USAID FROM THE AMERICAN PEOPLE | BOSNIA-HERZEGOVINA

IMPACT EVALUATION OF SOCIAL AND ECONOMIC PROGRAMS IN BOSNIA AND HERZEGOVINA:
IMPACT EVALUATION OF LIFE SKILLS CURRICULUM IMPLEMENTATION IN PRIMARY SCHOOLS IN BIH

JULY 2019, FINAL REPORT

This document is made possible by the generous support of the American people through the United States Agency for International Development (USAID). The contents are the responsibility of the authors and do not necessarily reflect the views of USAID or the United States Government.

MONITORING AND EVALUATION SUPPORT ACTIVITY (MEASURE-BiH)

IMPACT EVALUATION OF SOCIAL AND ECONOMIC PROGRAMS IN BOSNIA AND HERZEGOVINA:
IMPACT EVALUATION OF LIFE SKILLS CURRICULUM IMPLEMENTATION IN PRIMARY SCHOOLS IN BIH

July 2019, Final Report

Prepared under the USAID Monitoring and Evaluation Support Activity (MEASURE-BiH),
Contract Number AID-168-C-14-00003

Submitted to:
USAID MEASURE BiH
Submitted by:
Center for Development Evaluation and Social Science Research

TABLE OF CONTENTS

LIST OF EXHIBITS

ACKNOWLEDGMENTS

The authors would like to thank numerous individuals who have contributed to the implementation of the impact evaluation of the Life Skills curriculum implementation in primary schools throughout BiH and to the preparation of this report.

At USAID/Bosnia and Herzegovina (USAID/BiH), Elma Bukvić Jusić, the Contracting Officer's Representative (COR) of the MEASURE-BiH Activity, provided full support throughout evaluation.

We would like to extend our thanks to the MEASURE-BiH team, in particular to Marios Michaelides, Technical Director at MEASURE-BiH, who provided us with invaluable methodological guidance and insights, as well as to the Chief of Party Davorin Pavelić and Deputy Chief of Party Edis Brkić, who provided technical guidance for the successful implementation of this effort.

Furthermore, we thank the Pedagogical institute of Tuzla Canton and its director Nikola Čiča, PhD for recognizing the importance of monitoring and evaluation of public policies and providing us with the necessary approvals, as well as the administrative data of the schools from which data was collected.

Also, we would like to thank World Vision, who were interested in collaboration with CDESS in order to evaluate their efforts to improve the life skills of children in their partner schools within Tuzla Canton through their programs.

All of the above organizations and individuals have contributed tremendously to this report, but any remaining errors are the sole responsibility of the authors.

ACRONYMS

BiH – Bosnia and Herzegovina

CDESS – Centre for Development Evaluation and Social Science Research

FBiH – Federation of BiH

MoE – Ministry of Education

RS – Republika Srpska

TK – Tuzla Canton

USAID MEASURE-BiH – USAID/BiH Monitoring and Evaluation Support Activity

USAID/BiH – United States Agency for International Development, Office in BiH

WHO – World Health Organization

EXECUTIVE SUMMARY

This report provides results of the impact analysis of the project "Impact evaluation of Life Skills curriculum implementation in primary schools in BIH".

EVALUATION PURPOSE AND QUESTIONS

In order to provide evidence on the Life Skills program's contribution to an increase in life skills among children in program schools, CDESS conducted rigorous impact evaluation of the program in primary schools within Tuzla Canton.

The main evaluation question is as follows: *What is the effect of Life Skills program in primary schools on the psychological and behavioral responses of children?* All investigations undertaken as part of the evaluation aim to answer this.

EVALUATION METHODS AND LIMITATIONS

Due to the characteristics of the sample, the multilevel modelling approach (MLM) was chosen to take into account clustering by time, classroom and schools. As the preliminary results on interclass correlation indicated that ICC values were in all cases below 10% multilevel, and grouping between the data are not significant, for capturing the impact of the Life Skills program on psychological and behavioral response of children, the difference-in-difference (DID) method was used in addition to MLM.

The DID approach compares the before-after changes in outcomes between children in the treatment group and matched children in the comparison group. It is important to note that the DID does not require baseline (pre-intervention) conditions to be the same in treatment and comparison groups, as our baseline results indicate there is a statistically significant average difference between treatment and comparison groups when it comes to the following outcomes: Making decisions in everyday life (MD); Social responsibility (CR); Critical thinking in everyday life (CT); Social relationships (SR); and Emotional management (EM).

A key methodological limitation of the design of the evaluation was related to its quasi-experimental nature when compared to full experimental designs. Drawbacks were minimized by controlling for initial differences between treatment and comparison groups using propensity score matching and inverse probability weighting. Additional methodological limitations stemmed from the Theory of Change, whereby it is expected that life skills affect social change indicators, where it is possible that the level of social change indicators can have significant confounding variable influencing relationships between treatment and life skills.

FINDINGS AND CONCLUSIONS

Results indicate a statistically significant impact of the Life Skills program on some of the outcome indicators from the Cognitive Behavior list, such as somatic complaints or

anxiety and depression among children. These effects were confirmed from four different model estimation approaches. For other outcome indicators, we cannot find evidence suggesting a positive impact of the intervention. Moreover, for some indicators we can even find a negative impact of the intervention, although statistical significance of the intervention coefficient is not consistent across the model estimations.

1. INTRODUCTION

Recognizing the value, strength and high coverage of the education sector in the promotion and delivery of knowledge, attitudes, skills and behaviors important for short, medium and long term inclusion in society and the labor market, civil society organizations in BiH argued the importance of broadening educational curricula from their focus on academic knowledge to a social skills-oriented one. A number of initiatives have been accepted by education policy makers and "integrated" into mainstream curricula, including inclusive education. However, due to the obvious value of the education system as a channel for the development of children's skills, only a relatively small number of educational practices have been scrutinized using objective impact evaluation. Respective Ministries of Education (MoEs) have only begun to conduct evaluations of the secondary school curriculum as part of EU-funded programs, which were primarily focused on performance rather than on impact. The importance of conducting rigorous impact evaluations in such circumstances cannot be overstated.

As part of its efforts aimed at research and evaluation capacities and activities for more rigorous policy analysis in Bosnia and Herzegovina (BiH), USAID MEASURE-BiH published a Call for Proposals in summer 2017 and invited research organizations in BiH to submit proposals for rigorous impact evaluations of social and economic programs in BiH in any area of social and economic policy, which can inform decision-making and increase effectiveness of development programs in BiH. A further aim of the Call was to demonstrate to a broad audience of government, international organizations and research communities that low-cost impact evaluations are a powerful new tool for building scientific evidence about "what works" in social and economic spending. The Center for Development Evaluation and Social Science Research (CDESS) proposed conducting a rigorous impact evaluation of World Vision's Life Skills program. After a selection process, USAID MEASURE-BiH awarded CDESS a contract for conducting the proposed evaluation. This report presents the results of this impact evaluation.

In 2017, the Ministry of Education and Culture of Republika Srpska, and the Ministries of Education, Science, Culture and Sport of Tuzla Canton, Zenica-Doboj Canton and Bosnia Podrinje Canton, in cooperation with World Vision, initiated the development of a Life Skills Curriculum for primary schools. Life skills education is seen as an important vehicle to equip children for successful negotiation and mediation of challenges and risks in their lives, and the enabling of productive participation in society. Representatives of the MoEs recognized that provisions made in formal schooling – either in terms of curriculum content or teaching – for the development of life skills that would assist girls and boys in thriving solely within a school environment are not sufficient.

In order to provide evidence on the Life Skills program's contribution to increasing life skills among children in schools delivering the program, CDESS conducted rigorous impact evaluation of the program in primary schools in Tuzla Canton. Recognizing the importance of evidence-based decision making in the introduction of education policies, and lack of systems currently in place to allow adherence to a strict methodological

framework for the collection and analysis of quantitative and qualitative data, the Ministry of Education in Tuzla Canton[1] accepted an invitation to participate in the Impact Evaluation of the Life Skills Program.

The main objectives of the evaluation were:

- to provide information about changes in children's livelihoods attributable to the Life Skills program;
- to provide information about key components of the Life Skills program contributing to these changes;
- to develop a roadmap for evaluation practices in the education system to be used for the evaluation of future similar practices.

The main evaluation question was expressed as follows:

> What is the effect of the Life Skills program in primary schools on the psychological and behavioral responses of children?

The report is structured as follows. In the second part of the report we provide the background on the Life Skills program, its purpose, and the objectives of the evaluation. We explain the Theory of Change in detail, focusing on the outcomes that were measured. The third part of the report provides detailed information on the methodology used for sample selection, description of data collection instruments, and also a description of impact evaluation model and methods to be used for impact evaluation. The fourth part of the report provides the results of the impact analysis. The fifth and final part concludes and offers recommendations for continuation of the program, as well as for future impact evaluation in this area.

[1] Respective ministries of education in the other cantons and entity covered by this program did not provide approval for conducting impact evaluation in their schools in time, and therefore could not be included in this evaluation.

2. BACKGROUND AND PURPOSE OF EVALUATION

This section of the report provides background information pertaining to World Vision's Life Skills program, explains the Theory of Change underlying the introduction of the Life Skills program into primary schools, and describes the evaluation objectives and evaluation questions.

2.1. BACKGROUND TO WORLD VISION'S LIFE SKILLS PROGRAM IN BOSNIA AND HERZEGOVINA

In 2017, World Vision, along with its partners, the Ministry of Education and Culture of Republika Srpska and the Ministries of Education, Science, Culture and Sport from Tuzla Canton, Zenica-Doboj Canton and Bosnia-Podrinje Canton, initiated the development of the Life Skills program for primary schools.

The main objective of the Life Skills program has been to develop capacities of children in selected primary schools (World Vision's program partner schools) for making decisions and taking actions that positively impact their lives. The program is focused on critical thinking, effective communication, emotional management, interpersonal relationships and community responsibility. Such a program is expected to prepare children and youth for an independent life through development of those life skills required for understanding the potential use and consumption of available resources, identification of their roles and decision-making relating to the professional directions that they wish to pursue. The Life Skills program as such requires increased teaching efforts to develop a number of skills. It expects teachers to integrate additional educational content/dynamics into the teaching of regular subjects such as math, languages, sciences, etc. Additional educational content might include discussions about similarities between people, talking about identities, self-esteem, managing feelings, how to cope with anger and fury, assertive communication, negotiation skills, decision making, peer pressure, peer cooperation, needs, peer violence, etc. These themes need to be integrated into classroom-based teaching non-intrusively, without obstructing the education process. Teachers may further elect to organize special workshops specifically focused on those themes.

During the implementation of Life Skills program, the following activities were conducted:

1. Development of training material for teachers on integration of life skills themes into mainstream curriculums. Training material for teachers was prepared by the Ministries of Education / Pedagogical Institutes by 1 September 2017;
2. Conducting training for selected teachers on life skills integration into mainstream curriculum. In September 2017, World Vision initiated trainings with the teachers in project (treatment) schools;
3. Spill-over education to other teachers in program schools (from October 2018);
4. Integration of the Life Skills thematic program into regular school classes in 7th, 8th and 9th grades. Integration of the Life Skills program commenced as of 1November 2017 (activity was ongoing during the spring semester of the 2017-2018 academic year).

The plan for the first two years of the initiative was to implement the Life Skills program within the mainstream curriculum in 19 municipalities (covering 71 primary schools and pupils in targeted municipalities: Banja Luka, Banovići, Čelinac, Doboj, Doboj-Istok, Goražde, Han Pijesak, Kakanj, Kalesija, Kneževo, Kotor Varoš, Lukavac, Olovo, Pale, Petrovo, Sapna, Sokolac, Visoko, Zvornik) distributed across the Republika Srpska (RS) entity and three cantons in the Federation of BiH (FBiH) entity. The Ministry of Education in Tuzla Canton[2] accepted an invitation to participate in the Impact Evaluation of the Life Skills Program.

Based on the results of the two-year initiative, the Ministry of Education (through their Pedagogical Institutes) would propose integration of the Life Skills program within mainstream education curricula in all schools under the authority of the respective administrative unit. Apart from traditional assessment of initiative performance and qualitative analysis of the opinions of initiative partners on impact of the life skills curriculum on children's livelihoods, there was no commensurate effort relating to the systematic monitoring and evaluation of expected impact at individual level. This impact evaluation is expected to produce necessary evidence and contribute to the evidence-based decisions on the implementation of the program in other schools across Tuzla Canton.

2.2. LIFE SKILLS PROGRAM – THEORY OF CHANGE

Life skills can be defined as "the abilities for adaptive and positive behavior that enable individuals to deal effectively with the demands and challenges of everyday life" (WHO, 1997). The Theory of Change assumes that exposing children to a life skills program will develop different types of life skills and attitudes that will be beneficial both to the individual and to their immediate community and society.

The life skills curriculum's Theory of Change assumes that exposing children to social dilemmas (treatments – life skills program) will develop a variety of types of social skills (outcomes determined in the Manual for teachers on life skills curriculum, which included skills such as: decision making, problem solving, creative thinking, critical thinking, effective communication, interpersonal relationship skills, self-awareness, empathy, coping with emotions and coping with stress), as well as attitudes that will be beneficial both to individuals and to society as a whole. Expected benefits (outcomes) are separated into two groups: Psychological and Behavioral Response and Long Term outcomes.

Psychological and behavioral responses of children are defined in relation to self-esteem, anxiety, depression, somatic complaints, social problems, thought problems, attention problems, rule-breaking behavior, aggressive behavior and social status (popular children, rejected children, neglected children, average children and controversial children). Long-

[2] Respective ministries of education in other cantons/entity covered by this program did not provide their approval for conducting impact evaluation in their schools in time, so they could not have been included in this evaluation.

term outcomes include a reduced probability of a child becoming involved in conflict with the law when they are older, increased labor skills and employment rate.

This impact evaluation focuses on outcomes related to improvement of psychological and behavioral responses of children since these outcomes can be captured in the short term and within the timespan available for completing this evaluation. Increase in labor skills of children and their labor market performance (such as probability of gaining employment or earning higher wages due to increased productivity) in future stages of their lives are outcomes that can only be measured in the long-run and, therefore, are not considered in this evaluation.

Psychological and behavioral responses of children are related to the following three groups of skills:

1. Thinking skills. These are the skills that enhance the logical faculty of the brain using an analytical ability, thinking creatively and critically, and developing problem-solving skills and improving decision-making abilities.

2. Social skills. These include interpersonal skills, communication skills, leadership skills, management skills, advocacy skills, co-operation and team-building skills, etc.

3. Emotional skills. These skills involve knowing and being comfortable with oneself. They primarily relate to self–management, including managing/coping with feelings, emotions, stress and resisting peer and family pressures.

In recent studies, these changes attributable to Life Skills programs have been measured mainly through the children's academic performance and their psychological skills and behavior. Tella et al. (2008) found teachers' self-efficacy to be the best predictor of pupils' academic achievement. Kim and Bryan (2017) found evidence of significant relationships between parents' competence, self-determination, community belonging, community participation, and academic performance, which differentiate among parents based on racial/ethnic, linguistic, and socio-economic backgrounds.

As previous research suggests, it is expected that the Life Skills program will yield positive psychological and behavioral responses from children in all of the above-mentioned skills. For example, Yankey et al. (2012) found a positive impact of life skills programs on psychological, social and mental wellbeing of high school students, Esmaeilinasab et al. (2011) found improvement of self-esteem among students, Tahereh et al. (2011) reported increasing happiness, quality of life and regulation of emotions, while Nasser et al. (2010) found decreasing mental disorder symptoms, especially those of anxiety, depression and stress among students. The impact of life skills training upon psychological well-being has been shown to differ between boys and girls (Kumar et al., 2016). Further, life skills training has been shown to reduce psychological distress (Ghasemian and Kumar, 2017). Research focused on academic achievement has shown a significant correlation between student achievement and teacher experience, teacher guides and instructional time (Song, 2012).

The World Health Organization (WHO) provided guidelines for evaluating life skills programs where a positive outcome is captured and measured through ten life skills factors: self-awareness, empathy, critical thinking, creative thinking, decision making, problem solving, effective communication, interpersonal relationship, coping with stress and coping with emotion. Each of these factors is explained in greater detail below:

Self-awareness includes recognition of the 'self', our character, our strengths and weaknesses, desires and dislikes. Developing self-awareness can help children to recognize when they are stressed or feel under pressure. It is often a prerequisite to effective communication and interpersonal relations, as well as for developing empathy with others. It is expected that a Life Skills program will increase the levels of self-awareness of children.

Empathy: In order to have a successful relationship with our loved ones and society at large, we need to understand and care about other people's needs, desires and feelings. Empathy is the ability to imagine what life is like for another person. Without empathy, our communication with others will amount to one-way traffic. Even worse, we will act and behave according to our own self-interest, and are bound to run into problems. Empathy can help us to accept others who may be very different from ourselves. This can improve social interactions, especially in situations of ethnic or cultural diversity. Empathy can also help to encourage nurturing behavior towards people in need of care and assistance, or tolerance for persons with mental disorders, who may be stigmatized by the very people they depend upon for support. It is expected that a Life Skills program will increase empathy among children.

Critical thinking is the ability to analyze information and experiences in an objective manner. Critical thinking can help us to recognize and assess the factors that influence attitudes and behavior, such as values, peer pressure and the media. Creative thinking is a novel way of seeing or doing things that is characteristic of four components – fluency (generating new ideas), flexibility (shifting perspective easily), originality (conceiving of something new), and elaboration (building on other ideas). It is expected that a Life Skills program will increase children's critical and creative thinking abilities.

Decision making helps us to deal constructively with decisions about our lives. It can teach people to actively make decisions about their actions by means of a healthy assessment of different options and the effects that these different decisions are likely to have. It is expected that a Life Skills program will improve decision making among children.

Problem solving helps us to deal constructively with problems in our lives. Significant problems that are left unresolved can cause mental stress and give rise to accompanying physical strain. It is expected that a Life Skills program will increase the ability of effective problem solving among children.

Interpersonal relationship skills help us to relate in positive ways with the people with whom we interact. This may mean being able to form and maintain friendly relationships, which can be of great importance to our mental and social well-being. It may mean keeping

good relations with family members, who are an important source of social support. It may also mean being able to end relationships constructively. It is expected that a Life Skills program will improve interpersonal relationships among children.

