Limestone analytics

Youth Ready Impact Evaluation

Endline assessment for El Salvador and Honduras 2018-2019

Prepared for: World Vision Canada
Date: June 21, 2019

Suite 112, 945 Princess Street, Kingston ON, K7L 0E9 Canada
www.limestone-analytics.com

Executive Summary

Project Overview

Youth employment and educational opportunities are limited in El Salvador and Honduras. A large portion of the countries' youth population faces significant barriers to completing secondary and post-secondary schooling, and to engaging in formal employment in the labour force. To improve opportunities for youth, World Vision adapted its Youth Ready training program for implementation in El Salvador and Honduras. For participants, the first six months of the program involved regular program meetings and formal training on employability, entrepreneurship and life skills. The second six months of the program provided participants mentoring as they looked for work, engaged in entrepreneurial activities, continued in school, or pursued other life goals.

This report summarizes the results of a rigorous evaluation of the Youth Ready program on the cohort of participants that started the program in early 2018. The assessment evaluates the impact of the program on participant outcomes a little more than one year after beginning the program and just over six months after completing the program's training. It is the first time that the Youth Ready program model, in any form, has been subjected to a rigorous impact evaluation. We also discuss the challenges of the evaluation process and lessons regarding the suitability of the Youth Ready program model for results-based financing.

Methods

Data for the evaluation was collected in El Salvador and Honduras at the start of the program (baseline) in 2018 and at the end of the program (endline) in 2019. This data included a comprehensive survey of youth, and was supplemented by Youth Ready program data and administrative data from schools in which the program operated, both provided by World Vision. The assessment uses a difference-in-differences (DiD) analysis to compare changes in outcomes of program participants with changes in outcomes of similar youth who could not be admitted into the program at the same time due to limited cohort sizes.

The analysis had to adapt to several challenges that arose during project implementation and data collection, including the inability to accurately measure some activities of interest such as criminal participation, the program being adjusted to provide a greater share of the training through schools, and lack of strict adherence by implementers to the evaluator's random assignment of participants into treatment and control groups.

Key Findings

- The evidence consistently suggests that the Youth Ready program in El Salvador and Honduras succeeded at **increasing labour market opportunities for participants** following secondary school.
- Participants were **more likely to graduate from secondary school** than were non-participants.
- Overall, the program **significantly increased self-perceived leadership and empowerment**, as measured by the CARE Youth Leadership Index (YLI), with this increase being larger for male participants.
- The program **decreased the likelihood of wanting to pursue higher education** (university or higher). This is consistent with participants having greater opportunities in the labour market following secondary school.
- Participation in the program increased risk aversion among female participants.
- There was no evidence the program had an impact on measures of life outlook or time preferences.

Lessons for Program Design and Funding

- **The shift in implementation strategy to focus treatment on youth enrolled in school likely increased the impact of the program.**

  Our evaluation shows that the impact of the program on labour market outcomes is much stronger for in-school participants. This suggests that a program like Youth Ready may have the greatest impact as a preventative measure, reaching at-risk youth before they leave school and enter the formal labour market.

- **There is room to improve service provider capacity to implement programs in ways that facilitate evaluation.**

  We observe room for improvement in terms of improving the understanding of program staff in appreciating how changes on the ground during implementation may influence or undermine efforts to rigorously evaluate the program.

- **Using results-based financing (RBF) to fund Youth Ready is likely feasible, if the program and funding mechanism are well designed.**

  At this point in time, the Youth Ready program would be a strong candidate for a more-traditional RBF agreement, such as a grant-plus-bonus structure. Although the analysis suggests that we can be fairly confident that the program had a positive impact on some outcomes, there remains significant uncertainty about the magnitude of these impacts, and the likelihood of having similar impact in different

contexts. Because of this unresolved uncertainty, the program is unlikely to be a suitable candidate for pursuing an impact bond at this time.

- **The local contexts may pose unique challenges for evaluation design -- this can have implications for RBF that require impact evaluation.**

  The presence of gang activity in the local context of the El Salvador and Honduras program limited the ability to implement a randomized evaluation design and required special provisions for data collection that may be difficult to obtain in future iterations. Understanding how the local context may present barriers to evaluation is important before committing to impact targets under a RBF agreement.

Contents

Abbreviations & Acronyms

attr	Attrition
bl	Baseline
DiD	Difference-in-Differences
el	Endline
n	Sample size
OLS	Ordinary Least Squares
obs	Observations
oos	Out-of-school
RBF	Results-Based Financing
repl	Replacements
se	Standard error
YLI	Youth Leadership Index

Introduction

Project overview

Youth employment and educational opportunities are limited in El Salvador and Honduras. A large portion of the countries' youth populations face significant barriers to graduating from secondary and post-secondary education and participating in the labour force. These barriers are exacerbated by omnipresent criminal gangs which provide alternative pathways to young people, either by choice or by force. It is in this environment that young people in El Salvador and Honduras face limited and often undesirable opportunities to earn a livelihood as they reach the end of their primary and secondary education.

This report evaluates the impact of World Vision's Youth Ready program, which was implemented in El Salvador and Honduras in 2018[1]. The first six months of the program involved regular program meetings and employability, entrepreneurship and life skills training. The second six months involved mentoring and assistance in finding work, enrolling in school, or pursuing other life goals.

The Youth Ready program model was originally developed in Africa to provide training to out-of-school and unemployed youth. The implementation of the program in El Salvador and Honduras is the first time that the program has been implemented in Latin America and the first time it has been adapted to focus primarily on in-school youth. It is also the first time that the Youth Ready program model, in any form, has been subjected to a rigorous impact evaluation.

World Vision Canada engaged Limestone Analytics to design and conduct a randomized evaluation to assess the impact of the Youth Ready program in El Salvador and Honduras. The purpose of this assessment is to quantify the impact of the Youth Ready program on participants' lives and to determine the appropriateness of the Youth Ready program for Results-Based Financing (RBF) opportunities.

Structure of the report

This report summarizes the results of the Youth Ready program evaluation. The report begins with an overview of methods and is followed by a short section comparing summary statistics about the surveyed populations at baseline and endline. The main body of the report goes into detailed description of the findings of the evaluation. The report concludes with a discussion of lessons that were learned throughout the evaluation process, and recommendations based on the findings of the evaluation.

[1] The first offering of the program in El Salvador and Honduras was in 2017.

Methods

This report is informed by field data that was collected in El Salvador and Honduras at the start of the program (baseline) in 2018 and at the end of the program (endline) in 2019. This data was supplemented by Youth Ready program data and administrative data from schools, provided by World Vision. While the initial evaluation design involved use of a randomized control trial process, the method of analysis was changed to a difference-in-difference approach due to high attrition between baseline and endline data collection, as well as deviation from the randomization protocol during program implementation.

Methodology

Evaluation Approach

The evaluation of the intervention was *originally designed as a randomized control trial* (RCT), which would randomly assign out-of-work, unemployed youth to either an initial treatment group who were eligible to participate in the program immediately, or a control group who would be eligible to participate in the program during a second round of implementation one year later. When the program shifted focus to in-school participants, we adapted the randomization process for schools to assign classes to treatment and control groups in a way that balanced the initial observable characteristics across treatment and control groups, while maintaining randomization of out-of-school participants. The assignment process meant that the control group could provide an estimate as to what would have happened with the control group had they not received treatment in the first year of the program (a counterfactual), allowing the analysis to use difference-in-differences comparisons to determine how the Youth Ready program impacts outcomes of participants a year after enrolling in the program.

Unfortunately, there were some implementation issues at baseline that prevented adherence to the evaluator assignment of youth into treatment and control groups. For example, high attrition or low take up rates made it necessary to replace dropouts with control group participants or new participants. There is also some indication that the implementation team did not strictly adhere to the evaluator's assignment of individuals and classes to treatment and control groups. These issues are likely to introduce concerns regarding attrition and non-random assignment.

Attrition

Attrition is generally not random. Those who choose to leave the program (and become part of the control group) are likely to be fundamentally different than those who stay in the program. It is unclear whether the high attrition biased the estimates upward or downward. Participants may for example choose to leave because they are highly skilled and they find a job, in which case attrition would bias the estimates in a way that underestimates the positive effect of the intervention. Conversely, they may choose to leave because they are not sufficiently motivated to pursue the program, in which case attrition would bias the estimates in a way that overestimates the positive effect of the intervention.

Attrition is often a concern in the evaluation of youth training programs, especially those that target youth who are out of school by endline data collection. Because of this, any concern about attrition in the context of Youth Ready is shared by most assessments of training programs in international development.

Non-Random Assignment

The deviation from randomized assignment leads to treatment and control groups that are not necessarily similar to one another. The treated participants may differ from control participants in important ways that drive differences in outcomes that are not caused by the intervention itself. When these differences are driven by heterogeneity in unobservable characteristics, it is not possible to control for these factors. Their effect will then bias the overall estimate of the effect of the intervention as it will all get lumped together.

Instruments & Data Collection

Data was collected using a survey instrument which consisted of questions about participants' living situation, education, work experience, life outlook, leadership characteristics, and risk and time preferences. The survey was initially written in English by Limestone staff based off of the surveys that were used at baseline. The survey was then translated into Spanish by a professional translator. World Vision field staff reviewed the translated version, and both English and Spanish versions were coded into the Kobo ToolBox data collection platform by Limestone prior to enumerator training.

Training sessions were held with enumerators that focused on reviewing and updating the survey, discussing data collection protocols, and practicing collecting data with Kobo ToolBox.

Data collection took place over the course of approximately three weeks at the end of January and beginning of February 2019. Youth from the treatment and control groups were invited to a central location in each district to complete the survey. Some districts experienced lower turnout than expected. Various factors contributed to low turnout; one district coordinator reported that invited participants did not want to attend the data collection event because it was taking place at the participants' old school, and the participants were embarrassed to see their old teachers and peers. In other cases, participants may have been unwilling to travel from fear of needing to pass through dangerous areas. Other participants simply no longer lived in the area.

In order to overcome low turnout, some districts asked enumerators to reach out to participants over the phone to complete the survey. Enumerators from these districts called each absent participant up to three times.

Challenges Regarding Assignment Status

The assignment status (treatment or control) of each respondent is necessary to estimate the effect of the intervention. Two methods of determining assignment status were originally planned. The first was the randomized baseline assignment, and the second was through a question included in the endline survey verifying the respondent's participation. Unfortunately neither approach remained feasible. The first was made irrelevant by the

non-adherence to the randomized assignment, and the second was contaminated by the early kickoff of the second cohort, many of which were part of the control group in the first cohort.

Three other ways of identifying the assignment status were made possible by the supplementary data provided by the field team: the data from World Vision's "Encuesta de conocimiento" survey, the administrative school data, and the Youth Ready participant list. The first two were provided initially by World Vision when assembling the data necessary to perform the analysis. The Youth Ready participant list was provided at Limestone's request to shed some light on the discrepancies identified in the assignment constructed using the "Encuesta de conocimiento" data and the administrative school data.

The administrative school data provided the assignment for only a subset of in-school participants, and so it could not provide assignment for all observations in the data set. The "Encuesta de conocimiento", on the other hand, was administered to all participants, and so provided assignment for all treated participants. If there was no missing data in the "Encuesta de conocimiento" data set, all observations with no "Encuesta de conocimiento" data could be assumed to be control. In that sense, it provided assignment for all observations.

It was, however, found that 269 observations had no "Encuesta de conocimiento" data, and so were assumed to be in the control group, but were listed as part of the treatment group in the administrative school data. This could be explained by errors in the administrative school data or missing data in the "Encuesta de conocimiento" data set. This discrepancy is what led Limestone to request a comprehensive Youth Ready participant list for both the 2017 and 2018 cohorts. Including the list of participants in the 2017 cohort is important because the field team indicated that many of the replacements interviewed at endline were treated participants from the 2017 cohort.

