Non-Toxic Peers: Long-Run Returns from an Anti-Bullying Program

Tabea Braun* *University of Zurich*

Ana Costa-Ramón *University of Zurich*

Ana Rodríguez-González *University of Barcelona*

Ursina Schaede *Tufts University*

Christina Salmivalli *University of Turku*

January 15, 2026

Abstract

We study the long-run impacts of a randomized anti-bullying intervention, the KiVa program, in Finnish schools. We link the RCT survey data for 15,000 pupils attending grades 7-9 to comprehensive administrative records on educational and labor market outcomes in adulthood. Treated students experience gains in human capital attainment and labor market outcomes: they are more likely to enroll in academic high school, obtain a university degree, and earn higher wages by ages 27-29. We find that these gains accrue to all groups of pupils, irrespective of gender or social role in the classroom. Evidence suggests that reductions in bullying, particularly among boys, leading to a more positive learning environment for all groups of pupils are the main driver behind these effects.

JEL Codes: I20,I21

*Braun: University of Zurich (Department of Economics), email: tabea.braun@econ.uzh.ch. Costa-Ramón: University of Zurich (Department of Economics & Jacobs Center for Productive Child and Youth Development) & CEPR, email: ana.costa-ramon@econ.uzh.ch. Rodríguez-González: University of Barcelona (Department of Econometrics, Statistics and Applied Economics & AQR-IREA), email: anarodriguezgonzalez@ub.edu. Schaede: Tufts University (Department of Economics), email: ursina.schaede@tufts.edu. Salmivalli: University of Turku (Department of Psychology), email: christina.salmivalli@utu.fi. We would like to thank Lorenzo Casaburi, Roberto Weber, Ulf Zölitz, and Josef Zweimüller for helpful comments and suggestions. We thank seminar participants at UAM, ifo institute for education, JFest, CSIC-IAE, Uppsala University, Helsinki GSE, and University of Zurich for valuable feedback. Elisa Alonso and Andrea Hofer provided outstanding research assistance. We gratefully acknowledge financial support from the Yrjö Jahnsson Foundation. Tabea Braun acknowledges research support by the University of Zurich's Research Priority Program 'Equality of Opportunity'. Costa-Ramón acknowledges research support by the Jacobs Foundation.

1 Introduction

Bullying remains a widespread problem in schools with about one out of five students aged 12–18 being bullied during the school year across both the US and Europe (Irwin et al., 2024; UNESCO, 2018). Studies have shown that experiencing bullying is correlated with poorer mental and physical health, worse academic performance, and lower earnings in adulthood (Gorman et al., 2021; Sarzosa, 2024; Eriksen et al., 2014; Ponzo, 2013; Wolke and Lereya, 2015). However, establishing a causal link proves challenging as unobserved factors may drive both victimization and adverse outcomes.

In this paper, we analyze the long-run impacts of the KiVa program, launched in Finnish compulsory schools in the late 2000s as the largest randomized controlled trial of a school-based anti-bullying intervention. We link rich survey data of the original RCT to comprehensive administrative records, allowing us to observe students' long-term educational trajectory, labor market outcomes, and criminal activity until age 30. We find substantial long-term gains for pupils who participated in KiVa: the treatment group is more likely to pursue the academic track in upper secondary education, more likely to obtain a university degree, and exhibits higher earnings by their late twenties. Evidence suggests that the likely mechanism behind these results is a reduction in bullying in the classroom, primarily among boys, that facilitates an improved learning experience and delivers gains in socio-emotional well-being for all students. Consistently, the returns of the program accrue to pupils independent of gender or social role at baseline, suggesting that reducing harmful behavior benefits not just those directly impacted but the broader social group in which it occurs.

The KiVa anti-bullying program was developed by a team of education scientists and psychologists in Finland as a school-based intervention. It is based on the premise that bullying takes place in a group context and that successful mitigation requires active involvement of bystanders. The program consists of two main components: about 20 age-appropriate classroom lessons delivered to all pupils over the school year, and targeted interventions that address bullying incidents through a series of meetings between a team of specialized teachers and the different parties involved. KiVa was launched as a large-scale RCT with a school-level randomization design. The RCT was implemented between 2007 and 2009 for around 30,000 pupils in Finnish compulsory schools, and the program was subsequently rolled out nationwide. It has since been adopted in 23 countries around the world.

We leverage KiVa's original RCT design to study its long-term impact until early adulthood. We focus on the largest part of the RCT, implemented among middle school students (grades 7-9), as its control group provides a clean counterfactual. Due to the subsequent national roll-out, the control group for younger students was eventually treated. Our main analysis sample comprises around 15,000 pupils in grades 7-9 whom we can link to national register data by Statistics Finland. Our match rate is 90% and balanced across treatment and control. We further document balance between treated and control students based on baseline observables measured in the register data.

We start by examining pupils' educational trajectory immediately after middle school at age

16. While we do not find treatment effects on being enrolled in post-compulsory education at the extensive margin, treated students are 5 percentage points (ppt; 11% over the mean) more likely to attend an academic high school track rather than vocational training. Subsequently, treated students are 4 ppt more likely to have passed the nationally graded matriculation exam at age 19, which qualifies for university entry and is typically taken at the end of the three-year post-secondary level. These gains in academic attainment persist: by ages 27-29, treated students are 4 ppt or 9% more likely to hold a university-level Bachelor's degree.

Do these gains in educational attainment translate into returns in the labor market? We measure participation and earnings outcomes at the latest available year in the data, when our cohorts are between 27 and 29 years old, and control for cohort fixed effects. We show that treated students are 1.2 ppt more likely to be "active", defined as being either in employment or a student, and less likely to be inactive, i.e. unemployed or out of the labor force for reasons other than education. We further observe higher earnings: treated students earn about 4.3% more relative to the control group (unconditionally), and 2.5% conditional on labor market participation.