Effective communication means that we are able to express ourselves – both verbally and non-verbally – in ways that are appropriate to our cultures and situations. This means being able to express opinions and desires, and also needs and fears. And it may mean being able to ask for advice and help in a time of need. It is expected that a Life Skills program will improve effective communication among children.

Coping with stress means identifying the sources of stress in our lives, recognizing how this affects us, and acting in ways that help us control our levels of stress, by changing our environment or lifestyle and learning how to relax. It is expected that a Life Skills program will help children more effectively cope with stress.

Coping with emotions involves the recognition of emotions within us and others, being aware of how emotions influence behavior, and being able to respond appropriately to different emotions. Intense emotions like anger or sadness can have negative effects on us if we do not respond appropriately. It is expected that a Life Skills program will help children to cope with emotions more effectively.

The underlying assumption for achieving life skills outcomes is that children are exposed to at least a minimal quality of life skills teaching provided by teachers. Another assumption is that there will be no contamination of comparison groups by similar life skills-themed programs. As teachers come with different backgrounds, levels of teaching experience, different motivations for integration of a life skills curriculum and different competences for implementation of a life skills program, it is expected that project performance will largely depend on such characteristics of the teachers involved. Furthermore, children's levels of participation during classes might significantly differ, which could also influence project outcomes.

3. EVALUATION DESIGN

3.1. EVALUATION OBJECTIVES

In order to provide evidence on the effect of the Life Skills program on life skills among children in program schools, CDESS conducted a rigorous impact evaluation of the program in primary schools throughout Tuzla Canton.

The main objectives of the evaluation were:

- to provide information about changes in children's life attributable to the Life Skills program;
- to provide information about key components of the Life Skills program contributing to the change;
- to develop a roadmap for evaluation practices in the education system to be used for the evaluation of future similar practices.

3.2. EVALUATION QUESTIONS

The main evaluation question here can be specified as:

> **What is the effect of the Life Skills program in primary schools on the psychological and behavioral response of children?**

In addition, three supporting evaluation questions to assist the research were further specified, as follows:

> **1**: To what extent does the quality of the Life Skills program's implementation affect the outcomes?
> **2**: To what extent do the effects of the implementation of the Life Skills program differ between boys and girls?
> **3**: To what extent do the effects of the implementation of the Life Skills program differ between cohorts of children coming from households with higher and lower incomes?

In order to produce valid evidence as answers to the above questions, an evaluation approach was designed as described in the remainder of this chapter.

3.3. IMPACT EVALUATION MODEL

Based on the previous research, we developed an impact evaluation model to be tested with the proposed set of indicators, presented in Exhibit 1.

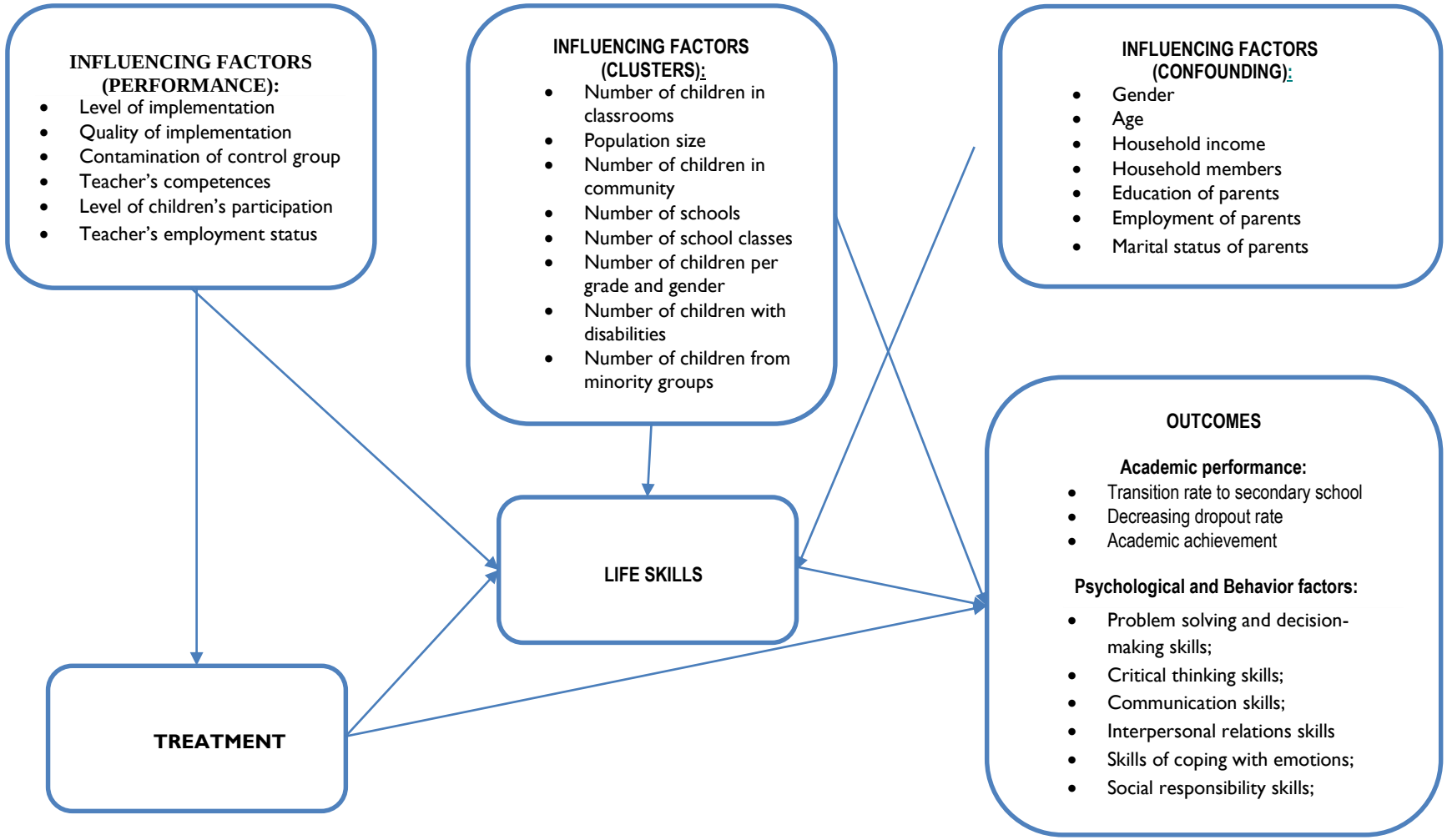


Exhibit 1: Life skills main evaluation model

The effect of the Life Skills program on the psychological and behavioral response of children was measured using a series of scales relating to the following:[3]

(1) Problem solving and decision-making skills;
(2) Critical thinking skills;
(3) Communication skills;
(4) Interpersonal relations skills;
(5) Skills of coping with emotions;
(6) Social responsibility skills;
(7) Self-esteem;
(8) Reducing incidence of violence.

Additionally, the effect of the Life Skills program on the psychological and behavioral response of children was also measured using World Vision's Developmental Assets Profile Scale (Search Institute, 2005) that capture development of children in the following eight asset categories: support, empowerment, boundaries and expectations, constructive use of time, commitment to learning, positive values, social competences and positive identity.

The underlying assumption for achieving life skills outcomes is that children are exposed to at least a minimal quality of life skills teaching provided by teachers, and established in the Teacher's Manual. Furthermore, children's levels of participation during classes might significantly differ, which could also influence project outcomes. To capture the quality of program implementation, it was measured in terms of frequency of program implementation, quality of child participation in the program, competences of the teachers, and quality of the relationships between children and teachers.

The Teachers' Manual was prepared by the Ministries of Education / Pedagogical Institutes by 1 September 2017, and in September 2017 trainings were initiated with the teachers in project schools. Integration of the Life Skills program commenced as of 1 November 2017. The program included training of trainer-teachers at the level of schools who will replicate trainings for all teachers within their school. The training of trainers was conducted by the Ministry of Education / Pedagogical Institute. The Life Skills curriculum was intended to be integrated within the regular curriculum for all subjects, and integrated by all teachers into their teachings. As teachers come with different backgrounds, levels of teaching experience, different motivations for integration of a life skills curriculum and different competences for implementation of a life skills program, it is expected that project performance will largely depend on the level of life skills curriculum implementation, quality of implementation, and the individual teacher's competencies. Furthermore, children's levels of participation during classes could differ significantly, and this may influence the relationship between treatment and life skills and consequently outcome variables.
Provision of treatment was at the level of the classroom, so all children in grades 7 to 9 in program schools received treatment from all teachers in subjects taught to them.

[3] Detailed explanations of the scales used for measuring impact are provided in Section 3.5 of this report.

3.4. DATA ANALYSIS METHODS

3.4.1 MULTILEVEL MODEL

The data collected from children and their parents was hierarchical or clustered in structure. The multilevel nature of data within the model arose from a repeated measures design (two times points) and children's belonging to the same classroom, school and municipality. Since the students are not placed within teachers' classrooms randomly, and student and teacher level predictors incorporated into two separate models provide better estimates of variance and predictors' effects, the most appropriate choice is to measure the effects of teacher and student level predictors using the multilevel model (MLM). According to many researchers, MLM can provide an appropriate framework for this type of analysis (Goldstein, 1995; Morris & Normand, 1992; Raudenbush & Bryk, 2002; Subedi, 2005). In this way, we can analyze residual components at each level across the hierarchy. The proposed evaluation design model enables assessment of residuals at least at the subject and classroom levels. Additionally, residuals at the school level (models including additional clusters) may be tested to assess improvements in model fit.

Education literature is increasingly acknowledging that researchers have often ignored the fact that student outcomes are affected by complex, multidimensional, interrelated factors that interact with each other within organizational settings, the most basic of which is the school classroom (Rowe & Hill, 1998; Rowe et al., 1995). For example, researchers implementing interventions at the school or classroom level commonly measure and analyze the effect of treatments at the student level, overlooking the nested structure of the data. This results in the biasing of parameter estimates and standard error estimations of the parameters of interest (e.g., group-level effects) as well as an increase in Type I error rates (see McCoach, 2010; McCoach & Adelson, 2010), even in the presence of low levels of intra-cluster correlation (i.e., low ratios of between-group variance to total variance). In contrast, other researchers (e.g. see Goldstein, 1997) have used the classroom as the unit of analysis and classroom means as the outcome. However, classroom analyses suffer from aggregation bias and do not provide information on whether the treatment was differentially effective for individuals or different subgroups of children because the students' characteristics cannot be used as covariates. MLM is preferred over these two alternatives in the case of nested data, since the assumption of independence of cases is not necessary in a multilevel analysis, due to the fact that the probable dependence of students in the same classroom is explored explicitly by means of nested data (Aitkin & Longford, 1986; Raudenbush & Bryk, 1988), and the aggregation bias is avoided because the HLM technique allows investigation of the effects of similar phenomena at multiple levels of aggregation (e.g. individual students' socio-economic status and class average socio-economic status). Multi-level techniques allow analyses to be conducted simultaneously at multiple levels of data, thereby allowing the partitioning of variance into within- and between-classroom components. Curriculum interventions are arguably a natural classroom-level variable, and therefore it is important that the classroom context as well as individual characteristics be considered both within- and between-classrooms when a treatment is being implemented. Modeling effects using HLM

provides for an effect size estimate that is generally larger and more precise than the estimates from individual-level or classroom-level analyses (Raudenbush & Bryk, 1988).

MLM allows us to test a hypothesis about whether the intervention is constant across groups and, if not, which group variables explain differences in intervention effects. We can expect both school and classroom effects on students' outcomes. While schools were randomly assigned to the intervention, implementation was carried out at the classroom level. It seems very likely that some schools and teachers would have been more enthusiastic about the interventions than others, and this is likely to have had a direct effect upon the success of the intervention.[4] Consequently, we expect to see both between-school and within-school-between-classroom variation in the outcomes, even after accounting for baseline level differences.

The MLM model is a three-level nested model, hierarchical design of the quasi experiment, since entire clusters (time, classrooms and schools) are included in the treatment. Taking into consideration the multilevel data structure (Child – level 1, Classroom – level 2, School – level 3) we use MLM for analysis of data that will ensure that parameter estimates incorporate the effects of hierarchies. Furthermore, we assess only one classroom per school, so the number of classrooms (level 2) and schools (level 3) is equal (higher power of the model at level 2 and 3). Multilevel modelling ensures that covariances can differ across levels of analysis and slopes at level 1 become the dependent measure at level 2. The main independent variable of interest is the interaction between treatment and time. Since the classrooms and schools are included as fixed effects, we do not include treatment and time variables separately in the estimation of the MLM model, since they would be colinear with the fixed effects.

A possible alternative to multilevel modelling that we will explore is the fixed effects model with clusters integrated as dummy variables. However, the effects of group-level predictors would be confounded with the effects of the group dummies, i.e. it is not possible to separate out effects due to observed and unobserved group characteristics. In a multilevel (random effects) model, the effects of both types of characteristics can be estimated through the use of random and fixed effects, as long as this improves model power.

3.4.2 DIFFERENCE-IN-DIFFERENCE APPROACH

Since our preliminary results on interclass correlation indicated that ICC values were in all cases smaller than 10% multilevel and grouping between the data is not significant (see section 4.2. Impact Analysis, Exhibit 14), suggesting that an MLM approach may not be

[4] For example, during the monitoring visits, it was discovered that some teachers and schools organized trainings for other teachers in the school who did not participate in trainings organized by World Vision as part of the intervention.

necessary, we decided to use the difference-in-difference (DID) method in addition to the MLM approach.

The DID approach compares the before-after changes between children in the treatment group and matched children in the comparison group. It is important to note that the DID does not require baseline (pre-intervention) conditions to be the same in treatment and comparison groups, as our baseline results indicate that there are statistically significant average differences between treatment and comparison groups when it comes to the following outcomes: Making decisions in everyday life (MD); Social responsibility (CR); Critical thinking in everyday life (CT); Social relationships (SR) and Emotional management (EM).

The psychological and behavioral response of children increases over time in both treatment and comparison groups, but the assumption is that outcomes increase/decrease (depending of the type of outcome) more for the treated children after the Life Skills program is implemented.

Specifically, we estimated the following multivariate regression model for outcome variables:

$$Outcome = \alpha + \beta T + \gamma F + \delta (T \cdot F) + \lambda X + \varepsilon. \quad (1)$$

The left-hand side of the equation is the outcome variable of interest, i.e. psychological and behavioral response outcomes.

The variables on the right-hand side include the following:

- A dummy variable T that is equal to 1 if the observation is in the treatment group and zero if otherwise. The estimate of captures the group effect. In other words, controls for any differences in the outcome variable that are associated with being in the treatment group.
- A dummy variable F that is equal to 1 for post baseline period and zero for the baseline period. The estimate of captures the time effect. In other words, controls for any changes in the outcome variable that occur over time and are common for treatment and comparison group.
- An interaction term T·F that is equal to 1 if the observation is in the treatment group and in post baseline period, and zero otherwise. The estimate of captures the impact of the Life skills program on the outcome variable—this is the parameter of interest.
- A vector X of other relevant explanatory variables that may be related to the outcome of interest and will help control for baseline characteristics of children (gender, mother education, father education, mother employment status, father employment status, number of household members, monthly income, child educational attainment, classroom and school membership).

For each regression model, we estimated the parameters α, β, γ, δ, and the elements of the vector λ. All things being equal, the positive parameter estimates indicate that the corresponding right-hand side variable is associated with an increase in the outcome measure. Likewise, negative parameter estimates indicate a negative association.

For the DID to be valid, the comparison group must accurately represent the change in outcomes that would have been experienced by the treatment group in the absence of the intervention. In other words, the key identifying assumption behind the DID is that trends (changes) in outcomes between the treatment and comparison groups should be similar. This is known as a "common trend assumption". It means that, even when we control for the baseline differences between the treated and comparison units, differences in their trends between baseline and endline may make these two groups incomparable and, as a consequence, bias the results. Since the common trend is not necessarily assured through selection of treatment and comparison schools, and since we lack data from data points before the baseline that may have allowed us to test the common trend assumption, we decided to match individual pupils from the treatment and control schools on the baseline levels of the outcome and other observable variables in order to correct for confounding bias by balancing on variables that are different in the treatment and control groups. Results from the above-explained estimation approach are presented in comparative tables in Chapter 4.

3.5. DATA SAMPLE AND DATA SOURCE

3.5.1. SAMPLE

Method of sample selection

In order to reach evaluation objectives, a quasi-experimental repeated measure study design with a comparison group was proposed. A fully randomized experiment was not a feasible option, as the initiative partners (the ministries and World Vision) had already selected municipalities/schools where the life skills curriculum would be implemented. Bearing in mind that the main outputs of the life skills curriculum were to be integrated within the mainstream curriculum (targeting all municipalities, schools and children) key randomization characteristics had to be at the level of school classes (the lowest level of the cluster including information about Municipality and School) rather than on a solely individual level, which was not feasible due to the intervention having taken place at the classroom level.

Selection of treatment and comparison classes of children was based on propensity scores and nearest-neighbor matching with no replacement. Information required for propensity score matching was obtained from the Pedagogical Institute.

In order to minimize unobserved differences between preselected clusters (municipalities, schools, classes) and comparison group clusters, and to ensure the obtainment of unbiased estimates of the intervention effects, the comparison group needed to be as similar as possible to the treatment (intervention) group with regards to the intervention-relevant variables; characteristics of the clusters (school classes) that can significantly affect treatment processes (dynamics, content, frequency) and consequentially relationships between treatments and life skills outcome variables.

On the level of municipalities, the process of selection was carried out in such a way that the comparison classes were selected from schools in non-treatment municipalities that are in the vicinity of municipalities included in the program. 'Vicinity' was defined as adjoining municipalities (having a shared border).

To control for intervention-relevant variables at the level of schools, sample classes from treatment schools were randomly selected and matched one-to-one with school classes from comparison schools using the propensity score based on:

1. number of children in the school,
2. number of children per school class,
3. number of boys per class, and
4. number of girls per class.[5]

Construction of the treatment and comparison groups

For the purposes of construction of the sample, data was obtained from two sources: World Vision Bosnia and Herzegovina and the Pedagogical Institute of Tuzla Canton. World Vision provided data on the implementation of the Life Skills program in Tuzla Canton (treated classes, the school and municipality that a class belongs to, data on teachers who participated in the program). The Pedagogical Institute of Tuzla Canton provided data on all 8th grade classrooms in Tuzla Canton, the number of boys and girls per class, and the class's grade teacher. All data referred to the 2017/18 school year.