Note on Interpretation of Results

Reference to the statistical significance of the estimates is used throughout the presentation of the results as it helps pinpoint the most insightful results. Estimates of the effect of the intervention on outcomes come with standard errors. These are measures of the precision of the estimates, and are based on the sample size and the variance of the underlying variables.. They can be used to compute t- or z-statistics and then p-values. The p-value indicates the probability of observing the computed estimate if in reality, there is no effect. The smaller the p-value, the less evidence there is that there is no effect. As is common practice in any statistical analysis, arbitrary cutoffs of 10%, 5% and 1% are used to determine the strength of the evidence as summarized in the following table:

	No significance	P-value is > 10% and the t or z statistic is < 1.65
*	Weak significance	P-value is > 5% and < 10% and the t or z statistic is > 1.65 and < 1.96
**	Moderate significance	P-value is > 1% and < 5% and the t or z statistic is > 1.96 and < 2.58
***	Strong significance	P-value is < 1% and the t or z statistic > 2.58

It should be kept in mind, however, what the p-values and significance categories do and do not tell us. They inform us, based on the analysis, how certain we should be that the results of the analysis "are not zero" and "don't go in the other direction". This insight can be useful, but it is not always the most meaningful indicator of how confident we are in any given result. For example, it does not tell us how confident we should be in the magnitude of any given result; however confidence intervals can help inform this.

Note also that the lack of statistical significance does not mean that there is no effect or even that the estimate is incorrect. The true effect could be similar to the estimated effect, it could be zero, or it could even go in the opposite direction. A larger sample would be required to estimate the effect with greater accuracy.

Summary Statistics

Table 1 summarizes key summary statistics of the beneficiary demographics at baseline and endline by assignment (control versus treatment). It includes participants at baseline who left the sample by endline, i.e. attrition, and replacements at endline who joined the program late and missed the baseline survey. Endline also includes replacements who were interviewed at endline specifically for the purpose of mitigating the high level of attrition. These include control group replacements who were chosen as they are expected to be similar ex-ante to the control group, as well as treatment group replacements who participated in the 2017 cohort (the intervention being evaluated is the 2018 cohort).

Table 1: Beneficiary Demographics at Baseline and Endline by Assignment

Characteristic		Baseline		Endline	
		Control	Treatment	Control	Treatment
Sex	Female	53%	51%	56%	50%
	Male	47%	49%	44%	50%
School Attendance	In-School	87%	85%	64%	55%
	Out-of-School	13%	15%	36%	44%
Average Age (Years)		17.0	17.1	17.9	18.4
Average # of Children		0.12	0.10	0.12	0.08
Married		3.6%	3.9%	4.1%	3.7%
Has at Least One Child		6.5%	6.8%	6.8%	5.6%
Family Member in Jail in the Past Year	Sibling	1.2%	1.5%	1.8%	1.5%
	Parent	2.3%	1.7%	1.3%	1.4%
Report a Disability		1.6%	3.1%	2.4%	2.4%

Overall, the summary statistics are quite similar at baseline and endline as well as across the control and treatment groups. The characteristics that were expected to change, including school attendance and average age, did so in the expected way. School attendance dropped from around 87% and 85% down to 64% and 55% in the control and treatment groups, respectively. Most of the drop was driven by those in their final year of high school who graduated by endline. The larger drop in the treatment group can likely be explained by the treated replacements from the 2017 cohort who make up over 20% of the endline treatment group.

Average age in years increased by 0.9 years, falling just short of the expected increase of one year. This is likely explained by the higher attrition rate among older participants which was identified in the Endline Data Quality Report. Conversely, average age increased by 1.3 years in the treatment group and is half a year higher than in the control group. Like the pattern observed in school attendance, it is likely driven by the treated replacements from the 2017 cohort who, as expected, are approximately a year older on average than those in the endline treatment group from the 2018 cohort as shown in **Table 2**.

Table 2. Average Age at Endline by Cohort in the Treatment Group

Treatment Group	Average Age at Endline
2017 Cohort	19.4
2018 Cohort	18.3

Findings

In this section we present the impact of the program on the following domains: youth leadership, life outlook, academic achievements, employment outcomes, risk and time preferences, and school attendance. We close the section by discussing how impact varied by female and male participants.

Youth Leadership Index

The 21 questions of the Youth Leadership Index were included in both the baseline and endline surveys to estimate the impact of the intervention on the soft skills that it covers. The score was rescaled into percentage to perform the analysis. To make the results more insightful, the questions were divided into three categories: agency, confidence and self-expression, and leadership. The questions, the methodology for calculating the YLI score, and how the questions are allocated into the three categories is described in **Appendix A**.

Overall, the aggregate YLI score increased from an average of 72.8% at baseline to 74.3% at endline across the whole sample. Similar patterns were observed when the YLI was broken down into the three categories listed above:

- Agency: average score from 80.3% at baseline to 81.9% at endline
- Confidence and self-expression: average score from 71.1% at baseline to 73.1% at endline
- Leadership: average score from 68.3% at baseline to 69.2% at endline

Figure 1 breaks it down by assignment. We find that the overall score increased from 73.0% and 72.1% at baseline to 73.1% and 76.4% at endline in the control and treatment groups, respectively. Similar patterns are observed again when the YLI is broken down into the three categories listed above, with the exception of leadership where the score decreases in the control group:

- **Agency:** average score went from 80.4% and 80.1% at baseline to 81.1% and 83.3% at endline in the control and treatment groups, respectively.
- **Confidence and self-expression:** average score from 71.1% and 71.0% at baseline to 71.8% and 75.4% at endline in the control and treatment groups, respectively.
- **Leadership:** average score from 68.8% at baseline to 66.5% at endline in the control group but from 67.7% at baseline to 71.8% at endline in the treatment group.

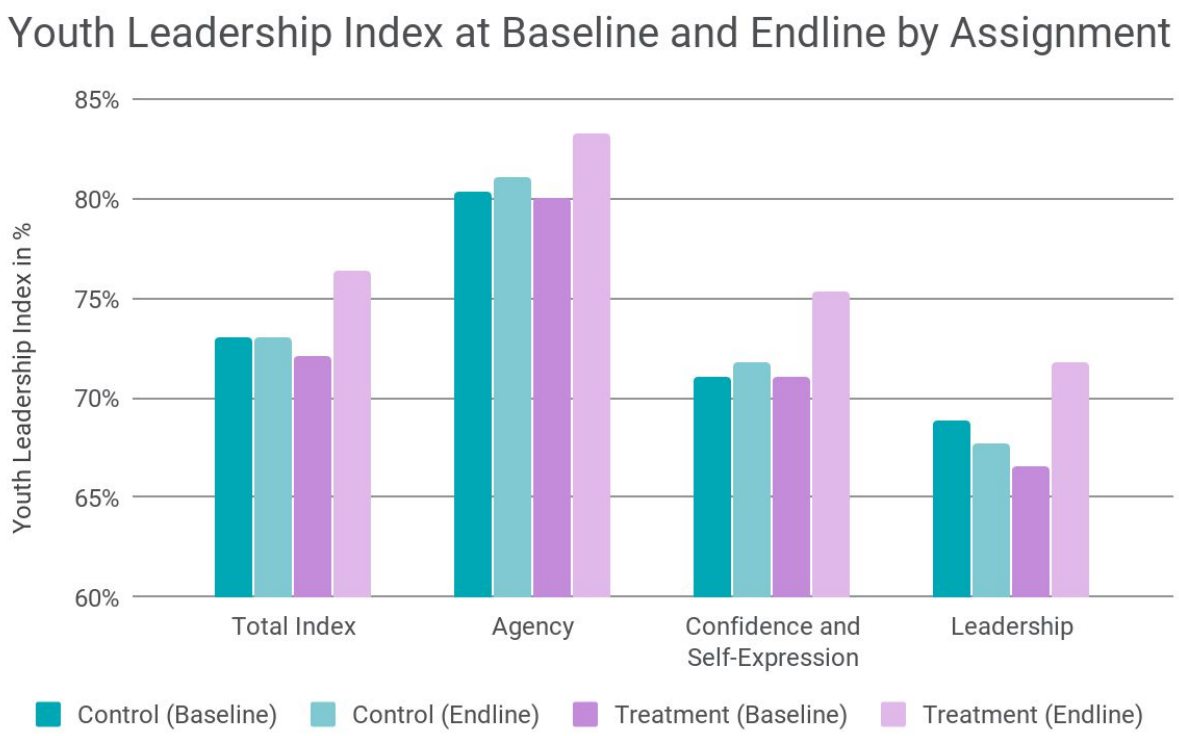

Figure 1. Youth Leadership Index at Baseline and Endline by Assignment

The differences in the changes between baseline and endline in the control and treatment groups are the DiD estimates, which are measures of the effect of the intervention. These are summarized in **Table 3**. The estimate for the total index is approximately 4.2 percentage point and is statistically significant at 1%. The p-value rounds to zero at three decimal points, which indicates that it would be very unlikely to observe such a high estimate if the true value was zero, i.e. if the intervention had no effect. This is strong evidence supporting a positive effect of the intervention on the skills evaluated by the YLI. In spite of the strength of the evidence, the magnitude of the effect is relatively small. The magnitude of 4.2 percentage point is approximately 35% of the baseline standard deviation of 11.9%.

Table 3. Estimate of the Effect of the Intervention on the Youth Leadership Index

Youth Leadership Index in %	Estimate (SE)	p-value	Confidence Interval	n	Methodology
Total Index	4.2022*** (0.9988)	0.000	(2.244, 6.161)	bl: 1,638 el: 1,386	• DiD OLS • Whole sample ○ w/attr. & repl. ○ in-school & oos
Agency	2.5456** (1.1611)	0.028	(0.269, 4.822)	bl: 1,650 el: 1,398	• DiD OLS • Whole sample ○ w/attr. & repl. ○ in-school & oos
Confidence and Self-Expression	3.5796*** (1.1725)	0.002	(1.281, 5.878)	bl: 1,648 el: 1,393	• DiD OLS • Whole sample ○ w/attr. & repl. ○ in-school & oos
Leadership	6.3859*** (1.1760)	0.000	(4.080, 8.692)	bl: 1,648 el: 1,399	• DiD OLS • Whole sample ○ w/attr. & repl. ○ in-school & oos

* p<0.10, ** p<0.05, *** p<0.01, Confidence Interval: 95%

We obtain similar positive estimates when breaking down the index into agency, confidence and self-expression, and leadership. The intervention is estimated to increase the agency score by 2.5 percentage points; the confidence and self-expression score by 3.6 percentage points; and the leadership score by a much larger 6.4 percentage points. All three estimates are statistically significant.

Comparing the magnitude of the three estimates to the baseline standard deviation for each of the three scores that ranges between 14.2% and 14.4%, we find that the estimates are approximately 18%, 25% and 44% of one standard deviation for agency, confidence and self-expression, and leadership. All three estimates provide strong evidence that the intervention has a positive effect on these skills.

Life Outlook

A number of questions on life outlook are included in both the baseline and endline surveys. These include questions on general life satisfaction today, and their expected life in five years. The questions on their expectations of their life in five years include their future general life satisfaction, the number of children that they expect to have, and whether they expect to live in the same neighbourhood, be in-school or have completed higher education, and have a job.