We further explore heterogeneity for our main outcomes based on gender and social role at baseline. We do not observe systematic differences in returns to KiVa by pupil gender. Strikingly, we also estimate similar educational gains for all pupils regardless of whether they were classified as a bully, victim, or bystander by their peers at baseline. The benefits of the KiVa program thus accrue to all groups of pupils.

How does KiVa generate those returns? To analyze the mechanisms behind our long-term outcomes, we leverage rich survey data from the original RCT for grades 8-9, and administrative records on criminal behavior and grades for the full sample. We first document the impacts of the program on pupils' perceptions of their teacher's actions as well as harmful behavior in the classroom and beyond. Treated students are substantially more likely to indicate that their teacher discussed bullying in the classroom (+ 29 ppt, 115% over control mean) and became active against bullying (+ 20 ppt, 49% over control mean). Consequently, harmful behavior in the classroom is reduced: both treated girls and boys are more likely to report a perceived reduction in bullying. When examining peer-reported social roles, we observe a decline in being nominated as a bully that is concentrated among boys.

Are these reductions in harmful behavior temporary or do they persist beyond the immediate intervention period? Using records on criminal behavior in adulthood, we find a similar pattern: boys in the treatment group are significantly less likely to have committed a crime in adulthood. While suggestive, these patterns point towards KiVa being successful in reducing negative social externalities beyond the immediate time window of the intervention.

We next examine how these reductions in harmful behavior, concentrated among boys, impact the learning environment for all groups of pupils. Treated students score more highly on an index that combines measures of a positive school and classroom climate. They also score higher on an index of academic self-concept, being intellectually curious and experiencing learning as a joyful activity. Regarding academic grades measured in administrative data, we measure positive but noisy treatment impacts at the end of middle school. Improvements in the

learning environment coincide with higher socio-emotional well-being: treated students exhibit higher social self-concept, having more positive thoughts about themselves, and score lower on the Beck Depression Scale. Similarly to the main results, these gains are present for all groups of pupils.

In contrast, we do not find that KiVa leads to more social cohesion or friendships in the classroom, suggesting that the program did not primarily operate via facilitating the transition to academic high school through a change in peer networks. Rather, these patterns are consistent with KiVa and classroom teachers being successful in shutting down toxic peer interactions, leading to all pupils having a more positive and joyful learning experience that fosters a continued progression along the academic track.

We contribute to three main strands of literature. First, prior literature has examined the causal impact of being surrounded by disruptive peers by relying on cross-cohort variation in exposure. Being around these types of classmates has detrimental impacts, both in the short term (Lavy and Yancu, 2025; Kristoffersen et al., 2015; Lavy et al., 2012; Lavy and Schlosser, 2011; Carrell and Hoekstra, 2010; Gould et al., 2009; Figlio, 2007) and long after the initial exposure (Carrell et al., 2018). While providing an important existence proof, this type of research design offers no policy solution as reallocating students across classrooms merely redistributes harm in a zero-sum manner. In this context, our results based on a large-scale RCT provide first-hand evidence that negative interactions between peers can be mitigated, and the concrete mechanisms through which such mitigation is successful.

Second, we highlight that an intervention that fosters non-cognitive components in social interactions among children who are close to the transition of adulthood delivers long-term educational and labor market returns. As such, our results underscore that the window of opportunity to address behavioral issues and shape non-cognitive components of learning may be open beyond the most early ages. A recent set of papers mostly focused on elementary school pupils has documented that skills such as patience (Alan and Ertac, 2018), grit (Alan et al., 2019), perspective taking (Alan et al., 2021), or coping strategies (Rege et al., 2025) can be learned, and may deliver improvements in classroom climate or academic outcomes in the short term. Fostering these types of skills early on also matters in the longer run: Algan et al. (2022) show improved labor market outcomes from teaching self-control for 250 at-risk boys in Montreal, while Sorrenti et al. (2025) document that first graders who received socio-emotional skills training in a Swiss city are more likely to pursue an academic education track in early adulthood.

Third, while research on the causes and consequences of bullying is abundant in disciplines spanning sociology, psychology, and education sciences, to date we lack comprehensive causal evidence on the impacts of bullying based on objective outcome measures. Prior papers examining the impacts of KiVa have documented reductions in bullying and improvements in self-reported well-being based on survey data (Kärnä et al., 2011; Kärnä et al., 2013). Anti-bullying programs that follow a similar approach also reduce self-reported bullying (Gaffney et al., 2021; Huitsing et al., 2020; Olweus and Limber, 2010; Jiménez-Barbero et al., 2016), as does an intervention by Cunha et al. (2023) that targets parental emotional involvement.

This paper is structured as follows. The next section details the KiVa anti-bullying program and the Finnish school system. Section 3 describes the data. Section 4 describes the experimental design, and Section 5 presents the results. Section 6 discusses mechanisms, and the final section concludes.

2 Background

2.1 The KiVa program

The KiVa program (in Finnish: Kiosaamista Vastaan - Against Bullying) was launched as an RCT in 2006 by the Finnish Ministry of Education and Culture and psychologists and educational researchers at the University of Turku to tackle bullying in Finnish schools through a common curriculum (Salmivalli et al., 2011). The approach of KiVa is based on research documenting that bullies' actions are often motivated by seeking higher social status within a peer group (Salmivalli and Peets, 2018). Successfully addressing a situation of bullying therefore requires mobilizing bystanders to signal their general disapproval of the bully's actions and to actively interfere on behalf of the victim (Salmivalli et al., 2009).

