Impact Evaluation of the Peacebuilding Educational Programme "Livingsidebyside"® in Kyrgyzstan

Revised Final Report

31 March 2017

Authors:[*1]

Anastasia Aladysheva*
Stockholm International Peace Research Institute
<aladysheva@sipri.org>

Gulzhan Asylbek kyzy
Stockholm International Peace Research Institute

Tilman Brück*
ISDC - International Security and Development Center
<tilman.brueck@isd-center.org>

Damir Esenaliev*
Stockholm International Peace Research Institute
<esenaliev@sipri.org>

Jamilya Karabaeva
former: Stockholm International Peace Research Institute

Winnie Leung
Stockholm International Peace Research Institute

Eleonora Nillesen*
The United Nations University – Maastricht Economic and Social Research Institute on Innovation and Technology
<nillesen@merit.unu.edu>

[*] Corresponding authors in this publication
[1] Authors are in alphabetical order

Acknowledgements

We gratefully acknowledge our partner research team at the University of Central Asia, Kyrgyzstan, for the implementation of this evaluation project and training courses for the data collection, namely Roman Mogilevsky, Kanat Tilekeyev, Bakhrom Mirkasimov, Aida Bolotbekova and Shoola Dzhumaeva. We would also like to thank Liina Kaldmäe, Yanming Kan and Francesca Pavan (former interns at the Stockholm International Peace Research Institute, SIPRI) for their excellent research assistance. In addition, we thank the staff members at the Center for Social and Economic Research (SOCECONIC) for their support in the collection of data, especially Shamsia Ibragimova, and the research assistants in the south of Kyrgyzstan who assisted in data collection and supervision, Gulsaadat Bayalieva and Mirbek Bayaliev.

We would like to thank our implementing partners at Legacy International, Marlene Ginsberg and Shanti Thompson, as well as Asel Kutubaeva, and Chinara Irisova at Centre Interbilim for the fruitful and patient collaboration.

Lastly, we would like to thank the International Initiative for Impact Evaluation, Inc. (3ie) for providing funding and a technical review of the project. The views expressed in this report are not necessarily those of 3ie or its members.

Executive Summary

a) Context and Relevance

Kyrgyzstan experienced a series of inter-ethnic violent clashes between Kyrgyz and Uzbek communities in 2010, mostly in the southern part of the country, that led to more than 400 people killed, over 400,000 civilians temporarily displaced and severe infrastructural damage with negative implications for the local economy. While the root causes of the conflict remain unclear it is evident that young people played a significant part in committing the violence. Following the clashes, there have been many peacebuilding interventions to restore trust, respect and cooperation between these two ethnic groups. This evaluation estimates the impact of one such initiative, a school-based peacebuilding educational training programme called "LivingSideBySide®" (LSBS) implemented in 2014 and 2015 in southern Kyrgyzstan.

LSBS has been developed and tested by LI for over three decades, covering inter-ethnic, racial and religious conflict-prone areas of Bosnia and Herzegovina, Nigeria and the United States. Yet none of these interventions has ever been rigorously evaluated. Legacy International (LI) and Center Interbilim (CI) implemented a pilot programme of LSBS with 753 students aged 13-18 in 10 treatment schools in Southern Kyrgyzstan. Each student participated in training program with 18 weekly sessions for 6-8 weeks. The LSBS programme in Kyrgyzstan is the first to be evaluated using a rigorous experimental design and an extensive set of quantitative and qualitative instruments to assess the programme's impact on promoting peace in Southern Kyrgyzstan.

b) Key Impacts of Interest

The rationale for the LSBS intervention is that increased, structured positive interactions between individuals of different groups can help people to increase their knowledge, to reduce their anxiety and their perception of threats. In addition the program helps participants to increase their empathy and facilitates taking "another perspective". This broadly relates to Allport's intergroup contact theory, where under certain 'optimal' conditions increased contact reduces prejudice.

We find some empirical support for this idea from our quantitative research. Intention-to-treat effects show that participation in the LSBS programme is robustly associated with increased levels of trust towards “seeing people for the first time”. We believe this outcome is most closely related to what the programme hopes to achieve: teaching young people to have an open mind towards anyone that may be of a different ethnic, religious or cultural origin. We also find that programme participants behave more cooperatively in a public goods game, where they played with both co- and non-co-ethnics.

At the same time, participants report to feel “less at home in Kyrgyzstan” than those in the control group. As the intervention also taught individuals to reflect on themselves and their environment and think about their own and others’ identity, this result suggests that the intervention “stirred up” something and possibly made young people much more aware of ethnic salience in every-day life and the potential challenges of dealing with this constructively.
Interestingly, some outcomes seem gender-specific. Participation for girls, for example, results in lower levels of self-efficacy for this group but also increased levels of cooperation, while there is no such effect for boys. The programme’s efforts to teach participants to become more reflective, take perspective and increase empathy may have been more effective for girls, making them realize how hard “being tolerant” really is, thereby reducing their self-efficacy while at the same time this realization may have increased their effort and willingness to cooperate in the games. When comparing results for the ethnic majority and minority groups we observe that results are typically stronger for the ethnic majority group.

We find much stronger support for a positive impact from the focus group discussions, where participants said they benefitted in terms of gaining more conflict-resolution skills, learning to accept different views and opinions, improved relations with family members or friends as well increased academic performance. One explanation for these more pronounced results compared to the survey and experimental evidence is that participants really felt that they learned a lot and that they did well in the programme, and that they hence expressed these views in the focus groups. Yet, our survey and experimental instruments measure ‘deeper’ outcomes related to beliefs and attitudes that may, on the one hand, be less amenable to short-term interventions like LSBS and, on the other hand, that are less obviously connected to the programme. At the same time, the programme may attenuate people’s tendency to give socially desirable answers, which is an expression of respect for others of sorts.

c) Recommendations

Our results from the analysis of the pilot intervention suggest that a 6 to 8-week training programme like LSBS is able to induce some intended impact among young people, inducing the children to reflect more about their own and others' identity. However, the programme maybe too short, or not comprehensive enough, to induce comprehensive and sustained changes in deeply held attitudes and behaviour. It is likely that such beliefs and attitudes are so deeply entrenched in children's upbringing, cultural and social norms, that they are less amenable to be changed by a short, school-based intervention. Addressing norm formation in the family, through the media and through teachers, for example, seem useful complementary activities when seeking to influence the attitudes and behaviour of young people.

Moreover, the programme seems to affect certain groups differently. Although the data does not permit us to investigate the underlying channels in depth, we posit that to make a large and sustained impact the programme needs to be targeted at specific vulnerable individuals that feel marginalised and may be particularly susceptible to intolerant behaviour. Due to budget limitations, the current programme curriculum was only translated into "neutral" Russian, instead of Russian, Kyrgyz and Uzbek as originally foreseen. This implied that the programme could only be implemented in Russian-speaking schools which are generally considered to be more ethnically diverse, with higher levels of ethnic tolerance and better educational resources than state-language schools in Kyrgyzstan. Implementing the programme in close-knit ethnic homogenous schools may have had very different effects. Moreover, as students self-selected into the programme, only those individuals that are interested in becoming (even) more tolerant towards other groups probably applied.

This suggests that the students who arguably stand to gain the most from a programme on peacebuilding – to become more self-confident, learn about others and thus lower their grievances towards other groups – are the least likely to be served. Incorporating the programme into the standard (national) curriculum may be one way to overcome some of these challenges.

A final comment relates to the intensity and duration of the programme. Changing attitudes and behaviour towards other groups may take a long time and it may also depend on changes in society at large. If the programme were scaled up, so should the evaluation of its impacts. This would also improve our understanding of how the programme impacted on different population sub-groups. One could think about using a large enough sample of schools that would allow testing various versions of the programme such as low versus high intensity training, short versus long programmes, intra- versus extra-curricular training, school-based only versus school- and family-based trainings, and numerous other variations in for example group size or composition.

In summary, this rigorous impact evaluation indicates that school-based peace-building interventions can have some positive impact. The study also suggests that future interventions should aim to be more comprehensive and more intense. Scaling up this pilot would offer a useful opportunity to learn about such more comprehensive approach. Given how little rigorous evidence there is in the field of peace-building for young people, subsequent studies should explore this topic in more detail.

Table of Contents

Acknowledgements 2

Executive Summary 3

Table of Contents 7

List of Figures 9

List of Tables 9

Abbreviations and Acronyms 10

1. Introduction 11

2. Intervention, Theory of Change and Research Hypotheses 12
- **2.1 Intervention** 12
- **2.2 Primary Outcomes and Impacts** 14
- **2.3 Theory of Change** 15

3. Context 18
- **3.1 Study Site and Target Group** 18
- **3.2 Country, Political, Social and Economic Context** 18
- **3.3 External Validity** 19
- **3.4 Representativeness of the Sample** 20

4. Timeline 20

5. Evaluation: Design, Methods and Implementation 21
- **5.1 Evaluation Strategy and Sampling** 21
- **5.2 Student Selection** 21
- **5.3 Sample Size** 22
- **5.4 Qualitative Sampling** 23
- **5.5 Data Collection** 24
- **5.6 Ethics** 26

6. Programme: Design, Methods and Implementation 26
- **6.1. Key Programme Elements and Programmatic Activities** 26
- **6.2. Monitoring System to Track Implementation Roll-out** 29
- **6.3. Recruitment Strategy and Take-up** 30
- **6.4. Unexpected or Adverse Events in Intervention** 30

7. Results 31
- **7.1 Baseline Characteristics** 31
- **7.2 Empirical Strategy** 33
- **7.3 Differences-in-means** 33
- **7.4 Sub-group Analysis** 35
- **7.5 Medium-term Impacts** 38
- **7.6 Results from the Focus Group Discussions** 38
- **7.7 Attrition Analysis** 40

8. Discussion 41

9. Conclusions 43

References 46

Appendix B. Sample Design **49**
Appendix H. Description of variables **50**
Appendix I. Results **53**
Appendix L. Timeline **58**

List of Figures

Figure B3: Random selection of schools 49
Figure B4: Random selection of students 49
Figure L5: Timeline 58

List of Tables

Table 1: Key outcome categories and sub-categories 15
Table 3: LSBS Programme Sample Breakdown 25
Table 4: LSBS Quantitative Panel Data Collection Breakdown 26
Table 5: Descriptive statistics at baseline (W1), control variables and outcome indicators that were measured at baseline. 32
Table 8: Programme impacts for girls using differences in means 35
Table 9: Programme impacts for boys using differences in means 35
Table 10: Programme impacts for ethnic majority 37
Table 11: Programme impacts for minor ethnic groups 37
Table 12: Probit regression results for attrition 41
Table H13: Description of variables 50
Table I14: Balance test (extended set of individual, household, and school characteristics) 53
Table I15: Programme impacts using difference-in-differences 54
Table I16: Programme impacts for boys 54
Table I17: Programme impacts for girls 55
Table I18: Programme impacts for the ethnic majority group 55
Table I19: Programme impacts for the ethnic minority group 56
Table I20: Medium-term programme impacts using difference-in-means 57
Table I21: Medium-term programme impacts using difference-in-differences 57

Abbreviations and Acronyms

CIB	Center Interbilim, Bishkek
FGDs	Focus Group Discussions
HH	Household
ISDC	ISDC - International Security and Development Center, Berlin
LI	Legacy International
LSBS	LivingSideBySide®
R1; R2; R3	Round 1; Round 2; Round 3
SIPRI	Stockholm International Peace Research Institute
SOCECONIC	Centre for Social and Economic Research, Bishkek
UCA	University of Central Asia, Bishkek
UNU-MERIT	United Nations University – Maastricht Economic and Social Research Institute on Innovation and Technology
USD	US Dollars
TOT	Training of Trainers
W1; W2; W3	Wave 1; Wave 2; Wave 3

1. Introduction

It is a necessary pre-condition for sustainable human development that people can live securely, free from fear of physical attack, political violence, or even war. However, in many countries this is not the case. Actual or possible political violence and violent conflict are, unfortunately, the reality for many ordinary citizens around the world, often expressing themselves as violent conflicts between different socio-economic or ethnic groups. Such conflict reduces people's welfare and even threatens their physical survival. Preventing the outbreak of political violence and building more peaceful societies is then both a valid goal of sustainable development in its own right and an important precondition for realizing shared prosperity and improved life satisfaction. Indeed, the Sustainable Development Goals include several key indicators (as part of its cluster 16 on peace, justice and strong institutions) for measuring progress towards a more secure and peaceful world. The innovation of these indicators is that they refer also to "peace" at the micro level. Hence, peacebuilding is not only the domain of macro-level or even global institutions like governments or the United Nations Security Council. Instead, peace is being recognized and operationalized as a concept with relevance for individuals and households.

However, preventing conflict and building peace among people is a very ambitious policy objective – and one for which significant knowledge gaps exist in the academic literature. Unlike in many other fields of development (like public health or education), there are no stylized facts prescribing how to build peace between people. In the absence of standardized policy prescriptions and in the light of the urgent need to strengthen fractured and fragile societies, practitioners have tried many approaches to strengthen peace between people.

One of the most popular approaches has been a variant of the contact theory (Allport 1954, Brück et al., 2016). This states that people from different (ethnic) groups may be alienated from each other, hence preparing the ground for possible violence between such groups. If members of the different groups are encouraged to get to know each other, for example by working together under positive circumstances, learn to take (a different) perspective then ignorance and prejudice can give way to knowledge and trust, reducing the probability of spontaneous or directed outbreaks of violence between the groups.

Such programmes have been implemented for decades with many different target groups by non-governmental organizations (NGOs), governments and international organizations in both developing and developed countries. However, the rigorous evidence for the effectiveness of such programmes is very thin.

The objective of the research on which we report here was hence to understand if and how a peacebuilding programme can be effective. Given the dearth of rigorous evidence in this field, even establishing if a programme has impact would be a valuable finding. Ideally, how programmes impact peace would be even more valuable to know.

The case we studied is of a youth programme in Southern Kyrgyzstan, a region which has experienced repeated outbursts of violent conflicts between ethnic Kyrgyz and Uzbek groups. The most recent wave of violence took place in 2010, leaving more than 400 people dead and 400 000 displaced. The issue of the relations between these two ethnic groups continues to be a sensitive topic in southern Kyrgyzstan, hence motivating two NGOs from Kyrgyzstan and the United States to implement an after-school youth training programme to bring young people from both ethnic groups into contact with each other. The programme met with great interest in Kyrgyzstan, both among local and national policy makers as well as bilateral and international agencies working in Kyrgyzstan. LI, CIB and the Kyrgyz Ministry of Education have developed plans to scale up the programme nationwide, pending funding and the results of this evaluation.

The research therefore contributes both to an emerging global debate on the effectiveness of peacebuilding programmes at the micro-level and to a specific Kyrgyz debate on how to build peace and prevent a recurrence of violent conflict in the country.

Specifically, the research project compared the knowledge, attitude and practices of ethnic Kyrgyz and Uzbek youth participating in a voluntary, after-school training programme implemented over several months in ten secondary schools in southern Kyrgyzstan. We compared the impact of the programme on the young people with a control group of students from the same ten schools. The programme was implemented in 2014 and the data was collected from 2014 till 2015. We collected survey data from the students before the programme, immediately afterwards and 12 months afterwards. We collected data on the characteristics of the students' families through household interviews and we conducted four lab-in-the-field experiments to gauge behavioural impacts related to altruism, cooperation, trust and risk-taking. We also collected a range of qualitative data through focus group discussions with students and teachers to interpret and triangulate findings from the quantitative analyses.