Table 4 summarizes the estimates of the effect of the intervention on life outlook. Their general life satisfaction today and in five years is measured using the life outlook ladder. The question on the life outlook ladder today is:

> "*Imagine a ladder with steps numbered from zero at the bottom to 10 at the top. The top part of the ladder represents the best possible life for you and the bottom is the worst possible life. On which step of the ladder do you feel you are at right now?*"

Table 4. Estimate of the Effect of the Intervention on Life Outlook

Outcome	Estimate (SE)	p-value	Confidence Interval	n	Methodology
Life Outlook Ladder (Today)	0.1509 (0.1451)	0.299	(-0.134, 0.435)	bl: 1,621 el: 1,384	• DiD OLS • Whole sample ○ w/attr. & repl. ○ in-school & oos
Life Outlook Ladder (in 5 years)	-0.1960* (0.1030)	0.057	(-0.398, 0.006)	bl: 1,553 el: 1,359	• DiD OLS • Whole sample ○ w/attr. & repl. ○ in-school & oos
Living in the Same Neighbourhood (in 5 years)	0.0380 (0.0403)	0.346	(-0.041, 0.117)	bl: 1,579 el: 1,358	• DiD Probit • Whole sample ○ w/attr. & repl. ○ in-school & oos
In-School or have Completed Higher Education (in 5 years)	-0.0123 (0.0328)	0.707	(-0.077, 0.052)	bl: 1,605 el: 1,359	• DiD Probit • Whole sample ○ w/attr. & repl. ○ in-school & oos
Have a Job (in 5 years)	0.0191 (0.0223)	0.391	(-0.025, 0.063)	bl: 1,617 el: 1,371	• DiD Probit • Whole sample ○ w/attr. & repl. ○ in-school & oos
Have Children (in 5 years)	0.0021 (0.0393)	0.957	(-0.075, 0.079)	bl: 1,623 el: 1,385	• DiD Probit • Whole sample ○ w/attr. & repl. ○ in-school & oos

* p<0.10, ** p<0.05, *** p<0.01, Confidence Interval: 95%

The estimate must be considered relative to the scale of the life outlook ladder which goes from zero to ten. Overall, the average response on the life outlook ladder today declined from 6.9 at baseline to 6.7 at endline. The difference-in-differences estimate of 0.15 therefore means that the average response declined by 0.15 less in the treatment group than in the control group. With a p-value of 0.3, the estimate is not statistically significant.

The average response on the life outlook ladder in 5 years increased from 9.0 to 9.1 between baseline and endline. Breaking this down by assignment, however, shows that the average response increased from 9.0 to 9.1 in the control group and decreased from 9.1 to 9.0 in the treatment group. The changes in opposite directions combine to form the difference-in-differences estimate of -0.20, which is weakly statistically significant at the 10% level. The p-value is just above 5%, and so although the evidence is relatively weak, it does point to a small decline in the treatment group in their expected future life satisfaction. Although significant, this change is small in magnitude.

The other estimates of future life outlook including their expectation of whether they will be living in the same neighbourhood, be in-school or have completed higher education, have a job, and have children, are all small and are not statistically significant. The evidence does not point to any effect of the intervention on those measures of life outlook.

Academic Achievements

The baseline survey was not originally designed to be administered to a large in-school population and so it contains relatively few measures of academic achievement. The most insightful one is the desired maximum amount of schooling. In this question, respondents are asked to choose the highest level of education that they want to achieve amongst a number of choices such as primary, secondary, trade school, university and graduate school.

To address this shortcoming, academic grades were collected in 2017 and 2018 for a subset of the in-school population. The effect of the intervention on academic grades can be a good proxy for long-run outcomes that cannot easily be measured in only one year, such as employment outcomes, especially when a large portion of the sample remains in-school at endline.

Table 5. Estimate of the Effect of the Intervention on Academic Achievements

Outcome	Estimate (SE)	p-value	Confidence Interval	n	Methodology
Plans to Complete Higher Education	-0.0918*** (0.0355)	0.010	(-0.161, -0.022)	bl: 1,636 el: 1,332	• DiD Probit • Whole sample ○ w/attr. & repl. ○ in-school and oos
Academic Grades in %	1.1850 (1.4893)	0.426	(-1.737, 4.107)	bl: 553 el: 556	• DiD OLS • Admin data sample ○ w/attr. & repl. ○ in-school only

* p<0.10, ** p<0.05, *** p<0.01, Confidence Interval: 95%

Table 5 summarizes the DiD analysis for two measures of academic achievements: plans to complete higher education (defined as university or above), and academic grades. The

analysis based on the desired maximum amount of schooling question focuses exclusively on the respondents' desire to complete higher education rather than looking at each option one-by-one because of translation issues at baseline that led to a set of choices for that question that were not representative of the respondents' educational options. The Spanish version of the education achievement choices for that question were corrected at endline during the enumeration training in consultation with the enumerators.

This unfortunately means that baseline and endline results cannot be seamlessly compared. Baseline and endline consistency is needed for the difference-in-differences methodology that is used to alleviate the assignment issues described above. This is achieved by combining education achievement choices into two categories: higher education and others. This allocation of the baseline and endline choices into the two categories is described in **Appendix B**.

The probability of wanting to achieve higher education increases from 68% and 72% at baseline to 80% and 75% at endline in the control and treatment groups, respectively. That is a 12 and 3 percentage points increase in the control and treatment groups, respectively. The difference is the difference-in-differences estimate of -0.092. The estimate is statistically significant at 1%. The low p-value of 0.01 shows that this is strong evidence that the intervention has a negative effect on the desire to achieve higher education. This is consistent with the positive effects on employment outcomes identified above.

The second estimate is the effect of the intervention on academic grades. The grades collected were out of 10, 100 or 1,000 depending on the school and geographic area. They were all rescaled into percentages for comparability. The DiD estimate measures the difference in the treatment and control group's change in academic grades between 2017 and 2018. Overall, academic grades declined in both groups on average. The DiD estimate means that grades declined by 1.19 percentage point less in the treatment group than the control group. The estimate is very small at just over 12% of the baseline standard deviation of academic grades of 9.6%.

However, the estimate is not statistically significant and we cannot be confident that the smaller decline was because of treatment or occurred only by chance. The p-value of 0.43 means that there is an over 40% probability of obtaining an estimate of 1.19 percentage point or larger (positive or negative) when there is in fact no effect, i.e. the true value is zero. Given the realized variance in academic grades within both the treatment and control groups, a larger sample size would be needed for an estimate of such a small magnitude to provide evidence of a positive effect of the intervention on academic grades.

Employment Outcomes

The baseline and endline surveys included a number of questions on employment to estimate the effect of the intervention on labour market outcomes. **Table 6** summarizes the estimates for whether those out-of-school at endline work for money, the earnings per

week of those who work for money, and whether those who work for money are self-employed or entrepreneurs. All the estimates compare the outcomes at endline without controlling for baseline differences like in difference-in-differences (DiD) analysis. The DiD approach is not used in this case because for the large in-school portion of the sample, it does not make sense to control for differences in post-graduation employment outcomes using baseline differences in part-time student jobs.

Table 6. Estimates of the Effect of the Intervention on Employment Outcomes

Outcome	Estimate (SE)	p-value	Confidence Interval	n	Methodology
Work for Money	0.0486 (0.0434)	0.262	(-0.036, 0.134)	el: 540	• Probit • Subsample ○ w/o attr. & w/repl. ○ oos at endline
Earnings per Week (USD)[2]	-7.8147 (-1.08)	0.283	(-22.123, 6.494)	el: 214	• OLS • Subsample ○ w/o attr. & w/repl. ○ oos & work for money at endline
Self-Employed/ Entrepreneur	-0.0333 (-0.66)	0.509	(-0.132, 0.065)	el: 233	• Probit • Subsample ○ w/o attr. & w/repl. ○ oos & work for money at endline

* p<0.10, ** p<0.05, *** p<0.01, Confidence Interval: 95%

The first estimate measures the probability of working for money among the endline out-of-school sample. Of this subsample, approximately 54% and 58% of those in the control and treatment group respectively work for money by endline. The difference of 5% is the estimated effect, but it is not statistically significant. The p-value of 0.26 indicates that there is a 26% probability of obtaining an estimate of 0.05 or larger (positive or negative) when the intervention has no effect.

The other two estimates compare the earnings per week and the proportion of entrepreneurs in the subsample of respondents who are out-of school and work for money at endline. The p-values are 0.28 and 0.51 respectively and so neither of the estimates are statistically significant. The sign of the estimates points to slightly lower earnings per week and a slightly lower share of entrepreneurs in the treatment group relative to the control group. The high p-values, however, makes it difficult to make this statement with any level of confidence.

[2] The earnings per week are reported in USD in El Salvador and Lempiras in Honduras. The values reported in Lempiras are converted to USD using the average Lempiras to USD exchange rate for 2018.

Risk and Time Preferences

Risk Preferences

Questions on risk preferences are included in both the baseline and endline surveys to estimate the effect of the intervention on risk preferences. The questions used to measure risk preferences are financial in nature. The respondents are offered the choice between a guaranteed payoff and a coin flip that yields $0 (L. 0) with 50% probability, and a higher payoff with 50% probability. There are three questions used to assess risk preferences that differ only in the non-zero risky payoff that is 2x, 3x and 4x the riskless payoff. The questions are shown in **Appendix C**.

The non-zero risky payoffs of 2x, 3x, and 4x the riskless payoff, respectively, yield an expected value of 100%, 150% and 200% of the risk-free payoff option. **Table 7** illustrates how respondents are assigned to the following categories of risk-aversion: risk-neutral, slightly risk-averse, risk-averse, very risk-averse, and inconsistent. Choosing to flip the coin on all questions is considered risk-neutral as the respondent would choose a risky payoff over a risk-free payoff even though they have the same expected value. The slightly risk-averse and risk-averse categories require risky payoffs that yield an expected value between 100% and 150% and between 150% and 200% of the riskless payoff respectively as compensation for the risk of the coin flip.

The very risk-averse category is the extreme case where even a coin flip that provides a risky payoff that is four times as large as the riskless payoff, yielding an expected value that is twice as large as the guaranteed option, is insufficient to compensate the agent for the risk of the coin flip. Finally, the inconsistent category covers any other combination of “Flip the coin” and “Take the money for sure”. These other combinations are considered inconsistent because they are cases where one would choose a risky payoff at a low expected values and a risk-free payoff at high expected values.

Table 7: Risk Aversion Categories by Answers to the Risk-Aversion Survey Questions

Risk-Aversion Categories	Risk-Free Payoff (x) vs Risky Payoff (0, 2x)	Risk-Free Payoff (x) vs Risky Payoff (0, 3x)	Risk-Free Payoff (x) vs Risky Payoff (0, 4x)
Risk-Neutral	Flip the coin	Flip the coin	Flip the coin
Slightly Risk-Averse	Take the money for sure	Flip the coin	Flip the coin
Risk-Averse	Take the money for sure	Take the money for sure	Flip the coin
Very Risk-Averse	Take the money for sure	Take the money for sure	Take the money for sure
Inconsistent	Any other combination of “Flip the coin” and “Take the money for sure”		

Figure 2 illustrates the probabilities of choosing each of the risk-aversion categories at baseline and endline broken down by assignment. Those with inconsistent risk preferences are excluded from the analysis. Overall, risk-aversion increases between baseline and endline in both the control and treatment group. The probability of being risk-neutral decreases significantly while the probability of being very risk-averse jumps up. The high probability of being very risk-averse at both baseline and endline — around 50-60% of the sample — suggests that the highest risky payoff of four times the riskless payoff is too low to obtain the most insightful results. This should be kept in mind when designing future surveys on similar populations. The risky payoffs should be chosen such that the probability of being in each risk-aversion category does not differ as much as it does here. The large share of respondents who fall in the very risk-averse category in this case likely includes a subgroup that require a moderately higher compensation for the risk of the coin flip, and another subgroup that is unwilling to undertake any risk for anything but an extremely high compensation.