In schools, the KiVa program is implemented in a two-pronged approach and aims to shift the group dynamics that sustain or tolerate bullying behavior (Kärnä et al., 2013): "Universal Actions" sensitize all students to bullying, while "Targeted Actions" are triggered in response to a bullying incident. The universal actions consist of 13-23 hours of structured lessons throughout the school year and a virtual learning environment. The lessons are led by class teachers and involve discussions, role play, and group exercises. For grades 7-9, these are centered around four main themes covering group dynamics and social pressure, types and consequences of bullying, and concrete strategies to counteract bullying. Figure A1 shows scenes from the teacher-led lessons and role-playing activities in the classroom. Additional KiVa materials include a parents' guide and visible school-wide symbols (e.g., posters and distinctive vests worn by supervising teachers) to keep anti-bullying themes salient. Targeted actions are implemented in response to a specific bullying incident and led by a designated team of KiVa teachers who iteratively meet with the bully, victim, and prosocial classroom peers, and ensure that the behavior stops.

Prior to implementation, the KiVa research team provided two days of face-to-face training for local school personnel. During the school year, networks of three schools met three times with a person from the KiVa program to exchange ideas and address potential obstacles in the program's implementation (Kärnä et al., 2011; Kärnä et al., 2013).

2.2 Prior Research and Roll-out of KiVa

The short-term effectiveness of KiVa in reducing bullying was evaluated through a large-scale RCT, which ran over two consecutive school years and involved over 200 schools. In 2007–08, the intervention was implemented in grades 4–6 (primary school, children aged 10–12). In 2008–09, the program expanded to grades 1–3 (ages 7–9) and grades 7–9 (secondary school, ages 13–15). Previous evaluations examined the impact of the program on survey measures collected

immediately at the end of the academic year. These studies report significant reductions in bullying, with the strongest effects observed among fourth graders and the weakest among students in grades 7–9 (Kärnä et al., 2011; Kärnä et al., 2013; Salmivalli and Poskiparta, 2012).[1] The program also improved subjective well-being, including a more positive attitude towards school, potentially extending the benefits beyond those directly involved in bullying, such as previous victims or perpetrators (Salmivalli et al., 2012). Between 2009 and 2011, KiVa was rolled out in all Finnish schools, with the KiVa curriculum targeted to grades 1, 4, and 7 (Kärnä et al., 2011). By 2011, 90 percent of all Finnish comprehensive schools were registered as KiVa users (Salmivalli and Poskiparta, 2012). KiVa has since been implemented across a total of 23 countries worldwide. Beyond Finland, the effectiveness of KiVa in reducing bullying among 8–10-year-olds has also been confirmed in a randomized evaluation in the Netherlands (Huitsing et al., 2020).

2.3 Finnish School System

During the period of our study, compulsory education in Finland starts at age 7 and spans grades 1 through 9. Pupils start in primary school for the first 6 years, and subsequently attend lower-secondary education (hereafter: middle school) for another three years. Students typically complete middle school at ages 15-16. The curriculum follows a nationwide standard, but municipalities are in charge of the detailed curriculum design. All students follow a common track during this period.

Upon completing grade 9, students may apply to non-compulsory, upper-secondary education, which spans three years and is divided into an academic and a vocational track. The academic track prepares students for tertiary education and culminates in the national matriculation examination (typically taken at age 19), a standardized exam required for university admissions. Vocational secondary school provides subject-specific occupational training and includes workplace-based learning. While the vocational track is primarily geared towards direct labor market entry, it also provides a pathway to higher education, typically to universities of applied sciences. Admissions to both upper-secondary tracks are based on grade point averages (GPA) from middle school. In particular, admission criteria of academic high schools typically place more weight on academic subjects like Finnish/Swedish and mathematics. Students submit ranked applications through a centralized system during grade 9, and are placed in the highest-preferred school for which they meet the GPA threshold and where seats are available (Huttunen et al., 2023).

For tertiary education, students are admitted to university studies based on their performance in the matriculation exam and institution-specific entry exams. Students who completed the upper-secondary vocational track may apply to universities of applied sciences, which offer professionally oriented higher education. Transitioning from the vocational track to academic universities is possible, but less common in practice.

[1] In grades 7-9, Kärnä et al. (2013) find that the program's effects were stronger for boys.

3 Data

3.1 Survey Data

The school year in Finland runs from mid-August to the end of May. During the KiVa RCT for grades 7–9, data were collected in three waves: May 2008 (academic year before the intervention), December 2008–February 2009, and May 2009 (endline survey). Baseline data was only collected for grades 8 and 9 due to the fact that, at the time of baseline data collection, grade 7 students were still in primary school and not in middle school. We use data from the baseline and endline surveys. Students filled out internet-based questionnaires in the schools' computer labs during regular school hours. The process was administered by the teachers, who were supplied with detailed instructions about two weeks prior to data collection. The teachers were told to act in such a way that the confidentiality of the response was secured to a maximum extent, and students were assured that their answers would not be revealed to teachers or parents.

The surveys collected information on student perceptions and peer dynamics, as well as some background characteristics of students, classes, and schools. We pursue a more agnostic and comprehensive approach and use data on three main domains: harmful behavior, students' socio-emotional wellbeing, and measures of social cohesion and relationships with peers. These are described in detail below. To reduce the number of statistical tests, we create separate indices for each outcome family and wave, following Anderson (2008).

Harmful behavior — The survey collected measures of pupils' perceived implementation of the program by classroom teachers, perceived changes in bullying in the classroom, and peer-reported bullying and victimization measures.

(a) Implementation of the program by classroom teachers. Students were asked about their perceptions of teachers' role with respect to bullying: "Has the teacher touched the issue of bullying during any lesson since last autumn?", "How does the teacher think of bullying?", "How much has the teacher done in order to decrease bullying since last autumn?", and "How much can the teacher do in order to decrease bullying?". They could answer from a scale of 0 (no/good thing/nothing) to 4 (more than 8 times/absolutely wrong/very much), with higher values thus indicating that they perceive teachers to talk about bullying more frequently, to have a more negative opinion of bullying, to have done more to fight bullying, and to have the ability to do more in order to tackle it. To understand the prevalence of the different practices, we construct the following indicators: a dummy equal to one if the teacher touched the topic of bullying more than once, an indicator equal to one if the teacher has a negative or very negative opinion of bullying, and indicators equal to one if the teacher took at least some action against bullying and is perceived to be able to do at least something to decrease bullying.