This report is structured as follows: Section 2 introduces the intervention, theory of change and research hypotheses. Section 3 discusses the context of the study and the country setting. Sections 4 and 5 present the timeline of the research and the evaluation design and methods, respectively. Section 6 introduces the details of the intervention. Sections 7 and 8 present and discuss the results, respectively, and section 9 provides the conclusions.

2. Intervention, Theory of Change and Research Hypotheses

2.1 Intervention

LivingSidebySide (LSBS) is an interactive peacebuilding training programme for youth aged 13-18. The programme has been developed and tested in the United States and Russia, and used with individuals and groups representing Northern Ireland, Iraq, Bosnia and Herzegovina, Georgia and Abkhazia, Nigeria, and other nationalities in previous years. Legacy International (LI), a non-profit organization in the United States collaborated with Center Interbilim (CIB) in Kyrgyzstan to implement LSBS in ten schools as an extra-curricular, after-school activity in

three southern *oblasts* (Osh, Jalalabad and Batken) in Kyrgyzstan.[2] The programme comprised two parts. The programme was implemented over three sequential rounds (R1, R2 and R3). Round 1 (**R1**) started in February 2014, Round 2 (**R2**) in April 2014 and Round 3 (**R3**) in October 2014, with 20 students trained per round per school.

Part one comprises a Training of Trainers programme (TOT). Interested teachers were requested to send their resumes to be considered for participation in the programme. Teachers were considered eligible if they (*a*) had a minimum of 3 years' teaching experience; (*b*) were fluent in at least 2 languages, including Russian; (*c*) had interest in conflict prevention; and (*d*) were computer literate. A selection committee comprising two staff members from the Bishkek and southern offices of CIB (the Project Manager and the Training Coordinator) reviewed 59 resumes and interviewed 56 teachers. After the interview, the committee selected 2 teachers of different ethnic origin per school (constituting 1 *pair* of teachers per school). This resulted in 20 teachers being selected in total for the programme. Selected teachers participated in the TOT for nine days. The materials for the training were designed and adapted to the local context by the implementing agencies. The teachers received a 9-days training course before R1 commenced and a 2-day retreat between R1 and R2, based on the feedback they received after training R1 students. In addition, the teachers' work was monitored by CIB throughout the programme. Teachers were trained in their new role as a facilitator and mentor, distinct from the more hierarchical position as a teacher transmitting knowledge to students. Facilitators are specifically being trained on empathizing with others, imparting values on how to honour and respect others, create an open, safe and equal environment and serve as a role model. Their training particularly focused on how to facilitate constructive and inclusive discussions and develop participants' abilities to verbalize thoughts, feelings and values; see things from a different perspective,; discover common ground; share personal experiences; train participants in considering new ways of thinking and behaving, and acquire new skills including for example disciplined listening skills – the abilities that may help in learning conflict prevention, resolution, and mediation. Trainers used a manual with a predefined list of activities and suggested duration for each workshop, but had some flexibility to adapt this list in accordance with the needs of the group. Trained teachers would receive additional input and feedback from CIB and LI throughout the year in which they were trained.

Part two comprises 18 workshops, each lasting two hours, were held for selected students from grades 9, 10 and 11. The programme ran from 6-8 weeks with 2-3 workshops per week. The workshops were usually held in groups averaging 20 participants, with an approximately equal share of different ethnic groups. The workshops were clustered around the following six themes: (1) building trust, identity and diversity; (2) communication and understanding; (3) tools for reducing prejudice; (4) learning how to resolve conflict; (5) mediation; and (6) cooperative action. These themes broadly relate to the least three commonly tested mediators that link intergroup contact to reduced prejudice including (i) knowledge, (ii) anxiety and (iii) empathy and perspective taking (Pettigrew and Tropp, 2008). On average three workshops were devoted to each theme. Each workshop would start with an opening remark where the facilitator would welcome everyone and explain the goals of the workshop. Teachers were encouraged to put up

[2] *Oblasts* are administrative units comparable to provinces.

a “quote of the day” as a group warm up. Fun interactive games with an instructional purpose (e.g. a group ball game that tests memory) were alternated with cooperative activities (e.g. creating a human pyramid), practical tasks (e.g. make a drawing of your community) and activities that required serious listening, reflection and discussion (e.g. ranking individual values and then comparing and discussing the differences). The final three workshops focused on developing a project plan to cooperatively conduct a community project, such as helping in an elderly home for a day, organizing a peace concert, helping handicapped children in a hospital, collecting books for a school library, making school decorations, or organizing a lecture on ethnic tolerance for the students of lower grades. In light of evidence (see for example Barlow et al., 2012) that suggests that negative intergroup contact (e.g. through everyday contact) effectively increases prejudice, it is important to note here how the intervention differs from the common daily exposure to outgroup members. Everyday intergroup contact among individuals in the control group could not be avoided as individuals of different groups are in the same school and possibly even the same class. Thus, the difference between the treatment and control groups is that the former received structured and specific guidance on interacting positively with members of different groups whereas the control individuals were only exposed to common everyday contact with out-group members.

The design of the programme stipulated that each round of the 6 to 8-week training was to be followed by a 1 to 2-day community project implementation phase. 553 students received certificates for completing the programme and 20 teachers received training certificates and a remuneration of 360 USD for their work.

2.2 Primary Outcomes and Impacts

The research team from Stockholm International Peace Research Institute (SIPRI), the United Nations University – Maastricht Economic and Social Research Institute on Innovation and Technology (UNU-MERIT), the International Security and Development Center (ISDC) and the University of Central Asia (UCA) are evaluating the impact of a peacebuilding programme among school-going adolescents in Kyrgyzstan. We are specifically interested in testing the intergroup contact hypothesis and possibly underlying mechanisms related to (*a*) knowledge *(b)* attitudes and beliefs and (*c*) student’s behaviour towards peers, teachers, outsiders, and family members, such as increased interaction with others (including those of different ethnicities, which implies higher levels of trust, acquaintanceship and conflict avoidance and mitigation).

We are seeking to assess whether these impacts differ with respect to gender and ethnic origin and whether certain impacts are more likely to spill over to non-treated persons than others. In addition, we hope to shed some light on the sustainability of the intervention: do (some) impacts sustain beyond the duration of the programme?

The outcome indicators fall into three groups: (*1*) knowledge, (*2*) attitudinal change and (*3*) behavioural change related to altruism, trust, cooperation and risk-taking. In addition, we are measuring possible changes in attendance rates and achievement as possible unintended consequences. A detailed list of outcome indicators is given in Table 1.

Table 1: Key outcome categories and sub-categories

	Key outcome category	Sub-category
1.	Knowledge	a. Knowledge about definitions b. Perceptions about intolerance
2.	Attitudes and beliefs	a. Beliefs b. Self-efficacy and locus of control
3.	Behaviour	a. Trust, Cooperation, Altruism, Risk-taking b. Fighting
4.	Unintended consequences	a. Self-reported grades b. Attendance rates

2.3 Theory of Change

Allport's Intergroup Contact Theory provides the broad theoretical basis underlying the LSBS intervention.[3] The original contact hypothesis states that intergroup tolerance may be created through increased interaction of people of different (racial or ethnic) groups within a community. The theory has, of to date, been tested in numerous applications in various settings, including two meta-analytical tests, supporting for the notion that intergroup contact reduces intergroup prejudice (Brück et al., 2016; Pettigrew and Tropp, 2006). Having established the key fact that intergroup contact works, new applications have tested to investigate the impact for groups that differ on characteristics other than race or ethnic origin and are looking into the underlying mechanisms: "how does intergroup contact reduce prejudice?" In another meta-analysis, Pettigrew and Tropp (2008) distinguish between three possible mechanisms; (i) knowledge; (ii) anxiety; and (iii) empathy and perspective taking, that have been most commonly identified and tested. Others include self-disclosure, extending one's view of the "ingroup", how important one considers the contact to be, and perceptions of a greater "outgroup" variability.

The LSBS intervention includes elements that appeal (in varying degrees) to these potential mediating channels, discussed in more detail below. The first three lessons (theme 1) are aimed to help students become more knowledgeable about themselves (i.e. their values, culture) and others (i.e. others' cultural heritage, values, norms) learn about possible (unbeknown) similarities, learn how to openly disclose their values, but also respect and appreciate intergroup differences and resulting cultural diversity. The various interactive "fun" activities that are part of each workshop session typically aim to reduce anxiety to interact with (relative) strangers generally and with people of different groups specifically. Lessons from theme (2) –(6) appeal more to empathy and perspective taking, where participants practice transformative thinking related to recognizing prejudice, communication and conflict resolution (e.g. participants are

[3] Note that for contact theory to be effective, Allport (1954) specifies four necessary conditions: equal status during the contact phase, common goals, intergroup cooperation and authority support. The LSBS education module is expected to fulfil these conditions through and throughout its programme design. First, as members of the same study group, the adolescents are assumed to be treated on an equal basis during the contact phase by their teachers. Cooperative projects that are implemented at the end of the study module provide common goals and opportunities for intergroup cooperation. Finally, authority support is assumed to result from the high quality of training received by the teachers, support from the school administration and parents, proper monitoring during the training process, and the youth's voluntary participation.

trained in mediation as an alternative to violent resolutions of conflict, and practice intergroup interaction in a safe, neutral environment). Although we acknowledge that the intervention would possibly appeal to multiple mediation process, we are unable to cleanly test which of these mechanisms matters most in our context, as the evaluation design only permits us to test the full programme that encompasses many multiple mediating elements. The main focus of the training, however, seems to have been on perspective taking; where students are actively trained on seeing things differently, to "transform" their thinking and propose new solutions to "old" problems

The ultimate goal of the programme was to provide a framework for positive attitudinal and behavioural change through intergroup contact, so that an individual chooses cooperation and understanding rather than violence or hatred towards individuals of different origin (gender or ethnicity). The connecting link between inputs and activities on the one hand and outputs and outcomes on the other hand is making people aware of their own (in)tolerance, comprehend the importance and need for tolerance, learn about their own identity, provide accurate (historical) information about other groups and increase empathy, encourage seeing things from a different perspective, but also provide hands-on practical skills on conflict mediation and effective communication. Naturally, this is not necessarily a sequential process, even though the LSBS programme is set up such that raising awareness, and learning about one's and others' identity precedes workshop sessions about how to respect and appreciate people's different ethnic or cultural heritage. Yet one could imagine multiple feedback loops from say learning conflict mediation skills to an increased awareness of the need to be tolerant towards other groups.

Also, we report on outcomes rather than final impacts. Final impacts would for example refer to reduced levels of violence at school or within the wider community, changing social norms, changes in social networks, different school curricula etc. Although these are clearly part of the theory of change, they are beyond the scope of this evaluation.

Finally, the programme may have had unintended (positive or negative) consequences on achievement and attendance of participating students. If the training makes students more self-confident, feel more at home at school, and motivated to achieve something, they may be more inclined to attend classes, do their homework and hence perform better. On the other hand, the training always took place after the ordinary school hours and could have possibly substituted for the time students would normally spent on their homework, thereby lowering achievement. Moreover, if the training alienated students from their school, they may become less motivated to attend classes. We investigate the possible unintended impacts of the programme on self-reported achievement and attendance rates.

2.4 Related literature

The literature on intergroup contact has become extensive in recent years and meta-analytic evidence supports a negative relationship between contact and prejudice towards other ethnic and racial groups with larger mean effects for experimentally designed studies (see Pettigrew and Tropp, 2006; 2008). Moreover, intergroup contact theory seems to be generally applicable

to a wide variety of field settings, including stigmatized outgroups like transgender (Broockman and Kalla, 2016) peace education programmes (Svensson and Brouneus, 2013; Kelleher and Ryan, 2012; Green and Wong, 2008) or attendance to a wilderness course (Green and Wong, 2009). Nigmatov (2013) is one of the few studies focusing on central Asia (Tajikistan and Kyrgyzstan) using peace theatre workshops to bring people in contact with people from different groups.

The study also contributes to the larger literature on prejudice reduction. Indeed, there are many other intervention types, most common in social psychology, aimed at reducing prejudice that do not necessarily rely on intergroup contact, but for example on (group) instruction and discussion, reading (e.g. Biton and Salomon, 2006; Pouezevara, Costello and Banda, 2013; Cameron and Rutland, 2006) media/entertainment, and cognitive training. Paluck (2009) for example investigated the role of mass media in reducing prejudice and stereotypes. Paluck (2011) shows that training peers to intervene when they observe prejudiced speech and behaviour among classmates, makes these peer trainers more likely to confront prejudice and helps spreading tolerant behaviour to friends and acquaintances. Also, the study by Zainal, Abu and Mohamad, 2010) does not focus on intergroup contact but whether an undergraduate course about ethnic relations changes student's attitudes and behaviour towards members of a different ethnic group. Yet, in contrast to the findings on evidence of intergroup theory, the effectiveness of other style interventions is less conclusive. A review by Paluck and Green (2009) of observational, laboratory and field experimental literature, where they review studies above shows that there are few studies that can causally infer an impact from the programme they evaluate. Recent research however is more promising. Devine et al., 2012 for example use a randomized laboratory setting that offers treatment students a three-months long programme that includes education about the linkage between implicit racial bias and discriminatory behaviour and provides training on various strategies that can help reduce such bias in daily life. Increasing contact with out-group members is (only) one of a variety of bias-reducing strategies that is offered. They find that their multifaceted intervention increases people's awareness of their own (racial) bias, makes them more concerned about racial discrimination in society in general and the overall deployment of strategies is associated with an implicit bias reduction.

Lastly, the study contributes to the literature on the effectiveness of prejudice reduction programmes targeted at children and adolescents. Here, the evidence seems even more mixed. Aboud et al. (2012) for example finds that some 40 per cent of studies evaluating the effects of interventions aimed at reducing ethnic prejudice and discrimination in children showed positive results, 50 per cent non-significant outcomes and 10 per cent negative effects. [45]

[4] The mixed findings in the broader literature on prejudice reduction and those specifically targeted at children are also plausibly related to the various research designs employed to study reduction in prejudice. Study designs vary from small mostly qualitative set-ups, to laboratory settings and full-fledged randomized field experiments. Yet small-scale qualitative studies often measure a very specific subset of participants and tend to suffer more from social desirability bias than anonymous surveys. These studies are typically less suited to draw on a valid comparison between participants and non-participants and have little to say about general patterns in the data. Fully controlled laboratory experiments on the other hand usually lack the necessary context. Field experiments are both capable of providing the relevant context and have a valid counterfactual, but are costly and often difficult to implement due to both ethical and practical considerations. Although now routinely used in some domains of development economics

This study thus contributes to the small but growing evidence base of the impacts of peace education targeted at children and adolescents and adds to the empirical literature on intergroup contact. In addition to the limited evidence available worldwide, the topic has surprisingly received little attention in research focusing on post-Soviet states despite a dramatic increase in ethnic tensions during and after the dissolution of the Soviet Union. We know of only one study that evaluates the impact of a youth "Theater for Peace" programme in Kyrgyzstan and Tajikistan (see Nigmatov, 2013). Yet the non-random assignment of treatment status complicates a rigorous assessment of the causal impact of the programme.