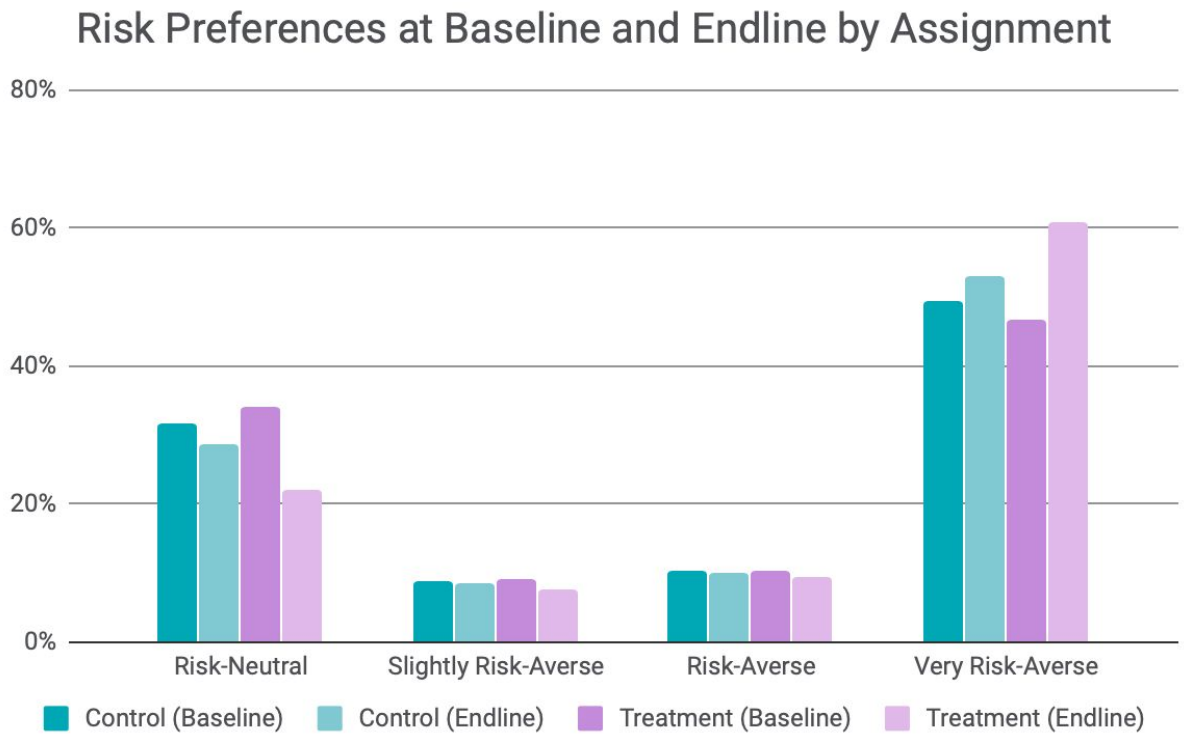

Figure 2. Predicted Probabilities (Ordered Probit) at Each Level of Risk Aversion at Baseline and Endline by Assignment

Table 8 describes the sample size and methodology used to estimate the effect of the intervention on risk-aversion. A difference-in-differences ordered probit is used to measure the difference between the control and treatment group in the change in the probability of choosing any specific level of risk-aversion.

Table 8. Sample Size and Methodology for Estimates of the Effect of the Intervention on Risk Preferences

n	Methodology

bl: 1,227
el: 1,129
	• Difference-in-Differences Ordered Probit
	• Whole sample
	  ○ with attrition and replacements
	  ○ in-school and out-of-school

Table 9 shows the difference-in-differences estimates of the effect of the intervention on risk-aversion. All estimates are found to be statistically significant at 5% and one at 1%; the p-values all hover around 0.01. The probability of risk-neutral, slightly risk-averse, and risk-averse declines respectively by 8.8%, 1.1%, and 0.6% more in the treatment group than the control group. Conversely, the probability of being very risk-averse increases by 10.5% more in the treatment group than the control group. All four estimates point towards a consistent pattern of increasing risk-aversion. The high level of statistical significance across all four estimates provides strong evidence for the intervention having a positive effect on risk-aversion.

Table 9. Difference-in-Differences (DiD) Estimates of the Effect of the Intervention on Risk Preferences with an Ordered Probit Model

Risk Preferences	DiD Change in Probability	Standard Error	p-value	Confidence Interval (95%)
Risk-Neutral	-0.0882**	0.0363	0.015	(-0.159, -0.017)
Slightly Risk-Averse	-0.0107***	0.0040	0.008	(-0.019, -0.003)
Risk-Averse	-0.0064**	0.0025	0.011	(-0.011, -0.002)
Very Risk-Averse	0.1054**	0.0420	0.012	(0.023, 0.188)

* p<0.10, ** p<0.05, *** p<0.01

Time Preferences

Time preferences are a measure of an individual's willingness to substitute consumption intertemporally. Patient individuals require a lower compensation, i.e. rate of return, to delay consumption while impatient individuals are only willing to delay consumption in exchange for a high compensation. Time preferences are assessed using questions offering the choice between a guaranteed payoff today and a higher guaranteed payoff a year from now. The three questions used to assess time preferences differ only in their future payoffs, which are 1.2x, 1.5x and 2x the payoff offered today. This is equivalent to a 20%, 50% and 100% rate of return, respectively. The questions are shown in **Appendix D**.

Table 10 illustrates how respondents are assigned to the following categories of time preferences: very patient, patient, impatient, very impatient, and inconsistent. Choosing to wait a year on all three questions is very patient as the respondent is willing to accept any compensation with as high as of a 20% rate of return to delay consumption by one year. Patient and impatient are the intermediate cases where a rate of return of 50% and 100%, respectively, are required to induce delaying consumption by a year, and the immediate payoff is chosen at lower rates of return.

The very impatient category is the extreme case where even a rate of return of 100% is insufficient to compensate the individual for delaying consumption by a year and the immediate payoff is chosen in all cases. Finally, the inconsistent category covers any other combination of "Wait a year" and "Take the money today". These other combinations are considered inconsistent because they are cases where one would choose to wait a year at low rates of return and to take the money today at high rates of return.

Table 10. Time Preference Categories by Answers to the Time Preference Survey Questions

Time Preference Categories	Payoff Today (k) vs Payoff in 1 Year (1.2k)	Payoff Today (k) vs Payoff in 1 Year (1.5k)	Payoff Today (k) vs Payoff in 1 Year (2k)
Very Patient	Wait a year	Wait a year	Wait a year
Patient	Take the money today	Wait a year	Wait a year
Impatient	Take the money today	Take the money today	Wait a year
Very Impatient	Take the money today	Take the money today	Take the money today
Inconsistent	Any other combination of "Wait a year" and "Take the money today"		

Figure 3 illustrates the probabilities of choosing each of the time preference categories at baseline and endline broken down by assignment. Those with inconsistent time preferences are excluded from the analysis. Overall, time preferences are very stable over time and do not vary much between baseline and endline. There is a slight increase in the probability of being very impatient and slight decreases in the probabilities of being very patient, patient and impatient.

The most significant result, however, is the very high probability — above 70% — of being very impatient. Like in the risk preference questions, this may be driven by survey design. Perhaps a rate of return of 100% is simply too low to shift consumption to the following year. This is certainly possible and likely explains a portion of those who are very impatient. It should be kept in mind when designing surveys intended for similar populations in the future. Another possibility is that the respondents do not have full confidence in the credibility of the future payoff. The survey clearly states that the payoff is guaranteed, but the respondents may still ascribe a higher probability of default to the future payoff than to

today's payoff. This may be especially true in the marginalized communities that this intervention is targeting.

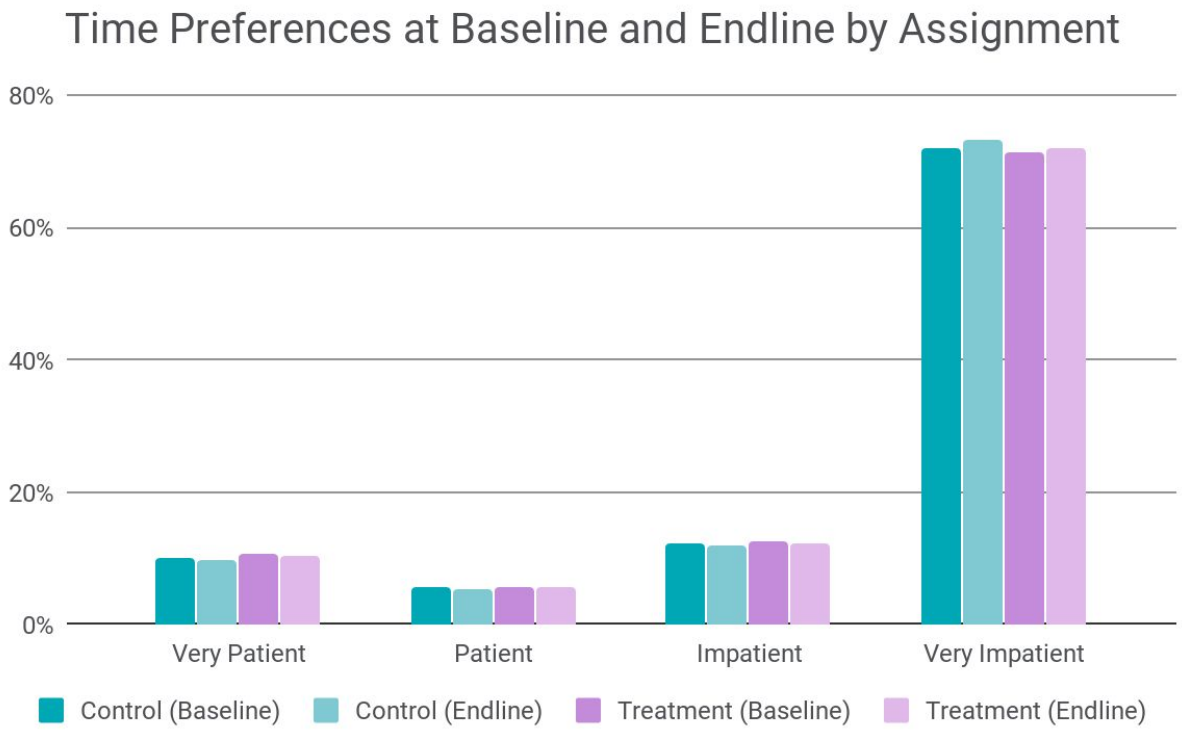

Figure 3. Predicted Probabilities (Ordered Probit) at Each Level of Time Preference at Baseline and Endline by Assignment

Table 11 describes the sample size and methodology used to estimate the effect of the intervention on time preferences. A DiD ordered probit is used to measure the difference between the control and treatment group in the change in the probability of choosing any specific level of time preference.

Table 11. Sample Size and Methodology for Estimates of the Effect of the Intervention on Time Preferences

n	Methodology

bl: 1,294
el: 1,157
	• Difference-in-Differences Ordered Probit
	• Whole sample
	  ○ with attrition and replacements
	  ○ in-school and out-of-school

Table 12 shows the DiD estimates of the effect of the intervention on time preferences. The estimates are all quite small and none are statistically significant. The overall pattern is a slightly lower increase in impatience and decrease in patience in the treatment group

relative to the control group. The very high p-values of around 0.90, however, shows that the estimates are essentially insignificant. The true value is very likely to be zero. There is hence no evidence that the intervention has any effect on time preferences.

Table 12. Difference-in-Differences (DiD) Estimates of the Effect of the Intervention on Time Preferences with an Ordered Probit Model

Time Preferences	DiD Change in Probability	Standard Error	p-value	Confidence Interval (95%)
Very Patient	0.0021	0.0202	0.916	(-0.037, 0.042)
Patient	0.0008	0.0071	0.909	(-0.013, 0.015)
Impatient	0.0013	0.0108	0.905	(-0.020, 0.022)
Very Impatient	-0.0042	0.0380	0.912	(-0.079, 0.070)

* p<0.10, ** p<0.05, *** p<0.01

Baseline School Attendance: In-School Participants

This section focuses on participants who were in-school at baseline. This lets us analyse the effect of the intervention on outcomes that are only relevant for in-school participants, such as high school graduation. It also allows the disaggregation of results on outcomes that have different pathways for in-school and out-of-school participants. The pathway to employment, for example, is likely to be very different for students graduating from high school than for those who began the Youth Ready program out of school.

Graduation and Employment Outcomes

Table 13 shows the estimates of the effect of the intervention on high school graduation and the employment outcomes for the participants who were in-school at baseline. The sample is also restricted to those out of school at endline as the outcomes are only relevant for those out of school.

This limits the analysis to the core sample — excluding attrition and replacements — as respondents must have participated in both the baseline and endline surveys for us to determine that they were in-school at baseline and out-of-school at endline. This restriction is not only necessary, but actually desirable. It intentionally excludes the replacements from the 2017 cohort who may have very different employment outcomes an extra year post-graduation.

