Walker, S.P., & Wasserman, G.A., *et al.* (2007). Child development: risk factors for adverse outcomes in developing countries. *Lancet*, **369** 9556, 145–157.

Grantham-McGregor, S., Cheung, Y.B., Cueto, S., Glewwe, P., Richter, L., & Strupp, B. *et al.* (2007). Developmental potential in the first 5 years for children in developing countries. *Lancet*, **369** 9555, 60–70.

IMF (2011). “World Economic Outlook Database 2011”. Washington D.C.: The International Monetary Fund [*website consulted on July 12th 2011*]

The International Rescue Committee. (2011)."Urwaruka Rushasha (New Generation): Improving the Well-Being of Vulnerable Girls and Boys in Burundi: The results of participatory activities with children to evaluate child well-being, family relations, child discipline, children's problems and support environment, Mid-Term Evaluation: Design and Findings.” The International Rescue Committee, Burundi.

Kaminski, J. W., Valle, L. A., Filene, J. H., & Boyle, C. L. A (2008). Meta-analytic Review of Components Associated with Parent Training Program Effectiveness. *Journal of Abnormal Child Psychology*, 36, 4, 567-589.

Lansford, J. E., Chang, L., Dodge, K. A., Malone, P. S., Oburu, P., Palmérus, K., Bacchini, D., Pastorelli, C., Bombi, A. S., Zelli, A., Tapanya, S., Chaudhary, N., Deater-Deckard, K., Manke, B., & Quinn, N. (2005). Cultural normativeness as a moderator of the link between physical discipline and children's adjustment: A comparison of China, India, Italy, Kenya, Philippines, and Thailand. *Child Development*, 76, 1234-1246.

Lund, C., De Silva, M., Plagerson, S., Cooper, S., Chisholm, D, Das, J., Knapp, M., & Patel, V. (2011). Poverty and mental disorders: breaking the cycle in low-income and middle-income countries. *Lancet*, 378, 1502-14.

Lustig SL, Kia-Keating M, Knight WG, et al. (2004). Review of child and adolescent refugee mental health. *J Am Acad Child Adolesc Psychology*, 43 (1), 24–36.

McKenzie, D. (2011). "Beyond Baseline and Follow Up. The Case for More T in Experiments." *Policy Research Working Paper 5639*. Washington DC: The World Bank.

Ozer, E. J., Fernald, L. C. H., Manley, J. G., & Gertler, P. J. (2009). Effects of a conditional cash transfer program on children's behavior problems. *Pediatrics.* Vol 123(4), 123, e630-e637.

Patel, V., Flisher, A. J., Nikapota, A., Malhotra, S. (2008). Promoting child and adolescent mental health in low and middle income countries. Journal of Child Psychology and Psychiatry, 49, 313-334**.**

Stewart R, van Rooyen C, Dickson K, Majoro M, de Wet T (2010) What is the impact of microfinance on poor people? A systematic review of evidence from sub-Saharan Africa. Technical report. London: EPPI-Centre, Social Science Research Unit, University of London.

Ravallion, M., Chen, S. and Sangraula, P. (2008) "Dollar a Day Revisited." *World Bank Policy Research Working Paper* 4620. Washington D.C.: The World Bank.

Rohner, R. P., & Khaleque, A. (Eds.). (2005). *Handbook for the study of parental acceptance and rejection* (4th ed.). Storrs, CT: Rohner Research Publications.

World Bank. (2011). *Burundi*. Retrieved from http://data.worldbank.org/country/burundi.