3. Context

3.1 Study Site and Target Group

Kyrgyzstan experienced a series of conflict clashes during April to June 2010, in light of the political turmoil since the removal of the President Bakiev from power. The violence was mostly manifested in Jalalabad and Osh cities in southern Kyrgyzstan. The largest violent conflict occurred in the city of Osh in early June 2010, in which around 470 people, mostly Uzbeks, were killed. About 400,000 fled temporarily from their homes, some to neighbouring Uzbekistan, and a large number of properties were destroyed. The actual trigger of these events is still unknown as the conflict did not appear to be spontaneous. It may have been prompted by a series of coordinated attacks carried out by separate groups of armed men. Yet we do know that the role of the youth in the 2010 conflict was rather pronounced. Young people widely participated in the violence in 2010.[6] This motivated the (inter) national policy community to design and introduce systematic conflict prevention programmes among youth and implement training programmes that promote conflict prevention, tolerance and inter-ethnical understanding.[7]

3.2 Country, Political, Social and Economic Context

The Kyrgyz Republic – a landlocked mountainous country in Central Asia with a multi-ethnic population of 6 million as of 2016 – is one of the poorest countries in Europe and Central Asia, with an income per capita of 1,103 USD in 2015.[8] Prior to the 2010 events, Kyrgyzstan

---

including health, education, agriculture and (micro)finance, there are few field experiments in peace education studies, especially in developing and post-conflict settings.

[5] Although non-experimental, Lustig (2003) evaluates a peace education programme in Israeli high schools targeted at teenagers aged (16-17) and finds that learning about 'another' conflict in Northern Ireland helps treatment students in perspective taking when reflecting on the Israeli-Palestinian conflict and connotate peace with more positive terms than non-participants.

[6] Neil Melvin, Promoting a Stable and Multiethnic Kyrgyzstan: Overcoming the Causes and Legacies of Violence, March 2011, Central Eurasia Project. The report is available at: http://www.opensocietyfoundations.org/sites/default/files/OPS-No-3-20110305.pdf

[7] Source: http://edu.gov.kg/images/report_nasilie.pdf

[8] Source: http://data.worldbank.org/country/kyrgyz-republic. Last visited: September 23, 2016.

experienced an impressive overall decline in poverty over the past decade. Aggregate poverty fell from 40 per cent to 32 per cent during 2006 – 2009. However, the 2010 violence led to a reversal of these gains due to a disruption in economic activities including trade with neighbouring countries, the destruction of business assets, and the loss of lives.

In response to the unrest in southern Kyrgyzstan in 2010, bilateral and multilateral donor agencies initiated a number of development aid programmes and interventions in the country. A range of peacebuilding programmes from inter-ethnic dialogue and conflict resolution to local economic development also began to emerge. However, these initiatives tend to be limited by short project durations and restricted geographic coverage; they mainly target the urban and peri-urban areas that were directly affected by the violence, disregarding the rural areas where poverty and social tensions are concentrated. [9]

Despite an increase in development aid and projects after the 2010 violence, little is known about the impact the aid money or programmes may have had. Our implementing partners, LI and CIB, by contrast were very keen on having their project rigorously evaluated and had extensive consultations with key staff in the Ministry of Education to discuss the design and implementation of the programme.

3.3 External Validity

LSBS is a standardized extra-curricular training programme that has been implemented with many individuals from various post-conflict countries over the past three decades. Although standardized in nature, the programme can be adapted somewhat to the local context. Indeed, LSBS shares many features with other peace education programmes (including LSBS programmes and other similar interventions applied outside of Kyrgyzstan), suggesting that there is scope to implement the programme in different settings, both within Kyrgyzstan and elsewhere.

The study area comprised three southern oblasts of Kyrgyzstan (Osh, Jalalabad, and Batken). These oblasts were selected to represent the areas with the highest percentage of minority group populations (Uzbek and Tajik) and those most affected during the ethnic conflicts of 1990 and 2010. The selected *oblasts* constitute over 50 per cent of the total population in Kyrgyzstan and the project covered both urban and rural areas. This allowed us to select a broadly representative sample from the population.

The schools that were considered eligible to receive the programme provided for a setting with a large pool of youths from different ethnic groups: Kyrgyz, Uzbeks, Tajiks, Uighurs, Russians and other ethnicities. Moreover, LI also required the teams of teacher-trainers per school to be ethnically mixed. We thus expect our full sample results to reflect knowledge, attitudes and behaviour that is broadly representative for the variety of ethnic groups present in Kyrgyzstan.

[9] SIPRI Working Paper, Evaluating peacebuilding interventions in southern Kyrgyzstan, 2014 [https://www.sipri.org/publications/2014/working-paper/evaluating-peacebuilding-interventions-southern-kyrgyzstan]

We of course acknowledge the fact that some or all results may be less, or more pronounced for specific groups.

Notwithstanding the relatively standardized curriculum, the representation of both boys and girls, and all ethnic groups present in Kyrgyzstan, we realize that the external validity of our findings may be comprised by the non-random selection of the study site, and the self-selection of students into our programme. The mixed method employed here facilitates the interpretation of our findings and allows us to shed some light on specific contextual factors that may underlie the results and predict the success of LSBS in future applications.

3.4 Representativeness of the Sample

The selected study sample is representative at the regional and national levels in terms of gender and ethnicity. However, due to budget constraints the curriculum could only be translated in one (Russian) instead of three (Kyrgyz, Uzbek and Russian) languages. After consultation with the Ministry of Education, it was decided to have the curriculum translated in Russian, as that was considered the most "neutral" of the three languages. Yet this implied the programme could only be implemented in schools where Russian is the main language of instruction. This inevitably provided a somewhat selective setting, also because it restricted the programme's appeal to those students fluent in Russian.

In terms of languages of instruction, our study sample of 20 schools represents: 12 schools with only Russian as the language of instruction, 1 school with both Russian and Uzbek as languages of instruction, and 7 schools with both Russian and Kyrgyz as languages of instruction.

The three selected oblasts in south Kyrgyzstan differ somewhat in comparison to the northern part of the country, which hosts a larger share of Russian speakers, and has a larger urban population (including the country's capital). State language schools in Bishkek account for only 4 per cent, for example, whereas this figure is substantially higher in Osh and Jalalabad. The Republic-wide testing (ORT) results report that schools with Russian as the main language of instruction have higher test scores (numeracy, literacy and life skills) than schools that use a different language. An insufficient supply of textbooks in Kyrgyz and other minority languages, an insufficient number of teachers available that are fluent in Kyrgyz and a lower level of professional capacity in general are believed to account for these differences.[10]

4. Timeline

[10] See more here https://www.opensocietyfoundations.org/sites/default/files/education_development.pdf.

Fieldwork data was collected from January 2014 to June 2015. This was done in accordance with the three main rounds (R) of the LSBS educational programmes: **R1, R2** and **R3,** i.e. the project interventions. For each round, we panel data on participants through different waves (W), namely: baseline (W1), follow-up (W2) and 1-year follow-up (W3), in order to capture short- and medium/longer-term impacts.

Each round of the LSBS educational programme lasted for a total of 6-8 weeks (R1: February - April; R2: April – June; and R3: October – December 2014). Data cleaning and code checking has been an on-going process throughout the study period (see Figure L3, in Appendix L).

Household (HH) baseline questionnaires for R1 and R2 were simultaneously conducted after the launch of the R1 and R2 interventions. This was unfortunate, as it potentially has led to some contamination of our R1 and R2 baseline values.

5. Evaluation: Design, Methods and Implementation

5.1 Evaluation Strategy and Sampling

The 10 schools where the LSBS programme was implemented were randomly selected from a non-random sampling frame of 31 schools. The list of 31 schools was compiled by CIB using the following selection criteria: (*a*) the school was based in a conflict-prone, multi-ethnic community (in one of the Osh, Jalalabad and Batken oblasts); (*b*) the school was multi-ethnic and public (requiring the presence of at least two ethnic groups); (*c*) the school had a sufficient number of senior grade students; and (*d*) the school's main language of instruction was Russian.

Since 10 schools had already been selected for treatment prior to the engagement of the research team in the study, we could not use an 'ex-ante' matched pairs randomized design.[11] In this report we therefore report estimation results for the sample of treatment schools only, where randomization took place at the individual level.
An alternative to restricting the estimation to treatment schools only is to use a difference-in differences design with matching, where matching between treatment and control students across schools is based on individual key variables. This is however not the focus of the current report.

5.2 Student Selection

The programme sought to select a group of adolescents mixed in terms of gender and ethnic origin. The programme organizers announced a call for applications for all students in grades 9-11 in treatment schools to participate in the training, during R1, R2 or R3. All students in grade 9-11 who submitted a complete application form were eligible to participate in the programme.

[11] Note that a cluster stratified randomized design in favour of an 'ex ante' pairwise matched design would not have been feasible due to the low number of clusters we could sample from.

Applicants were requested to provide basic personal information and to explain their motivation to participate in the programme. Applications were reviewed and allocated to one of four strata: (*1*) female Kyrgyz; (*2*) male Kyrgyz; (*3*) female non-Kyrgyz; and (*4*) male non-Kyrgyz. We randomly selected five students per stratum per school through a public lottery.

We also randomly selected 40 within-school control students in each school based on the same principles, to be able to measure potential within-school spill overs. Since R1 and R2 were implemented consecutively during Spring 2014, the application process was done jointly for these rounds with a few additional applications right before R2.

R1 and R2 controls were automatically added to the pool of applicants in the next round (R2 and R3 respectively) for which we again randomly selected our participants using the stratified sample. We ensured that the pool of control students was large enough such that we stayed with a sufficiently sized group that did not receive treatment during the study evaluation period. The selection of within-school control students in addition to control students from control schools explicitly allows us to estimate the potential presence and magnitude of spill over effects that may have occurred within schools. Survey questions related to violence and ethnic tolerance at home were used to gauge potential programme spill overs outside of the school.

By the time the research team started engaging in the research design discussions, the 10 treatment schools had already been randomly selected for treatment from a list of 31 schools that met the requirements of the programme. Another 10 selected control schools received the same LSBS promotion campaign as the treatment schools but were informed they would receive treatment only after the evaluation had been completed, and only if there was a sufficiently large number of applicants and available funds. As a token of appreciation for participating in the research, control schools would each receive a projector (worth about 500 USD). We considered this the only feasible way to motivate staff members of control schools to encourage pupils to apply and thus obtain a sample of similar types of applicants; the required "active" registration to be considered eligible for treatment precluded the use of a simple random sample of any 9^{th}-11^{th} grader in control schools. We sampled 20 students per control school, 5 per stratum for R1 and R2. For R3 we aimed to sample a new set of 20 students from each control school but some 25 per cent of the "new" applicants had already applied and were selected as controls during R1 or R2, leaving another 15 newly selected students for R3.

5.3 Sample Size

Power has been calculated to determine the minimum required number of students per schools. The intervention budget and ethical considerations did not allow us to select more than 10 schools: in each control school, we needed to make a credible promise that they would receive treatment after the study period (conditional on satisfying the criteria described above). We specify alpha=0.05, J=20 (the number of clusters), n=60 (students per school), rho (the intra-class cluster correlation), we vary at 0.05 and 0.15.

Recent meta-analyses by Wilson, Gottfredson and Najaka, (2001), Wilson et al. (2003) and Durlak et al. (2011) report average effect sizes of about 0.4 for school-based violence prevention programmes and school-based universal social and emotional learning programmes with effect sizes ranging from about 0.3 to 0.8. Given the fact that we use repeated measures for our main outcome variables (baseline and mid-or endline) data we are able to detect effects sizes between 0.36 and 0.56 (depending on the intra-class correlation) with power at 0.8.

5.4 Qualitative Sampling

We conducted focus group discussions (FGDs) with both teachers and students. FGDs with teachers took place in January and June 2014, thus before and after the TOT. All 20 teachers selected for the TOT were invited to the FGDs, and all but one teacher participated. The 19 participants were divided into groups of two. A research team member per group would lead the discussion by posing a question to the group and stimulate group members to openly discuss their thoughts and ideas.

Teachers were asked open-ended questions related to inter-ethnic conflict in their country, but also their possible experience with it in their own communities and schools.[12] Other questions pertained to values and social norms and concepts like 'ethnic tolerance'. The FGDs in June additionally probed for teachers' perceptions about the outcomes of the programme and its effectiveness. At the end of the FGDs, teachers were asked a number of individual questions related to teaching experience and the general education system.

The FGDs with youth were conducted in April 2015, during the 1-year follow–up of R1 and R2 participants. We held FGDs in three treatment and two control schools located in Osh and Uzgen. Students were randomly selected from treatment and control groups and we checked that all groups were balanced in terms of gender and ethnic representation. We had a total number of 68 students (43 control and 25 treatment students) participating in the FGDs, with the group size ranging from 6-12 students. We held separate discussions with both treatment and control students in treatment schools as well as control students in control schools. Yet, since our current analysis focuses on treatment schools only, we only report FGD results for the three treatment schools.

The language of all FGDs was mainly Russian, except for a few cases where some students felt more comfortable speaking Kyrgyz. Only one session was entirely conducted in Kyrgyz at the students' request. In order to facilitate a "round-table discussion" the team set up the students in a square shape. We had one member of the research team leading the discussion while a second member took notes.

The interviewers had two sets of focus group questions – one for treatment students and one for control students. Questions for the treated students asked about their opinion regarding the LSBS training programme and possible changes they experienced as a result of the programme; the community projects they implemented as a part of their assignment; future

[12] All FGDs questions are provided in the Appendix of the report.

career aspirations; personal perceptions about issues of ethnicity, religion; and trust and cooperation during the trainings. In the meantime, the interviewers asked the control students the same set of questions excluding those specifically related to the programme.

In order to analyse the data from the FGDs with teachers and students, we grouped the responses by the main subjects that we identified during the discussions. Note that due to the small number of observations (20), the TOT has been analysed based on the FGDs alone.

5.5 Data Collection

This study uses both primary and secondary data. We describe the relevant data collection activities below and refer to Appendix C for a copy of the survey format, FGDs guidelines and the protocol and answer sheet used for the behavioural games. The evaluation comprised three rounds of the intervention, with midline and 1-year follow-up data collected at different points in time, dependent on the intervention round (also see figure L1, in Appendix L).

After a competitive bidding procedure, we selected SOCECONIC (Center for Social and Economic Research), a survey firm based in the country's capital Bishkek, to collect the data in both treatment and comparison schools.

Table 2 presents the different data collection instruments.

Table 2: Data collection instruments

Data collection instruments	Baseline	Endline	1-year follow-up
Household survey	X	X	
Student survey	X	X	
FGDs students		X	
FGDs trainers	X	X	
Behavioural experiments		X	X
School administrative records			X
Network survey			X

Participating households would receive a box of tea to compensate for their time. Students would be compensated through participation in the behavioural experiments in which they could earn tokens that we would transfer into mobile phone credit at the completion of all games. Teachers were not given any additional compensation, except for the additional salary of 360 USD that they were paid by the implementing agency for participating in the intervention.

The data collection team was comprised of 24 professional enumerators, split into groups of eight that included one supervisor (typically a senior enumerator with many years' experience). We trained the enumerators for 3 days on survey and interviewing techniques, and provided detailed instructions and practice conducting behavioural experiments. The training included instruction sessions followed by informal tests for comprehension, and role-plays. Although we could not keep the treatment status secret from the supervisor of the research team (he or she would have to ask for official approval from principals of the schools by showing the letter from

the Ministry of Education), this information was kept from the rest of the team. Also, students were not informed about the intervention by anyone from the research team. To further minimize the possibility of Hawthorne effects, the research team made sure they would never be present at school on the same day as the implementing team. Questions were also framed neutrally and (except for mid-and endline data) never made any specific connection to the programme.

Two local representatives of the research team received daily updates from the survey supervisors during the data collection periods and would do random checks by visiting schools and households during one of the data collection phases to monitor the quality of the data collection team. In addition, they would perform random consistency checks to ensure high data quality. Two experienced senior researchers at SIPRI led the data cleaning process, and informed SOCECONIC regularly and systematically to cross-check inconsistent, implausible or missing data with the raw files.