Table 13.: Estimate of the Effect of the Intervention on Secondary School Graduation and Employment Outcomes for Participants In School at Baseline and Out of School at Endline

Outcome	Estimate (SE)	p-value	Confidence Interval	n	Methodology
Graduated from Secondary School	0.0938** (0.0445)	0.035	(0.007, 0.181)	bl & el: 192	• Probit • Core sample ○ w/o attr. & repl. ○ in-school at baseline & oos at endline
Work for Money	0.1512** (0.0710)	0.033	(0.012, 0.290)	bl & el: 191	• Probit • Core sample ○ w/o attr. & repl. ○ in-school at baseline & oos at endline
Earnings per Week[3]	-4.5684 (7.0813)	0.522	(-18.760, 9.623)	bl & el: 57	• OLS • Core sample ○ w/o attr. & repl. ○ in-school at baseline, oos at endline & work for money
Self-Employed/ Entrepreneur	-0.0963 (0.1109)	0.385	(-0.314, 0.121)	bl & el: 62	• Probit • Core sample ○ w/o attr. & repl. ○ in-school at baseline, oos at endline & work for money

* p<0.10, ** p<0.05, *** p<0.01, Confidence Interval: 95%

All estimates are computed with either a standard Probit or OLS regression, i.e. without the use of difference-in-differences, as it does not make sense to control for differences in post-graduation employment outcomes using baseline differences in part-time student jobs.

Of the in-school participants who are out-of-school by endline, 84% and 94% have graduated high school in the control and treatment group respectively. The difference of 9.4% is statistically significant at 5%. This is a large gap between the control and treatment group and provides reasonable evidence that the intervention may increase high school graduation rate among in-school participants.

[3] The earnings per week are reported in USD in El Salvador and Lempiras in Honduras. The values reported in Lempiras are converted to USD using the average Lempiras to USD exchange rate for 2018.

A similar effect is found on the probability of working for money. Of the in-school participants who are out-of-school by endline, 36% and 52% work for money in the control and treatment group, respectively. The difference of 15% is statistically significant at 5%. This result is consistent with the higher graduation rate observed in the treatment group. The large magnitude of the effect and low p-value of 0.03 is good evidence that the intervention may increase employment post-graduation among in-school participants.

Finally, the earnings per week and the self-employment/entrepreneurship status of the in-school participants who are out-of-school and work for money at endline are compared across the control and treatment group. Restricting the sample further to those who work for money shrinks the sample size to only about 60 observations. The difference in earnings per week is close to zero and the estimate is not statistically significant.

The share of respondents who are self-employed or entrepreneurs is 30% and 20% in the control and treatment group respectively. The difference of 10% is not statistically significant. In spite of the lack of statistical significance, the estimate is relatively large and is consistent with the intervention's effect on risk-aversion found in section 4.5. A larger sample size would be needed to more accurately determine the effect of the intervention on self-employment and entrepreneurship.

Self-Reported Academic Performance

The baseline and endline surveys ask respondents to rate their academic performance relative to their peers using one of five qualifiers ranging from "Poor (among the worst in the class)" to "Excellent (among the best in the class)". In addition to in-school participants, the question is also asked to out-of-school participants regarding their self-assessed performance in their last year of study.

Figure 4 illustrates the probabilities of reporting each of the five levels of academic performance at baseline and endline by assignment. The distribution has a large right tail with most respondents reporting average performance, but much larger shares of the respondents reporting above average and excellent than below average and poor.

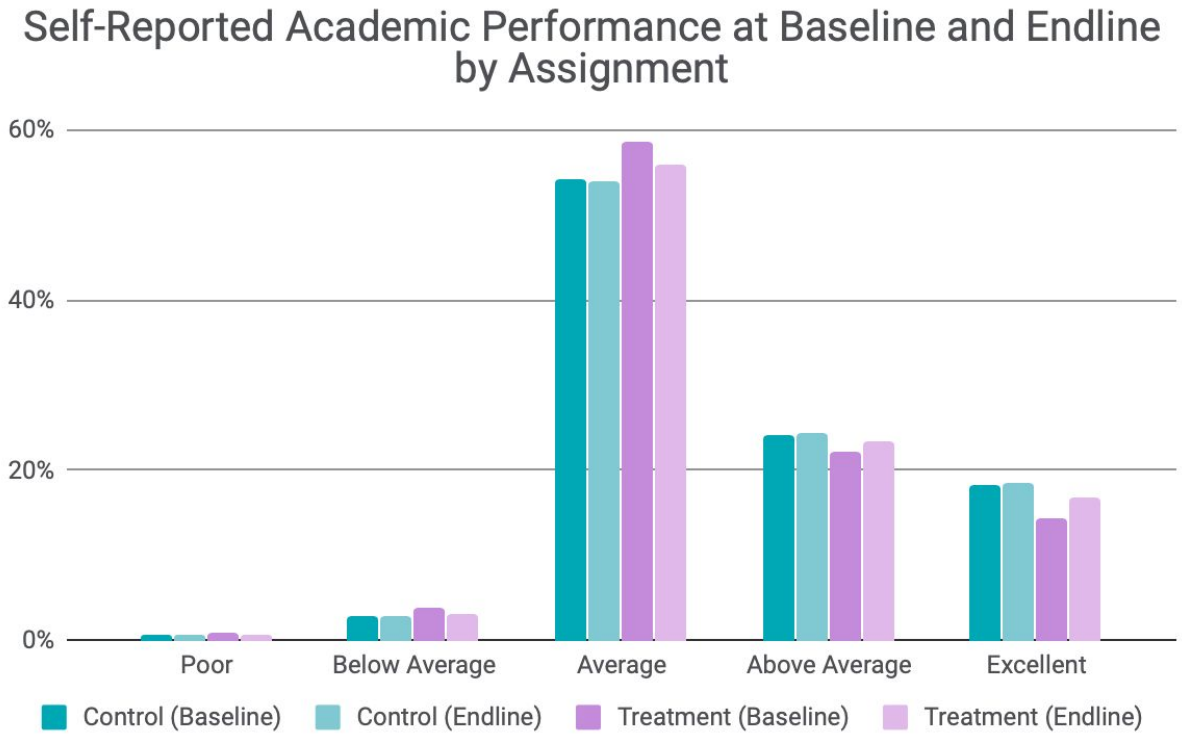

Figure 4. Predicted Probabilities (Ordered Probit) at Each Level of Self-Reported Academic Performance at Baseline and Endline by Assignment

Table 14 describes the methodology for estimating the effect of the intervention on self-reported academic performance. A difference-in-differences ordered probit is used to control for baseline differences in self-reported academic performance in the treatment and control groups. The analysis is restricted to the core sample as participation in the baseline survey is necessary to check school attendance at baseline. Only those in-school at baseline are included in the analysis as the only effect the intervention can have on self-reported academic performance of those who were out-of-school throughout the whole year of the intervention is on their perception of that performance and not their actual performance.

Table 14. Sample Size and Methodology for Estimates of the Effect of the Intervention on Self-Reported Academic Performance

n	Methodology

bl & el: 527	• Difference-in-Differences Ordered Probit
	• Core sample
	  ○ w/o attrition and replacements
	  ○ in-school at baseline

Table 15 shows the estimates of the effect of the intervention on self-reported academic performance. The signs of the estimates show that the probability of reporting above average and excellent increased by more in the treatment than the control group. Conversely, the probability of reporting poor, below average, and average decreased by more in the treatment than the control group. The difference is, however, very small and is not statistically significant. The large p-values of around 0.5 shows that the estimates are essentially insignificant. This result is consistent with the findings on academic grades iwhere no evidence of an effect on academic performance was found.

Table 15. Difference-in-Differences (DiD) Estimates of the Effect of the Intervention on Self-Reported Academic Performance with an Ordered Probit Model

Self-Reported Academic Performance	DiD Change in Probability	Standard Error	p-value	Confidence Interval (95%)
Poor	-0.0020	0.0029	0.482	(-0.008, 0.004)
Below Average	-0.0065	0.0094	0.486	(-0.025, 0.012)
Average	-0.0263	0.0423	0.534	(-0.109, 0.057)
Above Average	0.0132	0.0193	0.493	(-0.025, 0.051)
Excellent	0.0217	0.0351	0.538	(-0.047, 0.091)

* p<0.10, ** p<0.05, *** p<0.01

Baseline School Attendance: Out-of-School Participants

Further Education and Employment Outcomes

Table 16 shows the estimates of the effect of the intervention on further education and the employment outcomes from section 4.4 for the participants who were out-of-school at baseline. For employment outcomes, the sample is also restricted to those out of school at endline as they are only relevant for those out of school.

The analysis of endline outcomes with a restriction on baseline school attendance limits the analysis to the core sample — excluding attrition and replacements — as respondents must have participated in both the baseline and endline surveys for us to observe the endline outcome and to determine that they were in-school at baseline. This necessary exclusion of attrition and replacements combined with the already low share of participants who are out-of school at baseline and the high attrition rate among out-of-school participants results in a very small sample size. All estimates are produced with less than 100 observations and those that impose further restrictions go as low as 13 observations.

The further education outcomes are evaluated using a standard probit, i.e. without the use of difference-in-differences, as there are no baseline benchmarks against which the endline outcomes can be compared. The first further education outcome is whether out-of-school participants have returned to school by endline. 18% and 25% of the out-of-school participants in the control and treatment groups, respectively, have gone back to school by endline. The difference of 7% is not statistically significant. With a high p-value of 0.45, the possibility that the intervention has no impact on the out-of-school participants' decision to go back to school cannot be ruled out.

Restricting the sample further to those out-of-school who do not plan to go back to school, we estimate the effect of the intervention on the participants' decision to go obtain further education by looking at the share of those who have gone back to school or who plan to go back to school at endline. This restricts the sample size to only 15 observations. 44% and 83% of the out-of-school participants who do not plan to go back to school at baseline in the control and treatment groups respectively have either gone back to school or plan to go back to school at endline.

The difference of 39% is statistically significant at 10%. Given the low level of significance, this is relatively weak evidence that the intervention has an effect on this outcome. The weakness of the result is driven by the small sample size. Achieving statistical significance at all with only 15 observations speaks volumes about the magnitude of the observed difference. While the result is certainly exciting, however, it is important to be cautious as while the statistical analysis takes the small sample size into account — which is why such a large estimate is only weakly statistically significant — small sample analysis imposes additional assumptions on the normality of the underlying distribution, which may not hold. Therefore, in spite of the large observed difference, further evidence with a larger sample size would need to be collected before being able to express any level of confidence in this result.

Table 16. Estimate of the Effect of the Intervention on Further Education and Employment Outcomes

Outcome	Estimate (SE)	p-value	Confidence Interval	n	Methodology
Gone Back to School at Endline	0.0682 (0.0898)	0.448	(-0.108, 0.244)	bl & el: 84	• Probit • Core sample ○ w/o attr. & repl. ○ oos at baseline
Gone Back or Plans to Go Back to School at Endline	0.3889* (0.2249)	0.084	(-0.052, 0.830)	bl & el: 15	• Probit • Core sample ○ w/o attr. & repl. ○ oos at baseline & "Does not plan to go back to school"
Work for Money	-0.0833 (0.1682)	0.620	(-0.413, 0.246)	bl & el: 66	• DiD Probit • Core sample ○ w/o attr. & repl. ○ oos at baseline & endline
Earnings per Week[4]	1.3953 (21.4661)	0.948	(-41.956, 44.747)	bl: 13 el: 32	• DiD OLS • Core sample ○ w/o attr. & repl. ○ oos at baseline & endline, work for money
Self-Employed/ Entrepreneur	0.1813 (0.2234)	0.417	(-0.257, 0.619)	bl: 23 el: 36	• DiD Probit • Core sample ○ w/o attr. & repl. ○ oos at baseline & endline, work for money

* p<0.10, ** p<0.05, *** p<0.01, Confidence Interval: 95%

The last three estimates are the employment outcomes for this group. The first is whether the participants work for money. After restricting the sample to those who are out-of-school at baseline and endline, only 66 observations remain. The probability of working for money increased from 39% and 53% at baseline to 64% and 70% at endline in the control and treatment groups, respectively.