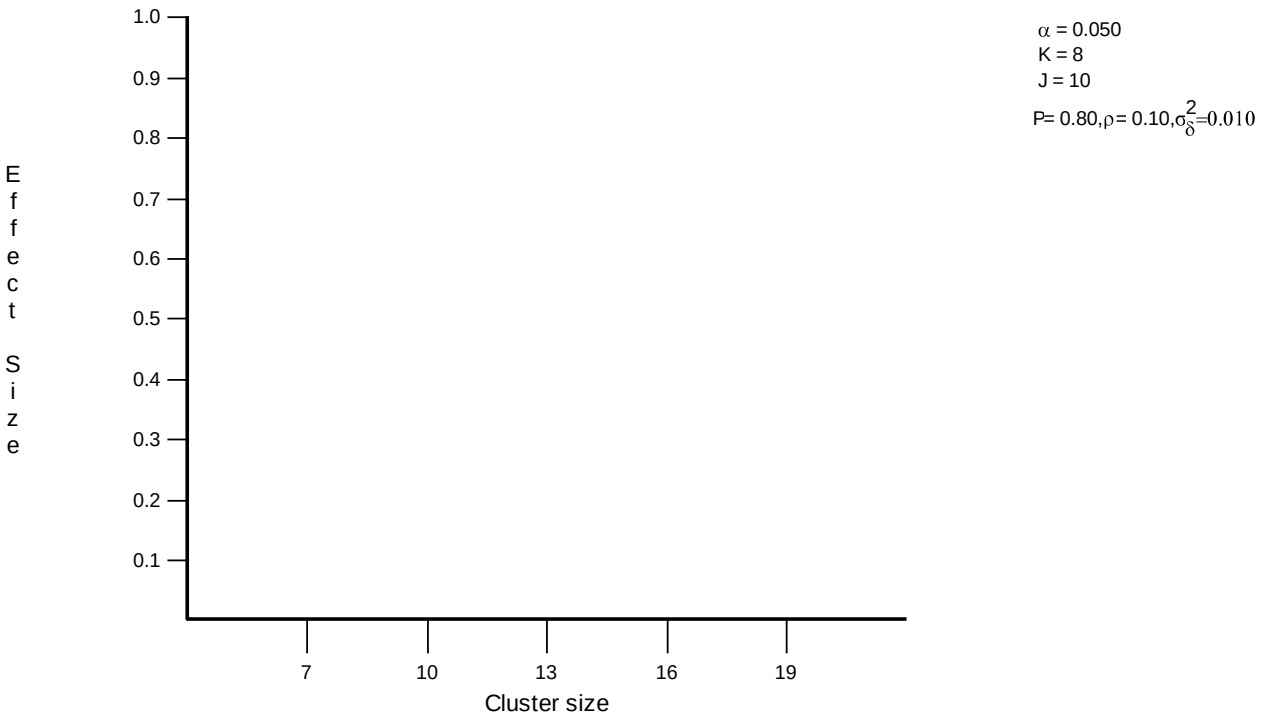
World Health Organization (2009). Preventing violence through the development of safe, stable and nurturing relationships between children and their parents and caregivers. Series of briefings on violence prevention: the evidence. Geneva: WHO Press.

Appendix A: Evaluation Design and Statistical Power

Statistical Power

To evaluate the effectiveness of Village Savings and Loans Associations, we used a three-level multi-site cluster randomized trial with household-level outcomes (see baseline report). "Three-level", because individual participants are nested within VSLA groups, which are nested within zones (zones are administrative divisions of communes). "Multi-site", because we have multiple zones (8 zones). And "cluster randomized" because we have randomized the VSLA groups rather than the individuals in the groups. 77 VSLA-groups participated in the randomization. They were randomly assigned to one out of three groups: VSLA-group (20 slots), VSLA Plus group (20 slots), or control group (37 slots). The first two groups (VSLA and VSLA Plus) are considered the treatment groups for the evaluation of the VSLA intervention. At baseline we had 40 treatment groups (800 households) and 37 control groups (800 households). To have a high powered design, we wanted our study to have a statistical power of 80%. Assuming a within-group correlation of 0.1 and 10 VSLA groups per site (zone), we calculated a minimum detectable effect size of just over 0.30, which can considered a small to medium effect (See Figure A1).