Table 3: LSBS Programme Sample Breakdown

	Treatment schools			*Control schools*	
	R1	**R2**	**R3**	**R1 & R2**	**R3**
Applied to LSBS	651	149 (651)	491		
Selected for LSBS	249	230	269		
Selected for control	402	321	222	251	228
Pure reserves	12	5	9		
Refused/Rejected LSBS at the beginning	36	0	67		
Stopped LSBS participation	23	45	4		
Completed LSBS	178	185	189		

Note: Pure reserve, rejected and stopped students are also included in the "Selected for LSBS category". Control students in R1 and R2 were added to the pool of new applicants to be considered for treatment in R2 or R3 respectively.

Table 4: LSBS Quantitative Panel Data Collection Breakdown

		Treatment schools				*Control schools*	
	Wave (W)	**R1 & R2**		**R3**		**R1 & R2**	**R3**
		Treatment students	**Control students**	**Treatment students**	**Control students**	**Control students**	**Control students**
Student Questionnaire	*W1: Baseline*	425	236	240	199	210	212
	W2: Follow-Up	370	179	n/a	n/a	198	n/a
	W3: 1-year Follow-Up	288	157			170	
Household Questionnaire	*W1: Baseline*	422	220	241	191	209	210
	W2: Follow-Up	370	179	211	182	198	185
	W3: 1-year Follow-Up	348	215			175	
Behavioural Experiments	*W2: Follow-Up*	306	117	214	148	183	159
	W3: 1-year Follow-Up	267	126			146	
Social Networks Questionnaire	*W3: 1-year Follow-Up*	288	552			170	
School records data		366	230	223	177	178	223

5.6 Ethics

Since our study includes surveying young people aged 13-18 years on the sensitive topics of conflict, ethnic and religious tolerance, we asked for and received ethical approval from the Canadian Institutional Review Board Services.[13] We also received ethical approval from the Ministry of Education of Kyrgyzstan and its regional offices. Following that, the Ministry of Education issued a letter for its regional offices and school principals to obtain their respective oral consent to undertake research activities in their catchment areas or schools. We obtained written consent from parents and youth to participate in the study.

We ensured that responses were confidential and explained that all data would be anonymized. Any information that could link a respondent's identity to their unique identification code was only accessible to the core research team and solely used to track the respective individual during monitoring and follow-up. The data that will be made publicly available does not contain this information.

6. Programme: Design, Methods and Implementation

6.1. Key Programme Elements and Programmatic Activities

Legacy International (LI), a U.S. not-for-profit, non-governmental organization established in 1979, is dedicated to promoting peace by strengthening civil society and fostering a culture of

[13] The Board's webpage: http://5ae.63e.myftpupload.com/

participation worldwide. LI trains community leaders, youth, professionals, and governmental and non-governmental administrators, helping them to develop and implement practical, community-based solutions to critical issues. The work of LI spans 105 countries to date.

Center Interbilim (CIB) is a Kyrgyz inter-regional, non-governmental public association created in 1994 in Bishkek. The mission of the Center is to build a just society in the Kyrgyz Republic through the promotion of democratic principles and capacity building of civil society and democratic institutions. CIB has worked on areas such as development, human rights, youth, and peace education. The Center has a branch office in the city of Osh in the south of the country.

The content of the programme or policy

Prior to the intervention, the training curriculum was prepared and adapted to the Kyrgyz cultural context by LI. Then it was translated from English into Russian, reviewed by cultural adaptation experts in Kyrgyzstan, and published with the help of CIB.

Programme implementation involved two training groups. Both implementing agencies selected 10 pairs of qualified trainers, one pair per school. Selected teachers had to be fluent in Russian and at least one other language spoken in Kyrgyzstan (Kyrgyz or Uzbek); be generally respected by students and fellow teachers; show a keen interest in peace education; and have the ability to commit during the requested period of time to the programme. Any previous experience with interactive learning methodology was preferred. However, very few of the 20 teachers had such experience before joining the programme.

Their training and support included the following elements: (*i*) selection and orientation, and a 9-day Training of Trainers (TOT) programme; (*ii*) a minimum of 12 hours of coaching and mentoring during the first year of implementation, refresher training events throughout the year, and an evaluation after which all teachers received their Basic Level Certification (which allowed them to implement the programme with teenagers). The research team attended some of the TOT sessions to observe the structure, contents and process of the training, but played no role in delivering the TOT. Since most of the teachers did not have prior experience with interactive methodologies, the TOT provided clear hands-on exercises to accommodate for this. CIB monitored the work of the teachers as they implemented the workshops via weekly written reports during R1 and in-person monitoring, coaching, and evaluation in all three rounds. LI, in turn, monitored all of CIB's work, via weekly Skype conferences, review and approval of draft documents and planned procedures, approval of personnel selection, and in-person supervision visits.

The training in Kyrgyzstan was delivered through trainer-trainee interaction in two-hour after-school interactive workshops in a classroom environment. Workshops took place two or three times per week, as the school calendar allowed. The training sessions consisted of structured interactive learning activities, such as games, discussions, teamwork challenges, reading, questionnaires, and skill practice exercises. The curriculum for each class of 20 students was

delivered by a pair of trainers representing two different ethnicities, who were selected from the teachers in their respective schools. At the end of each round of the training, the students were guided to develop and implement a school or community project, working in multi-ethnic groups and serving multi-ethnic audiences, to demonstrate and practice the skills learned during the training. School or community projects were initiated and implemented by students on a voluntary basis with no financial support from the implementing agency or grantee.

In total, 553 students implemented 60 community service projects in the 10 school communities. The projects covered a wide range of activities, such as cleaning and planting flowers in community and school areas; charity work and leisure programmes for orphanages and elderly homes; intercultural festivals and sports/ games to promote peace and universal values; and helping elderly people that live alone by cleaning their houses and gardens.

The partnership collaborated with the Ministry of Education and school administrations in three southern regions (Osh, Jalalabad, and Batken), in order to gain access to schools, select teachers and students, gain access to participant families and implement of local projects.

The setting in which the programme content was delivered to beneficiaries

The alpha version of the curriculum was in Russian; thus, treatment schools were selected among those that used Russian as the language of instruction, and which had Russian-Uzbek, Russian-Kyrgyz, or Russian-Tajik populations. The training sessions were organized in the students' own schools, after school hours. The organizers provided snacks during sessions and the trainees and trainers were supplied with all the necessary materials (e.g. manuals and stationary).

Since the intention of LI and CIB was to scale LSBS nationally, it was imperative for parents and the school administrations to be introduced appropriately and get their cooperation and support right from the start of the programme. CIB worked with the Ministry of Education to select appropriate schools for the implementation, and to provide a sound basis for sustainable cooperation with the school administrations. Staff from CIB's regional office in Osh and the national office in Bishkek met with key persons to introduce the programme and garner support from the administrations, teachers, and students.

In addition, CIB held a briefing meeting with parents to inform them of the goals and practicalities of the training sessions and obtain their consent to include the prospective student(s) in the programme. By informing the parents beforehand about the aim, CIB and LI hoped that parents would encourage their children to sign up for the programme and attend regularly. Snacks were also provided, to encourage participation, and to refresh students after a long day at school.

The materials or technologies required for the programme

In addition to teacher and student manuals, each pair of teachers received a Teachers' Kit, which included all the supplies required for the workshop activities: stationary (e.g. markers,

scotch tape, flipcharts, colour papers, colour pencils) and other objects (such as balls, whistle, rubber bands, thread, etc.). This was provided to support the innovative and experiential aspect of the training. Each pair of teachers also received a "snack kit" to serve hot drinks and provide small crackers, cookies, etc. During the recruitment phase, brochures and application forms were distributed to potential students and their parents to communicate information about the programme. Students were also provided with pens and notebooks.[14]

Whether any activities were proscribed among the treatment or comparison group

Although programme participation was voluntary throughout the duration of the training, trainers used a detailed protocol that prescribed the order and content of activities, and their suggested duration, and participants were expected to take part in every activity. The control group of students did not receive or participate in any activity other than the initial application to the programme.

6.2. Monitoring System to Track Implementation Roll-out

In general, the programme was implemented as it was initially designed. A minor modification of one lesson of the programme was made, splitting it into two sessions, once the delivery format and timeline for the students' project was worked out.

With the funding they had secured, LI and CIB's original plan was to do one TOT, followed by two rounds of training with students in grades 9-11, and then do a second TOT, followed by one round of training in five new schools. Due to the lack of additional financial support during the required time period, R3 training was conducted in the original 10 schools, with the same set of trained teachers, but with a mainly younger audience from a lower grade. On average, the students in the lower grade demonstrated more difficulty in understanding and using the concepts presented to them and hence may not have been able to take full advantage of the learning experience, whereas the grade 9-11 students were able to grasp these concepts more easily.

The programme protocols were prescriptive to ensure that every student in the programme was exposed to the same concepts and skills, and to participate in the same learning activities in the same conditions and circumstances, no matter what school they were in or who their trainers were. Minor variations in the experience from one workshop group to another were based not on different learning exercises, but on student interest and exercise results: different examples might be given by students; discussions of particular topics might be of varying lengths; different activity topics. In some cases, due to insufficient knowledge of the Russian language related to the more technical aspects of the curriculum, some teachers provided additional explanations in Kyrgyz and Uzbek.

[14] Additionally, presentations and brochures in English and in Russian were designed and printed to inform donors, government officials, and other NGOs interested in keeping abreast of the LSBS progress during the Treatment.

LI and CIB employed a team of experts to provide regular and keen mentoring and coaching to the new trainers, especially during R1 and R2. Some teachers were stronger than others in their performance; yet we noted a strong similarity of response and benefit among the students in the programme, which LI concludes is a result of the strength of the written curriculum. When one trained teacher had to withdraw from her team mid-programme for family reasons, her replacement (who had not had the TOT) was able to play a reasonably supportive role in the team due to her previous experience with interactive methodology.

The teachers who implemented LSBS, the participating students and the school administrations, in both treatment and control schools, were aware that there was an impact evaluation ongoing.

6.3. Recruitment Strategy and Take-up

CIB visited each treatment and control school, informing students about the programme with a 30-minute presentation, and distributing flyers and application forms. The application form was one page in length and required about 15-30 minutes to complete. It consisted of questions on personal information, and the motivation to attend the training. Interested students were requested to apply before the application deadline of each round of training. Students in both treatment and control schools were informed that there was a probability that they would not be selected into the programme because of the randomization.

After the application deadline, the research team received lists of applicants per school. Within each school, a research team member together with one or two enumerators organized a public lottery to select treatment students. The sample was stratified according to gender and ethnic origin, with stratification proportional to group size. For each round, 20 treatment and 20 control students were selected per school. The total number of applicants per school was always at least 60.

As was stated above, the teachers were compensated financially for delivering the workshops (not for participating in the TOT), which motivated them to stay in the programme.

The criteria for participation in the programme were clear: the students should come from grades 9, 10 and 11 in the programme schools, there should be equal number of boys and girls, there should be as wide a diversity of ethnicities as possible, there should be at least 6 students (of the 20 per group) of different ethnicity than the majority ethnic group. The beneficiaries clearly matched the intervention's target population.

6.4. Unexpected or Adverse Events in Intervention

There were no adverse events within the treatment groups involving student-to-student interaction.

There were two minor events involving the administration of the programme and the study.

- One case of a teacher/trainer pair having minor disagreements. LI mediated between the two, which resolved their issues and enabled them to continue working as a pair to complete the programme.
- One case where a teacher involved students from her own “homeroom” class in the service project, which reduced the leadership role the LSBS students were able to play. This case was dealt with from a supervisory perspective by CIB, and standards made clear for the next round.

Finally, despite the fact that Russian or semi-Russian schools were selected for all 3 rounds, we observed that some students (especially from Tajik and Uzbek communities) faced language challenges as they did not speak or understand Russian well enough. However, multilingual trainers provided translation to the local language if needed.

7. Results

Nine sources of data were used for the analysis of the results presented in this report. First, we held FGDs with our population of trainers (20 of which 19 participated) before (January 2014) and after (June 2014) the Training of Trainers. Second, we administered household and student-level baseline questionnaires to programme participants in both treatment and control groups prior to each round. Immediately after the programme ended, we did a first follow-up (midline) both for students and households, except for R3 students, where we only collected follow-up data as part of the household questionnaire that included a student’s module. In addition to the survey questions we conducted four types of behavioural experiments as part of the midline, including a dictator game, trust game, risk game and a public goods game. We conducted an endline survey for households and students and repeated the behavioural games for R1 and R2 participants one year after the programme completed. The endline also comprised a social network survey that was administered to all students in grades 9, 10 and 11 (irrespective of whether they applied to the programme or not). We also held FGDs with a randomly selected sample of students in three treatment. Finally, we used school administrative records to capture information on school-level demographic data and individual test scores. All outcomes are self-reported except for data used from the school records.

7.1 Baseline Characteristics

Sample means at baseline

Table 5 reports the basic summary statistics of main student characteristics that served as relevant strata for randomization and outcome indicators measured at baseline. Programme applicants are on average 15 years old and 41 percent are male. A little more than 50 per cent of programme applicants are Kyrgyz (which constitutes the major ethnic group) and the rest are minority ethnic groups, most of them Uzbek. Some 68 per cent of programme applicants have plans to pursue a university degree or study at vocational school. Self-reported grades are high, 4.3 on average scale from 2-5. Programme applicants display moderate levels of trust (a score

of 2.57 out of 4) towards people of the same ethnicity. Levels of trust are considerably lower towards the 'out-group' that includes people that belong to a different ethnic or religious group or people they meet for the first time. Programme applicants displayed high levels of self-confidence or self-efficacy, some 75 points out of 100. Yet the average score for an index that captured people's perception of "locus of control" was only 0.41 on a range from 0 to 1.

Turning to beliefs, many agreed rather than disagreed with the statements that "Kyrgyz language should be the only official language in Kyrgyzstan" (the average score is 3.7 out of 5) and "We need to protect our culture, religion and language from others" (the average score is 4 out of 5). On the other hand, personal perceptions of feeling at home and feeling secure at school rate high. The students mostly agree with the statement that "I feel myself at home in Kyrgyzstan" (the average score is 4.1 out of 5) and "My school creates safe and non-discriminatory environment" (the average score is 3.7 out of 5). Consistent with this are low levels of reported incidence of bullying, and fighting (5 and 8 percent respectively) although we cannot rule out that people have underreported here.

Table 5: Descriptive statistics at baseline (W1), control variables and outcome indicators that were measured at baseline.