The increase in the probability of working for money was 8% lower in the treatment than the control group leading to a negative estimate of -0.08. The result is not surprising given the large effect observed on further education. Given the high variances and the small

[4] The earnings per week are reported in USD in El Salvador and Lempiras in Honduras. The values reported in Lempiras are converted to USD using the average Lempiras to USD exchange rate for 2018.

sample size, the estimate is not statistically significant. With a large p-value of 0.62, there is no evidence of an effect on working for money for out-of-school participants.

The earnings per week and self-employment/entrepreneurship are evaluated only for out-of-school participants who work for money at endline. This shrinks the sample size further to around 10-20 observations at baseline and around 30 observations at endline. The estimate for earnings per week is essentially zero and is not statistically significant. The estimate for self-employment/entrepreneurship is relatively large at 18%, but is not statistically significant. The large estimate can really be explained by high ex-ante heterogeneity as the probability of being self-employed or an entrepreneur declined from 43% and 22% at baseline to 15% and 13% at endline in the control and treatment groups respectively. There is hence no evidence that the intervention has an effect on earnings per week or self-employment and entrepreneurship.

Gender Disaggregation: Female vs Male

This section presents the most insightful results produced in previous sections in a model that disaggregates the effects of the intervention by gender. Disaggregated estimates are included for the Youth Leadership Index, whether the respondent plans to complete higher education, risk preferences, and graduation and employment outcomes for in-school participants who are out-of-school at endline.

Youth Leadership Index

The Youth Leadership Index (YLI) measures a number of soft skills including agency, confidence and self-expression, and leadership. How the Youth Ready intervention affects those skills may differ across men and women, and so disaggregating the effect into two estimates allows some insights into how the mechanism of impact may differ. **Figure 5** and **6** show how the aggregate index, and its agency, confidence and self-expression, and leadership components evolved from baseline to endline among female and male participants, respectively.

A similar increase in the aggregate YLI score in the treatment group is observed for both genders, but the change in the control group differs significantly across male and females. The female control group sees an increase in the YLI score that partially matches the increase in the treatment group. Conversely, the YLI score in the male control group declines slightly between baseline and endline.

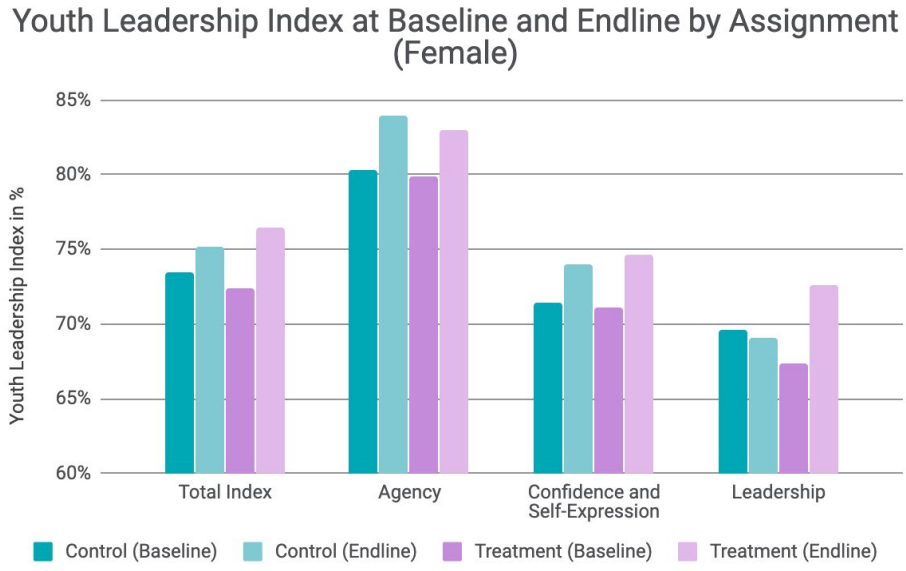

Figure 5. Youth Leadership Index at Baseline and Endline by Assignment (Female)

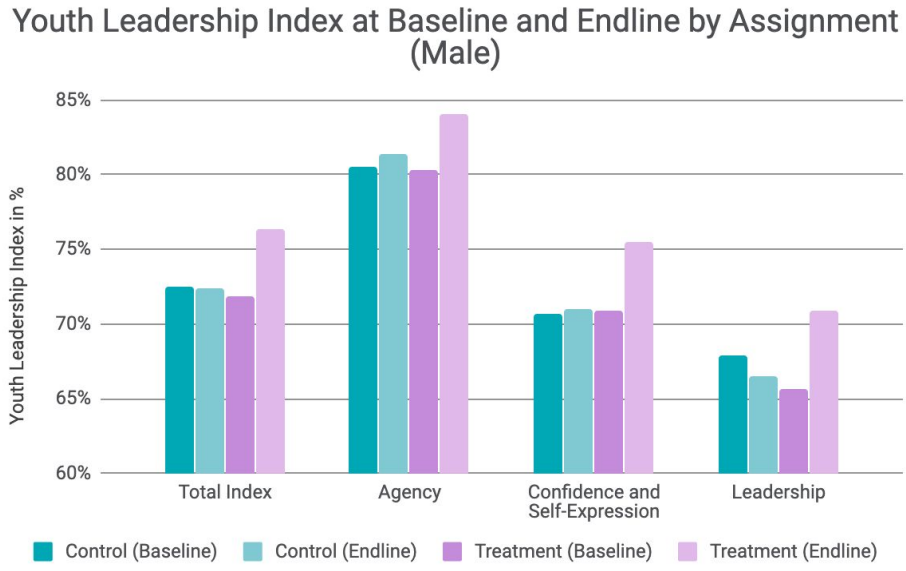

Figure 6. Youth Leadership Index at Baseline and Endline by Assignment (Male)

When breaking up the index into the agency, confidence and self-expression, and leadership components, we find the difference observed across male and female is mostly driven by agency, and confidence and self-expression. For those two components, the increase in their scores in the control group is nearly as high or even higher than the increase in the treatment group for females. For males, however, the pattern is similar

across the aggregate index and all its components with increases in the score in the treatment group while the score in the control group stays relatively flat.

Table 17 summarizes the gender-disaggregated difference-in-differences estimates of the effect of the intervention on the YLI scores. The estimate for males is approximately 4.6 percentage points and is statistically significant at 1%. For females, however, the magnitude is much lower with an estimated effect of 2.3 percentage points. It is not statistically significant, although it is close given the p-value of 0.12.

Breaking down the score into its subcomponents, we find that the estimate for the aggregate score for females is almost entirely driven by leadership, which has an estimate of 5.8 percentage points and is statistically significant at 1%. The estimates for agency, and confidence and self-expression for females are near zero and are not statistically significant.

As was shown above in Figure 5, the low estimates for those two components of the index for females is driven by increases in the score in the control group that are similar to the ones in the treatment group. The difference-in-differences methodology used to produce the estimates account for this, taking only the change in the score in the treatment group that is over and above the change observed in the control group as the estimate of the effect of the intervention.

The estimates for males closely track the ones for the whole sample derived. The estimates for agency, confidence and self-expression, and leadership are 2.9, 4.2 and 6.6 percentage points respectively. They are statistically significant at 10%, 5% and 1% respectively. Although the estimates are all slightly larger than the aggregated ones, their statistical significance is weaker because of the drop in sample size that comes with the gender-disaggregation.

Table 17. Estimate of the Effect of the Intervention on the Youth Leadership Index by Gender

Youth Leadership Index in %	Estimate (SE)	p-value	Confidence Interval	n	Methodology
Total Index					
Female	2.2610 (1.4551)	0.120	(-0.592, 5.114)	bl: 866 el: 547	• DiD OLS • Whole sample ○ w/attr. & repl. ○ in-school & oos
Male	4.5646*** (1.5112)	0.003	(1.601, 7.528)	bl: 771 el: 480	
Agency					
Female	-0.5697 (1.6933)	0.737	(-3.890, 2.751)	bl: 869 el: 553	• DiD OLS • Whole sample ○ w/attr. & repl. ○ in-school & oos
Male	2.9466* (1.7627)	0.095	(-0.510, 6.403)	bl: 780 el: 481	
Confidence and Self-Expression					
Female	0.9389 (1.7141)	0.584	(-2.422, 4.300)	bl: 870 el: 550	• DiD OLS • Whole sample ○ w/attr. & repl. ○ in-school & oos
Male	4.2061** (1.7833)	0.018	(0.709, 7.703)	bl: 777 el: 482	
Leadership					
Female	5.7639*** (1.7141)	0.001	(2.351, 9.177)	bl: 871 el: 551	• DiD OLS • Whole sample ○ w/attr. & repl. ○ in-school & oos
Male	6.5829*** (1.8066)	0.000	(3.041, 10.125)	bl: 776 el: 483	

* p<0.10, ** p<0.05, *** p<0.01, Confidence Interval: 95%

Plans to Complete Higher Education

An insightful result derived above is the probability that a respondent plans to complete higher education. It was found to increase between baseline and endline across both the control and treatment groups, but to increase less in the latter by over 9 percentage points. This result was statistically significant at 1%.

Table 18 shows the estimates broken down by gender. Overall, the effect of the intervention is quite similar across both males and females. The DiD estimates are -0.11 for females and -0.08 for males. The former is statistically significant at 5% and the latter is not statistically significant. Overall, the estimates for males and females are quite similar and there is no evidence that they differ. The weaker level of significance can be explained by the drop in sample size associated with the gender-disaggregation. The drop in sample

size combined with the slight drop in magnitude for the estimate for males relative to the aggregate estimate account for the lack of statistical significance.

Table 18. Estimate of the Effect of the Intervention on the Probability of Planning to Complete Higher Education

Outcome	Estimate (SE)	p-value	Confidence Interval	n	Methodology
Plans to Complete Higher Education					
Female	-0.1097** (0.0512)	0.032	(-0.210, -0.009)	bl: 864 el: 520	• DiD Probit • Whole sample ○ w/attr. & repl. ○ in-school and oos
Male	-0.0830 (0.0555)	0.135	(-0.192, 0.026)	bl: 771 el: 462	

* p<0.10, ** p<0.05, *** p<0.01, Confidence Interval: 95%

Risk Preferences

In the aggregated analysis, risk preferences were found very likely to have been influenced by the intervention. Overall, risk aversion has increased in the entire sample, and participants in Youth Ready were found to have gotten even more risk averse. **Figure 6** and **7** illustrate the evolution of risk-aversion in the control and treatment groups for female and male respectively.