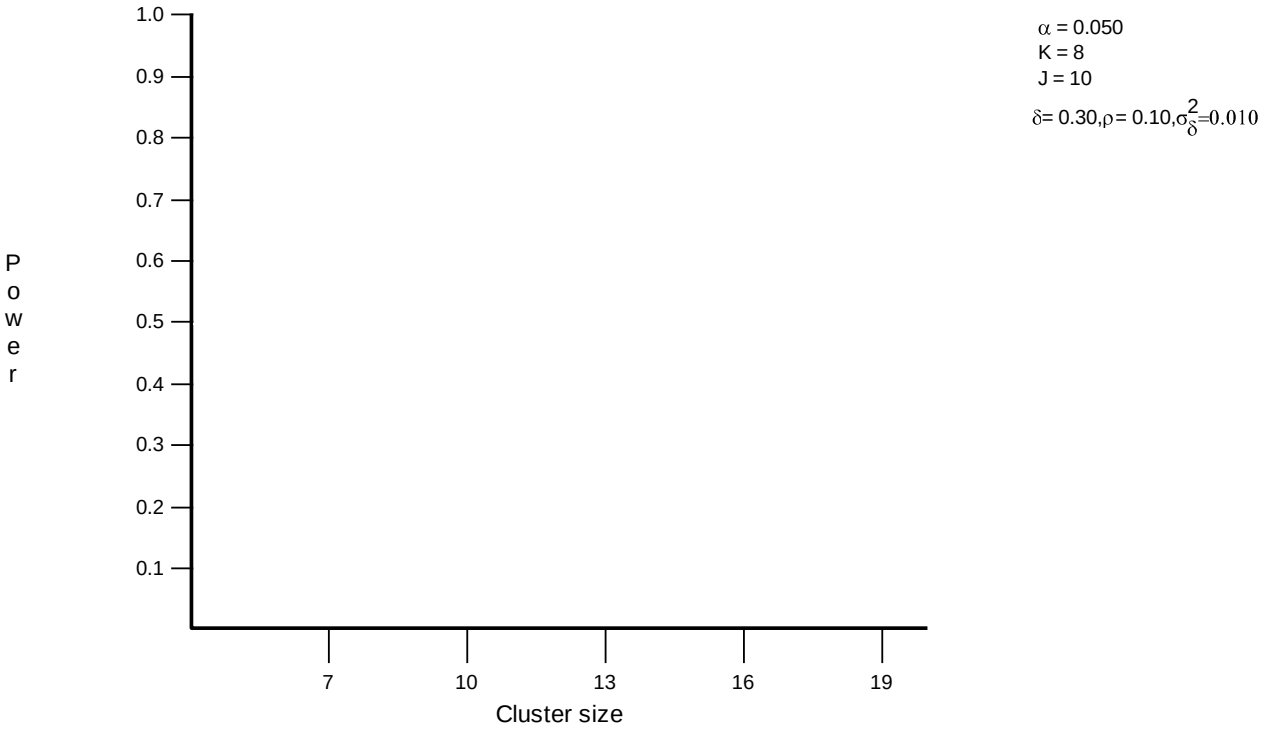
<u>Figure A1:</u> Effect Size vs. Cluster Size (# of people per VSLA) with a Statistical Power of 80%



At baseline, we had 1,600 members in 77 VSLA groups. Hence a cluster size (number of people per VSLA) of approximately 21. At mid-term, there are still 77 original VSLA groups, but with only 1,069 original members. Hence a cluster size of approximately 14. Figure A1 shows that the implication of reduced cluster size (caused by drop-out) is modest: Moving to the left on the horizontal axis increases the minimum detectable effect size with a power of 80% only marginally: From 0.31 with an average cluster size of 21 to 0.33 with an average

cluster size of 14. Holding the minimum detectable effect size constant at 0.3, statistical power amounts to 80% with an average cluster size of 21; with an average cluster size of 14, statistical power is still approximately 75% (see Figure A2). This means that our study has a 75% probability of detecting a small-to-medium effect size if in reality there was one.

Figure A2: Power vs. Cluster Size with a Minimum Detectable Effect Size of 0.3



Appendix B: Drop-Out and its Implications for the Comparability of Treatment and Control Groups

Of the 1,548 households interviewed at baseline, only 1,069 (69%) were still in the project at mid-term. The bulk of the drop-out in the treatment groups happened at the beginning of the cycle, when the VSLA participants were still being trained on the VSLA methodology[33]. At this stage, the drop-outs were replaced by other candidates in the community with the approval of the original group members. Since there is no baseline data on the persons who joined later, the project team decided not to interview them at mid-term. This decision was driven jointly by the limited budget and staff time and the modest consequences of statistical power: Since the original VSLAs were assigned to treatment or control groups, using a cluster-randomized design with person-level outcomes (see baseline report), the effect of individual drop-out on statistical power will be limited (Appendix A discusses the effect of drop out on statistical power).

Due to the specific design of the New Generation research component, drop-out in the control groups was managed differently to drop-out in the treatment groups. As shown in Figure A1, half of the first-cycle control group will become the second-cycle treatment group for the evaluation of the Healing Families and Communities discussion groups. As the research design requires a complete baseline for this second-cycle evaluation, all members of the control group had to be interviewed at mid-term, regardless of whether they were original members interviewed at baseline or new members that replaced drop-outs.

Table B1 shows an overview of the mid-term sample by location and whether the household was an original household (i.e. was already a member of a VSLA during baseline). 1,069 of the 1,369 interviewed households were original households that had already been interviewed at baseline. Three hundred households (169 in Bujumbura Rural and 131 in Makamba) are new households who replaced original households that dropped out of the control groups.