(b) Students' perceived change in bullying. Participants were asked about their perceived change in "the amount of bullying in your class compared to last autumn", using a five-point Likert scale. We construct a binary indicator equal to one if the respondent reports that bullying did not change, increased, or increased a lot, and zero otherwise.

(c) Peer-reported bullying and victimization. To measure bullying and victimization we rely on peer reports, where students nominate from a list of classmates those that bully others or

are victimized. Students see a list of their classmates' names and can mark none or several names of their peers whose behavior matches the different questions. For bullies, the questions are: "Starts bullying", "Gets others involved in bullying", "Always comes up with new ways to bully". For victims, "He/She is being pushed around and hit", "He/She is called names and mocked", "Nasty rumors are spread about him/her". The peer-reported measures are continuous variables which capture the share of responding classmates that nominate a child for each question. In order to measure the prevalence of bullying and victimization, we also dichotomize these variables and construct indicators for being a (peer-reported) bully or victim. These indicators take value one for students who, across the different questions, are on average nominated by at least 10% of their peers.[2]

Socio-emotional wellbeing — The survey collected measures of students' depression and anxiety symptoms, their social and academic self-concept, and their perceptions of the classroom and school environment.

The depression measure is a seven-item scale derived from the Beck Depression Inventory, a clinically validated instrument for assessing depressive symptoms (Beck et al., 1996). Students were asked to report how they had felt during the previous two weeks in response to the following questions: (i) How was your mood?, (ii) How do you look at the future?, (iii) How do you feel your life has been running?, (iv) How satisfied or dissatisfied do you feel yourself?, (v) How do you see yourself?, (vi) Do you feel senses of disappointment?, and (vii) How do you feel about your being and appearance?. Each item was answered on a five-point Likert scale ranging from 0 (fairly bright and good) to 4 (very depressed). Anxiety is measured using a nine-item scale capturing two dimensions of social anxiety: fear of negative evaluation (FNE) and social avoidance and distress (SAD). Students were asked to report how well each statement described them, using a five-point Likert scale ranging from 0 (not at all) to 4 (all the time). The FNE subscale consists of five items: (i) I'm worried about what the others think of me, (ii) I'm afraid the others won't like me, (iii) I'm worried about what the others talk about me, (iv) I'm worried that the others don't like me, and (v) If I have to argue about something, I'm afraid that the other won't like me. The SAD subscale consists of four items: (vi) I stay quiet when I'm in a group of people, (vii) I'm afraid of asking others to do things with me as they might turn me down and don't do things with me, (viii) I feel quite shy even among those mates I know well, and (ix) It's difficult for me to ask others to do things with me.

The questions used to measure students' social self-concept are based on the Rosenberg Self-Esteem Scale, adapted for children to report how they feel about themselves when around peers (Rosenberg, 1965; Salmivalli et al., 2005). Students indicated, on a scale from 0 to 4, how accurately each statement described their feelings about themselves among peers. The statements were: (i) I am more or less satisfied with myself, (ii) I feel I am not good enough for anything, (iii) I feel that I have a number of good qualities, (iv) I feel I do things as well as the others, (v) I don't feel I have much to be proud of, (vi) I sometimes feel really useless, (vii) I feel I am as valuable (as a person) as the others, (viii) I hope I could respect myself more, (ix)

[2]The survey also collected self-reported measures of bullying and victimization following Olweus' Bully/Victim Questionnaire (Olweus, 1996). Figure A2 shows that self-reported measures identify a substantially lower share of both bullies (7% vs 16%) and victims (9% vs 21%), suggesting significant under-reporting.

I consider myself a failure, and (x) I have positive thoughts of myself.

Similarly, pupils' academic self-concept is measured based on children's degree of agreement, on a scale from 0 to 4, with the following statements: (i) Learning brings me joy, (ii) I want to know and learn many different things, and (iii) I am doing fine at school.

Finally, we use information from several questions designed to measure students' perceptions of the classroom and school environment, as well as their general liking of school, to construct a composite measure of school climate. Students indicated their level of agreement with each statement on a scale from 0 (totally disagree) to 4 (totally agree). Perception of the classroom climate is measured using three statements: (i) There is a good atmosphere in my class, (ii) Helping others is common in our class, and (iii) I am happy to be in my class. Perception of the school climate is measured using five statements: (iv) I feel safe at school, (v) I am satisfied with the atmosphere of my school, (vi) Also those pupils who are different from the others are accepted at school, (vii) I am thought of highly at school, and (viii) I feel accepted as I am at school. Finally, students' general liking of school is measured using three statements: (ix) My schooldays are usually nice, (x) I like going to school, and (xi) I am happy with going to school in general.

Social cohesion — We measure social cohesion using three sets of questions capturing students' friendships, their general perceptions of peers, and their empathy toward victims.

Friendships (relations to classmates) are measured using three statements, answered on a scale from 0 (totally disagree) to 4 (totally agree): (i) I have good friends in my classroom, (ii) I have mates/friends in my own class, and (iii) I feel it easy to get along with my classmates. General perception of peers is measured using thirteen statements referring to students' agemates in general, answered on a scale from 0 (not true at all) to 4 (exactly true): (i) can really be relied on, (ii) really care about what happens to me, (iii) are there for me whenever I need help, (iv) shouldn't be trusted too much, (v) don't really care about me, (vi) only think about their own interest, (vii) betray one's trust whenever they get the chance, (viii) want to hurt me, (ix) can be confided in, (x) are honest with me, (xi) think bad things about me, (xii) usually have good intentions, and (xiii) are hostile. Empathy toward victims is measured using seven items, answered on a scale from 0 (never) to 3 (always): (i) When the bullied pupil is sad, I also feel sad, (ii) When the bullied pupil feels sad, I want to comfort him/her, (iii) When the bullied pupil starts to cry, I also feel bad, (iv) When someone is bullied, I start to get angry on his/her behalf, (v) I can understand how the bullied pupil must feel, (vi) I can see how the bullied pupil is feeling bad, and (vii) I can imagine how the bullied pupil must feel, even if he/she would not tell.