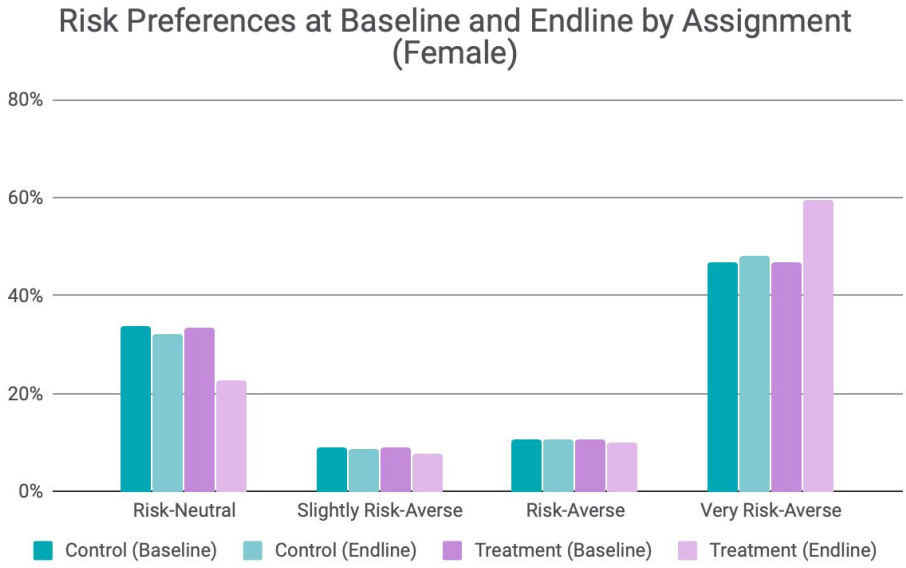

Figure 6. Predicted Probabilities (Ordered Probit) at Each Level of Risk Aversion at Baseline and Endline by Assignment (Female)

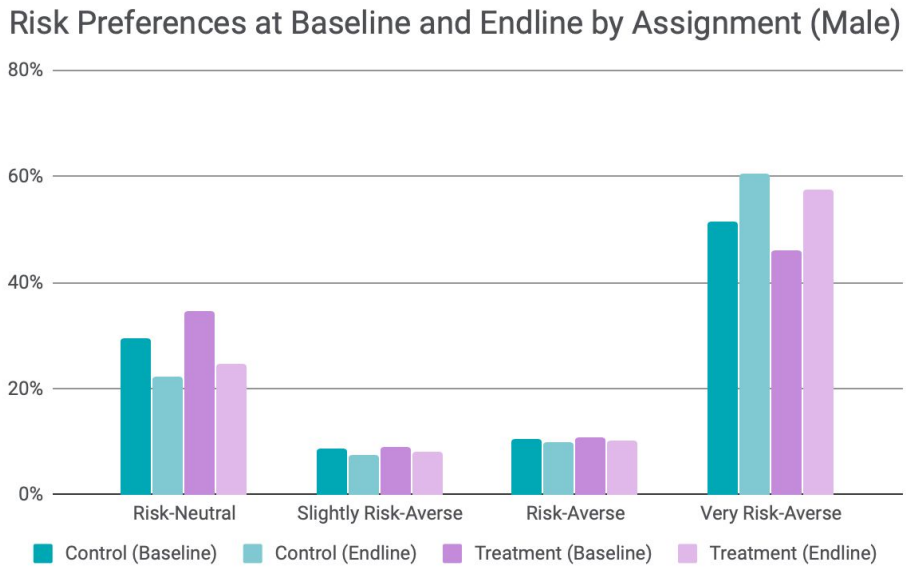

Figure 7. Predicted Probabilities (Ordered Probit) at Each Level of Risk Aversion at Baseline and Endline by Assignment (Female)

The difference in female and male participants is quite stark. The probability of being in each risk-aversion category is flat in the female control group, yielding a large relative increase in risk-aversion in the treatment group. Conversely, the increase in risk-aversion occurs in both the male control and treatment groups, and there is essentially no relative increase in risk-aversion in the treatment group.

Table 19 details the sample size and methodology for estimating the effect of the intervention by gender. The difference-in-differences estimates are shown in **Table 20**. As expected from the findings from the figures above, we find that the effect of the intervention on risk-aversion is driven almost entirely by the female subgroup. The estimates for the females are almost identical to the ones derived for the aggregate sample above. They are all statistically significant at 10%. Here again, the drop in statistical significance can be explained by the reduction in the sample size that comes with disaggregating results by gender. Unsurprisingly, none of the results for male are statistically significant as the estimates are small and close to zero.

Table 19. Sample Size and Methodology for Estimates of the Effect of the Intervention on Risk Preferences by Gender

n (Female)	n (Male)	Methodology

bl: 614
el: 432
	bl: 612
	el: 394
		• Difference-in-Differences Ordered Probit
		• Whole sample
		  ○ with attrition and replacements
		  ○ in-school and out-of-school

Table 20. Difference-in-Differences (DiD) Estimates of the Effect of the Intervention on Risk Preferences with an Ordered Probit Model by Gender

Risk Preferences	DiD Change in Probability	Standard Error	p-value	Confidence Interval (95%)
Risk-Neutral				
Female	-0.0915*	0.0555	0.099	(-0.200, 0.017)
Male	-0.0265	0.0535	0.620	(-0.131, 0.078)
Slightly Risk-Averse				
Female	-0.0111*	0.0059	0.060	(-0.023, 0.000)
Male	-0.0007	0.0064	0.914	(-0.013, 0.012)
Risk-Averse				
Female	-0.0066*	0.0036	0.064	(-0.014, 0.000)
Male	0.0012	0.0044	0.776	(-0.007, 0.010)
Very Risk-Averse				
Female	0.1092*	0.0636	0.086	(-0.016, 0.234)
Male	0.0259	0.0632	0.681	(-0.098, 0.150)

* p<0.10, ** p<0.05, *** p<0.01

In-School at Baseline: Graduation and Employment Outcomes

For in-school participants who are out-of-school by endline, the intervention was found to have a significant effect on the probability of having graduated from high school and of working for money in the aggregated analysis. **Table 21** breaks down these results by gender.

In both cases, the estimates are almost identical across both genders and to the ones derived above. Contrary to the aggregated estimates, none of them are statistically significant. The p-values vary from approximately 0.1 to 0.2. The loss of statistical significance in the estimates broken down by gender comes from splitting the already small sample size into two groups.

Only 192 observations could be used in the aggregated analysis as checking their baseline and endline school attendance status and observing the outcome at endline naturally restricts the sample down to the core sample of observations that appear at both baseline and endline. In addition, the young average age of in-school participants means that only a small fraction of those finish school by endline. Once the 192 observations are split by gender, the resulting sample size is less than or equal to 100 observations.

Table 21. Estimate of the Effect of the Intervention on Secondary School Graduation and Employment Outcomes for Participants In School at Baseline and Out of School at Endline by Gender

Outcome	Estimate (SE)	p-value	Confidence Interval	n	Methodology
Graduated from Secondary School					
Female	0.0907 (0.0709)	0.201	(-0.048, 0.230)	bl & el: 92	• Probit • Core sample ○ w/o attr. & repl. ○ in-school at baseline & oos at endline
Male	0.0899 (0.0553)	0.104	(-0.018, 0.198)	bl & el: 100	
Work for Money					
Female	0.1329 (0.0991)	0.180	(-0.061, 0.327)	bl & el: 92	• Probit • Core sample ○ w/o attr. & repl. ○ in-school at baseline & oos at endline
Male	0.1493 (0.0994)	0.133	(-0.046, 0.344)	bl & el: 99	

* p<0.10, ** p<0.05, *** p<0.01, Confidence Interval: 95%

Evaluation Summary

- Participants in the intervention group had higher high-school graduation rates and were more likely to report working for money following graduation than those in the control group.
- Participants in the intervention group had a lower probability of intending to pursue higher education compared to the control group. This is consistent with the higher opportunity cost of education given more of them are working after graduation, compared to the control group.
- Those in the intervention group had bigger positive changes in their Youth Leadership scores than the control group. The effects were more pronounced among male participants.
- Female participants in the intervention group became much more risk averse compared to the control group. There was no change among male participants.
- There was no evidence the program had an impact on measures of life outlook or time preferences

Evaluation Challenges

Evaluation is costly for a project, often requiring that choices are made during implementation that allow for rigorous evaluation. The implementation plans that allow for the most informative evaluation often can reduce the effectiveness of the implementation. But, evaluation is typically worth doing if the learning value is large enough, allowing future programs to be better designed and future funding to be better allocated.

In the case of this project, the hope of finding evidence that the relatively new Youth Ready program model has a significant impact on the outcomes of beneficiaries was enough to justify the evaluation. If the project was clearly shown to be beneficial, the results could be used to attract more resources to expand the Youth Ready program in other locations.

Several challenges arose during the evaluation of the Youth Ready program. These provide valuable lessons for future evaluations and have implications for the use of results based financing (RBF) to fund an expansion of the project, since evaluation is an essential component of RBF.

Challenge 1: There was a change in target participants, which would not have been an issue without evaluation, but did present issues for evaluation.

Limestone was initially engaged to design a randomized evaluation of the Youth Ready program in order to determine the impact of the program on the employment, entrepreneurship and short term behavior of out-of-school, unemployed youth. The evaluation design involved identifying potential program participants and then randomly assigning them to either an immediate cohort or a deferred cohort, which would be eligible for taking the program after one year. The first six months of participation involved regular program meetings and training, and the second six months involved mentoring and assistance in finding work, enrolling in school, or pursuing other life goals.

The timing of the program meant that the evaluation would measure the impact of the program six months after the completion of training. The goal was to measure the impact of the program on unemployment, school enrolment and other factors at this point in time. As a result, all of the instruments were initially developed to assess the economic outcomes of out-of-school, unemployed youth. Due to the delayed notification regarding the change in targeting, instruments were not adapted to track other potentially salient outcomes for in-school participants.

Challenge 2: Insufficient buy-in regarding the evaluation from the implementing team.

Evaluation imposes non-monetary costs on project as it can often interfere with implementation plans, leads to delays, and is sensitive to program changes. It is understood

that implementation teams are often busy and are working to achieve program targets, and so their role in the evaluation may not always seem clear or relevant.

However, it is important that they understand the importance of the evaluation and how choices made during the implementation of the program may affect the accuracy and feasibility of the evaluation. It is important that they are willing to implement the program in a way that does not undermine the evaluation.

In the case of Youth Ready, the apparent decision to disregard external randomization was an issue that undermined the interpretation of the results. Project changes, especially those involving changes to the beneficiary population, need to be better coordinated with evaluation design so that the surveys and assessments are appropriate for the beneficiary population.

Furthermore, more effort likely could have been undertaken to track and follow up with baseline survey participants. The attrition rate was possibly higher than it could have been. Of course, it is very difficult to engage and track out of school youth. More buy-in and direct engagement with the implementation team could have helped the evaluation team appropriately tailor the instruments in time and ensure the evaluation design was upheld.

Challenge 3: Sensitivities regarding criminal activity

The presence of gang activity and criminal participation among at risk youth in the locations of El Salvador and Honduras in which the program operated posed specific challenges for the evaluation.

First, although World Vision was interested in learning about the program's impact on criminal activity, accurately tracking such information through a survey is generally not feasible. Furthermore, our sample had a relatively small number of out-of-school and unemployed youths, who would be the group most likely to be engaging in such activities. The relatively small sample size for the population group that is most at-risk of criminal participation is too small to provide a representative picture of criminal activity within the group, even if we had a means of accurately measuring it. These factors mean that our analysis cannot speak to one of the potential benefits of the program.

Second, due to logistical and administrative concerns often related to the presence of criminal activity, randomization was possible in some, but not all locations. This also had implications on where and how enumerators could collect data. A number of strategies had to be adopted to ensure their safety and efficiency. It may be the case that future evaluation design may not be able to accommodate this type of data collection.

Lessons for Program Design & RBF

World Vision's Youth Ready program model was originally developed in Africa to provide employability, entrepreneurship and life skills training to out-of-school, unemployed youth. The implementation of the program in El Salvador and Honduras represents the first time that the program has been implemented in Latin America and the first time it has been adapted to focus primarily on in-school youth. It is also the first time that the Youth Ready program model, in any form, has been subjected to a rigorous impact evaluation.

A rigorous evaluation of the Youth Ready project is intended to assess both the impact of the program on beneficiaries in El Salvador and Honduras, and the general suitability of the program model for pursuing results based financing (RBF) opportunities, including impact bonds. In this section, we review the key insights that may affect World Vision's ability to adapt Youth Ready for RBF agreements.

Lesson 1: The shift in implementation strategy to focus treatment primarily on youth enrolled in school likely increased the impact of the program.

Although out-of-school, unemployed youth have the most-obvious, immediate use for life-skills and employability training, it can be difficult to enrol such a population segment in a formal training program. Many eligible participants are not interested in such an opportunity. Furthermore, the program might not do much to increase the labour market opportunities of those who do decide to enrol. This could be because those who enrol are already self-motivated people who would find other opportunities to improve their employability even if they did not have access to the program. It could also happen if it is already too late for training to improve labour market opportunities for those who have already been out-of-school and unemployed for a period of time.

Shifting focus to in-school youth appears to have significantly increased demand for the program, allowing the program to reach a larger participant base, and perhaps reach segments of the population that would not be willing to participate if the training was not incorporated into school curriculum.