[33] According to the VSLA officers, people in treatment groups dropped out when they realized that the project would not offer direct material benefits to the beneficiaries (the only thing the project offers is training in the VSLA methodology and family-based discussion modules). Drop-out in the control group can mainly be explained by impatience and the disappointment of being assigned to the control group (having to wait one year to receive the intervention).

Table B1: Number of Households Interviewed During the Mid-Term Survey, By Location and Whether Household was Already in Project At Baseline

	Overall	Original Households	New Households
Bujumbura Rural	699	530	169
Kabezi	*345*	*271*	*74*
Mutimbuzi	*354*	*259*	*95*
Makamba	670	539	131
Kayogoro	*371*	*305*	*66*
Nyanza-Lac	*299*	*234*	*65*
Total	**1369**	**1069**	**300**

Table B2 breaks down the 1,069 original households by province and treatment type. Four hundred and ninety one of the 1,069 households were control households, meaning that they did not benefit from any of the two interventions in 2010. Three hundred and three households benefited only from the VSLA intervention, and 275 households received both the VSLA intervention and the family-based discussion modules. Overall, 578 households received the VSLA intervention during the first cycle of the New Generation project.

Table B2: Breakdown of Original Households by Province and Treatment Status

	Control	VSLA	VSLA Plus	**Total**
Bujumbura Rural	227	171	132	**530**
Makamba	264	132	143	**539**
Total	**491**	**303**	**275**	**1069**

Table B3 shows drop out by location. Drop-out was higher in Makamba (32.7%) than in Bujumbura (29%). However, Makamba contains both the commune with the lowest (23.4% in Kayogoro) and the highest (41.9% in Nyanza Lac) level of drop-out.

Table B3: Drop-Out During the First Cycle of Urwaruka Rushasa (New Generation), by Location

	Interviewed at Baseline	Interviewed at Baseline and Mid-Term	Drop-out (%)
Bujumbura			
Rural	747	530	29
Kabezi	*363*	*271*	*25.3*
Mutimbuzi	*384*	*259*	*32.6*
Makamba	801	539	32.7
Kayogoro	*398*	*305*	*23.4*
Nyanza-Lac	*403*	*234*	*41.9*
Total	**1548**	**1069**	**30.9**

The high level of drop-out during the intervention (30.9%) has the potential to jeopardize the impact evaluation. Although assignment into treatment and control groups was done randomly, drop-out –if it is selective-could introduce bias into the design. The key feature and key benefit of randomization is that average treatment and control groups are similar on all observable and unobservable factors, such that any difference between groups after the intervention would necessarily be due to the intervention. Selective attrition would, however, introduce additional differences between the groups, resulting in a biased estimate of the treatment effect.

Table B4 shows the balance between treatment and control groups at baseline (just after randomization and before drop-out). As one would expect from randomization, there is a good balance: Important observables (such as education, literacy, assets, and expenditures) do not differ much between the treatment and the control group, as witnessed by the small values of the standardized differences (last column of Table B1).[34] Only two of the 26 variables presented in Table B1 show a standardized difference of more than *0.1*: The average age of the household's head is somewhat higher in the treatment (43.5 years) than in the control group (41.8 years), and a higher proportion of control households live in a house with iron roof sheeting (0.71) compared to treatment households (0.65). When looking at assets, control households seemed to be somewhat wealthier at baseline than treatment households: A higher proportion of control than treatment households lived in a house with iron roof sheeting and brick walls (superior construction materials in rural Burundi) and used charcoal for cooking (charcoal, in contrast to wood, has to be bought and hence can only be afforded by the "wealthier"). The average score on the asset index is somewhat higher for the control (*0.027*) than for the treatment households (-0.026), confirming the slightly better-off position at baseline of the control households (see Appendix C for information on the construction of the asset index).

[34] Following the argument of Altman (1985) and Bruhn and McKenzie (2008), t or F-statistics for differences between groups are not presented in Table B1: since groups were formed by random assignment any differences between groups are by definition due to chance. Hence, t- or F-statistics do not make sense.

In contrast to assets (which are indicative of longer-run wealth), there is no difference whatsoever in the baseline values of consumption expenditures (which are more indicative of the living standard of the household at the time of the survey). Per capita food expenditures and total expenditures were somewhat higher for the treatment than for the control households, though the differences are negligible (expressed in USD, monthly per capita expenditures in the treatment group are 12 cents higher than those in the control group).