Variable	N	Mean	Std.Dev	Min	Max
Age	1,636	15.30	0.90	13	18
Male	1,675	0.41	0.49	0	1
Majority ethnicity	1,672	0.52	0.50	0	1
Minority ethnicity	1,672	0.48	0.50	0	1
Average self-reported grade	1,428	4.28	0.63	2	5
Plans to study at university / vocational school	1,458	0.68	0.47	0	1
Trust to people of the same ethnicity	1,429	2.57	0.90	1	4
Trust to people of another ethnicity	1,431	2.28	0.88	1	4
Trust to people of other religion	1,427	2.09	0.90	1	4
Trust to people seeing 1st time	1,431	1.86	0.85	1	4
Average self-efficacy: confidence	642	75.42	11.83	43	100
Average locus of control	642	0.41	0.18	0	1
Kyrgyz language should be the only official language	1,429	3.73	0.87	1	5
We need to protect our culture, religion & language from others	1,428	4.04	0.76	1	5
I feel myself at home in Kyrgyzstan	1,427	4.12	0.86	1	5
My school creates safe & non-discriminatory environment	1,429	3.71	0.85	1	5
I was bullied in school	1,426	0.05	0.21	0	1
I did fight in last 12 months	1,381	0.08	0.28	0	1

Test of balance

Error! Reference source not found. presents the tests of balance between treatment and control students in treatment schools. Column (1) presents the mean value for the control group; column (2) presents the mean value for the treatment group and column (3) presents the t-statistic of a t-test for the equality of means. Baseline balance was not achieved for four variables: treatment students are more likely to plan to study at university; they have lower levels of trust in people that belong to a different religion, display lower levels of self-efficacy and are less likely to concur with the idea that Kyrgyz should be the only language. In all our

regressions, we include the variable 'plans to study at university' as an additional control to mitigate concerns bias. The other three unbalanced variables are outcome variables. As we would expect that the programme would increase trust towards members of the out-group, and boost self-confidence, the lower baseline values for the treatment group may then lead us to underestimate rather than over-estimate the impact of the programme on these outcomes. Yet this is different for the belief that "Kyrgyz should be the only language". Here any potential bias resulting from lower baseline values for the treatment group is expected to go in the same direction as the treatment effect.

Table 6: Balance test at baseline (W1)

Variable	Mean		t-stat of difference		N	
	Control	Treated			Control	Treated
Age	15.28	15.26	-0.25		492	723
Male	0.39	0.38	-0.36		507	743
Majority ethnicity	0.51	0.51	0.02		505	742
Minority ethnicity	0.49	0.49	-0.02		505	742
Average self-reported grade	4.30	4.30	0.00		397	644
Plans to study at university / vocational school	0.67	0.72	2.31	**	423	646
Trust to people of the same ethnicity	2.64	2.53	-1.25		415	632
Trust to people of another ethnicity	2.35	2.29	-1.03		416	630
Trust to people of other religion	2.16	2.03	-3.27	***	415	628
Trust to people seeing 1st time	1.87	1.79	-1.56		415	633
Average self-efficacy: confidence	77.10	75.18	-1.69	*	191	241
Average locus of control	0.41	0.41	0.00		191	241
Kyrgyz language should be the only official language	3.82	3.67	-3.94	***	398	644
We need to protect our culture, religion & language from others	4.04	4.02	-0.51		398	643
I feel myself at home in Kyrgyzstan	4.19	4.12	-1.08		397	643
My school creates safe & non-discriminatory environment	3.69	3.69	0.09		398	644
I was bullied in school	0.07	0.04	-0.91		398	643
I did fight in last 12 months	0.10	0.09	-0.27		376	632

Notes: Coefficient's significance level: *** p<0.01, ** p<0.05 *p<0.1

7.2 Empirical Strategy

We estimate the following equation:

$$Y_{ij} = \beta_1 T_i + \beta_2 X_{ij} + \delta_j + \varepsilon_{ij} \quad (1)$$

where Y_{ij} is an outcome for student i in school j, X_{ij} are individual controls, δ_j and are school dummies.

7.3 Differences-in-means

Table 7 presents the difference-in-means results.

Table 7: Difference in means

Outcome variable	Means			+ indiv.effects			+ school. effects		
	Coeff.	SE		Coeff.	SE		Coeff.	SE	
Self-reported average grade across subjects	0.09	0.05	*	0.04	0.04		0.04	0.04	
I was bullied in school	-0.03	0.02	*	-0.03	0.02	*	-0.03	0.02	*
I did fight in last 12 months	0.01	0.01		0.01	0.01		0.01	0.01	
Trust to people of the same ethnicity	0.10	0.07		0.12	0.07	*	0.17	0.06	***
Trust to people of another ethnicity	0.06	0.06		0.01	0.05		0.04	0.07	
Trust to people of other religion	0.05	0.04		0.02	0.04		0.03	0.07	
Trust to people seeing 1st time	0.21	0.09	**	0.26	0.10	***	0.32	0.09	***
Average self-efficacy: confidence	-1.85	1.31		-2.26	1.63		-2.50	1.55	
Average locus of control	0.00	0.02		0.00	0.02		0.00	0.02	
Kyrgyz language should be the only official language	-0.02	0.05		0.01	0.06		0.03	0.05	
We need to protect our culture, religion & language from others	-0.09	0.06		-0.07	0.07		-0.06	0.07	
I feel myself at home in Kyrgyzstan	-0.18	0.04	***	-0.18	0.06	***	-0.19	0.06	***
My school creates safe & non-discriminatory environment	-0.04	0.07		-0.01	0.07		-0.03	0.06	
Trust (experimental)	-0.08	0.11		0.02	0.13		0.00	0.11	
Cooperation in Game 1	0.20	0.15		0.25	0.15	*	0.27	0.15	*
Cooperation in Game 2	0.30	0.21		0.38	0.21	*	0.41	0.22	*
Cooperation in Game 3	0.15	0.16		0.16	0.20		0.16	0.20	
Altruism (experimental)	-0.04	0.15		0.05	0.17		0.10	0.17	
Answered correctly to what mediation skills are	0.02	0.03		0.03	0.05		0.00	0.03	
Answered correctly on behaviour of unequal treatment	0.03	0.07		0.09	0.08		0.04	0.08	

Notes: Each row presents a separate OLS regression of the outcome variable on the treatment variable only (column 2), the treatment variable plus individual controls (column 3) and the treatment variable, individual controls and school fixed effects. Standard errors are clustered at school level. Coefficient's significance level: *** p<0.01, ** p<0.05, p<0.1. ITT impacts are estimated.

The difference-in-means estimates for the full sample show that the programme seems to have had some effect on both in-group and outgroup-trust; programme participants increase their level of trust towards people of the same ethnicity by 7 percent and towards people they see for the first time by some 17 percent. Also, the incidence of bullying seems to decrease substantially by 3 percentage points (relative to a baseline mean of 5 percent). The programme also had a positive impact on cooperation as measured by contributions to a public goods game. Programme participation seems to affect at least one of the attitudinal outcomes negatively: programme students are less likely to agree to the statement that they feel at home

in Kyrgyzstan (some 5 percent change relative to the baseline mean). None of the other variables turn out significant at conventional levels.

7.4 Sub-group Analysis

As specified in our Theory of Change we expect outcomes to differ across gender and whether students belong to a major or minor ethnic group. We estimated (1) for girls and boys separately and for ethnic majorities and minorities. Table 8 and Table 9 report the results for the gender subsamples.

Table 8: Programme impacts for girls using differences in means

Outcome variable	Means			+ indiv.effects			+ school. effects		
	Coeff.	SE		Coeff.	SE		Coeff.	SE	
Self-reported average grade across subjects	0.07	0.03	*	0.00	0.04		-0.03	0.04	
I was bullied in school	-0.04	0.03		-0.04	0.02		-0.03	0.02	
I did fight in last 12 months	0.00	0.01		0.00	0.01		0.00	0.01	
Trust to people of the same ethnicity	-0.02	0.08		-0.04	0.08		0.07	0.07	
Trust to people of another ethnicity	0.11	0.10		0.00	0.10		0.00	0.13	
Trust to people of other religion	0.13	0.10		0.02	0.09		-0.01	0.11	
Trust to people seeing 1st time	0.18	0.15		0.25	0.16		0.32	0.14	**
Average self-efficacy: confidence	-2.25	1.48		-3.11	1.79	*	-3.31	1.83	*
Average locus of control	0.01	0.03		0.02	0.03		0.01	0.03	
Kyrgyz language should be the only official language	-0.06	0.08		0.00	0.08		0.00	0.07	
We need to protect our culture, religion & language from others	-0.10	0.09		-0.09	0.10		-0.05	0.10	
I feel myself at home in Kyrgyzstan	-0.21	0.07	***	-0.21	0.09	**	-0.19	0.10	**
My school creates safe & non-discriminatory environment	-0.07	0.10		-0.02	0.08		-0.02	0.08	
Trust (experimental)	-0.05	0.11		-0.06	0.11		-0.07	0.11	
Altruism (experimental)	0.07	0.14		0.14	0.16		0.17	0.17	
Cooperation in Game 1	0.30	0.20		0.23	0.17		0.28	0.18	
Cooperation in Game 2	0.37	0.23		0.45	0.18	**	0.50	0.20	**
Cooperation in Game 3	0.25	0.23		0.23	0.18		0.25	0.18	
Answered correctly to what mediation skills are	0.11	0.05	**	0.16	0.07	**	0.09	0.09	
Answered correctly on behaviour of unequal treatment	0.11	0.09		0.21	0.09	**	0.14	0.10	

Notes: Standard errors are clustered at school level. Coefficient's significance level: *** p<0.01, ** p<0.05, p<0.1. Intention-To-Treat (ITT) effects are estimated.

Table 9: Programme impacts for boys using differences in means

Outcome variable	Means		+ indiv.effects		+ school.effects	
	Coeff.	SE	Coeff.	SE	Coeff.	SE

Self-reported average grade across subjects	0.10	0.05	**	0.09	0.07		0.09	0.08	
I was bullied in school	-0.01	0.01		-0.01	0.02		-0.03	0.02	
I did fight in last 12 months	0.03	0.02	*	0.02	0.02		0.01	0.02	
Trust to people of the same ethnicity	0.26	0.15	*	0.37	0.17	**	0.30	0.15	**
Trust to people of another ethnicity	-0.01	0.15		0.04	0.18		-0.01	0.17	
Trust to people of other religion	-0.05	0.14		0.00	0.16		-0.07	0.14	
Trust to people seeing 1st time	0.26	0.13	*	0.30	0.15	**	0.30	0.15	**
Average self-efficacy: confidence	-1.24	1.34		-1.21	1.73		-1.52	1.25	
Average locus of control	0.00	0.02		-0.01	0.02		-0.01	0.02	
Kyrgyz language should be the only official language	0.04	0.10		0.04	0.12		0.08	0.11	
We need to protect our culture, religion & language from others	-0.07	0.04		-0.05	0.05		-0.10	0.06	
I feel myself at home in Kyrgyzstan	-0.12	0.08		-0.14	0.10		-0.23	0.11	**
My school creates safe & non-discriminatory environment	0.00	0.10		0.00	0.09		-0.07	0.09	
Trust (experimental)	-0.18	0.28		0.19	0.21		0.19	0.21	
Altruism (experimental)	-0.23	0.18		-0.10	0.22		-0.01	0.21	
Cooperation in Game 1	0.01	0.21		0.32	0.20		0.27	0.21	
Cooperation in Game 2	0.16	0.31		0.23	0.43		0.15	0.47	
Cooperation in Game 3	-0.04	0.30		0.01	0.45		-0.02	0.49	
Answered correctly to what mediation skills are	-0.09	0.07		-0.14	0.08	*	-0.14	0.10	
Answered correctly on behaviour of unequal treatment	-0.07	0.07		-0.09	0.10		-0.10	0.11	

Notes: Standard errors are clustered at school level. Coefficient's significance level: *** p<0.01, ** p<0.05, p<0.1.ITT effects are estimated.

Table 8 and Table 9 suggest the programme has some differential impacts on girls and boys. Boys are more likely to show higher levels of trust towards their ethnicity as well as to people they see for the first time. Girls on the other hand improve their knowledge about topics the programme specifically dealt with, display higher levels of cooperation, but also experience lower levels of self-confidence. The results are broadly similar for the difference-in-difference estimates (Table I16) (Table I17).

Results for the subsample of ethic majority students broadly follow the pattern observed in the full sample programme. The results are much weaker for the ethnic minority students (also see the DID results in Table (I18 and I19). Perhaps their status as ethnic minority was made more salient by the programme set-up that could potentially have impeded strong responses to the programme.

Table 10: Programme impacts for ethnic majority

Outcome variable	Means			+ indiv.effects			+ school.effects		
	Coeff.	SE		Coeff.	SE		Coeff.	SE	
Self-reported average grade across subjects	0.08	0.06		0.04	0.05		0.04	0.06	
I was bullied in school	-0.04	0.02	**	-0.04	0.03		-0.04	0.02	
I did fight in last 12 months	0.03	0.02	*	0.03	0.02	*	0.03	0.01	*
Trust to people of the same ethnicity	0.14	0.07	**	0.15	0.07	**	0.25	0.09	***
Trust to people of another ethnicity	0.14	0.09		0.12	0.10		0.18	0.12	
Trust to people of other religion	0.10	0.10		0.12	0.09		0.16	0.11	
Trust to people seeing 1st time	0.27	0.06	***	0.20	0.11	*	0.29	0.10	***
Average self-efficacy: confidence	-1.52	0.96		-1.28	1.03		-1.75	1.08	
Average locus of control	0.02	0.01	*	0.02	0.01		0.02	0.02	
Kyrgyz language should be the only official language	0.07	0.05		0.07	0.08		0.09	0.06	
We need to protect our culture, religion & language from others	-0.08	0.06		-0.05	0.08		-0.05	0.07	
I feel myself at home in Kyrgyzstan	-0.20	0.04	***	-0.17	0.05	***	-0.18	0.05	***
My school creates safe & non-discriminatory environment	0.04	0.09		0.06	0.09		0.01	0.06	
Trust (experimental)	-0.04	0.20		-0.08	0.24		-0.14	0.25	
Altruism (experimental)	-0.08	0.21		0.01	0.23		0.05	0.22	
Cooperation in Game 1	0.18	0.18		0.25	0.19		0.23	0.22	
Cooperation in Game 2	0.31	0.24		0.43	0.28		0.43	0.31	
Cooperation in Game 3	0.42	0.18	**	0.41	0.23	*	0.40	0.24	*
Answered correctly to what mediation skills are	0.00	0.04		0.02	0.04		-0.04	0.04	
Answered correctly on behaviour of unequal treatment	0.08	0.08		0.09	0.07		0.02	0.07	

Notes: Standard errors are clustered at school level. Coefficient's significance level: *** p<0.01, ** p<0.05, p<0.1. ITT effects are estimated.

Table 11: Programme impacts for minor ethnic groups

Outcome variable	Means		+ indiv.effects		+ school.effects	
	Coeff.	SE	Coeff.	SE	Coeff.	SE
Self-reported average grade across subjects	0.11	0.08	0.02	0.08	0.04	0.08
I was bullied in school	-0.02	0.01	-0.01	0.02	-0.01	0.02
I did fight in last 12 months	-0.01	0.01	-0.01	0.01	-0.01	0.01
Trust to people of the same ethnicity	0.07	0.15	0.06	0.17	0.09	0.17
Trust to people of another	-0.03	0.15	-0.15	0.17	-0.15	0.17

ethnicity									
Trust to people of other religion	-0.02	0.14		-0.13	0.17		-0.16	0.16	
Trust to people seeing 1st time	0.17	0.14		0.30	0.13	**	0.32	0.13	**
Average self-efficacy: confidence	-2.17	2.18		-3.29	2.82		-3.29	2.59	
Average locus of control	-0.02	0.04		-0.01	0.03		-0.01	0.03	
Kyrgyz language should be the only official language	-0.10	0.10		-0.06	0.10		-0.04	0.10	
We need to protect our culture, religion & language from others	-0.09	0.08		-0.08	0.08		-0.08	0.08	
I feel myself at home in Kyrgyzstan	-0.14	0.07	**	-0.21	0.09	**	-0.21	0.09	**
My school creates safe & non-discriminatory environment	-0.14	0.08		-0.09	0.09		-0.06	0.10	
Trust (experimental)	-0.13	0.11		0.12	0.11		0.08	0.09	
Altruism (experimental)	0.01	0.14		0.12	0.19		0.13	0.19	
Cooperation in Game 1	0.21	0.18		0.24	0.18		0.27	0.19	
Cooperation in Game 2	0.26	0.22		0.28	0.18		0.33	0.18	*
Cooperation in Game 3	-0.16	0.19		-0.15	0.28		-0.15	0.28	
Answered correctly to what mediation skills are	0.05	0.08		0.09	0.12		0.07	0.11	
Answered correctly on behaviour of unequal treatment	-0.03	0.07		0.08	0.12		0.05	0.12	

Notes: Standard errors are clustered at school level. Coefficient's significance level: *** p<0.01, ** p<0.05, p<0.1. ITT effects are estimated.