3.2 Administrative Data

We merge the original survey data from the KiVa RCT with rich administrative records maintained by Statistics Finland. The linked dataset allows us to track students over time and observe a wide range of demographic characteristics and long-run education and labor market outcomes, from compulsory schooling through early adulthood.

The primary source of background information is the Finnish Population Register (FOLK),

which provides detailed data on students' family and socioeconomic background, and measures of educational attainment and labor market performance. From this register, we construct baseline covariates such as gender, age, household composition, household income based on the occupation of the household head (as classified by Statistics Finland), and immigration background, all measured in 2007 (the year before program implementation).[3] We also obtain measures of students' educational attainment. First, we study pupils' enrollment in post-compulsory education: whether they attend an upper secondary school at age 16, and whether they are enrolled in the academic or vocational track at this age. We then follow their trajectories and measure if they passed the university entrance examination (which serves as an academic high school leaving certificate) by age 19, and whether they obtained a college degree by 2022, the last year we observe them in the data (at ages 27-29, depending on the cohort). We also study their labor market outcomes at the ages of 27-29: their labor force participation (whether they are employed or studying, unemployed, or out of the labor force) and their labor earnings (including wages and income from self-employment).

We supplement this dataset with additional registers. To measure academic performance in lower secondary school, we use data from the National Register of Applications to Upper Secondary Education. This register includes students' final (teacher-assigned) grades in grade 9, the final year of compulsory education. We focus on students' performance in core subjects—native language and mathematics—which play a key role in determining access to academic versus vocational tracks. We standardize grades to have a mean of zero and a standard deviation of one in the control group. Finally, we use data on Prosecutions, Sentences, and Punishments based on district court rulings up to 2023 and construct an indicator for being convicted of any crime in adulthood, from age 18 up to ages 28-30.

4 Empirical Strategy

To evaluate the long-run effects of KiVa, we leverage its randomized implementation in grades 7 to 9 during the 2008–09 school year. The program was scaled up nationally starting in the fall of 2009. By focusing on grades 7 to 9, we ensure a clean control group that was not exposed or only minimally exposed to the program.[4]

For the recruitment, the KiVa team sent letters in the fall of 2006 to all schools providing basic education in mainland Finland. These included both Finnish-language and Swedish-language schools. The volunteering schools (excluding special-education-only schools) were stratified by province and language and randomly assigned to the intervention. For grades 7-9 the final sample consisted of 38 intervention and 35 control schools.[5]

[3] Given that we cannot directly observe parental income, we proxy household income by the average disposable income of all adults in Finland (with children from the same cohorts) who have the same occupation as the household head, and construct an indicator for being above or below the median.

[4] Students in the control group during grade 9 were never exposed to the intervention. Some of the students in the control group during grades 7 and 8 could be exposed to some of the light-touch elements of the intervention, such as the school posters, when their school adopted the program as part of the national rollout. However, since in the scaled-up program KiVa lessons are only provided in grades 1, 4, and 7, these students were never exposed to these more intensive elements of the intervention.

[5] Originally, there were 39 schools in each treatment arm, but four control schools dropped out without

4.1 Empirical Model

We estimate the effect of the KiVa intervention using the following specification:

$$Y_i = \beta_0 + \beta_1 \text{Treat}_s + X_i'\gamma + \theta_s + \epsilon_i, \tag{1}$$

where Y_i is the outcome of interest and Treat_s is a binary variable indicating if the student's school s participates in the anti-bullying program. θ_s represents stratification-level fixed effects which account for the province of the school and the language of instruction. X_i is a vector of student i's pre-determined baseline characteristics. We use a post-double-selection (PDS) lasso to determine the set of controls (Belloni et al., 2014). As potential controls, we feed the model with a rich set of student characteristics observed in administrative data including students' age, gender, and socioeconomic background.[6] When studying outcomes that we observe at different ages for the different cohorts, such as earnings, we include school cohort fixed effects. Finally, where available (when studying mechanisms with survey data), we also control for the baseline measure of the outcome variable. Standard errors are clustered at the school level.

4.2 Sample and Balance

For our main analysis, our sample consists of all students who could be successfully linked to administrative data. We can merge the information of 90% of students that participated in KiVa to administrative records, and the matching rate is similar for the treatment and the control group. Our final sample consists of 15,088 students. The administrative records are largely complete. We observe secondary school attendance at age 16 for 15,087 students, whether they take the matriculation exam at age 19 for 15,068, and educational attainment and labor market outcomes at ages 27-29 for over 97% of the sample.

Columns 1–3 in Table 1 show that the control and treatment groups are well balanced across baseline characteristics. The share of female pupils is similar in the two groups (53% in the control group and 52% in the treatment group). In both groups, only a very small fraction of students have an immigrant background (1% and 2%, respectively), and the majority of students live in their region of birth (89%) and in urban areas (73% and 76%, respectively). Age-for-grade measures are identical across groups and family characteristics are also balanced: 17% of pupils live in a single-parent household in both groups, and 34% of pupils in the control group and 31% in the treatment group live in households with below-median disposable income. A joint test of orthogonality fails to reject the null hypothesis of no systematic differences between the groups.

To study mechanisms, we use survey data from the KiVa RCT. Our main mechanism sample consists of all students that participated in both the baseline and endline surveys. These are 8,154 out of the total of 15,088 students for whom we observe administrative outcomes. The main reason for the smaller sample size of the mechanism sample is the lack of baseline survey

providing any information, and one treatment school participated only in the baseline wave of data collection. Table 1 shows that the schools included in the final sample are similar in terms of characteristics.