Based on to the World Vision administrative data, the Life Skills program was implemented in 31 classes from 13 program schools located in 5 municipalities throughout Tuzla Canton (Sapna, Lukavac, Doboj Istok, Kalesija and Banovići). In order to construct the treatment group, 20 classes were randomly selected from the population of the treated classes in Tuzla Canton. To construct the comparison group, the municipalities adjoining the above-mentioned program municipalities were first selected. According to the administrative data from the Pedagogical Institute of Tuzla Canton, there were 74 classes in 32 schools in the selected non-treatment municipalities.

Using propensity score matching, 20 randomly selected treatment classes were matched to 20 comparison classes.

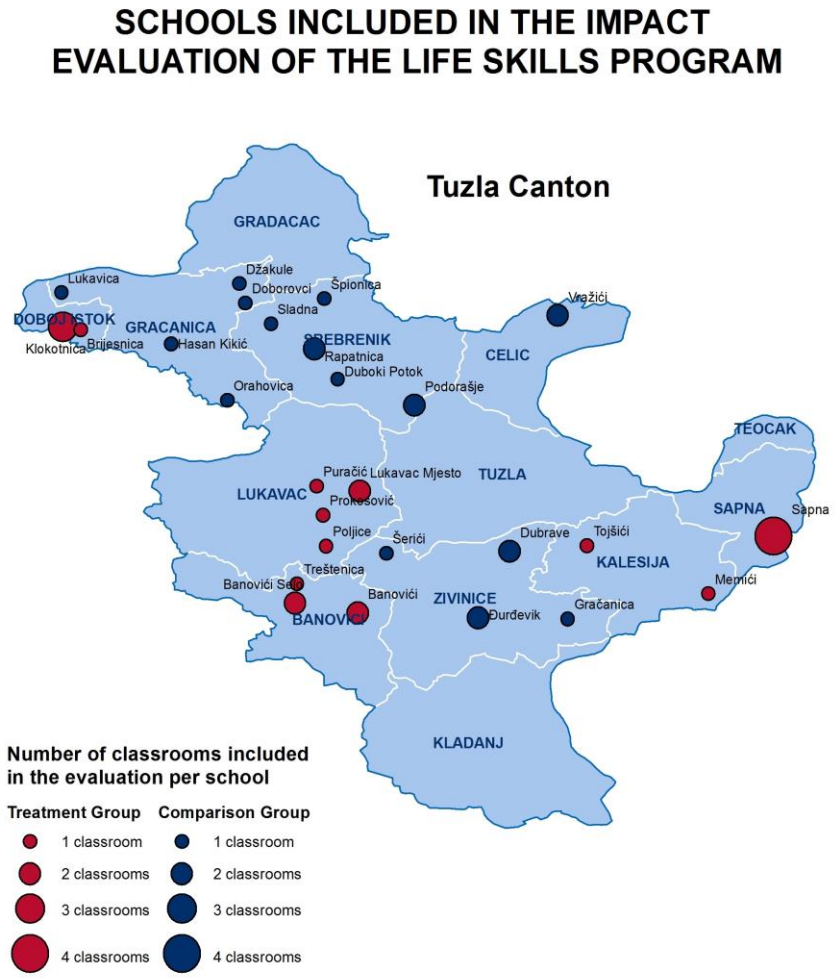
[5] According to the Work Plan, propensity score matching was supposed to additionally include the number of children with disabilities per school class and number of children from minority groups, e.g. Roma children. However, this data was not available.

Description of the sample

As explained above, 20 classes of 8th grade children from the treatment groups were included in the evaluation and compared against 20 classes of children from the comparison schools.

Overview of municipalities included in the evaluation is given in Exhibit 2.

Exhibit 2: Municipalities included in the impact evaluation of the Life Skills program



The list and distribution of treatment and comparison schools, as well as the information about the number of classrooms and pupils per school, are provided in Exhibit 3.

Exhibit 3: Distribution of treatment and comparison schools

No.	School	Municipality	Number of classrooms used in the	Group	Frequency	Percent

			evaluation per school		(number of pupils per school)	
1.	JU OŠ "Sapna"	Sapna	4	Treatment	72	8.8
2.	JU OŠ "Poljice"	Lukavac	1	Treatment	17	2.1
3.	JU OŠ "Puračić"	Lukavac	1	Treatment	22	2.7
4.	JU OŠ "Memići"	Kalesija	1	Treatment	22	2.7
5.	JU OŠ "Banovici Selo"	Banovići	2	Treatment	38	4.6
6.	JU OŠ "Brijesnica"	Doboj Istok	1	Treatment	22	2.7
7.	JU OŠ "Banovici" Banovići	Banovići	2	Treatment	51	6.2
8.	JU OŠ "Lukavac Mjesto"	Lukavac	2	Treatment	47	5.7
9.	JU OŠ "Tojšići"	Kalesija	1	Treatment	24	2.9
10.	JU OŠ "Klokotnica"	Doboj Istok	3	Treatment	44	5.4
11.	JU OŠ "Prokosovići"	Lukavac	1	Treatment	22	2.7
12.	JU OŠ "Treštenica"	Banovići	1	Treatment	19	2.3
13.	JU OŠ "Gračanica Živinice"	Živinice	1	Comparison	23	2.8
14.	JU OŠ "Džakule"	Gračanica	1	Comparison	30	3.6
15.	JU OŠ "Vražići"	Čelić	2	Comparison	31	3.8
16.	JU OŠ "Šerići"	Živninice	1	Comparison	24	2.9
17.	JU OŠ "Špionica"	Srebrenik	1	Comparison	27	3.3
18.	JU OŠ "Orahovica"	Gračanica	1	Comparison	18	2.2
19.	JU OŠ "Doborovci"	Gračanica	1	Comparison	19	2.3
20.	JU OŠ "Podorašje"	Srebrenik	2	Comparison	32	3.9
21.	JU OŠ "Duboki Potok"	Srebrenik	1	Comparison	23	2.8
22.	JU OŠ "Lukavica"	Gračanica	1	Comparison	17	2.1
23.	JU OŠ "Dubrave"	Živinice	2	Comparison	36	4.4
24.	JU OŠ "Đurđevik"	Živinice	2	Comparison	40	4.9
25.	JU OŠ "Hasan Kikić"	Gračanica	1	Comparison	24	2.9
26.	JU OŠ "Sladna"	Srebrenik	1	Comparison	20	2.4
27.	JU OŠ "Rapatnica	Srebrenik	2	Comparison	58	7.1
	Total	-	40	-	822	100.0

As can be seen from the above table, the initial sample consisted of a total 822 children from different schools across Tuzla Canton.

Overall, in the first cluster nest there are 822 children and 822 parents. In the second cluster there are 40 classrooms (20 treated classrooms and 20 comparison classrooms). In the third cluster, there are 27 schools (12 treated schools and 15 comparison schools). In the final (fourth) cluster, there are 9 municipalities (5 treated and 4 comparison).

3.5.2. DATA SOURCES

To address the evaluation questions, data was collected on-site in the treatment and comparison schools. Data was collected from children, parents, grade teachers and teachers who participated in the program. In the following section, we will present a brief overview of instruments used to collect primary data.

Data from children and parents was collected by the grade teachers and/or school pedagogues that were working with selected classes in selected schools. Before collection of data, training with grade teachers and/or school pedagogues was conducted. During the training, grade teachers:

- were introduced to the Child Protection Policy, requiring protection of children's privacy and its relationship the Ethical Code of Conduct during research with or about children in BiH, and then signed the Policy;
- became acquainted with the prepared instruments;
- came to understand that a parent's written permission was mandatory for a child's participation in the research;
- filled out instruments in the role of a teacher, parent and child, in order to ensure that all questions in the questionnaires / scales could be understood;
- were introduced to protocols for collection of data and its entry into the electronic system;
- were introduced to forms for data collection that were to be completed by themselves using the school's administrative data.

The ethical principles for this research were based on The Code of Ethical Research with Children and on Children in Bosnia and Herzegovina (2014) and the Child Protection Policy. All members of the research team read both sets of guidelines and signed a statement of agreement to comply with these during preparation of the research, during the research itself, and after the research had been conducted, in the phase of writing reports or publications stemming from the research. Parents of all children participating in the research had to provide written consent for participation of their child in the research only after receiving information on: (a) the specific purpose of the research; (b) benefits of the research; (c) possible risks; (d) type and duration of the research; (e) confidentiality of data derived from the research and protection of identity of all respondents; (f) implications the research may have; (g) voluntary consent and withdrawal from research. When described to children, the information was presented in a less technical format, using simple vocabulary. Parents were informed that while information about their child (identifiers) would be collected in school, and would subsequently be coded in all questionnaires. Information about the identifiers and their associated codes was to be stored separately in an encrypted file, to which only key evaluators had access.

3.6. DATA COLLECTION INSTRUMENTS

3.6.1 DESCRIPTION OF INSTRUMENTS AND VARIABLES

In order to assess achievement and attribution of any change to the implementation of the life skills curriculum, specific instruments for performance evaluation and impact evaluation were designed and applied. An overview of these instruments is given below.

Exhibit 4: Overview of performance and impact evaluation tools

Evaluation	*Instrument*	*Respondents*
Performance evaluation	*Child Participation Evaluation Form*	Children
	Life Skills Teacher's Competences Form	Life Skills Teachers
	Teacher's Assessment of the Life Skills Program	
Impact evaluation	*Life Skills Scale*	Children
	• Making decisions in everyday life	
	• Critical thinking in everyday life	
	• Communication scale	
	• Interpersonal relations skills	
	• Coping with emotions	
	• Social Responsibility Scale	
	Rosenberg Self-Esteem Scale	
	Violence Manifestation and Exposure Questionnaire	
	The Developmental Assets Profile	
	Cognitive Behavior Checklist	Parents
	Sociodemographic Questionnaire	
	Child Educational Attainment	Grade Teachers

These instruments are presented in Annexes 1 and 2.

A brief description of each instrument for impact evaluation is given below.

Life Skills Scale

The proposed Life Skills Questionnaire measures 6 skills:

(1) problem solving/decision making,
(2) critical thinking,
(3) communication skills,
(4) interpersonal relations skills,
(5) coping with emotions,
(6) social responsibility.

A description of these scales is given in Exhibit 5.

Exhibit 5: Description of the Life Skills Constructs

Construct	Instrument	Description
PROBLEM SOLVING/DECISION MAKING	Making Decisions in Everyday Life (MDEL); Source: Youth Life Skills Evaluation Project at Penn State.	This 20-item scale examines the frequency of use decision-making skills (defining the problem,

	Instrument also cited by the CYFAR Life Skills Project at Texas A&M University; Developers: C. C. Mincemoyer and D. F. Perkins; Year: 2001.	identifying alternatives, risks, and consequences, selecting an alternative, evaluation); Target Audience(s): Youth aged 12 to 18; Internal consistency: .89
CRITICAL THINKING	Critical thinking in everyday life (CTEL); Source: Youth Life Skills Evaluation Project at Penn State. Instrument also cited by the CYFAR Life Skills Project at Texas A&M University; Developers: C. Mincemoyer, C. Perkins, D.F. and C. Munyua; Year: 2001.	This 20-item scale examines frequency of use of the following skills: reasoning, enquiry, analysis/information processing, flexibility, evaluation; Target Audience(s): Youth aged 12 to 18; Internal consistency: .72
COMMUNICATION SKILLS	Communication scale (CS); Source: Youth Life Skills Evaluation Project at Penn State. Instrument also cited by the CYFAR Life Skills Project at Texas A&M University; Developers: S. Barkman and K. Machtmes; Year: 2002	This 23-item scale assess youth's ability to communicate by examining the frequency of use of the various skills that are needed to use effective communication practices; Target Audience(s): Youth aged 12 to 18; Internal consistency: .79
INTERPERSONAL RELATIONS SKILLS	Social relationships scale (SRS); Source: Quasi-experimental Evaluation Research Report on Life skills Model, The Regional Education Action Learning (REAL) Program, World Vision Bosnia and Herzegovina; Year: 2016	This 8-item scale examines frequency of use of the various skills that are needed to manage effective interpersonal relationships; Target Audience(s): Youth aged 10 to 16; Internal consistency: .77
COPING WITH EMOTIONS	Emotional management scale (EMS); Source: Quasi-experimental Evaluation Research Report on Life Skills Model, The Regional Education Action Learning (REAL) Program, World Vision Bosnia and Herzegovina; Year: 2016	This 6-item scale assess emotional management skills; Target Audience(s): Youth aged 10 to 16; Internal consistency: .72
SOCIAL RESPONSABILITY	Social responsibility scale (SRS); Source: Quasi-experimental Evaluation Research Report on Life Skills Model, The Regional Education Action Learning (REAL) Program, World Vision Bosnia and Herzegovina; Year: 2016	This 7-item scale focuses on assessing social responsibility; Target Audience(s): Youth aged 10 to 16; Internal consistency: .78

In order to assure that Life Skill scales are appropriate for 8th and 9th grade children, the questionnaire were pilot tested.[6]

[6] In order to construct Life Skills Scale, an in-depth analysis of scales used in previous researches was conducted. Items and factors from different scales were extracted and compared with the intervention logic of the project. All items that had a direct or indirect logical connection with the program intervention logic

Rosenberg Self-Esteem Scale

Rosenberg Self-Esteem Scale – This is a tool for assessing global self-esteem. Ten statements are included in the self-reporting measure that pertain to self-worth and self-acceptance. These are measured using a four-point scale ranging from "strongly agree" to "strongly disagree". The items were selected as a Guttman scale with 7 "contrived items". The Rosenberg Self-Esteem Scale presented high ratings in reliability areas; internal consistency was 0.77, minimum Coefficient of Reproducibility was at least 0.90 (Rosenberg, M., 1965, 1987). A varied selection of independent studies – each using samples ranging between parents, men over 60, high school students, and civil servants – showed alpha coefficients ranging from 0.72 to 0.87 (Hatcher, Jennifer; Lynne & Hall, 2009).

The Developmental Assets Profile

The Developmental Assets Profile (DAP) uses a 58-item questionnaire as a standard method of eliciting and quantifying information on the Developmental Assets as seen by the adolescents themselves. The same DAP is used for boys and girls ages 11 to 18. For the purpose of this impact evaluation, we employed Asset View as way of scoring and portraying reported assets. Asset View portrays scores on eight different asset categories: Support, Empowerment, Boundaries and Expectations, Constructive Use of Time, Commitment to Learning, Positive Values, Social Competencies and Positive Identity (Search Institute, 2005).

Cognitive Behavioral Check List (CBCL)

The Cognitive Behavioral Check List (CBCL) offers a comprehensive approach to assessing adaptive and maladaptive functioning (standardized test). Developed through decades of research and practical experience to identify actual patterns of functioning, providing professionals tools to measure anxiety and depression [AD], somatic complaints [SC], social problems [SP], thought problems [TP], attention problems [AP], rule-breaking behavior [RBB], aggressive behavior [AB] with children up to the age of 18. The Alpha Cronbach coefficient of checklist reliability varies from 0.82-0.92 for different factors (Achenbach, T.M. & Rescorla, L.A., 2001). The questionnaire developed additionally integrates items measuring difficulties in learning and behavior [MA], relationship with peers [MB], relationship between child and teachers [MC], coping with problems and locus of control [MD], social anxiety [ME] and hyperactive behavior [MG]. Additionally, it includes extracts of factors from the Matrix for Early Recognition of Rule Breaking Behavior Risk for Children that is being piloted in Canton Sarajevo and Republika Srpska as of the 2017/18 academic year by the Criminal Policy Research Center in cooperation with the respective Ministries of Education.

---

became a constituent part of proposed Life Skills Scale. The psychometric characteristics of the Initial Life Skills Scale were assessed during the pilot test, looking into the construct validity of the scale (number of orthogonal or covariate factors extracted, communality and explanation of the scale items by the extracted factors) and reliability of the factors (Alpha and Omega). Based on the pilot test results, items with a strong relationship to the extracted interpretable factors were retained, while others were revised and/or removed from the scale. A summary of the instruments' pilot run is presented in Annex 3.

Child Educational Attainment

The Academic checklist – was completed by the grade teacher of each child participating in the evaluation. It contained information about academic achievement: average school grades (or even mark per subject), grade related to school (mis)conduct, number of justified absences during the school year, number of unjustified absences during the school year, existence of disciplinary measures (reprimand, changing school classes, changing schools), reported occurrences of violence. The academic checklist also contained information about a child's participation during classes.

Sociodemographic Questionnaire

Additional **socio-economic data** about children was collected: gender, age, disability status, average household income and number of household members. This questionnaire was completed by the parents.

As per the Code of Ethical Research with Children and on Children in Bosnia and Herzegovina (2014) written consent by parents was a prerequisite for the children's participation. Before collecting data from children, parents' meetings will be organized, where consent letters were to be collected. Only children with a written consent letter were allowed to enter the room where data collection from children was to take place.

The following table gives an overview of main variables used in the analysis.

Exhibit 6: Overview of variables

Scale	Variable name	Label
Life Skills Scale	MD	Making decisions in everyday life
	CR	Social responsibility scale
	CT	Critical thinking in everyday life
	SR	Social relationships scale
	CS	Communication scale
	EM	Emotional management scale
Cognitive-Behavior Check List	AP	Attention problems
	SP	Social problems
	TP	Thought problems
	AB	Aggressive behavior
	MB	Relationship with peers
	MC	Relationship between child and teachers
	CBCL_MD	Coping with problems and locus of control
	ME	Social anxiety
	MG	Hyperactive behavior
	MA	Difficulties in learning and behavior
	SC	Somatic complaints
	RBB	Rule-breaking behavior
	AD	Anxiety and depression
Rosenberg self-esteem scale	ROS	Rosenberg self-esteem
Violence Manifestation and Exposure Questionnaire	IZL	Violence exposure
	ISP	Violence manifestation
The Developmental Assets Profile	DAP_SUP1	Support

	DAP_EMP2	Empowerment
	DAP_BE3	Boundaries and expectations
	DAP_CUT4	Constructive use of time
	DAP_CL5	Commitment to learning
	DAP_PV6	Positive values
	DAP_SC7	Social competencies
	DAP_PI8	Positive identity
	DAP_EAC	External asset side (SUP1+EMP2+BE3+CUT4)
	DAP_IAC	Internal asset side (CL5+PV6+SC7+PI8)
	DAP_TOTAL	Total

Estimation of the four different models, as explained above, was performed on each of the alternative dependent variables presented in the table above. Results are presented in comparative tables in Chapter 4.