7.5 Medium-term Impacts

Finally, we estimated medium-term impacts using data for our sample of students for which we had 1-year follow-up data (see Appendix I Table I20 and Table I21) Here we notice that increased levels of trust towards people of a different ethnicity (and different religion) come out significant. Programme participation leads to an increase from 2.57 at baseline to 2.65 on a scale from 1-4 in average trust towards non-co-ethnics (and an increase from 2.09 to 2.22 again on a scale from 1-4 towards people from a different religion in the DID results). Also, altruism and general trust (measured experimentally) increased in the medium-term. This suggests some of the anticipated effects of increased tolerance and lower levels of prejudice needed some time to "sink in". Somewhat strikingly we find that self-reported average grades were lower for programme participants a year after the programme ended relative to their peers in the control condition.

During the 1-year follow-up we also asked about people's social network. Interestingly, programme participants seem to have more friends and have a larger proportion of non-co-ethnics in their social network. These questions were however only asked during the final round of data collection.

7.6 Results from the Focus Group Discussions

Foundation Tolerance International investigated the role of youth in the 2010 conflict in southern *oblasts* of Kyrgyzstan in 2011 and reported that multi-ethnic schools tend to have a lower incidence of conflict along ethnic lines.[15] This assumption is somewhat supported by the findings from our student focus groups discussions (FGDs) conducted in Spring 2015 as part of the current impact evaluation project. The average response to the programme is positive. Students remarked that the programme typically was fun, innovative and effective. Yet, they also noted that several students were facing difficulties in grasping somewhat abstract concepts like identity, stereotypes and mediation, often combined with language barriers.

Throughout the course of FGDs, it seemed that the perception of the programme by students (even though mainly positive) was conditional upon other factors like the language of communication between students off-class and the level of existing diversity at schools. Schools where Russian was both the main language of instruction in-class and the main language of communication between students off-class tend to have a more diverse student body. Students in these types of schools were more likely to accept differences and more tolerant to issues of inter-religious communication than students in more homogenous schools. The section below summarizes the responses from the main topics discussed during the FGs.

Perceived effect of the LSBS programme

In general, students seemed to be very enthusiastic about the training and their instructors. The objective of the training in the students' understanding was to promote the concepts of "friendship of people" and "living in harmony". Treatment students generally admitted that they gained a new set of conflict-resolution and mediation skills that allowed them to improve their relations with the family members and other people. The programme specifically taught them a new skill of mediating disputes, which may happen between other people, while staying neutral without taking any sides.

Treatment participants said that the number of conflicts in schools decreased. Some of the students admitted that they learned to accept different opinions, refrain from "useless" arguing or judging other people based on their looks without getting to know them on a personal level. They understood the need to make an effort to get along with other people. "If you feel that you are wrong, get over yourself and apologize" is one of the quotes from one of the students in the FGDs. It was also noted by some students that the programme encouraged them to excel academically and help other students with their studies. The programme also helped some of these students to reflect better on their own behaviour and that of others.

Implemented community-development projects

Students discussed how the community projects, part of the LSBS programme, affected their perception of their role in the society. Most of their projects addressed social issues: organization of a charity concert, charity auction, charity for local orphanage, lecture for students of junior grades on the danger of tobacco, team-building football matches, flash mobs

[15] http://fti.org.kg/ru/publications/analytical-documents/assessments/93-conflict-factors-in-the-south

etc. Interviewed students admitted that participation in the community projects made them feel useful and encouraged them to continue helping other less fortunate people.

Trainers

Treatment students noted that teaching methods changed as a result of the TOT. Teachers' attitude towards students was considered to have become more equal and respectful.

Personal outlooks

Apart from the LSBS training, we also discussed students' perceptions of religion, ethnicity, migration, and individual career aspirations. Some students seemed to be more open and liberal than others when it came to religion, but these differences were more prominent between schools rather than between treatment and control students per se. Treatment students reported that the programme helped them to rethink their attitudes towards differences (ethnic and religious differences) although this was not obvious from their answers to specific questions on this topic.

Religious rather than ethnic differences seem to be particularly salient when it comes to intergroup marriage. Students did not mind marrying someone of another ethnicity as long as that person practices the same religion. Specifically, a future spouse should understand their wife's or husband's religious views and values. For this reason, they are likely to marry only a person with the same religious views and practices. Parents are also unlikely to permit intergroup marriage.

Bonding effect

Participation in the training programme reportedly turned out to be a good bonding link between the treatment students. Many students said they made new friends as a result of this shared experience. We however did not hear that treatment students started to interact more with members from a different religious or ethnic group that were not part of the treatment group.

7.7 Attrition Analysis

We estimated overall and differential attrition for our sample using three variables (i.e. self-reported grade belief of fitting in school, and having bullied). These variables were available for all rounds and all waves. Thus, a person is defined as attriting from the survey if none of these questions were answered. This is a stricter definition of attrition or non-response than when only considering whether a student was tracked for the midline survey. Overall attrition rates amount to 27 percent and differential attrition is 7 percentage points (24 percent in the treatment group and 31 percent in the comparison group). We estimated a probit model with attrition as depend variable and all relevant outcomes variables as controls. We observe that attrition is uncorrelated with treatment status but that students who were planning to study at post-secondary studies and the students who were more likely to be engaged in fighting were less likely to be included in the midline. Also, people that expressed lower levels of trust towards members of the same ethnic group and those that were more trusting towards members of

another ethnicity were more likely to attrite. We have dealt with this non-random attrition in two ways. First, we have included all variables that turned out significant as controls in all regressions, except for the regression where the variable served as an outcome variable (the results are not reported but available on request). Second, we follow Gerber and Green (2012) and filled in extreme values for the missing variables to estimate lower and upper bounds for our treatment effects (results not reported but available on request).

Table 12: Probit regression results for attrition

	Attrited (=1)	Robust SE	
Treatment dummy	0.00	0.10	
Age	0.09	0.11	
Male	-0.02	0.13	
Minority ethnic group	0.19	0.20	
Average self-reported grade	0.04	0.15	
Plans to study at university / vocational school	0.18	0.06	***
Trust to people of the same ethnicity	-0.20	0.09	**
Trust to people of another ethnicity	0.15	0.07	*
Trust to people of other religion	0.05	0.08	
Trust to people seeing 1st time	0.04	0.08	
Kyrgyz language should be the only official language	0.00	0.08	
We need to protect our culture, religion & language from others	-0.04	0.06	
I feel myself at home in Kyrgyzstan	0.10	0.06	*
My school creates safe & non-discriminatory environment	0.03	0.07	
I was bullied at school	0.00	0.00	
I did fight in the last 12 months	0.44	0.19	**
Constant	-3.14	2.22	

Notes: p<0.10, **p<0.05, ***p<0.01. Robust standard errors in parentheses

8. Discussion

The results of our research suggest that the overall impact of the LSBS training programme is modest. Knowledge and various behavioural outcomes did improve as a result of the programme. However, the evidence is less conclusive when it comes to attitudes and beliefs. While an increase in out-group trust appears to be one of the most robust impacts of the programme, both in the short and medium term, beliefs related to “feeling at home” seem to have moved in the opposite direction. This may suggest that the programme, as intended, teaches people how to reflect more on themselves and others, trains them how to take perspective, and this may make a person more aware of his or her own group initially, and think about its place in Kyrgyz society. Interestingly, one year later we see a positive impact emerging (although the result is not robustly significant across specifications), which may suggest that the training initially “stirred up” emotions and feelings, which, after a while, result in positive attitudinal changes. It is also interesting to note that treatment students report to have made

more friends and have a higher proportion of non-co-ethnics in their network, suggesting that the programme may have had some impact beyond the treated individual.

Sub-group analyses indicate that girls are affected somewhat differently than boys. Specifically, girls seem to gain more in terms of increasing their knowledge about new concepts the programme specifically dealt with and they also cooperate more in the public goods game. The programme's efforts to teach participants to become more reflective, take perspective and increase empathy may have been more effective for girls, making them realize how hard "being tolerant" really is, thereby reducing their self-confidence. At the same time, this realization may have increased their effort and willingness to cooperate in the games. Likewise, students belonging to the country's ethnic majority on average seem to gain somewhat more from the programme than those who belong to an ethnic minority.

Medium term impacts show that out-group trust remains higher among treatment students, and some of behavioural outcomes measured in the experimental games (trust and altruism) actually only emerge after one year, whereas cooperation for a public good does not sustain.

The absence of strong quantitative results contrasts somewhat with findings from the qualitative part of the study, in which students self-reported to have benefitted from the programme in various ways. Perhaps students indeed felt the programme was useful and they really enjoyed participating in it, especially when asked immediately afterwards and in a direct manner. Yet, our survey and experimental instruments measure 'deeper' outcomes related to beliefs and attitudes that are less obviously connected to the programme, which may attenuate people's tendency to give socially desirable answers. Also, true beliefs and attitudes may be less amenable to interventions like LSBS as positive feelings and bonding may not be enough to change possibly deep-rooted beliefs about one-self and others.

The quantitative design, on the other hand, suffered from at least two major shortcomings that may also explain why we do not identify strong and robust treatment effects there. The first limitation pertains to the small number of schools in our sample, coupled with the non-random assignment of treatment across schools. Hence results are confined to treatment schools only, where treatment was randomly assigned at the individual level. Ethical concerns and budget constraints prevented us from collecting data on all 31 schools in the sampling frame. Although 31 schools is still a small number, such a sample could have enabled us to either ex-ante or ex-post create balanced school pairs. The problem with estimating treatment effects within treatment schools only is that potential spill-overs can plausibly contaminate the control group. This may then account for the modest effects we tend to find across our full set of outcomes. To the extent that this issue played out in the treatment schools, we may actually underestimate the actual impact.

The second limitation has to do with the timing of the baseline data collection. It was in fact started when the programme had already commenced. Theoretically, it would thus have been possible that participant's beliefs, attitudes and behaviour had already changed by the time the baseline data was collected, although we would have expected within-school balance to be

much worse than it was, unless there had been large spill overs right from the start. Again, this concern may have led to an underestimate of the actual programme impact. One way to mitigate the possible concern of having baseline information that was only collected once the programme started is by running the models separately for students in the third round of data collection only. This is however not a straightforward comparison, as some outcome indicators were collected in earlier rounds but not in later rounds and vice versa.

Despite these difficulties, our evaluation provides some insights that may be useful for both the implementing partner and the Kyrgyz Ministry of Education when considering if to scale up this programme or not. First, we see that trust towards people they see for the first time seems to improve for the full sample and across almost all subgroups. Although this is the only robust outcome among various variables, it is arguably one of the most important indicators, and most closely related to the programme's objective to teach adolescents to be open and non-judgmental towards strangers of a different ethnic, religious or cultural background. We also believe that such outcomes are amenable to change as a result of a peace education programme.

Second, this intervention was set in specific Russian-speaking schools that generally have a more ethnically diverse population and, possibly, higher levels of ethnic tolerance than other schools in Kyrgyzstan. When implementing the programme in non-Russian speaking schools one may therefore find that the programme has a much larger impact than in our sample of schools, if tolerance levels in our schools where much higher to begin with. On the other hand, schools that are less ethnically diverse may also be more hesitant to facilitate the programme in their school (i.e. demand for and take-up of the programme may be much lower).

A final point to note here is that we only estimated effects for students that voluntarily signed up for extra-curricular programme. This begs the question to what extent we captured the real target population, namely vulnerable young people who feel marginalized and excluded and who express their dissatisfaction through intolerance and, possibly, violent acts. Such students either may chose not to apply for the intervention offered in their schools – or their families may forbid them from participating. If such students participated in a peacebuilding programme (for example because it was an obligatory part of the school curriculum), then the impact of the programme might have been different.

9. Conclusions

Our research findings have implications at three levels: policy, practice and research.

Concerning policy implications, we note that the peacebuilding programme we study has some intended impact, especially on trust towards members of the out-group and on some behavioural measures like bullying and fighting. Moreover, these results are still identified a year after the programme ended, which provides some confidence in the sustainability of programme impacts in the medium term. Hence, we established in principle that a peacebuilding

programme for youth of different ethnicities can have positive impacts. Interestingly, results seem to be mainly driven by the changed outcomes of the female participants. This may suggest girls are more receptive to a programme like LSBS and are better able to reflect on what tolerance really implies, and how that affects the perception of one-self and others.

Our analysis also indicates that peacebuilding at the individual level is determined by a variety of factors. It is likely that only some of these determinants can be shaped, if at all, by a short-term, school-based intervention lasting a few weeks or months. Based on our research, we wonder if being and acting in a peaceful manner is actually a very large-scale and comprehensive behavioural shift, which is unlikely to respond to smaller, shorter or more focused interventions. In situations where violent conflict may flare up again, a larger, longer and broader approach to peacebuilding will be required, addressing drivers of peace at school, at home, in the media, in politics, in civil society and beyond and doing so over many years (if not at all times). It remains to be tested if such a concerted effort, the main-streaming of peacebuilding, has stronger or longer lasting impacts or is more cost effective. But it is a question that is worthwhile to address and that is relevant for policy makers to come to terms with until the evidence base has been built in a rigorous and conclusive fashion.

Concerning practice, we recommend to NGOs, governments and other agencies implementing peacebuilding programmes that their programmes either draw on or build their own very specific conflict analysis and related theory of change. There is a risk that a “one size fits all” approach leads to a programme design, which is not properly adapted to the local context. For example, inter-ethnic relations may be very different across the same country depending on the population share of the different groups across localities, depending on which ethnic group holds relatively speaking more power, and depending on who is perceived to be the likely driver of potential conflict (the majority group, the minority group or both). For example, it could be that one programme has a positive impact on the dominant majority group but no impact on a minority group, or vice versa. In situations where groups compete on a more equal footing for power a programme may have to bridge the gaps in a different way yet again. From our experience in this research programme, tailoring interventions to such differences is very important.

We also advise practitioners to differentiate clearly between peacebuilding knowledge, attitudes and practices on the one hand and more general academic or life skills on the other hand. It may be that better educated students or more confident students are more peaceful. However, “mainstreaming peace” does not imply that every or any action or thing can be interpreted to be a peacebuilding one. Helping students develop academically or gain better life skills may be worthwhile aims in themselves, perhaps with the added benefit of having positive unintended consequences for peace in the community.

In addition, we recommend that the theory of change be very explicit about the expected outcomes and how to measure them. “Peace” is all too often understood as a macro-level concept, implying the absence of war at the country level. Defining in the context of a specific context what peace means for programme participants and how this can be measured is a

challenge. If done properly, it can help future monitoring, evaluation and learning activities and hence the continued improvement of programming (Brück et al., 2016).