However, the figures in Table B4 present the balance before drop-out (attrition). To examine the nature of the attrition during the intervention, Table B5 examines whether treatment households who dropped out were different than control households who dropped out. If the nature of attrition is similar across groups, then attrition is unlikely to bias the results. However, if attrition followed a different pattern in the treatment than in the control group, then any estimate of the treatment effect would be biased.

Unfortunately, Table B5 shows a number of important differences in attrition between treatment and control groups, notably concerning education and asset holdings. Treatment households that dropped out were on average less educated than control drop-outs (in the sense that they were less often headed by an educated household head), and treatment drop-outs also had lower scores on the asset index than control drop outs. Both differences are statistically significant at conventional levels (respectively at the 5% and 1% level). Treatment households that dropped out were also more likely to be headed by a woman, although the difference is not statistically discernable from zero. The difference in consumption expenditures follows the opposite pattern, with treatment drop-outs having higher expenditures than control drop-outs. The difference is, however, not statistically significant.

Table B4: Balance between Treatment and Control Group at Baseline (before drop-out)

	Treatment	Control	Std. Diff
Household Size	5.79	5.72	-0.03
% of Children Less than Five	0.18	0.19	0.1
% of Children between Five and Nine	0.15	0.15	0.04
% of Children between Nine and 14	0.11	0.1	-0.03
% of Adults (15-59)	0.49	0.48	-0.08
% of Elderly (60 and older)	0.07	0.07	-0.01
Age Head of Household	43.5	41.8	-0.13
% Female Headed	0.27	0.27	0
Household Head Educated (% Yes)	0.45	0.47	0.04
% of Literate Adults (% Yes)	0.58	0.57	-0.02
Owns Radio	0.33	0.32	-0.01
Owns Bicycle	0.22	0.22	-0.01
Owns Watch	0.13	0.11	-0.03
Owns Mobile Phone	0.17	0.18	0.02
Owns Bed	0.56	0.59	0.05
Owns Matress	0.13	0.14	0.04
Lives in House with Brick Walls	0.43	0.48	0.09
Lives in House with Iron Roof Sheeting	0.65	0.71	0.12
Lives in House with Concrete Floor	0.03	0.03	0.005
Number of Rooms in House	2.27	2.18	-0.09
Uses Charcoal for Cooking	0.065	0.073	0.03
Owns Land	0.53	0.52	-0.03
Number of Tropical Livestock Units	0.106	0.092	-0.05
Asset Index	-0.026	0.027	0.05
Food Consumption per Capita	17566	17332	-0.015
Total Expenditures per Capita	18886	18726	-0.01
N	785	763	

The finding that treatment drop-outs were on average less educated and less wealthy (lower score on the asset index) is concerning. It means that the treatment households that remained in the sample will be on average the more educated and better-off, which would provide a positive bias for the estimate of the treatment effect. In Table B6, the FGM (Fitzgerald-Godschalk-Moffitt) method was performed for selective

attrition in panel data to examine in more detail the nature of attrition in this study.[35] Here, the probability of dropping out was regressed on the treatment dummy and the interaction of the treatment dummy with important baseline characteristics. If the coefficients on the interaction terms are statistically indiscernible from zero, attrition followed similar patterns in the treatment and control group (and hence there would not be a problem in terms of biasing the treatment effect). However, if the interaction terms are not equal to zero, attrition could be deemed selective.

Table B5: Testing for Selective Attrition (1)

	Treatment	Control	Mean Difference
Household Size	5.34	5.47	0.12
	[0.142]	[0.158]	[0.212]
% of Children Less than Five	0.21	0.21	0
	[0.012]	[0.012]	[0.017]
% of Children between Five and Nine	0.15	0.16	0.01
	[0.01]	[0.01]	[0.014]
% of Children between Nine and 14	0.09	0.09	0.005
	[0.009]	[0.009]	[0.012]
% of Adults (15-59)	0.49	0.47	-0.02
	[0.013]	[0.015]	[0.021]
% of Elderly (60 and older)	0.07	0.07	-0.003
	[0.011]	[0.011]	[0.016]
Age Head of Household	43.3	44.4	1.1
	[1.2]	[1.2]	[1.7]
% Female Headed	0.313	0.256	-0.057
	[0.028]	[0.030]	[0.041]
Household Head Educated (% Yes)	0.449	0.541	0.092**
	[0.032]	[0.032]	[0.046]
% of Literate Adults (% Yes)	0.564	0.563	-0.001
	[0.025]	[0.025]	[0.035]
Asset Index	-0.257	0.066	0.324***
	[0.065]	[0.063]	[0.090]
Food Consumption per Capita	19355	17383	-1972
	[944.4]	[1103.8]	[1450.0]
Total Expenditures per Capita	20861	19020	-1841
	[1145.5]	[1009.3]	[1524.5]
N	236	242	

[35] See Fitzgerald, Godschalk and Moffitt (2000).