[6] Full list of variables in Panel A in Table 1.

data for grade 7 (see Section 3.1). Columns 4–6 in Table 1 compare the baseline characteristics of the treatment and control groups for those in the mechanism sample. Both groups are also similar in background characteristics we can measure in administrative records. Panel B of Table 1 further shows that the treatment and control groups are also well balanced with respect to baseline characteristics measured in survey data. Rates of bullying and victimization at baseline are very similar across groups. In the control group, 15% of students are identified as bullies and 22% as victims by their peers. Similarly, in the treatment group, 15% are identified as bullies and 20% as victims. Summary indices of bullying and victimization at baseline are likewise balanced. Class size is also similar across treatment arms, with an average of around 20 students (of which we can observe approximately 17 in our data).

5 Results

5.1 Educational Attainment

In Panel (a) of Figure 1, we show that the intervention has a lasting impact on students' educational attainment by tracing pupils' trajectories through the key educational stages of the Finnish school system. By age 16, almost all students attend upper secondary school after compulsory education (89% in the control group), and we do not find a treatment effect on the extensive margin of overall enrollment. However, there is a shift in their trajectory path: the share of students enrolling in academic high school track (instead of vocational training) increases significantly by 5.2 percentage points—an effect of 11% relative to the control mean.

The impacts of the intervention persist over time, with a remarkably stable relative effect size. By age 19, treated students are 4.1 percentage points more likely to graduate from academic high school and pass the matriculation exam, which qualifies for university entry and is usually taken at the end of the three-year upper secondary education. This effect corresponds to a 9.6% increase over the control group mean. Finally, in the last year we observe our cohorts in the administrative records, when they are aged 27-29 years old, the treatment group is 3.7 percentage points or 8.8% more likely to hold a university degree. These findings suggest that exposure to the KiVa program effectively shifts students' educational trajectories towards the academic track, making them more likely to ultimately complete a university degree.

5.2 Labor Market Outcomes

We now turn to labor market outcomes at ages 27–29. Panel (b) in Figure 1 shows that treated students are slightly more likely (by 1.2 percentage points or 1.4%) to be active in the labor market, either employed or studying, at this age, and 0.9 percentage points (13%) less likely to be unemployed.

Panel (c) shows effects on labor income. KiVa participation raises students' annual earnings by around 1,200 euro, which represents an increase of 4.3% relative to the control group mean. The last row in this panel shows that, conditional on labor force participation, treated students earn 830 euro or 2.5% more than students in the control group. These results indicate that

the higher annual earnings of the treatment group likely reflect both higher wages and a higher degree of labor market attachment.

5.3 Heterogeneity by Gender and Social Role

Tables A1 and A2 explore gender heterogeneity in our main results. We see that the treatment effects are similar for male and female students—for none of the outcomes can we rule out the null hypothesis of equal effects. In Tables A3 and A4 we further explore heterogeneity by whether students are identified as bullies, victims, or bystanders in the baseline survey, using the indicators described in Section 3.1. We see positive effects on educational outcomes for all three groups, again suggesting that the intervention benefited everyone in class. In terms of labor market outcomes, we see slightly larger increases in earnings for bystanders.

The fact that the long-run impacts of the program are similar for different groups suggests that its benefits also extend to those students who were not directly involved in bullying dynamics. To delve further into how the intervention affected the average student in the classroom, we next study the effects of KiVa on the the incidence of harmful behavior, on students' socio-emotional well-being, on their social connections, and on their academic performance while in school.

6 Mechanisms

The results presented in Section 5 show that the KiVa program has a long-lasting impact on students' outcomes, increasing their educational attainment and their earnings in early adulthood. In order to understand the mechanisms through which the intervention may have shifted their trajectories, we exploit the rich survey data collected during the RCT. Throughout this section, we control for each outcome with its baseline value and therefore rely on our "Mechanism Sample" that consists of pupils attending grades 8 and 9. Students in grade 7 did not participate in the baseline survey as this survey was conducted at the end of the prior academic year, i.e., while grade 7 students were still in primary school.

6.1 Reduction of Harmful Behavior

Implementation of the program by classroom teachers — We start by examining how pupils perceive the implementation of the program and the role of their classroom teacher in the RCT's endline survey. In Figure 2, we see that treated pupils are substantially more likely to report that their classroom teacher has become active regarding bullying: we observe an increase of 29 ppt, or 115% over the control group mean increase in the teacher having discussed bullying at least twice during the academic year. Treated students are also about 50% more likely to report that their teacher has taken actions in order to decrease bullying (61% vs 41%). Treated students also perceive their teachers as somewhat more capable to do something against bullying with a 13% increase in perceiving that their teacher has at least some influence to reduce bullying and an 12% increase in perceiving that their teacher opposes bullying. Taken together, while there are modest treatment impacts on students' perception of how teachers feel about and might be

able to intervene against bullying, we see large perceived increases in teachers' concrete actions against bullying.

Short-term harmful behavior — We next examine changes in overall and peer-reported bullying. Panel (a) in Figure 3 shows treatment impacts on pupils' perceived change in bullying. The outcome is an indicator equal to one if students report that bullying has either stayed constant or increased over the past academic year. We see that both treated boys and girls are about 4 ppt (8% over the mean) less likely to agree that bullying has stayed the same or increased, indicating a perceived decrease in the occurrence of bullying.