3.6.2 INTERNAL CONSISTENCY

Before going any further with the analysis of the baseline results, it is necessary to conduct a reliability analysis, i.e. to examine the reliability of the instruments employed. To see how well the selected instruments measure what they really should, we used Cronbach's alpha, α (or *coefficient alpha*), which is a measure of reliability that ranges from 0 to 1, with values of .60 to .70 deemed the lower limit of acceptability (Hair, Black, Babin & Anderson, 2014). Cronbach's alpha is sensitive to the number of items in a scale. So, a larger number of items can result in a larger – and a smaller number of items a smaller – Cronbach's alpha. The results are presented in Exhibit 7.

Exhibit 7: Internal consistency

	Scale	**PRE**	**POST**
Subconstruct	**Life Skills Scale**	**Cronbach's α**	
MD	Making Decisions in Everyday Life	0.91	0.93
CR	Social responsibility scale	0.77	0.78
CT	Critical thinking in everyday life	0.93	0.94
SR	Social relationships scale	0.81	0.84
CS	Communication scale	0.90	0.92
EM	Emotional management scale	0.59	0.61
Subconstruct	***Cognitive-Behavior Check List***		
AP	Attention problems	0.74	0.76
SP	Social problems	0.65	0.67
TP	Thought problems	0.59	0.58
AB	Aggressive behavior	0.81	0.81
MB	Relationship with peers	0.59	0.61
MC	Relationship between child and teachers	0.48	0.62
CBCL_MD	Coping with problems and locus of control	0.41	0.45
ME	Social anxiety	0.61	0.59
MG	Hyperactive behavior	0.70	0.67
MA	Difficulties in learning and behavior	0.51	0.53

SC	Somatic complaints	0.72	0.74
RBB	Rule-breaking behavior	0.51	0.64
AD	Anxiety and depression	0.78	0.77
Construct	**Rosenberg self-esteem scale**		
ROS	Rosenberg self-esteem	0.78	0.81
Subconstruct	**Violence Manifestation and Exposure Questionnaire**		
IZL	Violence exposure	0.87	0.86
ISP	Violence manifestation	0.86	0.84
Subconstruct	***The Developmental Assets Profile***		
DAP_SUP1	Support	0.72	0.73
DAP_EMP2	Empowerment	0.66	0.70
DAP_BE3	Boundaries and Expectations	0.73	0.75
DAP_CUT4	Constructive Use of Time	0.37	0.41
DAP_CL5	Commitment To Learning	0.78	0.81
DAP_PV6	Positive Values	0.79	0.81
DAP_SC7	Social Competencies	0.73	0.72
DAP_PI8	Positive Identity	0.68	0.71
DAP_EAC	External asset side (SUP1+EMP2+BE3+CUT4)	0.79	0.79
DAP_IAC	Internal asset side (CL5+PV6+SC7+PI8)	0.72	0.86
DAP_TOTAL	Total	0.66	0.83

Source: Own calculations using Stata 14

As can be seen from the Cronbach's alpha results presented in Exhibit 7, the reliability of eight instruments is below the lower level of acceptability: Emotional management scale (α=0.59); Thought problems (α=0.59); Relationship with peers (α=0.59); Relationship between child and teachers (α=0.48); Coping with problems and locus of control (α=0.41); Difficulties in learning and behavior (α=0.51); Rule-breaking behavior (α=0.51) and Constructive Use of Time (α=0.37)). Cronbach's alpha is sensitive to the number of items in an instrument. Since Constructive Use of Time has only four items, the result of a low Cronbach's alpha does not come as a surprise.

3.7. EVALUATION DESIGN LIMITATIONS

The key methodological limitation of the evaluation design is related to its quasi-experimental design when compared to a full experimental design. Drawbacks are minimized by controlling for initial differences between treatment and comparison groups using propensity score matching and inverse probability weighting.

An additional methodological limitation comes from Theory of Change – whereby it is expected that life skills affect social change indicators – where it is possible that the level of social change can act as a significant confounding variable influencing the relationship

between treatment and life skills. The evaluation design furthermore takes into account the fact that some natural maturing and developmental processes can produce considerable change independently of the program. In order to ensure a reliable link between progress and impact, these maturing-related and developmental processes have to additionally be taken into consideration by the evaluation team.

The Theory of Change, when specifically applied to the Life Skills program, envisages long-term changes through reduction of the probability of a child coming in conflict with the law, increased labor skills, increased employment rate and improved chances of continued education; characteristics which are not possible to follow during the evaluation timeframe. These would, however, form a good basis for the development of follow-up evaluation project(s) in the future.
Some of the short-term outcomes, such as drop-out, and some of the confounding variables such as disability status of the children and their belonging to minority groups are rare events, and the evaluation methodology lacks a sample size capable of tracking the progress or influence of such rare events.

Estimated produced from data about parents and data about children may significantly differ. Interclass reliability coefficients will be calculated in order to determine differences between estimates. Separately, a regression model will be calculated for the estimates provided by children and parents.

Information about classroom, school, municipality, level of program implementation and teacher competence will only be available at the cluster level, and therefore multilevel regression analysis will have to be used to estimate the effect of the treatment on the program outcome.

4. FINDINGS

4.1. DATA ANALYSIS

This chapter of the report contains the results of baseline data analysis, i.e. the results of descriptive statistics analysis; results of reliability analysis and results of the average treatment effect at the baseline level.

4.1.1. DESCRIPTIVE STATISTICS

In this section, we first present baseline differences between the treatment and comparison children and households in the context of their demographic and other characteristics used as independent variables in the model. Following this, we present results of matching the two groups on the values of their baseline characteristics. Finally, we present baseline differences between the treatment and comparison children in the values of outcome characteristics used as dependent variables in the model.

Baseline differences between treatment and comparison group in independent variables

The following two tables give a brief overview of the household and individual socio-demographic characteristics of children included in the sample.

Exhibit 8: Household characteristics

Characteristic	**Group**			
	Treatment		**Comparison**	
Average total monthly income in the family				
	Freq.	Percentage	Freq.	Percentage
Up to 300 BAM	54	13.8	51	12.7
From 301 to 800 BAM	156	40.0	131	32.6
From 801 to 1600 BAM	145	37.2	170	42.3
From 1601 to 3000 BAM	32	8.2	41	10.2
Above 3000 BAM	3	0.8	9	2.2
Total	390	100.0	402	100.0
Household size				
	Mean	St. Dev.	Mean	St. Dev.
Average number of household members	4.27	0.98	4.46	1.09
Mother's education				
	Freq.	Percentage	Freq.	Percentage
No school	8	2.0	12	2.9
Primary school	153	38.3	181	43.6
Secondary school	206	51.5	204	49.2
Higher school	11	2.8	5	1.2
Faculty (4 or 5 years of education)	21	5.3	11	2.7
Master of science or PhD	1	0.3	2	0.5
Total	400	100.0	415	100.0
Father's education				

	Freq.	Percentage	Freq.	Percentage
No school	2	0.5	9	2.2
Primary school	61	15.3	78	18.8
Secondary school	293	73.4	302	72.6
Higher school	21	5.3	15	3.6
Faculty (4 or 5 years of education	17	4.3	10	2.4
Master of science or PhD	5	1.3	2	0.5
Total	399	100.0	416	100.0
Mother's employment status				
Employed	95	23.8	92	22.3
Temporarily employed	16	4.0	21	5.1
Unemployed	285	71.3	298	72.2
Retired	3	0.8	0	0.0
Other	1	0.3	2	0.5
Total	400	100.0	413	100.0
Father's employment status				
Employed	278	70.2	307	74.3
Temporarily employed	23	5.8	26	6.3
Unemployed	70	17.7	66	16.0
Retired	20	5.1	12	2.9
Other	5	1.3	2	0.5
Total	396	100.0	413	100.0

Source: Own calculations using Stata 14

Exhibit 9: Individual characteristics of the children included in the sample

Characteristic	**Group**			
	Treatment		**Comparison**	
Gender	Freq.	Percent	Freq.	Percent
Male	195	48.8	206	49.4
Female	205	51.3	211	50.6
Total	400	100.0	417	100.0
Family situation of child				
	Freq.	Percent	Freq.	Percent
Child lives with both parents	373	93.3	390	93.3
Child lives with one parent (parents are divorced)	22	5.5	18	4.3
Child lives with one parent (one parent deceased)	4	1.0	6	1.4
Child lives in alternative care institutions (foster care, institution)	0	0.0	2	0.5
Child lives with the relative (grandparents, cousins, etc.)	1	0.3	2	0.5
Total	400	100.0	418	100.0
Developmental difficulties of child	Freq.	Percent	Freq.	Percent
Child's development was as expected, equal as his/her peers	354	97.0	391	97.8
Child is having developmental difficulties (diagnosed - categorized)	7	1.9	8	2.0
Child is not categorized but we suspect they have some developmental difficulties	4	1.1	1	0.3
Total	365	100.0	400	100.0
Child educational attainment				
	Mean	St. Dev.	Mean	St. Dev.
Average school mark at the end of 2016/17	3.56	1.28	3.58	1.21

Behavior mark at the end of 2016/17	4.96	0.27	4.88	0.44
Total # of absences at the end of 2016/17	17.77	21.72	16.23	21.12
Total # of unjustified absences at the end of 2016/17	0.62	3.45	0.53	2.78
Total # of disciplinary measures imposed on a child in the school at the end of 2016/17	0.08	0.38	0.12	0.38

Source: Own calculations using Stata 14

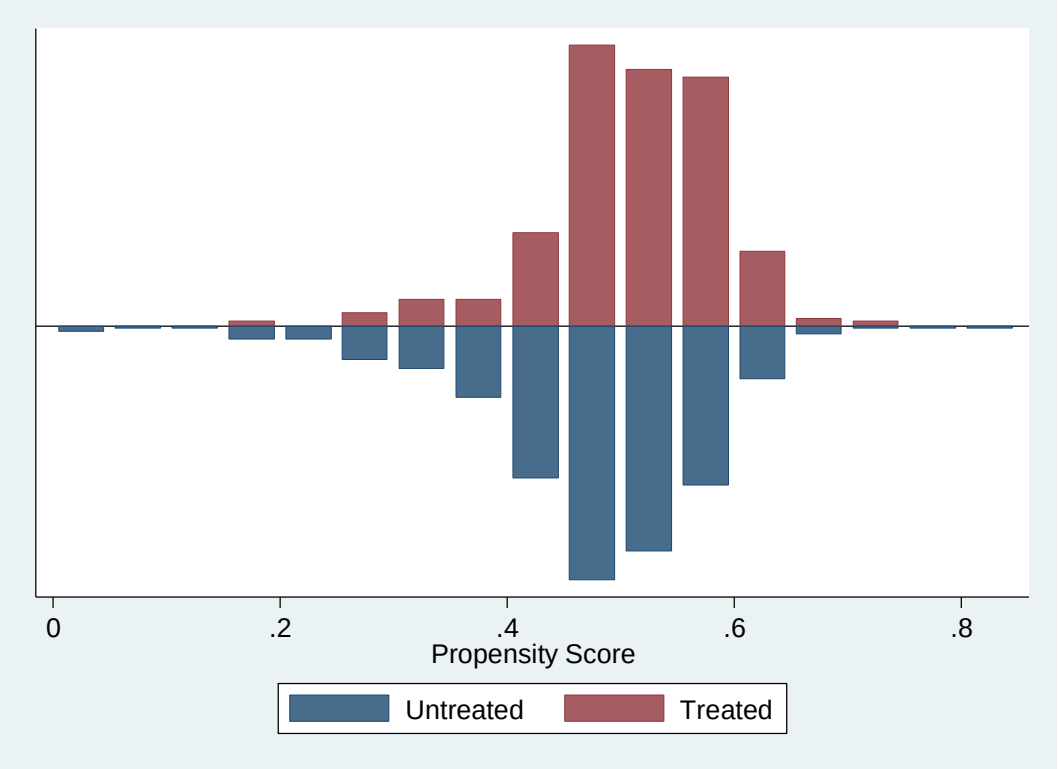
4.1.2 BALANCE OF COVARIATES AFTER WEIGHTING THE SAMPLE BY A PROPENSITY SCORE

In addition to propensity score matching for schools, we also tested balancing of the sample on the level of the first cluster (level of children). The inverse probability of treatment weights with the propensity score was used for balancing. The following covariates were used for balancing after weighting the sample by a propensity score:

- Gender;
- Household size;
- Average total monthly household income;
- Child's educational attainment.

The overlap of the distribution of the propensity scores across treatment and comparison groups is displayed in Exhibit 10.

Exhibit 10: Distribution of Propensity Score across Treatment and Comparison



Source: Own calculations using Stata 14

The results of balance in covariates across treatment and comparison groups after using inverse probability of treatment weights with the propensity score are presented in Exhibit 11.

Exhibit 11: Balance in covariates across treatment and comparison groups after using inverse probability of treatment weights with the propensity score

Covariate	**Mean in treated**	**Mean in untreated**	**Standardized difference**
Gender			
Boy	0.49	0.49	-0.01
Girl	0.51	0.51	0.01
Household size	4.37	4.36	0.01
Average total household monthly income			
Up to 300 BAM	0.14	0.14	0.00
From 301 to 800 BAM	0.36	0.36	-0.01
From 801 to 1600 BAM	0.39	0.39	0.00
From 1601 to 3000 BAM	0.09	0.09	0.01
Above 3000 BAM	0.01	0.02	-0.01
Child's educational attainment			
Average school mark at the end of 2016/17	3.59	3.57	0.02
Behavior mark at the end of 2016/17	4.93	4.92	0.02

Total # of absences at the end of 2016/17	16.88	17.06	-0.01
Total # of unjustified absences at the end of 2016/17	0.68	0.80	-0.04
Total # of disciplinary measures imposed on a child in a school at the end of 2016/17	0.10	0.10	-0.00

Source: Own calculations using Stata 14

The results show that the treated and comparison groups differ by less than 1 standard deviation. The densities of individual continuous covariates in the comparison and treatment groups are presented in Annex 5.

4.1.3 BASELINE DIFFERENCES BETWEEN TREATMENT AND COMPARISON GROUP IN DEPENDENT VARIABLES

Before presenting other results, we provide the summary statistics of the baseline characteristics for the treatment and comparison groups in Exhibit 12.

Exhibit 12: Basic descriptive analysis of the constructs and subconstructs

	Scale	**PRE**				**POST**			
		Group				**Group**			
		Treatment		**Comparison**		**Treatment**		**Comparison**	
Subconstruct	**Life Skills Scale**	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
MD	Making decisions in everyday life	3.89	.59	3.73	.62	3.89	.64	3.84	.62
CR	Social responsibility scale	4.03	.61	3.92	.64	3.95	.62	3.92	.61
CT	Critical thinking in everyday life	3.93	.64	3.79	.65	3.92	.65	3.88	.66
SR	Social relationships scale	4.22	.56	4.11	.61	4.12	.62	4.20	.59
CS	Communication scale	3.61	.55	3.54	.55	3.62	.58	3.62	.59
EM	Emotional management scale	3.37	.52	3.26	.53	3.38	.54	3.38	.56
Subconstruct	***Cognitive-Behavior Checklist***	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
AP	Attention problems	.31	.29	.34	.30	.26	.28	.28	.27
SP	Social problems	.16	.19	.17	.21	.13	.18	.16	.19
TP	Thought problems	.11	.14	.11	.14	.09	.14	.10	.13
AB	Aggressive behavior	.14	.18	.17	.20	.12	.19	.13	.16
MB	Relationship with peers	.14	.23	.14	.24	.12	.22	.12	.21
MC	Relationship between child and teachers	.09	.21	.09	.22	.08	.23	.08	.20
CBCL_MD	Coping with problems and locus of control	.20	.25	.20	.26	.16	.25	.17	.24
ME	Social anxiety	.19	.26	.20	.27	.17	.24	.16	.23
MG	Hyperactive behavior	.27	.38	.31	.44	.21	.36	.22	.35
MA	Difficulties in learning and behavior	.18	.26	.21	.29	.15	.25	.19	.28
SC	Somatic complaints	.15	.21	.16	.20	.12	.19	.16	.20
RBB	Rule-breaking behavior	.08	.10	.09	.11	.081	.11	.09	.12
AD	Anxiety and depression	.25	.25	.27	.26	.21	.24	.24	.24

Construct	**Rosenberg self-esteem scale**	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
ROS	Rosenberg self-esteem	3.12	.47	3.07	.45	3.15	.50	3.18	.45
Subconstruct	**Violence Manifestation and Exposure Questionnaire**	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
IZL	Violence exposure	.41	.41	.37	.38	.37	.39	.30	.36
ISP	Violence manifestation	.28	.34	.29	.35	.27	.32	.22	.29
Subconstruct	***The Developmental Assets Profile***	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
DAP_SUP1	Support	24.63	4.50	24.07	4.83	24.04	4.83	24.30	4.54
DAP_EMP2	Empowerment	22.47	4.68	21.88	4.95	22.66	4.84	22.64	4.78
DAP_BE3	Boundaries and expectations	23.29	4.45	23.06	4.82	22.99	4.52	23.55	4.69
DAP_CUT4	Constructive use of time	18.44	6.28	18.00	6.37	17.72	6.38	17.98	6.21
DAP_CL5	Commitment to learning	19.78	5.57	19.08	5.59	19.96	5.60	19.89	5.49
DAP_PV6	Positive values	24.10	4.15	23.74	4.12	24.24	3.88	24.00	4.23
DAP_SC7	Social competencies	22.53	4.81	22.00	4.69	22.42	4.38	22.72	4.69
DAP_PI8	Positive identity	21.51	5.06	20.87	4.91	21.92	4.65	22.05	5.07
DAP_EAC	External asset side SUP1+EMP2+BE3+CUT4	222.33	38.90	217.75	41.79	218.79	39.41	221.380	41.38
DAP_IAC	Internal asset side (CL5+PV6+SC7+PI8)	220.01	41.99	214.44	40.27	221.59	38.29	221.94	41.63
DAP_TOTAL	Total	442.34	74.43	432.19	77.00	440.38	70.74	443.32	77.61

Source: Own calculations using Stata 14

It can be seen that, at the baseline level, children from the treatment group, on average, have higher mean scores across all constructs.