Notes: Figures in the Table show sample means for the 478 households that dropped out between baseline and mid-term. The last column shows the results of a t-test for differences in means. ***: Statistically significant at the 1%-level; **: Statistically significant at the 5%-level.

Table B6: Testing for Selective Attrition (2)

Dependent Variable: 1 if Household Dropped Out	(1)	(2)
Treatment Household (1 if Yes)	0.047	0.027
	[0.582]	[0.590]
Household Size*Treatment Dummy	0.063	0.073
	[0.056]	[0.056]
Age of Household Head*Treatment Dummy	-0.013*	-0.013*
	[0.007]	[0.007]
Female Headed*Treatment Dummy	0.174	0.134
	[0.343]	[0.316]
Household Head Educated*Treatment Dummy	-0.385	-0.414*
	[0.265]	[0.248
Per Capita Expenditures*Treatment Dummy	0.122	0.121
	[0.081]	[0.079]
Asset Index*Treatment Dummy	-0.386**	-0.392**
	[0.177]	[0.176]
Dummies for Zones	No	Yes
Pseudo R-Squared	0.026	0.058
N	1545	1545

Notes: Dependent variable takes on 1 if the household dropped out between baseline and mid-term survey. Results from a logit regression of dropout on the treatment dummy and its interaction with important baseline characteristics. *: Statistically significant at the 10%-level; **: Statistically significant at the 5%-level.

Results from regression (1) in Table B6 show that two interaction effects are statistically different from zero: Age of the household's head and the household's score on the asset index. Older household heads in treatment groups were less likely to drop out than older heads in control groups. If age is correlated with observables and unobservables that influence economic outcomes, this can bias the estimate of the treatment effect. More importantly, however, wealthier households (as measured by a higher score on the asset index) were less likely to drop out in treatment groups than in control groups. This means that the households who stayed in the treatment group were on average wealthier than those who stayed in the control group, creating an upwards bias in the estimation of the treatment effect.

The second analysis in Table B6 repeats the first analysis but adds dummies for the strata (zones). The interaction of the treatment dummy with age of the household's head and household wealth remains significant. In addition, the interaction between the treatment dummy and education of the household's head becomes marginally significant at the 10%-level: Households with an educated head were less likely to drop out of the treatment group, thus confirming the general finding that better-off households in terms of education and wealth were more likely to remain in the treatment group.

What does this mean for the comparability of post-intervention treatment and control groups? The finding that the less wealthy and less educated were more likely to drop out of the treatment group means that the post-intervention treatment group is a privileged sample of the pre-intervention treatment group (in the sense that they are relatively wealthier and more educated). This means that a simple post-intervention comparison between treatment and control group would result in a positive effect on the treated, without this being due to the actual intervention (but rather to the selective drop-out). The fact that there is baseline data alleviates this concern, as differences between groups in relation to pre-intervention differences can be evaluated post-intervention. Nevertheless, the actual econometric analysis will use sensitivity analysis to gauge the importance of this selective attrition.

Appendix C. Construction of the Asset Index

The asset index is constructed along the lines proposed by Filmer and Prittchet (1998). This methodology consists of constructing an index of household economic status based on a series of individual asset indicators. To construct the index, a principal component analysis (PCA) is performed on the individual asset indicators to derive the weight of each indicator on the index. The score of the households on the main factor resulting from the PCA is used as the households' asset score. By construction, the higher the asset score the wealthier the household and vice versa.

Table C1 shows the construction of the baseline asset index. 14 asset indicators are included in the index: Ownership of a radio, watch, mobile phone, bicycle, bed, mattress, land (1 if yes), whether the household lives in a house with brick walls, dung walls, a concrete floor, an aluminum roof (1 if yes), whether the household uses charcoal for cooking (1 if yes), the number of separate rooms in the household's dwelling and the number of tropical livestock units (which consists of 5 different livestock species).