Regarding peer nominations for social roles, we construct an index for each pupil based on their peers' nominations as a bully or victim.[7] Panel A in Figure 3 documents treatment impacts on these indices for peer-reported bullying and victimization outcomes. On average, there is a non-significant decrease in bullying of .04 SD, and a non-significant decrease in victimization of .07 SD. When examining effects by gender, there is little change in bullying rates among female students. However, we find significant reductions in bullying of around 10 percent of a standard deviation for males. Among boys, the estimated reduction in bullying corresponds to closing roughly 20 percent of the baseline gender gap. In terms of victimization, we find a statistically significant reduction only for boys; nonetheless, we observe decreases for both groups and cannot reject the null of equal effects for males and females. These results suggest that the program was effective in changing the behavior of those more likely to harass other students—boys. These changes in problematic behavior seem to benefit girls and boys alike.

Long-term harmful behavior — Do these changes in behavior persist over time? To answer this question, we use administrative data from the crime register and study the probability of committing any crime up the last available year in our data, when individuals in our sample are aged 28–30 years old. The results are presented in Panel B of Figure 3. We find that males in the treatment arm are 2.7 percentage points less likely to commit any crime. This represents a 17.2% decrease compared to the average crime rate of males in the control group (15%).

Overall, these findings indicate that the intervention had a significant impact on the incidence of problematic behavior among boys while in school, and that these changes persisted over time. We find no changes in bullying or crime rates among females, who have a much lower probability of taking part in these activities.

6.2 Socio-Emotional Well-being

The first bar in Figure 4 shows treatment effects on an index that captures students' perceptions of the school and classroom climate with a battery of questions covering pupils' liking of their school and classroom as well as perceptions around safety. Treated students score 6 percent of a standard deviation higher on this index, mostly driven by students' increased enjoyment of school. Improved school and classroom climate is directly tied to improvements in students' own well-being. While we do not observe statistically significant reductions in a (social) anxiety

[7] For bullies, peers are asked to nominate students who "start bullying", "make the others join in the bullying", or "always finds new ways of harassing the victim". For victims, peers are asked to nominate students who are "pushed and hit", "called with nasty names or made fun of" or "usually talked about with bad tone".

index, we see a significant decrease of .6 SD in depressive symptoms, as measured by Beck et al. (1996) depression scale. This increase in students' well-being is also visible in their self-esteem in relation to others and their attitudes towards learning. The fourth bar of Figure 4 shows that treated students have higher levels on an index measuring social self-concept among their peers (Rosenberg, 1965). We further observe positive treatment impacts on an index of academic self-concept that elicits students' joy and curiosity for learning (rather than performance in terms of grades). Overall, these results suggest that the intervention created a more positive classroom and school environment in which students feel better about themselves and are able to pursue their academic curiosity and experience learning as a joyful activity.

6.3 Social Connections and Empathy

Does KiVa create new friendships and social connections, or operate via having more empathy and helping each other out? We find little evidence that speaks for such a mechanism. First, in the first bar in Figure 5, we examine pupils' self-assessed relationships with their classmates in an index that combines three questions on having good friends and relationships with classmates. We see no treatment impact on this dimension. Similarly, KiVa does not change students' perceptions of either good (e.g. reliable, honest, support, helpful) or bad (e.g. not trustworthy, hostile, egoistic) qualities of their peers, as measured with an index in the second bar of Figure 5. The positive impacts of KiVa are also unlikely to operate via increased empathy for victims: we do not find a treatment impact on an index that captures feeling for and empathizing with a bullied classmate in the third bar of Figure 5.

6.4 Academic Performance

Finally, Figure 6 shows treatment effects on teacher-assigned, standardized grades in grade 9 (the last year of compulsory school) for core subjects: native language and mathematics. While the point estimates for treated students are positive (Math: .06 SD, native language: .09 SD) we fail to reject that coefficients are statistically different from zero. We do not detect significant improvements in students' self-assessed academic performance either, with treated students equally likely to report that they face no difficulties across a range of academic subjects. While we cannot rule out that small increases in grades contribute to higher enrollment in the academic upper secondary track, these improvements go hand-in-hand with higher academic motivation and a positive perception of learning, as outlined above.

6.5 Discussion of Mechanisms

Taken together, these different pieces of evidence suggest that KiVa was effective in reducing harmful behavior in the classroom and beyond, primarily among boys. The benefits from reducing negative peer interactions extend to everyone in class: we see an improved perception of the classroom environment, increases in self-reported well-being while in school, and more curiosity and joy in learning. While impacts on academic grades, a key evaluation criteria for the transition to post-compulsory education, are noisy, the point estimates are positive. Taken

together, these patterns are consistent with a Lazear (2001) model of education, where reducing disruptions by a few students mitigates negative externalities for the whole class.

7 Conclusions

This paper provides the first causal evidence on the long-run effects of a large-scale, school-based anti-bullying intervention. Leveraging experimental variation from a randomized controlled trial in Finnish secondary schools and linking rich survey data to administrative records, we show that exposure to the KiVa program at ages 13-15 has lasting effects on students' socio-economic trajectories. Treated students are significantly more likely to attend academic secondary school and obtain a university degree, and they earn higher wages at ages 27-29.

We find that these effects are likely driven by reductions in harmful behavior in the classroom and beyond, especially among male students, which create a more positive learning environment and improve the socio-emotional well-being for all groups of pupils. As such, we observe similar educational and labor market gains for all groups of students, including those not directly involved in bullying dynamics. In contrast, we not find evidence that these effects operate through other changes in the social context, such as facilitating new friendships or fostering empathy.

Our findings contribute to a broader literature on the returns to non-cognitive interventions during adolescence, highlighting that programs aimed at stopping negative peer dynamics can improve socio-emotional well-being and yield meaningful labor market returns.

References

Alan, S., C. Baysan, M. Gumren, and E. Kubilay (2021). Building Social Cohesion in Ethnically Mixed Schools: An Intervention on Perspective Taking. *Quarterly Journal of Economics 134*(4), 2147–2194.

Alan, S., T. Boneva, and S. Ertac (2019). Ever failed, try again, succeed better: Results from a randomized educational intervention on grit. *The Quarterly Journal of Economics 134*(3), 1121–1162.