Concerning research and research policy, we note that this impact evaluation aims to contribute to what is a very small field of rigorous academic research indeed. No single study can do it all. Instead, all this study can hope to do is to provide yet one more puzzle piece to the larger picture of how to build peace between people. Future studies can build on our experiences by testing mechanisms in more detail, preferably with larger programmes and larger samples. Having completed this assignment, we continue to see the need for a substantial expansion of rigorous evidence in the field of peacebuilding and peace education for young people.

References

Aboud, F.E. et al., 2012. Interventions to reduce prejudice and enhance inclusion and respect for ethnic differences in early childhood: A systematic review. *Developmental Review*, 32(4), pp.307–336.

Allport, G., 1954*. The nature of prejudice*, Reading: Addison-Wesley.

Anon, 2009. Supplemental Material for Reducing Intergroup Prejudice and Conflict Using the Media: A Field Experiment in Rwanda. *Journal of Personality and Social Psychology*, pp.1–14.

Barlow, F. K., Paolini, S., Pedersen, A., Hornsey, M. J., Radke, H. R., Harwood, J., Rubin, M, & Sibley, C. G. (2012). The contact caveat: Negative contact predicts increased prejudice more than positive contact predicts reduced prejudice. *Personality and Social Psychology Bulletin*, *38*(12), 1629-1643.

Biton, Y. & Salomon, G., 2006. Peace in the Eyes of Israeli and Palestinian Youths: Effects of Collective Narratives and Peace Education Programme. *Journal of Peace Research*, 43(2), pp.167–180.

Blattman, C., Hartman, A.C. & Blair, R.A., 2014. How to Promote Order and Property Rights under Weak Rule of Law? An Experiment in Changing Dispute Resolution Behavior through Community Education. *American Political Science Review*, 108(01), pp.100–120.

Broockman, D. and Kalla, J., 2016. Durably reducing transphobia: A field experiment on door-to-door canvassing. *Science*, *352*(6282), pp.220-224.

Brück, T., N. T. N. Ferguson, V. Izzi & W. Stojetz (2016). “Jobs Aid Peace: A Review of the Theory and Practice of the Impact of Employment Programmes on Peace in Fragile and Conflict-affected Countries”. 16 September, Report for ILO, PBSO, UNDP and WBG, Berlin.

Bruhn, M. & McKenzie, D., 2009. In Pursuit of Balance: Randomization in Practice in Development Field Experiments. *American Economic Journal: Applied Economics*, 1(4), pp.200–232.

Burde, D., Middleton, J. & Samii, C., 2016. The Assessment of Learning Outcomes and Social Effects of Community-Based Education: A Randomized Field Experiment in Afghanistan*,* pp.1-58.

Cameron, L. & Rutland, A., 2006. Extended Contact through Story Reading in School: Reducing Children’s Prejudice toward the Disabled. pp.1–21.

Devine, P.G., Forscher, P.S., Austin, A.J. and Cox, W.T., 2012. Long-term reduction in implicit race bias: A prejudice habit-breaking intervention. *Journal of experimental social psychology*, *48*(6), pp.1267-1278.

Durlak, J.A. et al., 2011. The Impact of Enhancing Students' Social and Emotional Learning: A Meta-Analysis of School-Based Universal Interventions. *Child Development*, 82(1), pp.405–432.

Fearon, J., Humphreys, M. & Weinstein, J., 2008. Community-Driven Reconstruction in Lofa County. pp.1–47.

Gottfredson, D.C. & Wilson, D.B., 2003. Characteristics of Effective School-Based Substance Abuse Prevention. *Prevention Science*, 4(1), pp.27–38.

Gerber, A.S. and Green, D.P., 2012. *Field experiments: Design, analysis, and interpretation*. WW Norton.

Green, D.P. & Wong, J.S., 2009. Tolerance and the Contact Hypothesis: A Field Experiment. *The Political Psychology of Democratic Citizenship*, pp.228–246. Available at: http://books.google.com/books?hl=en&lr=&id=8LD6j1IcFNMC&oi=fnd&pg=PA228&dq=Tolerance+and+the+contact+hypothesis&ots=nssDlEOR-_&sig=bPBt7GS8XJwqnR3-K8vtrU_8se0.

Kelleher, A. & Ryan, K., 2016. Successful Local Peacebuilding in Macedonia: Sustained Dialogue in Practice. *The Canadian Journal of Peace and Conflict Studies*, pp.1–33.

Lustig, I., 2003. *The influence of studying foreign conflicts on students' perceptions of the Israeli Palestinian conflict*, Unpublished masters thesis.

Malhotra, D., 2005. Long-Term Effects of Peace Workshops in Protracted Conflicts. *Journal of Conflict Resolution*, 49(6), pp.908–924.

Megoran, N. et al., 2014. Evaluating peacebuilding interventions in southern Kyrgyzstan. pp.1–8.

Melvin, N., 2011. Promoting a Stable and Multiethnic Kyrgyzstan: Overcoming the Causes and Legacies of Violence. pp.1–64.

Nigmatov, V., 2012. Youth Theater for Peace (YTP) Programmes in Kyrgyzstan and Tajikistan. pp.1–107.

Paluck, E.L., 2011. Peer pressure against prejudice: A high school field experiment examining social network change. *Journal of Experimental Social Psychology*, 47(2), pp.350–358.

Paluck, E.L. & Green, D.P., 2009. Prejudice Reduction: What Works? A Review and Assessment of Research and Practice. *Annual Review of Psychology*, 60(1), pp.339–367.

Paluck, E.L. & Shepherd, H., 2012. The salience of social referents: A field experiment on collective norms and harassment behaviour in a school social network. *Journal of Personality and Social Psychology*, 103(6), pp.899–915.

Pettigrew, T.F. & Tropp, L.R., 2006. A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, 90(5), pp.751–783.

Pouezevara, S., Costello, M. & Banda, O., 2013. Abe/Link Malawi Teacher Professional Development Support Draft Malawi Reading Intervention Early Grade Reading Assessment (EGRA). pp.1–41.

Salomon, G., 2004. Does peace education make a difference in the context of an intractable conflict? *Peace and Conflict: Journal of Peace Psychology*, *10*(3), p.257.

Schonert-Reichl, K.A. et al., 2011. Promoting Children's Prosocial Behaviors in School: Impact of the "Roots of Empathy" Programme on the Social and Emotional Competence of School-Aged Children. *School Mental Health*, 4(1), pp.1–21.

Svensson, I. & Brouneus, K., 2013. Dialogue and interethnic trust: A randomized field trial of "sustained dialogue" in Ethiopia. *Journal of Peace Research*, 50(5), pp.563–575.

Wilson, D.B., Gottfredson, D.C. & Najaka, S.S., 2001. School-Based Prevention of Problem Behaviors: A Meta-Analysis. *Journal of Quantitative Criminology*, 17(3), pp.247–272.

Zainal, K., Abu, D.T. & Mohamad, Z., 2010. The effect of ethnic relations course on the students perceptions towards ethnic relations among first year students of one public university in Malaysia. *Procedia - Social and Behavioral Sciences*, 2(2), pp.3596–3599.

Appendix B. Sample Design

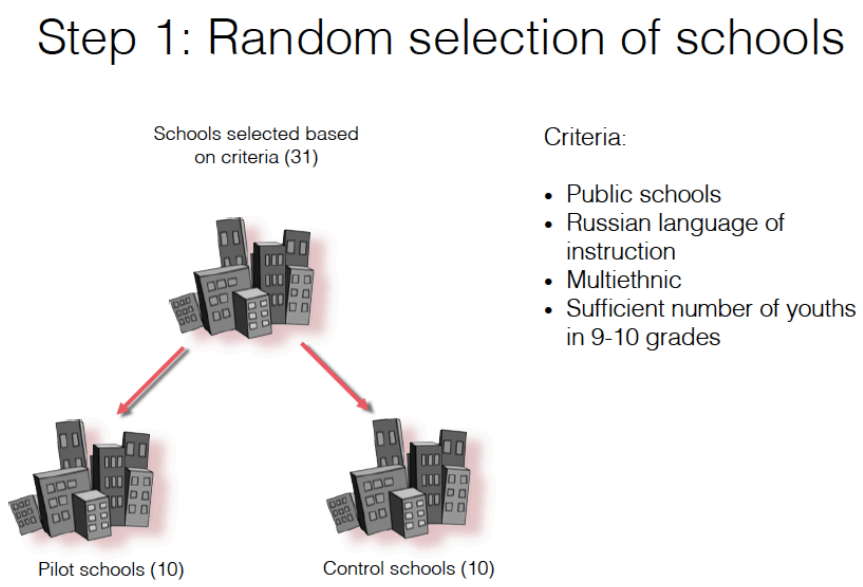

Figure B1: Random selection of schools

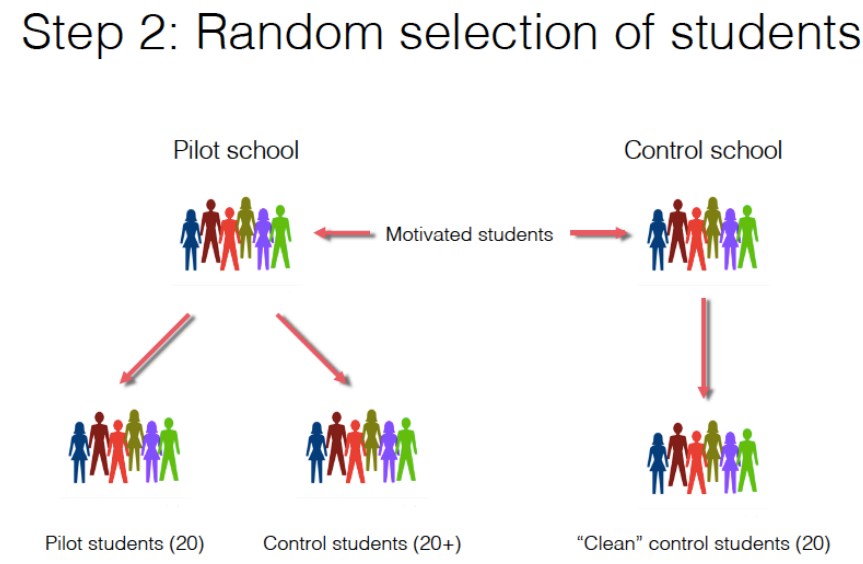

Figure B2: Random selection of students

Appendix H. Description of variables

Table H13: Description of variables

Variable Description	Coding
Demographic	
Male	Binary: 1 if male, 0 if female
Age	Scale: 13 to 18
Major ethnic group	Binary: 1 if yes, 0 if no
Minor ethnic group	Binary: 1 if yes, 0 if no
# of Siblings	Open-ended scale
School and Out-of-School Activities	
Plans to study at a university/vocational school	Binary: 1 if yes, 0 if no
Average grade (self-reported)	Scale: 2 (lowest) to 5 (highest)
Average grade from school records data (includes grades from Kyrgyz language, Russian language, Kyrgyz literature, Russian literature)	Scale: 2 (lowest) to 5 (highest)
Mathematics grade	Scale: 2 (lowest) to 5 (highest)
Social Beliefs	
“Kyrgyz language should be the only official language in Kyrgyzstan”	Scale: 1 (strongly disagree) to 5 (strongly agree)
“We need to protect our culture, religion and language from others”	Scale: 1 (strongly disagree) to 5 (strongly agree)
“I feel myself at home in Kyrgyzstan”	Scale: 1 (strongly disagree) to 5 (strongly agree)
“My school creates safe and non-discriminatory environment”	Scale: 1 (strongly disagree) to 5 (strongly agree)
Conflict Involvement	
"I was bullied in school"	Binary: 1 if yes, 0 if no
I did fight in the last 12 months	Binary: 1 if yes, 0 if no
Trust	
People of the same ethnicity People of another ethnicity; People of another religion People you meet for the first time	Scale: 1 (do not trust at all) to 4 (trust completely)
Self-Efficacy	
"Average Self-Efficacy," i.e. confidence to: 1. Resist peer pressure to do things that get me into trouble; 2. Make and keep friends of opposite sex; 3. Make and keep friends of same sex; 4. Make and keep friends of same ethnic group; 5. Make and keep friends of another ethnic group; 6. Make and keep friends of same religious group;	Scale: 0 (no confidence) to 100 (full confidence)

7. Make and keep friends of another religious group; 8. Work well in a group; 9. Express my opinions when other classmates disagree with me; 10. Stand up for myself when I feel I am being treated unfairly; 11. Get others to stop annoying me or hurting my feelings; 12. Stand firm to someone who is asking me to do something unreasonable or inconvenient; 13. Get a friend to help me when I have social problems; 14. Get myself to study when there are other interesting things to do; 15. Finish my homework assignments by deadlines; 16. Plan my schoolwork for the day	
Locus of Control	
"Average locus of control" is indicated by: 1. Do you think your school grades are mostly affected by accidental happening? 2. Do you think teachers are often unfair to students? 3. Do you believe that most problems will solve themselves if you just don't fool with them? 4. Do you feel that most of the time it doesn't pay to try hard because things never turn out right anyway? 5. Do you feel that one of the best ways to handle most problem is just not to think about them? 6. Do you feel that when someone your age decides to hit you there's little you can do to stop him or her?	Binary: 1 if yes (no control), 0 if no (control)
7. Do you believe that high-test scores are a matter of hard work; luck has little or nothing to do with it? 8. Do you feel that you have a lot of choice in deciding who your friends are? 9. Do you believe that whether or not people like you depends on how you act?	Binary: 0 if yes (no control), 1 if no (control)
Behavioural Experiment Outcomes	
Trust is indicated by the number of apples given to player 2	Scale: 0 (do not trust at all) to 5 (trust completely)
Cooperation is indicated by the number of stones donated to the common pot	Scale: 0 (no cooperation at all) to 5 (full cooperation)
Altruism is indicated by the number of apples donated	Scale: 0 (not altruistic at all) to 5 (very altruistic)
Knowledge	
Why do we need mediation skills?	Binary: 1 if answered correctly (e.g. To be able to resolve conflicts; to stress your own point of view and try to persuade others), 0 otherwise
What would you call a behaviour that treats people unequally because of their group membership?	Binary: 1 if answered correctly (e.g. discrimination), 0 otherwise
Social Network	

Count of all friends (“agreeable” ratings given to fellow students; friends are defined by a given rating of +1, +2 or +3 to signify a positive relationship)	Open-ended scale
% of non-co-ethnic friends (ratings given to fellow students of a non-co-ethnic background at either +1, +2 or +3)	0% (none of my friends have a different ethnic background to myself) to 100% (all my friends are of a different ethnic background to myself)

Appendix I. Results

1. Balance test: extended set of controls and outcomes

Table I14: Balance test (extended set of individual, household, and school characteristics)