Table C1: Construction of the Asset Index

Asset	Factor Loading	Mean	Std. Dev.	Poorest 40 percent	Middle 40 percent	Richest 20 percent
Owns Radio	0.48	0.33	0.47	0.16	0.29	0.71
Owns Watch	0.25	0.13	0.39	0.06	0.11	0.29
Owns Mobile Phone	0.51	0.17	0.38	0.04	0.14	0.51
Owns Bicycle	0.45	0.22	0.42	0.09	0.17	0.58
Owns Bed	0.51	0.57	0.49	0.30	0.68	0.90
Owns Mattress	0.55	0.14	0.34	0.01	0.08	0.48
Owns Land	0.19	0.53	0.58	0.46	0.54	0.67
House with Brick Walls	0.56	0.45	0.50	0.10	0.62	0.81
House with Mud Walls	-0.55	0.37	0.48	0.70	0.19	0.07
House with Concrete Floor	0.31	0.03	0.17	0.00	0.02	0.12
House with Aluminum Roof	0.47	0.68	0.47	0.41	0.81	0.94
Uses Charcoal for Cooking	0.18	0.07	0.25	0.04	0.06	0.15
Number of Rooms in Dwelling	0.50	2.22	0.94	1.68	2.46	2.80
Tropical Livestock Units	0.37	0.10	0.27	0.04	0.08	0.26
Asset Index		0.00	1.00	-0.95	0.17	1.50

For all but one variable in the asset index, a higher value indicates a better-off position (for instance having a mobile phone –value 1- is probably better than not having one –value 0). Only for "mud walls" does a higher value indicate a worse position (1 if mud walls, 0 otherwise). In the factor loadings (the weights or the importance of the individual assets in composing the index), this translates into positive weights for all

variables except the “mud walls”: Living in a house with mud walls diminishes the score on the asset index, having any of the other assets increases the household’s score on the asset index.

By construction the average score on the asset index is zero. The average asset score amounts to -0.95 for the poorest 40% of households (according to the asset index), 0.17 for the middle 40% of households and 1.5 for the top 20%. Economically, the asset index makes a lot of sense: Ownership of individual assets increase in each higher wealth group, while the only indicator of poverty (mud walls) falls with each higher wealth group.

Appendix D: Mean Treatment Effects under Various Missing Data Assumptions

Table D1: Mean Treatment Effects under Various Missing Data Scenarios

	Lower bounds			Unadjusted treatment effect	Higher bounds		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Per Capita Consumption Expenditures	-11095.6***	-1006.1	2955.5**	3443**	5537.6***	9872.1***	22402.3***
	(1287.9)	(1177.7)	(1165.8)	(1337.7)	(1177.7)	(1250.5)	(1779.1)
Score on Asset Index	-0.491***	-0.032	0.148***	0.222***	0.279***	0.483***	0.926***
	(0.069)	(0.053)	(0.051)	(0.078)	(0.052)	(0.055)	(0.070)

Annex E: Quarterly Monitoring and the Incidence of "Desirable" Answers

In line with Urwaruka Rushasa (New Generation)'s M&E plan, a sample of treatment households was surveyed every quarter since the start of project activities to monitor the impact of the project. About 25% of first-cycle beneficiary households (approximately 200 households) were administered a short questionnaire every quarter by project staff. The goal of this quarterly monitoring is to track progress in three key areas: Food consumption, child discipline and child well-being. Between the baseline survey (January-March 2010) and the mid-term survey (April – May 2011), the household selected for monitoring were interviewed twice: The first quarterly monitoring took place in August 2010 (3 months after start of project activities) and the second monitoring in November-December 2010 (6 months after the start of project activities).

Repeatedly administering the same modules to the same households (four times: baseline and mid-term survey and two monitoring surveys) might induce desirable answers. The participants know very well that New Generation is about promoting child well-being and are likely to know which kinds of answers are "preferred" by the IRC. To see whether the frequency of data collection has impacted the way in which respondents answer the questions, Table 11 shows the incidence of child discipline techniques across monitoring and non-monitoring beneficiary households. Both types of households received the intervention during the first cycle of the project. The only difference is that the monitoring households were included in the quarterly data collection (and hence answered the same questions every time) while the other households were not.

Table 11 shows that at baseline (first part of the table), monitoring and non-monitoring households were relatively similar when it comes to child discipline practices. Notable exceptions are "calling the child dumb, lazy or another name like that", which had a higher incidence among households involved in monitoring (54%) than other households (44%) and "believing a child has to be physically punished in order to be raised properly" (8% for monitoring households vs. 2% for other households). Only the latter difference is statistically significant.