More precisely, children from the treatment group have, on average, better life skills mean scores compared to those from the comparison group.

At the baseline level, when controlling for gender, household size, average total monthly household income, and child educational attainment, we found evidence of statistically significant differences between treatment and comparison groups in the following outcomes: Making decisions in everyday life (MD); Social responsibility (CR); Critical thinking in everyday life (CT); Social relationships (SR); and Emotional management (EM). This means that the schools selected for treatment initially performed better with regard to decision making skills (MD) by an average of 0.13, compared to the average score of 3.73 for children in the comparison group. Similarly, social responsibility skills (CR) scores in treatment schools were, on average, 0.09 better than the 3.93 average recorded for children from the comparison group. For children in treatment schools, the average critical thinking (CT) score was higher by an average of 0.12 than the average of 3.79 within the comparison group. Treatment schools were also found to have social relationships (SR) scores an average of 0.09 higher than for children in the comparison group, where this score was 4.12. Finally, the emotional management (EM) scores of the children in treatment schools at the baseline was, on average, 0.11 higher than the average of 3.25 of children from the comparison group.

4.2. IMPACT ANALYSIS

4.2.1 RESULTS FROM THE MULTILEVEL MODEL

The impact analysis is conducted in order to detect potential statistically significant differences in outcomes between children from the treatment and comparison groups, when controlling for the following covariates: gender; household size; average total monthly household income; and child educational attainment.

Before applying the multilevel model approach to our data, we must first conduct preliminary testing to determine whether or not a multilevel model is necessary at all. Specifically, this preliminary testing is focused on the so-called Interclass Correlation (ICC), which depicts the ratio of the between-cluster variance to the total variance. The ICC shows the proportion of the total variance in a depended-on variable that is accounted for by the clustering. According to commonly used rule-of-thumb, ICC values smaller than 10% suggest that multilevel and grouping between the data are not significant, and therefore the use of simple regression is preferred to multilevel modeling. ICC results for all constructs used in this evaluation are presented in the following table.

Exhibit 13: Interclass correlations

	INTERCLASS CORRELATIONS			
CONCSTRUCTS	2 level null model	2 level null model	3 level null model	
	L1-Child, L2-School	L1-Child, L2-Classroom	Level 2 ICC	Level 3 ICC
AP	0.021	0.009	0.077	0.008
SP	0.021	0.017	0.021	0.018
TP	0.014	N/A	N/A	N/A
AB	0.021	0.014	0.020	0.014
MB	0.004	0.008	0.004	0.008
MC	0.022	0.023	0.023	0.025
CBCL_MD	0.010	N/A	N/A	N/A
ME	0.021	0.004	0.022	0.004
MG	0.009	0.005	0.009	0.006
MA	0.020	0.012	0.022	0.015
WD	0.022	0.011	0.021	0.010
MD	0.039	0.010	0.040	0.011
RBB	0.019	0.006	0.018	0.045
AD	0.021	0.004	0.021	0.005
SC	0.021	0.005	0.022	0.007
CR	0.031	N/A	N/A	N/A
CT	0.050	0.011	0.050	0.009
SR	0.035	0.000	N/A	N/A

CS	0.037	0.007	0.037	0.006
EM	0.056	0.022	0.054	0.015
ROS	0.058	0.007	0.058	0.007
IZL	0.033	0.012	0.032	0.010
ISP	0.055	0.025	0.056	0.024
DAP_SUP1	0.028	0.005	0.028	0.006
DAP_EMP2	0.029	0.001	0.029	0.003
DAP_BE3	0.031	N/A	N/A	N/A
DAP_CUT4	0.033	N/A	N/A	N/A
DAP_CLP5	0.059	0.009	0.058	0.005
DAP_PV6	0.041	N/A	N/A	N/A
DAP_SC7	0.030	0.000	0.031	0.003
DAP_PI8	0.053	N/A	N/A	N/A
DAP_IAC	0.058	N/A	N/A	N/A
DAP_EAC	0.032	0.006	0.032	0.006
DAP_TOTAL	0.044	0.000	0.044	0.001

Source: Own calculations using Stata 14

Although the ICC values were smaller than 10% in all cases, suggesting that use of a single-level model is preferred to multilevel modeling, we still decided to estimate the model as a multilevel model, since the above test is not necessarily sufficient in and of itself in allowing an ultimate decision about the appropriateness of this approach. Later, we will estimate the model using a difference-in-difference (DID) approach, and the results from these two models will be compared in the final section of this chapter.

Since we have 33 different dependent variables, presenting results with estimated values of all coefficients from each regression would create confusing and unreadable tables. For the sake of clarity and focus, we present only the values and statistical significances of coefficients for the interaction between time and treatment from each regression result. Such summary results of the multilevel model approach are presented in the table below. The direction of expected influence, assuming that the intervention should result in improvements, is also indicated in the table for each outcome variable.

Exhibit 14. Estimated coefficient for the interaction between treatment and time (DID) from the multilevel model

Dependent variable			Coeff. for did	p value
Sub-construct	**Life Skills Scale**			
MD	Making decisions in everyday life	positive	-0.012	0.777
CR	Social responsibility scale	positive	-0.083	0.041
CT	Critical thinking in everyday life	positive	-0.032	0.468
SR	Social relationships scale	positive	-0.109	0.007
CS	Communication scale	positive	-0.015	0.689

EM	Emotional management scale	positive	-0.021	0.572
Sub-construct	**Cognitive Behavior Checklist**			
AP	Attention problems	negative	-0.010	0.561
SP	Social problems	negative	-0.014	0.245
TP	Thought problems	negative	-0.001	0.950
AB	Aggressive behavior	negative	0.000	0.975
MB	Relationship with peers	negative	-0.007	0.632
MC	Relationship between child and teachers	negative	0.019	0.130
CBCL_MD	Coping with problems and locus of control	negative	-0.015	0.344
ME	Social anxiety	negative	-0.014	0.384
MG	Hyperactive behavior	negative	-0.022	0.355
MA	Difficulties in learning and behavior	negative	-0.008	0.596
SC	Somatic complaints	negative	-0.026	0.043
RBB	Rule-breaking behavior	negative	0.010	0.166
AD	Anxiety and depression	negative	-0.033	0.035
Construct	**Rosenberg self-esteem scale**			
ROS	Rosenberg self-esteem	positive	-0.029	0.350
Sub-construct	**Violence Manifestation and Exposure Questionnaire**			
IZL	Violence exposure	negative	-0.009	0.735
ISP	Violence manifestation	negative	0.004	0.857
Sub-construct	**The Developmental Assets Profile**			
DAP_SUP1	Support	positive	-0.433	0.177
DAP_EMP2	Empowerment	positive	-0.108	0.742
DAP_BE3	Boundaries and expectations	positive	-0.114	0.721
DAP_CUT4	Constructive use of time	positive	-0.593	0.173
DAP_CL5	Commitment to learning	positive	-0.707	0.041
DAP_PV6	Positive values	positive	-0.111	0.682
DAP_SC7	Social competencies	positive	-0.429	0.162
DAP_PI8	Positive identity	positive	-0.026	0.936
DAP_EAC	External asset side (SUP1+EMP2+BE3+CUT4)	positive	-3.155	0.258
DAP_IAC	Internal asset side (CL5+PV6+SC7+PI8)	positive	-3.239	0.219
DAP_TOTAL	Total	positive	-6.407	0.204

Source: Own calculations using Stata 14

As we can see from the table above, only the effect of the intervention on the social responsibility scale and social relationship scale is statistically significant; however, the signs of this are opposite to those expected. This suggests that the impact of the intervention was generally non-existent, except for the two abovementioned scales, where treatment schools even experienced decreases compared to the comparison schools.

With regards to the outcome variables related to cognitive behavior, where a decrease in the average value in the treatment group schools compared to the comparison group schools indicates an improvement that can be attributed to the intervention, we can see that the results above do not support the claim that such an impact materialized, as the effect is not statistically significant, with the exception of somatic complaints and anxiety and depression constructs. Still, even these two indicators suggest possible evidence of the impact of the intervention, as well as the signs of coefficients for other dependent variables, which are negative, as expected.

The results for constructs related to violence manifestation and exposure also do not suggest any evidence of the impact of the intervention upon these outcomes.

Finally, for the development assets profile constructs, the values of coefficients for the effect of the intervention have statistically insignificant values, with exception of one construct; that measuring children's commitment to learning. However, the value suggests a negative impact, therefore implying that the situation in the treatment schools has even worsened during the intervention when compared to non-treated schools.

As a preliminary conclusion based on the first results from the MLM model estimation, we can that the Life Skills program has had a very limited impact on some of the outcome indicators, with no statistically significant effect on the vast majority of these. It goes without saying, however, that these results must be compared to results from other estimation approaches before drawing final conclusions.

4.2.2 RESULTS FROM THE DIFFERENCE-IN-DIFFERENCE MODEL

The ICC values estimated and presented in the previous section – which were in all cases below 10% – suggest that multilevel and grouping between the data is not significant and, therefore, using single level modeling is preferable to multilevel modeling. Given these results, we decided to estimate the model with a difference-in-difference (DID) approach in addition to the originally proposed multilevel modeling, as explained in the Methodological Chapter. Two different models will be estimated. First, we will estimate the DID model by including a set of covariates that were also included in the MLM model, but without matching at the individual level. Second, since the common trend is not necessarily assured through selection of treatment and comparison schools, and since we don't have data from data points prior to the baseline that would potentially allow us to test the common trend assumption, we decided to match individual pupils from the treatment and control schools on the baseline levels of the outcome and other observable variables in order to correct for confounding bias by balancing on variables that differ between the treatment and control groups. Results from the DID both with & without matching are presented in the following table and compared below. The purpose is to compare them and identify the influence of matching on the results.

Exhibit 15: Results of DID with covariates

Sub-construct	**Life Skills Scale**		DID	p value	DID with PSM	p value
MD	Making decisions in everyday life	positive	-0.111	0.082	-0.161	0.030
CR	Social responsibility scale	positive	-0.093	0.141	-0.003	0.967
CT	Critical thinking in everyday life	positive	-0.111	0.099	-0.135	0.080
SR	Social relationships scale	positive	-0.174	0.005	-0.211	0.004
CS	Communication scale	positive	-0.061	0.297	-0.036	0.589
EM	Emotional management scale	positive	-0.110	0.054	-0.156	0.023
Sub-construct	**Cognitive Behavior Checklist**					
AP	Attention problems	negative	0.020	0.473	-0.015	0.636
SP	Social problems	negative	-0.005	0.780	-0.026	0.236
TP	Thought problems	negative	0.006	0.702	-0.006	0.734
AB	Aggressive behavior	negative	0.024	0.201	-0.005	0.806
MB	Relationship with peers	negative	0.008	0.718	0.023	0.384
MC	Relationship between child and teachers	negative	-0.009	0.639	0.030	0.183
CBCL_MD	Coping with problems and locus of control	negative	0.010	0.699	-0.008	0.781
ME	Social anxiety	negative	0.019	0.456	-0.010	0.741
MG	Hyperactive behavior	negative	0.032	0.414	0.004	0.923
MA	Difficulties in learning and behavior	negative	-0.013	0.597	-0.030	0.255
SC	Somatic complaints	negative	-0.022	0.278	-0.021	0.359
RBB	Rule-breaking behavior	negative	0.002	0.861	0.007	0.579
AD	Anxiety and depression	negative	-0.009	0.726	-0.019	0.516
Construct	**Rosenberg self-esteem scale**					
ROS	Rosenberg self-esteem	positive	-0.074	0.118	-0.114	0.044
Sub-construct	**Violence Manifestation and Exposure Questionnaire**					
IZL	Violence exposure	negative	0.019	0.649	0.023	0.633
ISP	Violence manifestation	negative	0.041	0.230	-0.006	0.867
Sub-construct	**The Developmental Assets Profile**					
DAP_SUP1	Support	positive	-0.521	0.293	-1.242	0.036
DAP_EMP2	Empowerment	positive	-0.578	0.253	-0.936	0.115
DAP_BE3	Boundaries and expectations	positive	-0.479	0.328	-0.671	0.239
DAP_CUT4	Constructive use of time	positive	-0.608	0.362	-0.533	0.485

DAP_CL5	Commitment to learning	positive	-0.731	0.170	-0.734	0.228
DAP_PV6	Positive values	positive	-0.169	0.686	0.123	0.800
DAP_SC7	Social competencies	positive	-0.822	0.084	-0.380	0.491
DAP_PI8	Positive identity	positive	-0.949	0.061	-1.337	0.026
DAP_EAC	External asset side (SUP1+EMP2+BE3+CUT4)	positive	-5.475	0.201	-8.421	0.091
DAP_IAC	Internal asset side (CL5+PV6+SC7+PI8)	positive	-6.682	0.098	-5.803	0.213
DAP_TOTAL	Total	positive	-12.157	0.116	-14.224	0.114

Source: Own calculations using Stata 14

According to the results presented in the table above, we can first see that the values of coefficients from the two estimation approaches differ, but not to any significant extent. Moreover, the signs and statistical significance of the vast majority of coefficients are the same on both models, with the exception of several coefficients such as Rosenberg's self-esteem scale or four indicators in the development assets profile construct (support, social competences, and external and internal assets side). When it comes to the direction of influence, i.e. the signs of the coefficients, these are, in the majority of cases, opposite to that which would be expected (although admittedly, these are mainly statistically insignificant). The results worthy of further interpretation are several output indicators from the Life Skills Scale construct, namely the indicators related to making decisions in everyday life and critical thinking, and the social relationship and emotional management scales. Values of coefficients for these indicators are statistically significant in both models; however their sign is negative, suggesting that the difference that the intervention has made in treatment schools is less than the 'natural' improvement witnessed over time in the comparison schools.

Another statistically significant coefficients is that of the the Rosenberg's self-esteem scale, which is only statistically significant in the DID with matching model, albeit with a negative sign, while a positive was to be expected. Finally, for the development assets profile construct, there are several statistically significant coefficients (i.e. support, social competences, and external and internal assets side), although only in one of the two models, and even then with a negative sign of influence, which – once again – is the opposite of what should have been expected as a result of the treatment. For all the coefficients for these outcome indicators, results suggest that the difference that the intervention has made in treatment schools is less than the 'natural' improvement witnessed over time in the comparison schools.

5. CONCLUSIONS AND RECOMMENDATIONS

The main objective of the Life Skills program was to develop the capacities of children in selected primary schools (World Vision program partner schools) for making decisions and taking actions that would positively impact their lives. The Life Skills program is focused on critical thinking, effective communication, emotional management, interpersonal relationships, and community responsibility. In order to provide evidence as to the Life Skills program's contribution to the raising of life skills among children in program schools, CDESS conducted a rigorous impact evaluation of the program in primary schools in Tuzla Canton, the results of which are presented in this report.

The model that was developed for the purpose of measuring the impact of the intervention was estimated through the use of three different estimation approaches, namely MLM, DID without matching, and DID with matching.

Results indicate the statistically significant impact of the Life Skills program on some of the outcome indicators from the Cognitive Behavior list, such as somatic complaints or anxiety and depression among children. These effects were confirmed via four different model estimation approaches. For other outcome indicators, no evidence could be found to suggest a positive impact of the intervention. Moreover, for some indicators a negative impact of the intervention could be interpreted, although the statistical significance of the intervention coefficient is not consistent across the model estimations.

Possible explanations for statistical insignificance in some outcomes include the possibility that the treatment in its very nature was not capable of addressing all the issues that could be measured by the data collection instruments employed. If this is the case, these results can be informative for future designs of similar interventions, where comparison of characteristics and impacts captured between similar interventions may suggest possible correlations between some of the sub-activities and sub-outcomes. Moreover, it is possible that, for some of these variables, the timeframe covered by this impact evaluation was not sufficient to capture evidence of a statistically significant impact, which may well manifest itself at a later point in time. In order to ascertain this, both further investigation of such constructs and a repeated impact evaluation at a future point(s) in time would be helpful and informative for the design of future impact evaluations of similar interventions.

Another possible explanation for the lack of robust evidence on the impact of the Life Skills program can be found in limited information provided by the teachers implementing the program who we interviewed during the monitoring visits. Their perception of the program was that it did not considerably differ to what they had been implementing prior to its introduction. Since no information regarding the schools' needs assessments that led to design of the Life Skills program was made available, it can only be stated that such responses by teachers deserve more careful investigation. If there is some truth to their statements, and if there was no detailed needs assessment undertaken prior to the implementation of the program, this could go some way to explaining why no significant

impact of the program compared to comparison schools was recorded; the Life Skills program merely replaced an existing program, whereas this existing program continued to run in the comparison schools. This would also serve as a good lesson for future designs of similar interventions.

A key methodological limitation of the design of the evaluation was related to its quasi-experimental nature when compared to full experimental designs. Drawbacks were minimized by controlling for initial differences between treatment and comparison groups using propensity score matching and inverse probability weighting. Additional methodological limitations stemmed from the Theory of Change, whereby it is expected that life skills affect social change indicators, where it is possible that the level of social change indicators can have significant confounding variable influencing relationships between treatment and life skills.

6. BIBLIOGRAPHY

Achenbach, T. M., & Rescorla, L. A. (2001). *Manual for the ASEBA School-Age Forms & Profiles*. Burlington, VT: University of Vermont, Research Center for Children, Youth, & Families.

Anuradha, K. & Lokeswari, V. (2016). Impact of life skills education & self-esteem in adolescence – A revisit. *EPRA International Journal of Multidisciplinary Research (IJMR)*, 2(2), pp. 14-17.