Furthermore, our evaluation of the El Salvador and Honduras data suggests that the evidence of impact of the program on labour market outcomes is much stronger for in-school participants. This suggests that a program like Youth Ready may have the greatest impact as a preventative measure, providing training to at risk youth before they leave school and enter the formal labour market. It is worth considering a revision of the program model to prioritise training for youth before they leave school. Since this is the population segment for which our analysis shows the strongest evidence of impact, focusing on this population would help reduce the risks inherent in RBF agreements and impact bonds.

Lesson 2: There is room to improve service provider capacity to implement programs in ways that facilitate evaluation.

Although the evaluation provides several meaningful insights into the impact of the Youth Ready program, it could have been even more informative had the evaluation team been aware of certain implementation changes sooner. Not being immediately informed of the program's decision to shift focus to in-school trainings did not allow the evaluation team to fully adapt the evaluation instruments and design to fully measure the impact of the program on in-school participants.

Furthermore, the program appears not to have followed the evaluation strategy's assignment of participants into treatment and control groups. This was never discussed with the evaluation team and we only discovered it due to inconsistencies when reviewing data collected at endline. We discuss this concern elsewhere in this report, but the overall issue is that it is important for an external evaluator to be able to determine how people were selected into treatment and comparison groups and to ensure that such selection would not undermine the evaluation.

In the current context, such communication failures are unfortunate and lead to a less-informative evaluation than may have otherwise occurred. The same failures would have had severe financial implications if the project had been funded through a RBF agreement or impact bond. Changes to the target population that requires a reconsideration of key outcome variables, or a failure to follow the evaluation's assignment of treatment and control groups would have led the project to not achieve the required impact targets to receive a payment for success.

Specific suggestions for improving implementer capabilities to facilitate evaluation include the following:

- Ensure that program managers have a fundamental understanding of monitoring and evaluation processes, and how implementation decisions can undermine evaluation design.
- Ensure that service provider leadership and financing groups understand how rigorous evaluation imposes significant costs on the project and may limit its impact on beneficiaries.
- Ensure engagement of front-line implementers and in-country teams when designing programs and evaluation strategies at earlier stages.

Lesson 3: Using RBF to fund Youth Ready is likely feasible in certain contexts, if the program and funding mechanism are well designed.

The evaluation of the Youth Ready program in El Salvador and Honduras has shown that the program has likely increased school completion rates and improved labour market opportunities of participants following secondary school, and that it offers benefits in terms of improvements to the Youth Leadership Index measuring self perceptions of

confidence, engagement and agency. Although the analysis suggests that we can be fairly confident that the program had a positive impact on these outcomes, there remains significant uncertainty about the magnitude of these impacts, and the likelihood of having similar impact in different contexts.

Because of this unresolved uncertainty, the program is likely not a suitable candidate for pursuing an impact bond at this point in time. Impact bonds require attracting an investor to the project in order to provide up-front funding. In the event that the project succeeds in achieving a series of agreed upon outcome targets, the investor is paid back plus a reasonable rate of return from an outcome funder who has committed to pay for success. An impact bond involving the Youth Ready program would require an investor to take on large amounts of risks, due to uncertainty about whether the program would be able to achieve reasonable performance targets in different environments, and whether implementation of the program will facilitate evaluation.

At this point in time, the Youth Ready program would be a strong candidate for a more-traditional RBF agreement, such as a grant-plus-bonus structure in which a funding organization provides initial funding to cover most or all of the costs of implementation, plus a bonus payment when the implementation of the program achieves certain performance targets. The bonus payment can still provide incentives to the service provider to focus on outcomes, facilitate rigorous evaluation, and adapt the program to better achieve results. But, such a payment structure provides these incentives without transferring risk to an investor (and ultimately needing to compensate the investor for taking on such risks).

Lesson 4: The local contexts may pose unique challenges for evaluation design -- this can have implications for RBF focused on impact

In the case of Youth Ready in El Salvador and Honduras, the presence of criminal activities limited the ability to implement a randomized evaluation design and required special provisions for data collection that may be difficult to obtain in future iterations.

RBF mechanisms that intend to pay out on measures of impact (such as impact bonds) however require rigorous evaluation. Therefore if an impact evaluation is likely to be compromised by external factors, and there are limited options for mitigating these factors, an impact bond may be a risky financing mechanism to pursue. This however does not mean other types of RBF mechanisms cannot pursued, particularly if the mechanism rewards outputs. However it does mean that if a type of evaluation design required to verify results under a given RBF mechanism, cannot be implemented, then a different financing model should be pursued.

Conclusion

The evidence consistently suggests that the Youth Ready program in El Salvador and Honduras succeeded at increasing labour market opportunities for participants. The evidence is most substantial for youth who participated in the Youth Ready program while still enrolled in secondary school. The analysis shows that these youth were more likely to graduate from secondary school and were less likely to be unemployed after graduation.

Overall, the program also significantly influenced measures of leadership and empowerment, as measured by the Youth Leadership Index as well as participant's attitudes towards risk.

Although the evaluation shows how the program likely increases post-graduation employment of participants, much uncertainty remains about the magnitude of the impact on employment and education outcomes. Results are stronger on proximal outcomes like the impact of the program on the YLI.

Because there remains substantial uncertainty about the impact of the Youth Ready program in new environments, the program is not a strong candidate for certain types of RBF such as impact bonds at this time. The financial risk to the investor under such an arrangement is likely too high for such a funding mechanism to be feasible. However, other types of RBF, such as a grant + bonus mechanism is likely feasible. Such an agreement may provide incentives to the project implementers to carefully plan the implementation in a way that allows rigorous evaluation. RBF has to potential to improve incentives for better evaluation, more consistent implementation, and more streamlined implementation design, particularly given the noted evaluation challenges that arose in El Salvador and Honduras, and may arise again in other environments.

Appendix A: Youth Leadership Index

The Youth Leadership Index (YLI) is constructed using answers to 21 survey questions on participant's "self-confidence, their decision-making, problem solving and organizational skills, their sense of voice, and their ability to motivate others" (CARE USA, 2014). It was designed by CARE International to "measure changes in self-perceptions of leadership among youth" over time (CARE, 2014).[5]

The index was tested through a pilot on 2,750 youth in Bangladesh, Burundi and India. It was then used as part of a rigorous evaluation of the "Improving Girls' Access Through Transformative Education" program in Zimbabwe that was implemented by a consortium led by World Vision UK (CARE USA, 2014). The index was found to perform well across these interventions and produce scores within a consistent range (CARE USA, 2014).

Table A.1 shows the 21 questions that make up the Youth Leadership Index. Each question may be answered with one of the following four options: 1. Rarely 2. Sometimes 3. Most of the Time 4. Almost Always. The table also shows how the questions were allocated to the following categories: agency, confidence and self-expression, and leadership[6].

Table A.1: Youth Leadership Index Questions and Allocation to Categories

#	Question	Category
1	I like to try new activities that I may not know how to do.	Confidence and Self-Expression
2	My friends ask me for advice.	Leadership
3	I recognize when people have different skills to contribute to a task.	Leadership
4	I am comfortable when a teacher or boss calls on me to answer a question.	Confidence and Self-Expression
5	I contribute ideas to discussions at home even if they are different from others' ideas.	Confidence and Self-Expression
6	I ask questions at school or work when I don't understand something.	Confidence and Self-Expression
7	I can describe my thoughts to others.	Confidence and Self-Expression
8	The things I do set a good example for my peers.	Leadership
9	I consider possible outcomes of my decisions before making them.	Agency
10	I accept responsibility for the outcomes of my decisions.	Agency
11	I recognize when choices I make today can affect my life in the future.	Agency

[5] In 2012, the YLI replaced CARE's previously developed Girls' Leadership Index as their preferred means of assessing self-perceived leadership attributes among girls and other youth aged 10-17. The index has been used to measure youth leadership in several settings, both by CARE and others (CARE, 2014).
[6] In a future analysis, these domains will be tested for homogeneity using factor analysis.

12	I can show what is important to me with my actions.	Confidence and Self-Expression
13	If someone does not understand me, I try to find a different way of saying what is on my mind.	Confidence and Self-Expression
14	I encourage others to join together to help my community.	Leadership
15	I cooperate with others to get things done at home.	Leadership
16	If someone treats me unfairly at school or work, I am comfortable telling a teacher, parent or boss.	Confidence and Self-Expression
17	I work hard to achieve my dreams.	Agency
18	I am better able to finish a task when I plan ahead.	Agency
19	When I have the opportunity, I can organize my peers to do an activity.	Leadership
20	I am interested in being a leader at my school or work, or in my community.	Leadership
21	I try to understand the cause of a problem before trying to solve it.	Agency

The index is calculated by assigning 1, 2, 3 or 4 points for each of the respective choices (1. Rarely 2. Sometimes 3. Most of the Time 4. Almost Always). The index is out of 84 points (21 questions $\times$ 4 points). The score is divided by 84 and multiplied by 100 to obtain a percentage, which is what is used in the analysis.

Similarly, the index for each of the three categories of questions is calculated by summing the 1-4 points awarded to each of the relevant questions and dividing by the total number of points available in that category. The score is then multiplied by 100 to obtain a percentage.

Appendix B: Desired Maximum Amount of Schooling

Translation issues with the survey question on the desired maximum amount of schooling has made it necessary to aggregate the responses into a binary measure of the desire to achieve higher education. Table B.1 details how the baseline and endline education level choices are allocated to "higher education" and "others".

Table B.1: Allocation of Desired Maximum Amount of Schooling into a Binary Measure of the Desire to Achieve Higher Education

Desired Maximum Amount of Schooling	Baseline	Endline
Higher Education	• Técnico universitario • Licenciatura o Ingeniería • Maestría o Postgrado	• Universidad • Maestría o postgrado • Doctorando
Others	• Primaria • Secundaria • Bachillerato general • Bachillerato técnico • No estoy seguro(a)	• Primaria • Secundaria • Bachillerato general • Instituto • No estoy seguro(a)

Appendix C: Risk Preferences

The following three questions in the baseline and endline surveys are used to assess risk preferences:

1. Imagine that you have the two following options:

   (i) - You receive $6 (L. 240) for sure OR
   (ii) - You flip a coin; if it is tails you get 0 and if it is heads you get $18 (L. 720).

   Which option would you choose?

   1. Take the money for sure
   2. Flip the coin
   3. Not sure
   4. N/A

2. Imagine that you have the two following options:

   (i) - You receive $6 (L. 240) for sure OR
   (ii) - You flip a coin; if it is tails you get 0 and if it is heads you get $24 (L. 960).

   Which option would you choose?

   1. Take the money for sure
   2. Flip the coin
   3. Not sure
   4. N/A

3. Imagine that you have the two following options:

   (i) - You receive $6 (L. 240) for sure OR
   (ii) - You flip a coin; if it is tails you get 0 and if it is heads you get $12 (L. 480).

   Which option would you choose?

   1. Take the money for sure
   2. Flip the coin
   3. Not sure
   4. N/A

Appendix D: Time Preferences

The following three questions in the baseline and endline surveys are used to assess time preferences:

1. Now I'm going to propose two new options. Imagine that you have to make a choice between either one:

   (i) - You receive $60 (L. 1200) today OR
   (ii) - You receive $90 (L. 1800) for sure in a year from now.

   Which option would you choose?

   1. Take the money today
   2. Wait a year
   3. Not sure
   4. N/A

2. Now I'm going to propose two new options. Imagine that you have to make a choice between either one:

   (i) - You receive $60 (L. 1200) today OR
   (ii) - You receive $120 (L. 2400) for sure in a year from now.

   Which option would you choose?

   1. Take the money today
   2. Wait a year
   3. Not sure
   4. N/A

3. Now I'm going to propose two new options. Imagine that you have to make a choice between either one:

   (i) - You receive $60 (L. 1200) today OR
   (ii) - You receive $72 (L. 1440) for sure in a year from now.

   Which option would you choose?

   1. Take the money today
   2. Wait a year
   3. Not sure
   4. N/A