Variable	Mean		t-stat of difference		N	
	Control	Treated			Control	Treated
Age	15.28	15.26	-0.25		492	723
Male	0.39	0.38	-0.36		507	743
Major ethnic group	0.51	0.51	0.02		505	742
Minority ethnic groups	0.49	0.49	-0.02		505	742
Average grade self-reported	4.30	4.30	0.00		397	644
Plans to study at university / vocational school	0.67	0.72	2.31	**	423	646
Russian is the main language of teaching	0.85	0.87	1.06		423	646
Hours spend for home assignments	2.16	2.16	0.08		414	630
Number of school clubs a student attends	0.38	0.41	1.17		437	667
Number of out-of-school clubs a student attends	0.90	0.93	0.40		365	572
Father has university or higher degree	0.49	0.51	0.60		335	521
Mother has university or higher degree	0.48	0.48	-0.02		374	592
Number of siblings	2.44	2.59	1.28		417	638
HH owns a car (reported by adult)	0.58	0.60	0.39		398	641
HH has a computer at home	0.69	0.72	0.58		414	636
Monthly food expenditures, 000'Soms	6.23	6.44	1.13		395	642
A HH member was exposed to the conflict in June 2010	0.07	0.09	0.83		207	401
Student / teacher ratio (school data)	49	50	0.61		507	743
Average scores at National Test	121	121	-0.79		507	743
Kyrgyz language should be the only official language	3.82	3.67	-3.94	***	398	644
We need to protect our culture, religion & language from others	4.04	4.02	-0.51		398	643
I feel myself at home in Kyrgyzstan	4.19	4.12	-1.08		397	643
I fit in my school	0.99	0.99	-0.25		398	643
My school creates safe & non-discrim. environment	3.69	3.69	0.09		398	644
I was bullied in school	0.07	0.04	-0.91		398	643
I did fight in last 12 months	0.10	0.09	-0.27		376	632
Trust to people of the same ethnicity	2.64	2.53	-1.25		415	632
Trust to people of another ethnicity	2.35	2.29	-1.03		416	630
Trust to people of other religion	2.16	2.03	-3.27	***	415	628
Trust to people seeing 1st time	1.87	1.79	-1.56		415	633
Average self-efficacy: confidence	77.10	75.18	-1.69	*	191	241
Average locus of control	0.41	0.41	0.00		191	241

2. Difference-in-differences

a. Main results

Table I15: Programme impacts using difference-in-differences

Outcome variable	No controls			+ indiv.effects			+ school effects		
	Coeff.	SE		Coeff.	SE		Coeff.	SE	
Average grade self-reported	0.09	0.06		0.03	0.07		0.03	0.07	
I was bullied in school	0.00	0.04		-0.04	0.02	*	-0.04	0.02	*
I did fight in last 12 months	0.02	0.02		0.02	0.03		0.02	0.03	
Trust to people of the same ethnicity	0.21	0.13		0.21	0.13	*	0.21	0.13	
Trust to people of another ethnicity	0.13	0.09		0.06	0.08		0.06	0.09	
Trust to people of other religion	0.19	0.06	***	0.12	0.07	*	0.12	0.08	
Trust to people seeing 1st time	0.29	0.09	***	0.27	0.09	***	0.27	0.09	***
Average self-efficacy: confidence	0.07	1.48		-0.82	1.60		-1.81	1.67	
Average locus of control	0.00	0.03		0.02	0.03		0.03	0.03	
Kyrgyz language should be the only official language	0.13	0.07	*	0.14	0.08	*	0.14	0.08	*
We need to protect our culture, religion & language from others	-0.06	0.05		-0.05	0.06		-0.05	0.06	
I feel myself at home in Kyrgyzstan	-0.10	0.06		-0.12	0.06	*	-0.12	0.07	*
My school creates safe & non-discrim. environment	-0.05	0.09		-0.06	0.09		-0.05	0.09	

b. Sub-sample impacts by gender

Table I16: Programme impacts for boys

Outcome variable	No controls			+ indiv.effects			+ school.effects		
	Coeff.	SE		Coeff.	SE		Coeff.	SE	
Average grade self-reported	0.18	0.08	**	0.14	0.10		0.16	0.10	
I was bullied in school	0.02	0.05		-0.05	0.04		-0.05	0.04	
I did fight in last 12 months	0.05	0.04		0.03	0.04		0.03	0.04	
Trust to people of the same ethnicity	0.48	0.23	**	0.59	0.25	**	0.59	0.25	**
Trust to people of another ethnicity	0.16	0.20		0.19	0.24		0.19	0.24	
Trust to people of other religion	0.13	0.17		0.11	0.18		0.12	0.18	
Trust to people seeing 1st time	0.37	0.16	**	0.39	0.17	**	0.41	0.16	**
Average self-efficacy: confidence	2.88	2.22		2.33	2.20		1.55	2.15	
Average locus of control	-0.01	0.04		0.01	0.03		0.02	0.03	
Kyrgyz language should be the only official language	0.16	0.10		0.13	0.13		0.13	0.13	

We need to protect our culture, religion & language from others	-0.08	0.09		-0.09	0.08		-0.10	0.09	
I feel myself at home in Kyrgyzstan	-0.25	0.10	**	-0.32	0.11	***	-0.32	0.12	***
My school creates safe & non-discrim. environment	-0.01	0.14		-0.05	0.11		-0.06	0.12	

Table I17: Programme impacts for girls

Outcome variable	No controls			+ indiv.effects			+ school.effects		
	Coeff.	SE		Coeff.	SE		Coeff.	SE	
Average grade self-reported	0.02	0.08		-0.05	0.08		-0.05	0.08	
I was bullied in school	-0.01	0.04		-0.03	0.03		-0.03	0.03	
I did fight in last 12 months	0.00	0.03		0.01	0.03		0.01	0.03	
Trust to people of the same ethnicity	0.01	0.13		-0.06	0.11		-0.06	0.10	
Trust to people of another ethnicity	0.11	0.11		-0.03	0.09		-0.02	0.10	
Trust to people of other religion	0.23	0.11	**	0.13	0.12		0.13	0.13	
Trust to people seeing 1st time	0.23	0.12	*	0.18	0.15		0.16	0.16	
Average self-efficacy: confidence	-1.55	1.55		-3.00	1.69	*	-4.03	1.82	**
Average locus of control	0.01	0.04		0.04	0.03		0.03	0.04	
Kyrgyz language should be the only official language	0.11	0.12		0.14	0.14		0.14	0.14	
We need to protect our culture, religion & language from others	-0.06	0.05		-0.03	0.07		-0.02	0.07	
I feel myself at home in Kyrgyzstan	-0.01	0.07		0.00	0.08		0.01	0.07	
My school creates safe & non-discrim. environment	-0.07	0.12		-0.06	0.13		-0.05	0.13	

c. Sub-sample impacts by ethnicity

Table I18: Programme impacts for the ethnic majority group

Outcome variable	No controls			+ indiv.effects			+ school.effects		
	Coeff.	SE		Coeff.	SE		Coeff.	SE	
Average grade self-reported	0.09	0.07		-0.01	0.08		-0.01	0.07	
I was bullied in school	0.03	0.06		-0.04	0.02	*	-0.04	0.02	*
I did fight in last 12 months	0.03	0.03		0.03	0.03		0.03	0.03	
Trust to people of the same ethnicity	0.14	0.13		0.20	0.13		0.22	0.14	
Trust to people of another ethnicity	0.16	0.10		0.17	0.12		0.18	0.13	
Trust to people of other religion	0.23	0.11	**	0.25	0.13	**	0.28	0.14	**
Trust to people seeing 1st time	0.26	0.09	***	0.22	0.12	*	0.24	0.11	**
Average self-efficacy:	1.14	1.97		-0.68	1.27		-1.63	1.68	

confidence									
Average locus of control	0.01	0.04		0.05	0.03		0.04	0.04	
Kyrgyz language should be the only official language	0.14	0.05	***	0.14	0.06	**	0.14	0.06	**
We need to protect our culture, religion & language from others	-0.05	0.06		-0.03	0.08		-0.01	0.08	
I feel myself at home in Kyrgyzstan	-0.18	0.09	*	-0.20	0.09	**	-0.18	0.10	*
My school creates safe & non-discrim. environment	0.01	0.14		-0.02	0.16		-0.01	0.16	

Table I19: Programme impacts for the ethnic minority group

Outcome variable	**No controls**			**+ indiv.effects**			**+ school.effects**		
	Coeff.	**SE**		**Coeff.**	**SE**		**Coeff.**	**SE**	
Average grade self-reported	0.09	0.09		0.07	0.10		0.08	0.10	
I was bullied in school	-0.03	0.03		-0.03	0.03		-0.03	0.03	
I did fight in last 12 months	0.01	0.03		0.01	0.03		0.01	0.03	
Trust to people of the same ethnicity	0.30	0.23		0.24	0.21		0.21	0.21	
Trust to people of another ethnicity	0.09	0.19		-0.08	0.19		-0.11	0.19	
Trust to people of other religion	0.12	0.15		-0.08	0.15		-0.10	0.15	
Trust to people seeing 1st time	0.34	0.12	***	0.36	0.13	***	0.34	0.13	***
Average self-efficacy: confidence	-0.58	1.93		-1.44	2.70		-2.31	2.59	
Average locus of control	0.00	0.03		0.01	0.03		0.01	0.03	
Kyrgyz language should be the only official language	0.12	0.14		0.13	0.15		0.13	0.15	
We need to protect our culture, religion & language from others	-0.07	0.08		-0.07	0.09		-0.08	0.09	
I feel myself at home in Kyrgyzstan	-0.01	0.10		-0.04	0.12		-0.04	0.12	
My school creates safe & non-discrim. environment	-0.11	0.09		-0.10	0.09		-0.10	0.09	

3. Medium-term impacts

Table I20: Medium-term programme impacts using difference-in-means

Outcome variable	Means			+ indiv.effects			+ school effects		
	Coeff.	SE		Coeff.	SE		Coeff.	SE	
Self-reported average grade across subjects	-0.12	0.05	**	-0.13	0.06	**	-0.15	0.05	***
I was bullied in school	-0.03	0.03		-0.05	0.04		-0.03	0.03	
I did fight in last 12 months	0.00	0.00		-0.01	0.00		0.00	0.00	
Trust to people of the same ethnicity	0.09	0.06		0.27	0.08	***	0.22	0.10	**
Trust to people of another ethnicity	0.16	0.08	**	0.23	0.07	***	0.19	0.08	**
Trust to people of other religion	0.14	0.11		0.18	0.11		0.15	0.12	
Trust to people seeing 1st time	-0.01	0.05		0.02	0.08		0.01	0.09	
Kyrgyz language should be the only official language	-0.05	0.17		0.05	0.21		0.04	0.21	
We need to protect our culture, religion & language from others	-0.10	0.11		-0.16	0.11		-0.15	0.11	
I feel myself at home in Kyrgyzstan	0.14	0.09		0.19	0.10	*	0.13	0.08	
My school creates safe & non-discrim. environment	0.05	0.09		-0.04	0.10		-0.09	0.11	
Trust (experimental)	0.36	0.22	*	0.53	0.24	**	0.46	0.25	*
Altruism (experimental)	0.27	0.13	**	0.22	0.17		0.12	0.17	
Cooperation in Game 1	0.27	0.21		0.46	0.24	*	0.30	0.24	
Cooperation in Game 2	0.28	0.25		0.00	0.00		0.00	0.00	
Cooperation in Game 3	0.16	0.19		0.45	0.36		0.28	0.36	

Table I21: Medium-term programme impacts using difference-in-differences

Outcome variable	Means			+ indiv.effects			+ school effects		
	Coeff.	SE		Coeff.	SE		Coeff.	SE	
Self-reported average grade across subjects	-0.12	0.04	***	-0.14	0.06	**	-0.19	0.07	***
I was bullied in school	0.00	0.04		-0.05	0.04		-0.05	0.04	
I did fight in last 12 months	0.00	0.02		0.01	0.03		0.02	0.03	
Trust to people of the same ethnicity	0.20	0.10	**	0.35	0.15	**	0.32	0.16	**
Trust to people of another ethnicity	0.22	0.08	***	0.28	0.09	***	0.26	0.09	***
Trust to people of other religion	0.27	0.12	**	0.28	0.13	**	0.27	0.13	**
Trust to people seeing 1st time	0.06	0.08		0.05	0.12		0.03	0.12	
Kyrgyz language should be the only official language	0.09	0.18		0.15	0.23		0.12	0.23	
We need to protect our culture, religion & language from others	-0.08	0.11		-0.16	0.11		-0.14	0.11	
I feel myself at home in Kyrgyzstan	0.22	0.14		0.23	0.17		0.20	0.15	
My school creates safe & non-discrim. environment	0.04	0.11		-0.07	0.13		-0.08	0.13	

Appendix L. Timeline

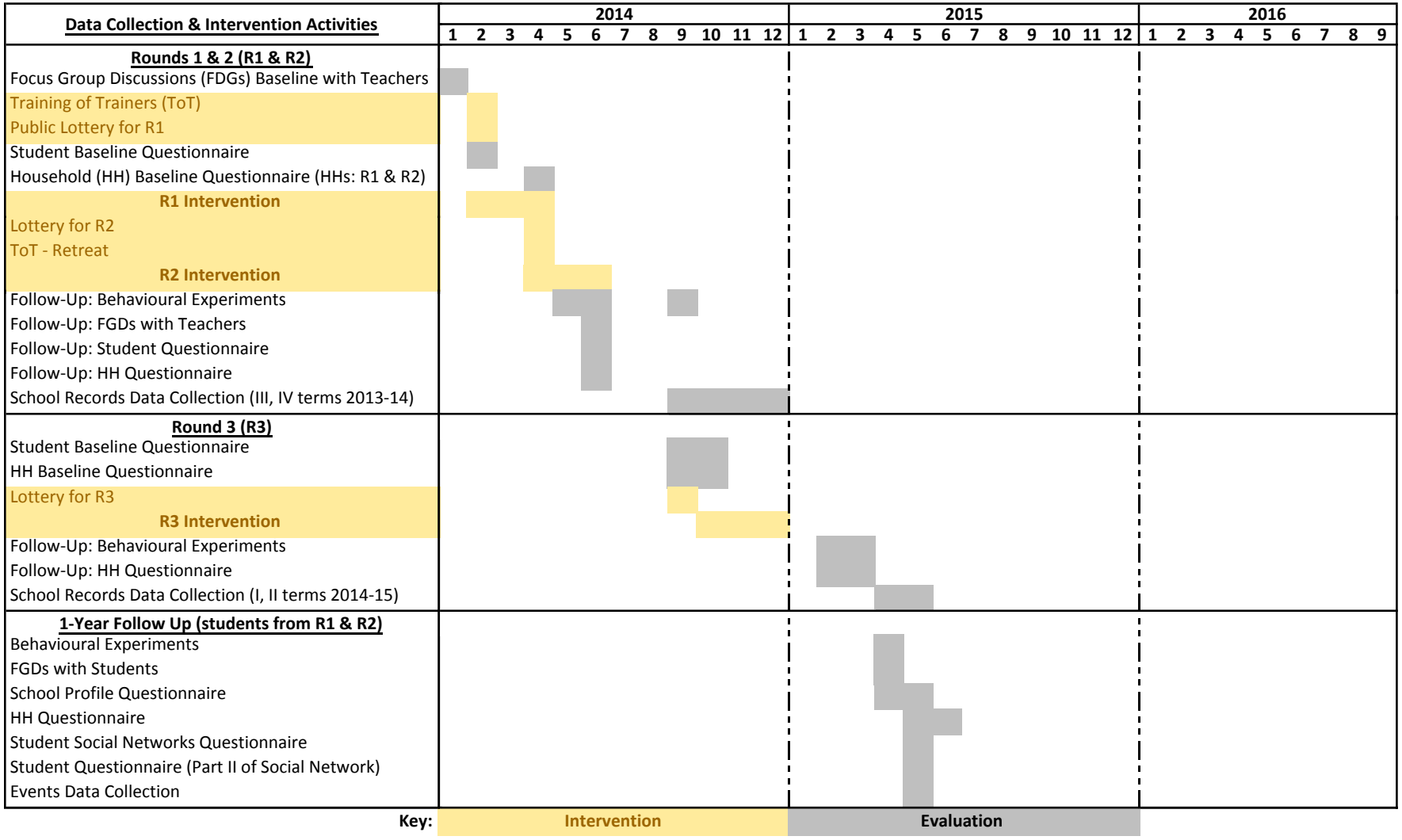

Figure L3: Timeline