Botvin, G. J. & Griffin, K. W. (2014). Life skills training: Preventing substance misuse by enhancing individual and social competence. *New Directions for Youth Development*, 14(Spring 2014), pp. 57-65.

Esmaeilinasab, M., Malek Mohamadi, D., Ghiasvand, Z. & Bahrami, S. (2011). Effectiveness of life skills training on increasing self-esteem of high school students. *Procedia - Social and Behavioral Sciences*, 30, pp. 1043 – 1047.

Ghasemian, A. & Kumar, G. V. (2017). Effect of life skills training on psychological distress among male and female adolescent students. *Indian Journal of Health and Wellbeing*, 8(4), pp. 279-282

Gorman, D. M. (2011). Does the Life Skills Training program reduce use of marijuana? *Addiction Research and Theory*, 19(5), pp. 470–481.

Gorman, D. M., Conde, E. & Huber, J. C. (2007). The creation of evidence in 'evidence-based' drug prevention: a critique of the Strengthening Families Program Plus Life Skills Training evaluation. *Drug and Alcohol Review*, 26(November 2007), pp. 585 – 593.

Grawe, R. W., Hagen, R. Espeland, B. & Mueser, K. T. (2007). The Better Life Program: Effects of group skills training for persons with severe mental illness and substance use disorders. *Journal of Mental Health*, 16(5), pp. 625-634.

Hair, J. F., Black, W. C., Babin, B. J. & Anderson, R. E. (2014*). Multivariate Data Analysis*. Harlow: Pearson Education Limited.

Hatcher, J. & Hall, L. (2009). Psychometric Properties of the Rosenberg Self-Esteem Scale in African American Single Mothers. *Issues in Mental Health Nursing*, 30(2), pp.70-77.

Hedges, L.V. & C. Rhoads. (2010). Statistical Power in Analysis in Education Research, NCSER, U.S. Department of Education

Kim, J. & Bryan, J. (2017). A First Step to a Conceptual Framework of Parent Empowerment: Exploring Relationships Between Parent Empowerment and Academic Performance in a National Sample. *Journal of Counseling & Development*, 95(April 2017), pp. 168-179.

Kumar, V. S. K. & Krishnamurthy, A. R. (2016). The impact of life skills training on psychological well-being: The role of gender differences. *Indian Association of Health*, 7(2), pp. 193-197.

Malatest, R. A. & Associates (March 2014). Final Evaluation Report of the Life Skills Training Program (LST). Submitted to Public Safety Canada.

Nasser & Rajabi., (2010): Efficacy of life skills training on increase of mental health and self-esteem of the students, Procedia Social and Behavioral Sciences, 5, 1818–1822

Pick, S., Givaudan, M, Sirkin, J. & Ortega, I. (2007). Communication as a protective factor: Evaluation of a life skills HIV/aids prevention program for Mexican elementary-school students. *AIDS Education and Prevention*, 19(5), pp. 408–421.

Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton, NJ: Princeton University Press.

Search Institute (2005). *Developmental Assets Profile – User Manual*. Minneapolis: Search Institute.

Song, S. (2012). Influences on Academic Achievement of Primary School Pupils in Cambodia. *Excellence in Higher Education*, 3, pp. 79-87.

Tella, A. (2008). Teacher Variables As Predictors of Academic Achievement of Primary School Pupils Mathematics. *International Electronic Journal of Elementary Education*, 1(1), pp. 16-33.

Tahereh., (2011): The Effectiveness of life skills training on happiness, quality of life and emotion regulation", Procedia - Social and Behavioral Sciences, 30, 407–411.

Visser, M. J. (2005). Life skills training as HIV/AIDS preventive strategy in secondary schools: evaluation of a large-scale implementation process. *SAHARA-J: Journal of Social Aspects of HIV/AIDS*, 2(1), pp. 203-216.

Werunga, K. R., Musera, G. & Sindabi, O. (2011). Factors Affecting Transition Rates from Primary to Secondary Schools: The Case of Kenya. *Problems of Education in the 21st Century*, 23, pp. 129-139.

World Health Organization (1997) Life Skills Education for Children and Adolescents in Schools, WHO-MNH-PSF-93.7A.Rev.2

Yankah, E. & Aggleton, P. (2008). Effects and effectiveness of life skills Education for HIV prevention in young people. *AIDS Education and Prevention*, 20(6), pp. 465–485.

7. ANNEXES

ANNEX I: DATA COLLECTION INSTRUMENTS USED FOR IMPACT EVALUATION

LIFE SKILLS SCALE (to be filled by children)

Making decisions in everyday life (MDEL)

The following statements describe how you might make a decision in everyday life. Select the appropriate response that best corresponds to how often or how likely you did what is described in the last 30 days.

	When I have a decision to make . . .	**Never**	**Rarely**	**Sometimes**	**Often**	**Always**
MDEL1.	I easily identify my problem.	1	2	3	4	5
MDEL2.	I think about the problem before I take action.	1	2	3	4	5
MDEL3.	I look for information to help me understand the problem.	1	2	3	4	5
MDEL4.	I ask others to help me identify my problem.	1	2	3	4	5
MDEL5.	I think about ways of dealing with my problem.	1	2	3	4	5
MDEL 6.	I think before making a choice.	1	2	3	4	5
MDEL 7.	I discuss choices with my friends before making a decision.	1	2	3	4	5
MDEL 8.	I discuss choices with my parents before making a decision.	1	2	3	4	5
MDEL 9.	I look for positive points of possible choices.	1	2	3	4	5
MDEL 10.	I look for negative points of possible choices.	1	2	3	4	5
MDEL 11.	I consider the risks of a choice before making a decision.	1	2	3	4	5
MDEL 12.	I consider the benefits of a choice before making a decision.	1	2	3	4	5
MDEL 13.	I make decisions based on what my parents tell me.	1	2	3	4	5
MDEL 14.	When faced with a decision, I realize that some choices are better than others.	1	2	3	4	5

MDEL 15.	I make a decision by thinking about all the information I have about the different choices.	1	2	3	4	5
MDEL 16.	I prioritize my choices before making a decision.	1	2	3	4	5
MDEL 17.	Before making another decision, I think about how the last one turned out.	1	2	3	4	5
MDEL 18.	I do think of past choices when making new decisions.	1	2	3	4	5
MDEL 19.	If I experience negative consequences, I change my decision the next time.	1	2	3	4	5
MDEL 20.	Decision-making is easy for me.	1	2	3	4	5

Critical thinking in everyday life (CTEL)

The following statements describe how you might think about certain things in your daily life. Select the answer that corresponds to how often you have done what is described in the last 30 days.

		Never	**Rarely**	**Sometimes**	**Often**	**Always**
CTEL1.	I think of possible results before I take action.	1	2	3	4	5
CTEL 2.	I get ideas from other people when having a task to do.	1	2	3	4	5
CTEL 3.	I develop my ideas by gathering information.	1	2	3	4	5
CTEL 4.	When facing a problem, I identify options.	1	2	3	4	5
CTEL 5.	I can easily express my thoughts on a problem.	1	2	3	4	5
CTEL 6.	I am able to give reasons for my opinions.	1	2	3	4	5
CTEL 7.	It is important for me to get information to support my opinions.	1	2	3	4	5
CTEL 8.	I usually have more than one source of information before making a decision.	1	2	3	4	5
CTEL 9.	I plan where to get information on a topic.	1	2	3	4	5
CTEL 10.	I plan how to get information on a topic.	1	2	3	4	5
CTEL 11.	I put my ideas in order by importance.	1	2	3	4	5
CTEL 12.	I back my decisions by the information I got.	1	2	3	4	5

CTEL 13.	I listen to the ideas of others even if I disagree with them.	1	2	3	4	5
CTEL 14.	I compare ideas when thinking about a topic.	1	2	3	4	5
CTEL 15.	I keep my mind open to different ideas when planning to make a decision.	1	2	3	4	5
CTEL 16.	I am aware that sometimes there are no right or wrong answers to a question.	1	2	3	4	5
CTEL 17.	I develop a checklist to help me think about an issue.	1	2	3	4	5
CTEL 18.	I can easily tell what I did was right or wrong.	1	2	3	4	5
CTEL 19.	I am able to tell the best way of handling a problem.	1	2	3	4	5
CTEL 20.	I make sure the information I use is correct.	1	2	3	4	5

Communication scale (CS)

Select the answer that best corresponds to how often you did what is described in the last 30 days.

		Never	**Rarely**	**Sometimes**	**Often**	**Always**
CS1.	I use my tone of voice to reinforce what I am trying to say.	1	2	3	4	5
CS2.	I don't hear everything a person is saying, because I am thinking about what I want to say.	1	2	3	4	5
CS3.	When talking to someone, I try to maintain eye contact.	1	2	3	4	5
CS4.	My body language reinforces what I am trying to say.	1	2	3	4	5
CS5.	I interrupt other people to say what I want to say before I forget it.	1	2	3	4	5
CS6.	I recognize when two people are trying to say the same thing, but in different ways.	1	2	3	4	5
CS7.	I try to watch other people's body language to help me trying to say.	1	2	3	4	5
CS8.	I recognize when people are using their hands to reinforce what they are saying.	1	2	3	4	5
CS9.	I recognize when a person is listening to me, but not hearing what I am saying.	1	2	3	4	5
CS10.	I use my own experiences to let my friends know that I understand what they are going through.	1	2	3	4	5
CS11.	When I am listening to someone, I try to understand what they are feeling.	1	2	3	4	5

CS12.	I try to see the other person's point of view.	1	2	3	4	5
CS13.	I change the way I talk to someone based on my relationship with them (i.e., friend, parent, teacher, etc.).	1	2	3	4	5
CS14.	I try to respond to what someone is saying, rather than just reacting to their tone of voice.	1	2	3	4	5
CS15.	To help a person understand me, I change the way I speak based on how the other person is talking to me.	1	2	3	4	5
CS16.	I find it easy to get my point across.	1	2	3	4	5
CS17.	I use my hands to illustrate what I am trying to say.	1	2	3	4	5
CS18.	I organize thoughts in my head before speaking.	1	2	3	4	5
CS19.	I use body language to help reinforce what I want to say.	1	2	3	4	5
CS20.	I make sure I understand what another person is saying before I respond.	1	2	3	4	5
CS21.	I rephrase what another person said, to make sure that I understood them.	1	2	3	4	5
CS22.	When someone gets mad, I change my tone of voice to help calm them down.	1	2	3	4	5
CS23.	I find ways to redirect the conversation when people rattle on and on.	1	2	3	4	5

Social relationships scale (SRS)

Select the answer that best corresponds to how often you experience what is described in the last 30 days.

		Never	**Rarely**	**Sometimes**	**Often**	**Always**
SRS1.	I noticed that other people believe in my competencies	1	2	3	4	5
SRS 2.	I share my important problems with the family members	1	2	3	4	5
SRS 3.	I would request help from teacher and/or school pedagogue/ psychologist to help me with some serious problem	1	2	3	4	5
SRS 4.	I can rely on my friends when having difficult times	1	2	3	4	5
SRS 5.	I love to encourage others to do what is good for the group	1	2	3	4	5
SRS 6.	I am satisfied with the relationship within my family	1	2	3	4	5
SRS 7.	I am satisfied with the relationship with teachers	1	2	3	4	5

SRS 8.	I am satisfied with the relationship with peers	1	2	3	4	5

Emotional management scale (EMS)

Select the answer that best corresponds to how often you experience what is described in the last 30 days.

		Never	Rarely	Sometimes	Often	Always
EMS1.	I feel anxious and I am not aware of the reason	1	2	3	4	5
EMS 2.	I avoid talking with friends about my feelings	1	2	3	4	5
EMS 3.	It happens to me to become so furious that I am not aware of my actions	1	2	3	4	5
EMS 4.	I do not feel comfortable when others are talking about their feelings	1	2	3	4	5
EMS 5.	I need a lot of time to calm myself if someone offend me	1	2	3	4	5
EMS 6.	It happens to me to not be aware of my own feelings	1	2	3	4	5

Social Responsibility Scale (SRS)

Select the answer that best corresponds to how likely you will experience what is described in the last 30 days.

		Never	Rarely	Sometimes	Often	Always
SRS1.	I believe that ecological actions (cleaning school yard, river band etc.) are just waste of time	1	2	3	4	5
SRS 2.	In a case if I find out that a fight is agreed, I would report it to some adult	1	2	3	4	5
SRS 3.	When I promise something, I fulfill it	1	2	3	4	5
SRS 4.	I provide help to older people in my community	1	2	3	4	5
SRS 5.	I am ready to make an effort to improve school environment for the better	1	2	3	4	5

SRS 6.	I would rather get a lower academic score, compared to copy from other student	1	2	3	4	5
SRS 7.	I would gladly provide help to children with disabilities in my community	1	2	3	4	5

ROSENBERG SELF-ESTEEM SCALE (To be filled by children)

Name: _____ Classroom: _____

School:_____ Place: _____

Below is a list of statements dealing with your general feelings about yourself. Please indicate how strongly you agree or disagree with each statement by selecting appropriate response (circle your response).

1. On the whole, I am satisfied with myself.
Strongly Agree Agree Disagree Strongly Disagree

2. At times I think I am no good at all.
Strongly Agree Agree Disagree Strongly Disagree

3. I feel that I have a number of good qualities.
Strongly Agree Agree Disagree Strongly Disagree

4. I am able to do things as well as most other people.
Strongly Agree Agree Disagree Strongly Disagree

5. I feel I do not have much to be proud of.
Strongly Agree Agree Disagree Strongly Disagree

6. I certainly feel useless at times.
Strongly Agree Agree Disagree Strongly Disagree

7. I feel that I'm a person of worth, at least on an equal plane with others.
Strongly Agree Agree Disagree Strongly Disagree

8. I wish I could have more respect for myself.
Strongly Agree Agree Disagree Strongly Disagree

9. All in all, I am inclined to feel that I am a failure.
Strongly Agree Agree Disagree Strongly Disagree

10. I take a positive attitude toward myself.
Strongly Agree Agree Disagree Strongly Disagree

VIOLENCE MANIFESTATION AND EXPOSURE QUESTIONNAIRE (To be filled by children)

Below is a list of situations that any child could find him/her selves in. Please rate how often you have been in situations described and in the left part of the questionnaire enter how often you have been exposed to such situation and in the right side enter how often you have expressed described behavior against other children during the last 6 months.

0 = never 1 = once or twice 2 = several times (more than twice) 3 = on daily basis

EXPOSED TO					MANIFESTED			
0	1	2	3	Physical fight	0	1	2	3
0	1	2	3	Pushing	0	1	2	3
0	1	2	3	Throwing	0	1	2	3
0	1	2	3	Ignoring	0	1	2	3
0	1	2	3	Verbal Assault	0	1	2	3
0	1	2	3	Threatening	0	1	2	3
0	1	2	3	Gossiping	0	1	2	3
0	1	2	3	Unwanted touch	0	1	2	3
0	1	2	3	Exposed to inappropriate material	0	1	2	3
0	1	2	3	Stealing and/or abducting	0	1	2	3
0	1	2	3	Destroying property	0	1	2	3
0	1	2	3	Sexual violence	0	1	2	3
0	1	2	3	Recording with mobile without approval	0	1	2	3
0	1	2	3	Distribution of recorder video/photo without approval	0	1	2	3
0	1	2	3	Extortion of money	0	1	2	3

WRITTEN CONSENT (to be filled by parent before child inclusion in the evaluation)

Ministry of Education in cooperation with World Vision is integrating Life Skills program to mainstream curriculum. Life skills education is seen as an important vehicle to equip children to negotiate and mediate challenges and risks in their lives, and to enable productive participation in society. Life skills curriculum's Theory of Change assume that exposing children to social dilemmas will develop different type of social skills (decision making, problem solving, creative thinking, critical thinking, effective communication, interpersonal relationship skills, self-awareness, empathy, coping with emotions and Coping with stress) and attitudes that will be beneficial both to individual and to nearest society.

In order to assess quality of integrated Life Skills program, Ministry of Education and World Vision in partnership with the Center for Development Evaluation and Social Science Research (CDESS) are conducting a research on the level of social skills and its association with cognitive, behavioral, social, emotional and academic response of children.

Currently Life Skills program is being implemented in some schools in Republika Srpska, and the classroom of your child has been preselected to participate in implementation of the Life Skills Program or is preselected to participate in the research on life skills programme and respective cognitive, behavioral, social, emotional and academic response of children as part of the comparison group (group of children who do not receive Life Skills Programme). It is of utmost importance to assess the effects of the programme in order to ensure evidence for further replication of the models to all schools.

Hereby, we request your approval for your child to participate in the research. If you allow your child to participate in the research he or she will be asked by the grade teacher to fill the questionnaire on different Life skills, Self-esteem and Social status.

Furthermore, we would like to ask you to fill some questionnaire on behalf of your child. Be aware, that the research will collect data at the beginning of the school year (currently collecting data) and at the end of 2017/18 academic year, and therefore in order to ensure coherence between data collected now and in May, we are requesting from research participants to enter their names. However, results of all children will be protected in a way that only your grade teacher and research team leader will have access to full data. All other analysts, statisticians and interested stakeholders might receive access to data that are cleaned of all personal information as per The Code of Ethical Research with Children and on Children in Bosnia and Herzegovina (2014) and Word Vision Child Protection Policy.

Hereby, I declare that I am familiar with the research being conducted to evaluation the effects of Life Skills Programme in Primary schools. I understand that all personal data that can reveal identity of the child will be secured. I understand that I do not expect material benefits and my eventual refusal to participate in the research (own participation and participation of the child) will not affect the quality of education programme that my child is attending.

Hereby, I accept participation in the research on Life Skills for me and my child.