Table E1: Child Discipline For Households Selected for the Quarterly Monitoring and Other Households

	Baseline			Mid-Term		
	Monitoring	Non-Monitoring	*Diff*	Monitoring	Non-Monitoring	*Diff*
Shouted, Yelled or Screamed at the Child	0.65	0.69	*0.04*	0.54	0.57	*0.03*
Shook the Child	0.29	0.25	*0.04*	0.16	0.16	*0*
Spanked, Hit or Slapped the Child with the Bare Hand	0.21	0.21	*0*	0.11	0.15	*0.04*
Hit the Child on the Bottom or Elsewhere on the Body with a Belt, Stick or Other Hard Object	0.06	0.08	*0.02*	0.03	0.07	*0.04*
Call the Child Dumb, Lazy or Another Name like That	0.54	0.44	*-0.1*	0.38	0.33	*-0.05*
Hit or Slapped the Child on the Face, Head or Ears	0.03	0.05	*0.02*	0.03	0.05	*0.02*
Hit or Slapped the Child on the Hand, Arm or Leg	0.17	0.18	*0.01*	0.13	0.15	*0.02*
Beat the Child Up, Hit the Child Over and Over Again	0.02	0.04	*0.02*	0	0.04	*0.04*
Believe that a Child Has to be Physically Punished for a Good Education	0.08	0.02	*-0.06***	0.02	0.02	*0*
Score on the Discipline Scale	2.05	1.97	-0.08	1.4	1.55	0.15

Notes: All households included in the calculations received either VSLA or VSLA Plus. The monitoring households are the households included in the quarterly data collection. **: Statistically significant at the 5% level

If monitoring households were more likely to give desirable answers during the mid-term survey, we would expect to see:

(1) Increasing differences between monitoring and other households in those child discipline practices with a similar incidence at baseline (since monitoring households would over-report their improvement)

(2) Decreasing differences between monitoring and other households in those child discipline practices with a higher incidence amongst monitoring households at baseline (since monitoring households would over-report their improvement)

Comparing the baseline and mid-term differences in Table 11 shows some evidence, albeit weak, of desirable answers. Similar to the situation at baseline, the incidence of most child discipline practices does not differ much between the two groups of households at mid-term. For most discipline practices, however, the improvement between baseline and mid-term is somewhat higher for the monitoring households than for the other households, but not significantly so. The only notable difference is the evolution in the proportion of respondents who believe a child has to be physically disciplined: This dropped from 8% to 2% for monitoring households, and remained steady at 2% for the other households.

The score on the aggregate discipline scale improves considerably for both groups of households. The improvement is, however, larger for the monitoring households (reduction in discipline score of 31%) than for the other households (reduction of 21%). This difference could be due to desirable answers, but could of course also be due to a genuine bigger reduction amongst the monitoring households.

Next to the discipline module, the monitoring questionnaire also included the module on child well-being. To see whether monitoring households were more likely to give desirable answers, Table 12 compares the answers to the child well-being questions for monitoring and non-monitoring households at baseline and mid-term. Overall, the percentage of parents/caretakers who responded "Never" to the listed child well-being items is pretty similar across monitoring and non-monitoring households, both at baseline and at mid-term.

Table E2: % of Respondents that Responded "Never" to the Child Well-Being Item

	Baseline			Mid-Term		
	Monitoring	Non-Monitoring	*Diff*	Monitoring	Non-Monitoring	*Diff*
Eaten when Hungry	56.1	55.9	*-0.2*	23.8	20.2	*-3.6*
Felt like s/he Was Well Dressed	66.7	67.2	*0.5*	42.9	40.9	*-2*
Studied without Problems	20.1	24.5	*4.4*	14.3	15.4	*1.1*
Been in Good Health	25.8	20.1	*-5.7*	6.3	6.9	*0.6*
Had a Good Behavior	15.2	18.1	*2.9*	4.8	6.6	*1.8*
Felt Happy	30.3	25	*-5.3*	4.8	6.9	*2.1*
Received Support when s/he Needed it	40.9	47.1	*6.2*	20.6	23.9	*3.3*
Score on the Discipline Scale	4.7	4.9	0.2	7.5	7.5	0

The aggregate score on the well-being scale was somewhat lower for the monitoring households (4.7) than for the other households (4.9) at baseline. At mid-term, the child well-being score for both types of households had increased to 7.5 (on a total of 14). Although the improvement in child well-being is slightly bigger for monitoring (from 4.7 to 7.5) than for other households (from 4.9 to 7.5), the difference is so small that we

cannot discern it from zero. As such, it does not seem that desirable answers are likely to bias the results we find of the impact of our interventions on child well-being.