Parents signature: _____ Date: _____

SOCIODEMOGRAPHIC QUESTIONNAIRE (to be filled by parents)

Please provide following information about your family

Name of the child: _____ Classroom: _____

School:_____ Place of school: _____

Gender of the child: ☐Male ☐Female Child date of birth:__/__/____ (dd/mm/yyyy)

1a. Mother's education (circle):
- ☐ no school
- ☐ primary school
- ☐ secondary school
- ☐ higher school
- ☐ faculty (4 or 5 years of education
- ☐ master of science or PhD

1b. Father's education (circle):
- ☐ no school
- ☐ primary school
- ☐ secondary school
- ☐ higher school
- ☐ faculty (4 or 5 years of education
- ☐ master of science or PhD

2a. Mother's employment status (circle):
- ☐ employed
- ☐ temporarily employed
- ☐ unemployed
- ☐ retired
- ☐ other: _____ (describe)

2b. Father's employment status (circle):
- ☐ employed
- ☐ temporarily employed
- ☐ unemployed
- ☐ retired
- ☐ other: _____ (describe)

3. What description best describes family situation of the child (circle):

☐ child lives with both parents
☐ child lives with one parents (parents are divorced)
☐ child lives with one parents (one of the parents deceased)
☐ child lives in alternative care institutions (foster care, institution)
☐ child lives with the relative (grandparents, cousins, etc.)

4. What is the number of household members living in the apartment with you and your child?

_____

5. Select all that applies to the development of your child?
☐ child's development was as expected, equal as his/her peers
☐ child is having developmental difficulties (diagnosed - categorized)
☐ child is not categorized but we suspect to have some developmental difficulties

5.a If a child is categorized or there is a suspicious on the developmental difficulties, please select difficulty child is supposed to have:
☐ visual impairments
☐ hearing impairments
☐ physical difficulties (cerebral paralysis, pledgees, muscular dystrophy, etc.)
☐ cognitive difficulties (Down Syndrome, Autistic Spectre Disorder, below average cognitive abilities)
☐ language and speech difficulties (articulation, dysgraphia, dyslexia, etc.)

6. What is the average total monthly income in your family
☐ up to 300 BAM
☐ from 301 to 800 BAM
☐ from 801 to 1600 BAM
☐ from 1601 to 3000 BAM
☐ above 3000 BAM

COGNITIVE-BEHAVIOR CHECK LIST (to be filled by parents)

Below is a list of items that describe children and youths. For each item that describes your child now or within the past 6 months, please circle the 2 if the item is very true or often true of your child. Circle the 1 if the item is somewhat or sometimes true of your child. If the item is not true of your child, circle the 0. Please answer all items as well as you can, even if some do not seem to apply to your child.

Be sure to answer all items.

0 = Not True (as far as you know) **1 = Somewhat or Sometimes True**
2 = Very True or Often True

0	1	2	Acts too young for his/her age [AP1]	0	1	2	Feels or complains that no one loves him/her [AD33]
0	1	2	Drinks alcohol without parents' approval [RBB2]	0	1	2	Feels others are out to get him/her [SP34]
0	1	2	Argues a lot [AB3]	0	1	2	Feels worthless or inferior [AD35]
0	1	2	Fails to finish things he/she starts [AP4]	0	1	2	Gets hurt a lot, accident-prone [SP36]
0	1	2	There is very little he/she enjoys [WD5]	0	1	2	Gets in many fights [AB37]
0	1	2	Comparing to peers, he/she has difficulties in learning in school [MA6]	0	1	2	Destroys things belonging to his/her family or others [AB38]
0	1	2	Don't do homework [MA7]	0	1	2	Hangs around with others who get in trouble [RBB39]
0	1	2	Can't concentrate, can't pay attention for long [AP8]	0	1	2	Hears sound or voices that aren't there [TP40]

0	1	2	Dropping school classes [MA9]	0	1	2	Impulsive or acts without thinking [AP41]
0	1	2	Can't sit still, restless, or hyperactive [AP10]	0	1	2	Would rather be alone than with others [WD42]
0	1	2	Clings to adults or too dependent [SP11]	0	1	2	Lying or cheating [RBB43]
0	1	2	Complains of loneliness [SP12]	0	1	2	Breaks school rules [MA44]
0	1	2	Confused or seems to be in a fog [AP13]	0	1	2	Nervous, highstrung, or tense [AD45]
0	1	2	Cries a lot [AD14]	0	1	2	Nervous movements or twitching [TP46]
0	1	2	Cruel to animals [AB15]	0	1	2	Nightmares [SC47]
0	1	2	Not accepted by peers in the classroom [MB16]	0	1	2	Not liked by other kids [SP48]
0	1	2	Coming into verbal conflicts with peers [MB17]	0	1	2	Constipated, doesn't move bowels [SC49]
0	1	2	Deliberately harms self or attempts suicide [TP18]	0	1	2	Too fearful or anxious [AD50]
0	1	2	Is being mocked by peers [MB19]	0	1	2	Feels dizzy or lightheaded [SC51]
0	1	2	Destroys his/her own things [AB20]	0	1	2	Feels too guilty [AD52]
0	1	2	Gets teased a lot [SP20]	0	1	2	Complains on the behavior of peers [MB53]
0	1	2	Disobedient at home [AB22]	0	1	2	Overtired without good reason [SC54]
0	1	2	Disobedient at school [AB23]	0	1	2	Peers complains on his/her behavior [MB55]
0	1	2	Had disciplinary measures in schools [MC24]				Physical problems without known medical cause:
0	1	2	Doesn't get along with other kids [SP25]	0	1	2	Aches or pains (not stomach or headaches) [SC56.a]
0	1	2	Doesn't seem to feel guilty after misbehaving [RBB26]	0	1	2	Headaches [SC56.b]
0	1	2	Easily jealous [SP27]	0	1	2	Nausea, feels sick [SC56.c]
0	1	2	Breaks rules at home, school, or elsewhere [RBB28]	0	1	2	Rashes or other skin problems [SC56.d]
0	1	2	Fears certain animals, situations, or places, other than school [AD29]	0	1	2	Problems with eyes (not if corrected by glasses) [SC56.e]
0	1	2	Fears going to school [AD30]	0	1	2	Stomachaches [SC56.f]
0	1	2	Fears he/she might think or do something bad [AD31]	0	1	2	Vomiting, throwing up [SC56.g]
0	1	2	Feels he/she has to be perfect [AD32]	0	1	2	Other (describe): _____ [SC56.h]

Be sure you answered all items. Then see other side.

Continuation from the previous page. Be sure to answer all items.
0 = Not True (as far as you know) 1 = Somewhat or Sometimes True
2 = Very True or Often True

0	1	2	Physically attacks people [AB57]	0	1	2	Sulks a lot [AB88]
0	1	2	Picks nose, skin, or other parts of body [TP58]	0	1	2	Suspicious [AB89]
0	1	2	Plays with own sex parts in public [TP59]	0	1	2	Swearing or obscene language [RBB90]
0	1	2	Plays with own sex parts too much [TP60]	0	1	2	Talks about killing self [AD91]

0	1	2	Poor school work [AP61]	0	1	2	Talks or walks in sleep [TP92]
0	1	2	Poorly coordinated or clumsy [SP62]	0	1	2	Is difficult for cooperation with teachers [MC93]
0	1	2	Prefers being with older kids [RBB63]	0	1	2	Teases a lot [AB94]
0	1	2	Prefers being with younger kids [SP64]	0	1	2	Temper tantrums or hot temper [AB95]
0	1	2	Refuses to talk [WD65]	0	1	2	Thinks about sex too much [RBB96]
0	1	2	Entering in verbal conflicts with teachers [MC66]	0	1	2	Threatens people [AB97]
0	1	2	Runs away from home [RBB67]	0	1	2	Don't seek help and support from adults [MD98]
0	1	2	Screams a lot [AB68]	0	1	2	Smokes, chews, or sniffs tobacco [RBB99]
0	1	2	Secretive, keeps things to self [WD69]	0	1	2	Trouble sleeping [TP100]
0	1	2	Sees things that aren't there [TP70]	0	1	2	Truancy, skips school [RBB101]
0	1	2	Self-conscious or easily embarrassed [AD71]	0	1	2	Underactive, slow moving, or lacks energy [WD102]
0	1	2	Sets fires [RBB72]	0	1	2	Avoids talking with others [ME105]
0	1	2	Sexual problems [RBB73]	0	1	2	Unusually loud [AB104]
0	1	2	Blame others for his/her inappropriate behavior in school [MD74]	0	1	2	Believes that others are determining his/her life [MD105]
0	1	2	Too shy or timid [WD75]	0	1	2	Unhappy, sad, or depressed [WD106]
0	1	2	Sleeps less than most kids [TP76]	0	1	2	Vandalism [RBB107]
0	1	2	Believed that he/she can't influence the future [MD77]	0	1	2	Uses drugs for nonmedical purposes (don't include alcohol or tobacco) [RBB108]
0	1	2	Inattentive or easily distracted [AP78]	0	1	2	When talking, express discomfort [ME109]
0	1	2	Speech problem [SP79]	0	1	2	Don't establish eye contact [ME110]
0	1	2	Stares blankly [AP80]	0	1	2	Withdrawn, doesn't get involve [WD111]
0	1	2	Steals at home [RBB81]	0	1	2	Worries [AD112]
0	1	2	Steals outside the home [RBB82]	0	1	2	Don't shows interest about others [ME113]
0	1	2	Stores up too many things he/she doesn't need [TP83]	0	1	2	Is significantly quieter and calmer comparing to other children [ME114]
0	1	2	Strange behavior [TP84]	0	1	2	Having trouble concentrating [MG115]
0	1	2	Looks fearful [ME85]	0	1	2	Can't sit tight in one place [MG116]
0	1	2	Stubborn, sullen, or irritable [AB86]	0	1	2	Easily district attention [MG117]
0	1	2	Sudden changes in mood or feelings [AB87]				

Be sure you answered all items.

CHILD EDUCATIONAL ATTAINEMENT (to be filled by grade teacher)

Data to be filled for all school groups of eight graders selected into treatment or comparison group. Data to be entered for each branch school separately. Total number of boys and girls in school will be equal for all school groups of eight graders within same

school building. Please be aware that data are needed for each school building irrespective of the center of school administration.

School: _____; Class group: _____

Info on [enter period] 2016/17 – for collection of baseline data 2017/18 – for collection of post treatment data Info on [enter period] cell of this table will be deleted before submission of the table to teachers Name of the child	Average school mark at the end of _____ [enter period]	Behavior mark at the end of _____ [enter period]	Total # of absent classes during the _____ [enter period]	Total # of unjustified absent classes during the _____ [enter	Total # of disciplinary measures imposed on a child in a school in _____ [enter period]	Quality of relationship with the child parents (1-no relationship; 2-mediocre relationship; 3 – good relationship)	Quality of child participation during school lectures (1 – child is withdrawn, no quality participation; 2 – mediocre participation; 3 – excellent

ANNEX II: DATA COLLECTION INSTRUMENTS USED FOR MONITORING

LIFE SKILLS TEACHER'S COMPETENCIES FORM (to be filled by teachers)

A. During the last six months of your teaching of the class group _____ [ENTER CLASSROOM IDENTIFIER] (during the second half of the academic 2016/17 year) how

often have you integrated elements, in your regular teaching practice, with the objective to increase stated competencies of children from that class group. Please indicate the level of integration that was consciously and as part of the plan being integrated into your teaching curriculum.

0 = never 1 = once or twice 2 = three to five times 3 = six to ten times 4 = more than 10 times

0	1	2	3	4	To increase capacity for critical thinking of the children
0	1	2	3	4	To increase problem solving and decision making skills
0	1	2	3	4	To increase effective communication between children
0	1	2	3	4	To increase quality of interpersonal relationship skills of children
0	1	2	3	4	To increase capacities to cope with emotions and stress of children
0	1	2	3	4	To increase community responsibility of children

B. How competent do you find yourself in providing teaching to children that incorporate elements focused on development of competencies stated below into regular curriculum.

0 = no competencies at all

1 = only competencies that are inseparably linked to my teaching practice

2 = have some additional competencies that are basic level

3 = have competencies that are advanced level for implementation with children

4 = have competencies that are advanced level and could provide mentorship for other teachers on these subjects

0	1	2	3	4	To increase capacity for critical thinking of the children
0	1	2	3	4	To increase problem solving and decision making skills
0	1	2	3	4	To increase effective communication between children
0	1	2	3	4	To increase quality of interpersonal relationship skills of children
0	1	2	3	4	To increase capacities to cope with emotions and stress of children
0	1	2	3	4	To increase community responsibility of children

C. During the last six months of your teaching lessons with above class group], how many specifically planned workshops / school classes have you conducted with children with the objectives stated below. Please indicate number of events (workshops, seminars, regular school classes) that you have participated in, and not the one you only heard about. On the rights side please indicate an average duration of such events.

Number of events	Objective	Average duration of event
	To increase capacity for critical thinking of the children	
	To increase problem solving and decision making skills	
	To increase effective communication between children	
	To increase quality of interpersonal relationship skills of children	
	To increase capacities to cope with emotions and stress of children	
	To increase community responsibility of children	

D. Are you teaching children in more than one school? ☐ Yes ☐ No

If you selected YES, please write the name and place of all primary schools you are working with:

CHILD PARTICIPATION EVALUATION FORM (to be filled by children during monitoring visits)

Name of the child: _____ Classroom: _____

School:_____ Date: __/__/____ (dd/mm/yyyy)

Gender of the child: ☐Male ☐Female

This questionnaire asks several questions about lecture you have just attended. Could you please answer the questions honestly in order to use your answers while improving content and ways of its delivery to children by teachers.

1. In your opinion what are the main goals of the school lecture you have just attended? (please list at least three objectives you find most relevant for the lecture). At the right side please indicate to what extend the objectives was met (in your opinion) during the lecture!

0 = it is not met at all; 1 = partially met; 2 = met to full extent (circle as appropriate)

Objective			
A -	0	1	2
B -	0	1	2
C -	0	1	2

2. To what extent the lecture was interesting to you (color the smiley that best describe your opinion)?

It was not interesting at all. I was bored all during entire lecture

It was partially interesting. During some parts of the lecture I was bored, but some other parts were very interesting.

It was exceptionally interesting

3. To what extent you have had opportunities to express your opinion during the lecture on the relevant topic (color the smiley that best describe your opinion)?

I did not express mine opinion at all during the lecture

During the lecture I have expressed my opinion at least one but it did not had significant effect on the

During the lecture I had expressed my opinion several times, and I had a feeling that the opinion is

continuation of the lecture	respected by other participants and the teacher

Turn the page ➔➔➔

4. To what extent you have received an answer on your posed questions related to lecture (color the smiley that best describe your opinion)?

 If questions are posed, teacher did not provide answer to them	 If questions are posed, teacher provided the answer that was not sound plausible	 If questions are posed, teacher provided sound answers. Answers were aligned with the question and it was clear that teacher knows what he/she talks about

5. Assess the quality of teacher's communication during the lecture (color the smiley that best describe your opinion)!

 Quality of communication is very low. Teachers rarely ask as for opinion and he/she has done most of talking.	 Quality of communication is moderate. Teacher has established contacts with children, however, children were mainly repeating previously said.	 Quality of communication is exceptional. Teacher is capable to convey messages to children by tailoring discussions in certain directions.

6. Please assess to what extent the lecture was contributing to achievement of objectives stated below.

0 = do not reach objective

1 = objective is slightly achieved
2 = objective is achieved
3 = objective is exceptionally achieved

Objective	0	1	2	3
To increase capacity for critical thinking of the children	0	1	2	3
To increase problem solving and decision making skills	0	1	2	3
To increase effective communication between children	0	1	2	3
To increase quality of interpersonal relationship skills of children	0	1	2	3
To increase capacities to cope with emotions and stress of children	0	1	2	3
To increase community responsibility of children	0	1	2	3

7. Please describe two related themes to the theme of the lecture that you would like to learn more about it during the next 12 months?

Thank you!
Please, return the filled questionnaire to the teacher

TEACHERS' ASSESSMENT OF THE LIFE SKILLS PROGRAM (to be filled by teachers)

Name of the teacher: _____ Classroom: _____

School:_____ Date: __/__/____ (dd/mm/yyyy)

Gender of the teacher: ☐Male ☐Female Date of birth of the teacher: __/__/____ (dd/mm/yyyy)

A. Please assess the quality of Life Skills Programme you are implementing in schools in cooperation with Pedagogical Institute and World Vision by selecting the answers that best describes the level of your agreement with the statement below.

1 – disagree to full extent; 3 – neither agree nor disagree; 5 – agree to full extent

1.	Life skills program provide clear objectives of the lectures	1	2	3	4	5
2.	Life skills program provide clear didactic and methodic instructions for realization of the lectures	1	2	3	4	5
3.	Life Skills program provide unambiguous instructions for realization of lectures	1	2	3	4	5
4.	Life skills objectives and techniques are appropriate for the children	1	2	3	4	5
5.	During the lectures there were questions on behalf of children that I was unable to competently answer	1	2	3	4	5
6.	I would recommend that all children in primary school participate in Life Skills education program	1	2	3	4	5

7.	I believe that Life Skills program should be integrated in mainstream curriculum as of age of 6	1	2	3	4	5
8.	Life skills program is easily integrated with planned education outcomes	1	2	3	4	5
10.	Life skills program increase capacity for critical thinking of the children	1	2	3	4	5
11.	Life skills program increase problem solving and decision making skills	1	2	3	4	5
12.	Life skills program increase effective communication between children	1	2	3	4	5
13.	Life skills program increase quality of interpersonal relationship skills of children	1	2	3	4	5
14.	Life skills program increase capacities to cope with emotions and stress of children	1	2	3	4	5
15.	Life skills program increase community responsibility of children	1	2	3	4	5

B. Please describe additional topic that you would like to learn more during the next 12 months and are related to Life skills program!

Thank you!

ANNEX III: PILOT FINDINGS SUMMARY

The pilot research was conducted on a sample of 8th grade children from the Čengić Vila I and Grbavica II elementary schools in Sarajevo. The key purpose of the pilot was to test children's and parents' understanding of questions, confusions that may arise in providing responses, gauging of the initial responses of parents to the involvement of their children in the research, the time required to complete the questionnaire, both for parents and children, understanding of the instructions and general response of parents to initiatives such as this one. In total, 43 children (19 from Čengić Vila I, 24 from Grbavica II) and 12 parents participated. Only one parent forbade their child from participating in the pilot, with their justification being that research instruments should be submitted to parents prior to their engagement in providing informed consent. Because of this, all teachers were instructed to share the children's instruments with parents prior to the signing of an informed consent. On average, children needed approximately 22 minutes to complete the questionnaires and parents needed approximately 16 minutes on average.

Table 1 Time needed to complete the questionnaires

	Čengić Vila I	**Grbavica II**
CHILDRE	Min. time: 12:58 Max. time: 28:26 Others: 14:22, 15:00, 15:36, 16:58, 17:00, 17:52, 20:38, 21:16, 22:35, 22: 41, 24:02, 25:07, 26: 05, 26:27, 27:41, 28:00	Min. time: 18:07 Max. time: 31:00 Others: 18:14, 22:27, 22.36, 24:12, 25:08, 25:16, 27.00, 27:11, 27:39, 27: 47, 27:58, 29.10, 29:40, 29:50, 30:40
PAREN	10:28; 10:31; 12:42; 14:18; 14:32; 15:10; 17:22; 17: 44; 17:46; 19:02; 20:13; 20:14	

Children's feedback

The child code, which required parents and children to write 4 letters based on their first and last names was seem as flawed, as the second letter of a girl's name is often A, and hence a number of identical codes were produced within the same classroom. In order to avoid this, the instruction was revised to request third letter within a name, which leads to less frequent occurrences of generation of identical codes. Furthermore, on one occasion a girl asked what to do if she is has two names. Researchers were instructed that children should use their first name as written in the school register. Some of the draft instructions for the Making decision (MD) questionnaire were identified as incomplete. For example, children recognized that some questions are focused on problem solving instead of decision making. These instruction were revised accordingly in the final instruments. Furthermore, there was a number of questions that were not clear to some children. Examples of such questions are: I stand for those I believe in; I positively relate to those things making me unhappy, I am helping others in my community (don't know what the community is); I am involved in religious groups and activities; my family set clear rules; my body speech emphasizes what I am trying to express etc. During the pilot, and after children submitted the filled

questionnaire, they have been inquired about questions that they did not understand, as well as asked about possible other ways for specifying the same questions. The suggested revisions were then integrated into the final instruments. Three questions in the instruments piloted, were identified as unnecessary because they were perceived equal to some other questions in the same instrument. For instance, the question 48 and 55 in the first questionnaire to children were specified as: "I have good neighbors that takes care of me" and "My neighbors look after me", respectively. Such repetitive questions was excluded from the instruments, while keeping only one of the questions that children selected as more appropriate.
Items CT9 and CT10, and MD7 and MD12 were recognized as different in meaning but because they were placed next to each other they confused the children. In the final instruments, these questions were moved further apart within the questionnaire.

Several children protested against question: I avoid smoking, drinking alcohol and drugs, as it was seen to be a leading question. This was adjusted so as to say 'I do/do not …[partake in this behavior].

6 questionnaires were submitted incomplete. Most of these lacked responses on the exposure to violence and manifestation of violence. During discussions, children indicated that they did not notice this group of questions, as they were on the final page of the questionnaire. Because of this, additional lines were introduced to questionnaires indicating that interviewees should turn the page and that there are more questions that must be responded to. Researchers were informed about this problem and asked to direct the attention of the interviewees to this problem and to check each questionnaire on submission. Children also indicated that they primarily returned socially desirable answers to the questions within the questionnaires on managing emotions.

Parents' feedback

Several parents complained that the introductory part is long and confusing to them. They suggested it be changed (shortened) and also that it be made clear that the child code has to be inserted on introductory page. The same problem with the child's code as with the children's questionnaires (second letter being the same across many names) was identified, stressing the need for a revision of the coding scheme. Several questions on the CBCL were identified as being similar or the same: AP3 and MG1, MG2 and MG3, RBB11 and AP4, MA3 and RBB10. However, as CBCL is a standardized instrument, identical or similar items have remained in the final questionnaire. Some questions were not clear to parents. Most of them had a problem with the question: "All children sometimes lie". This question was reformulated to "[The child] lies and cheats". Question: "He/She thinks that they can't influence the future" was assessed as being imprecise and it was recommended that this be changed to "He/She thinks that they can't influence *their own* future". Parents requested a grammar change for two questions, which was accepted as a suggestion. Questions about somatic issues have their own instruction section within the instrument and thus created confusion among parents. This was primarily related to whether they need to select an answer on the instruction part or not, so a recommendation was given that it be emphasized to parents that they don't need to select anything in the instruction part (on the basis of the fact that scores are not offered next to instructions). Socially desirable answers were given

by several parents to questions relating to sexual problems, cruelty against animals, childish behavior for their age, as well as lying and cheating.

ANNEX IV: PROPENSITY SCORE MATCHING MATRIX (TO BE FILLED IN BY THE MINISTRY OF EDUCATION)

Data to be filled for all school groups of eighth-graders. Data to be entered for each branch school separately. The total number of boys and girls in school will be equal for all school groups of eighth-graders within the same school building. Please be aware that data is required for each school building irrespective of the center of school administration.

Name of the school / branch school	Urban / rural area	Class Group	Total # of boys in school	Total # of girls in school	# of boys in school group	# of girls in school group	# of boys with disabilities	# of girls with disabilities	# of boys from Roma ethnic group	# of girls from Roma ethnic group

ANNEX V: COMPARISON OF DENSITY OF COVARIATES IN THE COMPARISON AND TREATED GROUP

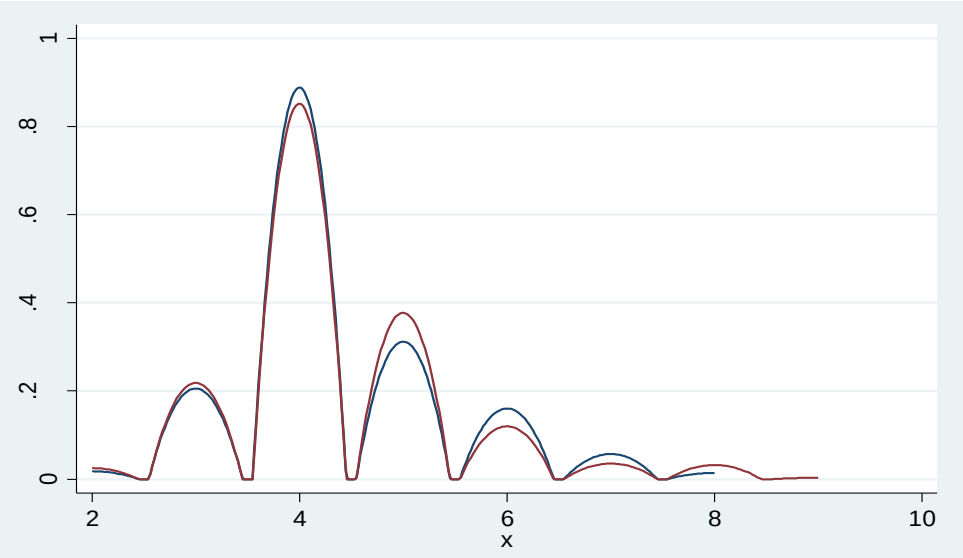


Figure 1 – Comparison of density of household size in the comparison and treated groups

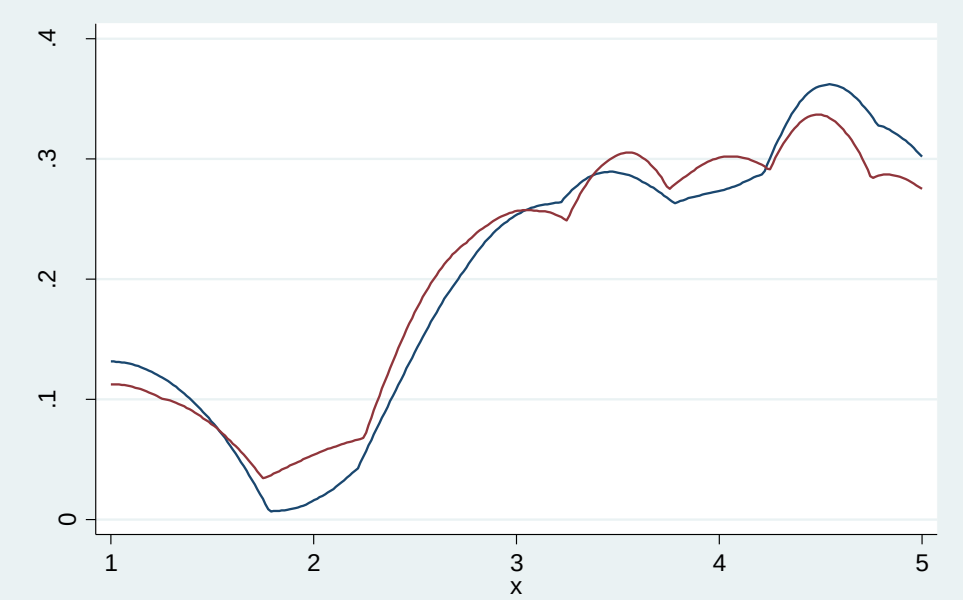


Figure 2 – Comparison of density of ACH1 in the comparison and treated groups

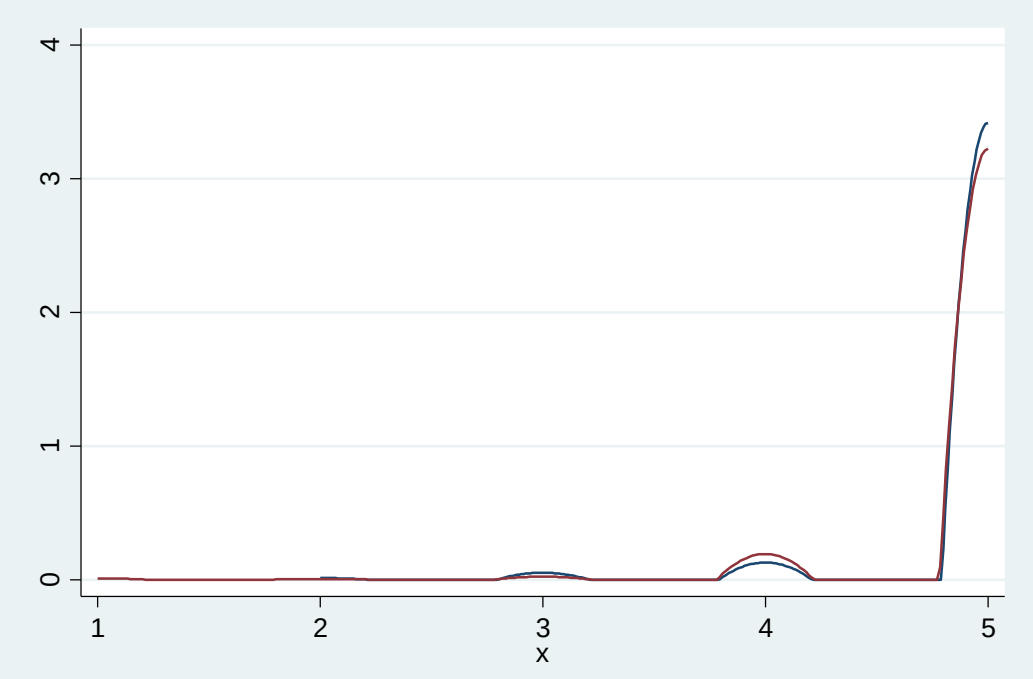


Figure 3 – Comparison of density of ACH2 in the comparison and treated groups

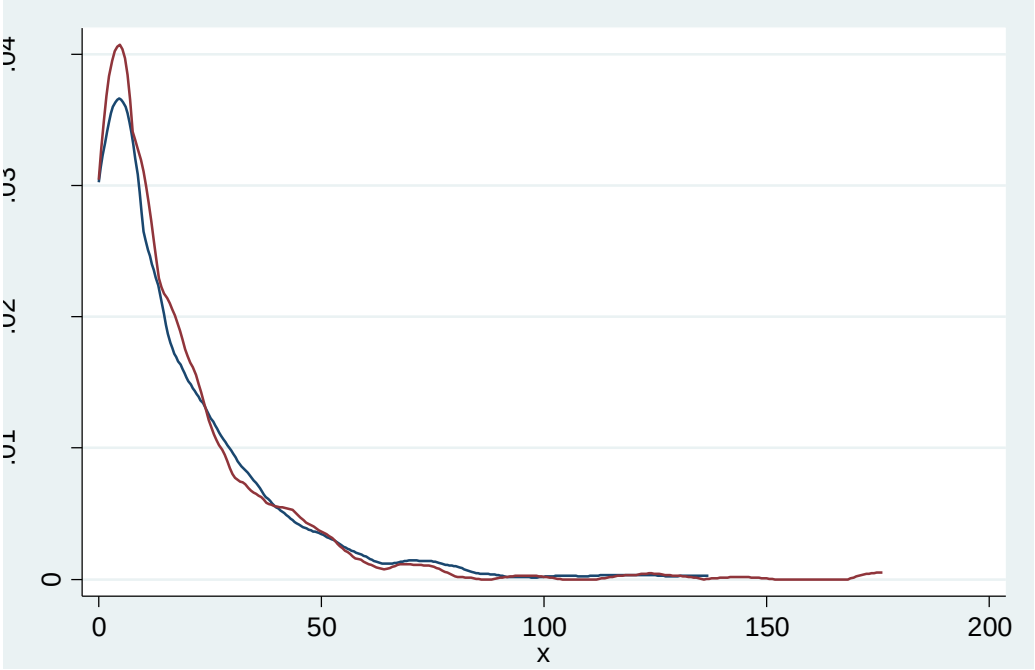


Figure 4 – Comparison of density of ACH3 in the comparison and treated groups

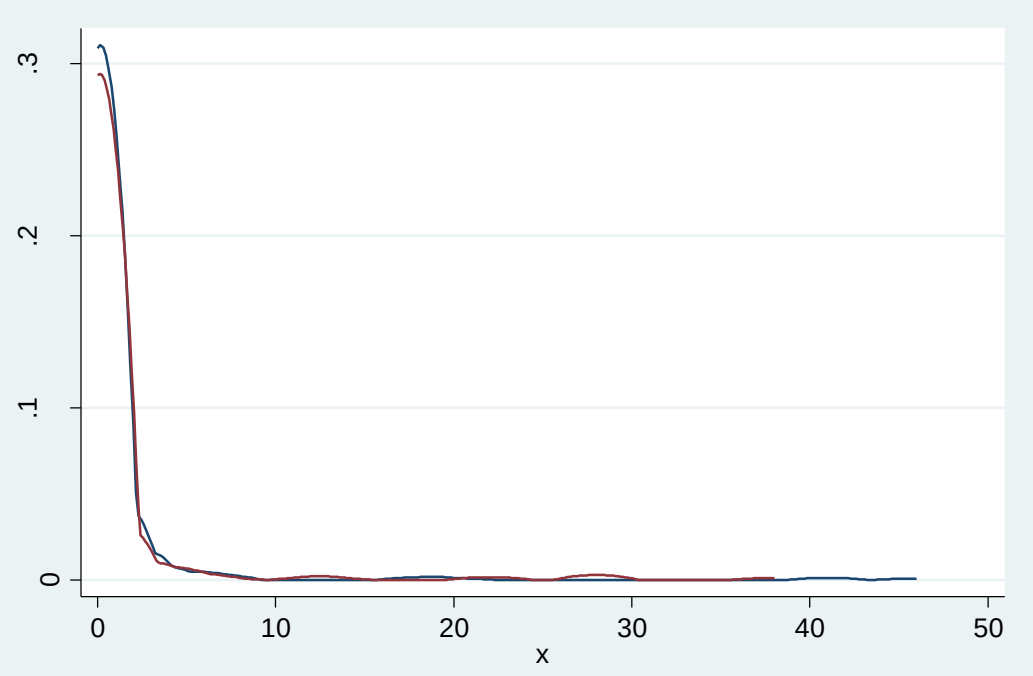


Figure 5 – Comparison of density of ACH4 in the comparison and treated groups

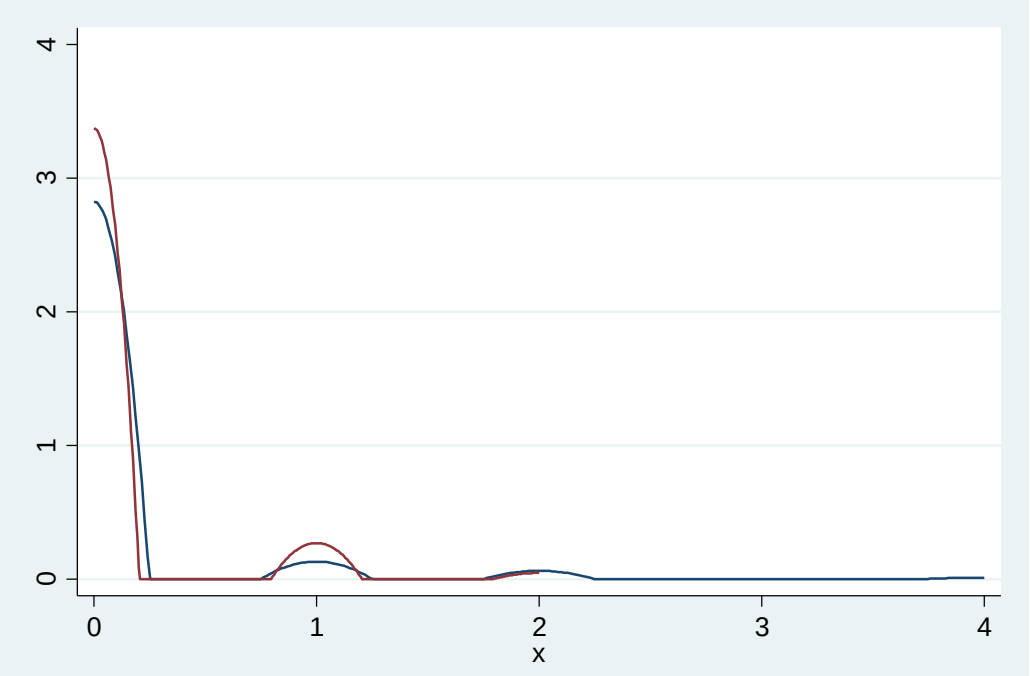


Figure 6 – Comparison of density of ACH5 in the comparison and treated groups

MONITORING AND EVALUATION SUPPORT ACTIVITY (MEASURE-BiH)

Fra. Anđela Zvizdovića 1
UNITIC Tower B, Floor 21
71000 Sarajevo
Bosnia and Herzegovina
PHONE: + (387) 033 941 676
contacts@measurebih.com
www.measurebih.com