Alan, S. and S. Ertac (2018). Fostering patience in the classroom: Results from randomized educational intervention. *Journal of Political Economy 126*(5), 1865–1911.

Algan, Y., E. Beasley, S. Côté, J. Park, R. E. Tremblay, and F. Vitaro (2022, August). The Impact of Childhood Social Skills and Self-Control Training on Economic and Noneconomic Outcomes: Evidence from a Randomized Experiment Using Administrative Data. *American Economic Review 112*(8), 2553–2579.

Anderson, M. L. (2008). Multiple inference and gender differences in the effects of early intervention: A reevaluation of the abecedarian, perry preschool, and early training projects. *Journal of the American Statistical Association 103*(484), 1481–1495.

Beck, A. T., R. A. Steer, and G. Brown (1996). *Beck depression inventory–II.* APA PsycTests.

Belloni, A., V. Chernozhukov, and C. Hansen (2014, apr). Inference on Treatment Effects after Selection among High-Dimensional Controls. *The Review of Economic Studies 81*(2), 608–650.

Carrell, S. E., M. Hoekstra, and E. Kuka (2018, November). The long-run effects of disruptive peers. *American Economic Review 108*(11), 3377–3415.

Carrell, S. E. and M. L. Hoekstra (2010, January). Externalities in the classroom: How children exposed to domestic violence affect everyone's kids. *American Economic Journal: Applied Economics 2*(1), 211–28.

Cunha, F., Q. Hu, Y. Xia, and N. Zhao (2023). Reducing Bullying: Evidence from a parental involvement program on empathy education. *NBER Working Paper 30827*.

Eriksen, T. L. M., H. S. Nielsen, and M. Simonsen (2014). Bullying in elementary school. *Journal of Human Resources 49*(4), 839–871.

Figlio, D. N. (2007). Boys named sue: Disruptive children and their peers. *Education Finance and Policy 2*(4), 376–394.

Gaffney, H., M. M. Ttofi, and D. P. Farrington (2021). Effectiveness of school-based programs to reduce bullying perpetration and victimization: An updated systematic review and meta-analysis. *Campbell Systematic Reviews 17*(2), e1143.

Gorman, E., C. Harmon, S. Mendolia, A. Staneva, and I. Walker (2021). Adolescent school bullying victimization and later life outcomes. *Oxford Bulletin of Economics and Statistics 83*(4), 1048–1076.

Gould, E. D., V. Lavy, and M. Daniele Paserman (2009). Does immigration affect the long-term educational outcomes of natives? Quasi-experimental evidence. *The Economic Journal 119*(540), 1243–1269.

Huitsing, G., G. M. Lodder, W. J. Browne, B. Oldenburg, R. Van der Ploeg, and R. Veenstra (2020, jul). A Large-Scale Replication of the Effectiveness of the KiVa Antibullying Program: a Randomized Controlled Trial in the Netherlands. *Prevention Science 21*(5), 627–638.

Huttunen, K., T. Pekkarinen, R. Uusitalo, and H. Virtanen (2023). Lost boys? Secondary education and crime. *Journal of Public Economics 218*, 104804.

Irwin, V., K. Wang, J. Cui, and A. Thompson (2024). Report on indicators of school crime and safety: 2023. nces 2024-145/ncj 309126. *National Center for Education Statistics*.

Jiménez-Barbero, J. A., J. A. Ruiz-Hernández, L. Llor-Zaragoza, M. Pérez-García, and B. Llor-Esteban (2016). Effectiveness of anti-bullying school programs: A meta-analysis. *Children and Youth Services Review 61*, 165–175.

Kärnä, A., M. Voeten, T. D. Little, E. Alanen, E. Poskiparta, and C. Salmivalli (2013). Effectiveness of the Kiva antibullying program: Grades 1-3 and 7-9. *Journal of Educational Psychology 105*(2), 535–551.

Kärnä, A., M. Voeten, T. D. Little, E. Poskiparta, E. Alanen, and C. Salmivalli (2011). Going to scale: A nonrandomized nationwide trial of the kiva antibullying program for grades 1–9. *Journal of Consulting and Clinical Psychology 79*(6), 796–805.

Kärnä, A., M. Voeten, T. D. Little, E. Poskiparta, A. Kaljonen, and C. Salmivalli (2011). A large-scale evaluation of the KiVa antibullying program: Grades 4-6. *Child Development 82*(1), 311–330.

Kristoffersen, J. H. G., M. V. Krægpøth, H. S. Nielsen, and M. Simonsen (2015). Disruptive school peers and student outcomes. *Economics of Education Review 45*, 1–13.

Lavy, V., M. D. Paserman, and A. Schlosser (2012). Inside the black box of ability peer effects: Evidence from variation in the proportion of low achievers in the classroom. *The Economic Journal 122*(559), 208–237.

Lavy, V. and A. Schlosser (2011, April). Mechanisms and impacts of gender peer effects at school. *American Economic Journal: Applied Economics 3*(2), 1–33.

Lavy, V. and A. Yancu (2025, November). Violent peers at school: Impacts and mechanisms. Working Paper 34482, National Bureau of Economic Research.

Lazear, E. P. (2001). Educational production. *The Quarterly Journal of Economics 116*(3), 777–803.

Olweus, D. (1996). *The revised Olweus bully/victim questionnaire*. Bergen: University of Bergen, Research Center for Health Promotion.

Olweus, D. and S. P. Limber (2010). Bullying in school: Evaluation and dissemination of the O lweus bullying prevention program. *American Journal of Orthopsychiatry 80*(1), 124–134.

Ponzo, M. (2013). Does bullying reduce educational achievement? An evaluation using matching estimators. *Journal of Policy Modeling 35*(6), 1057–1078.

Rege, M., E. Bru, I. F. Solli, M. W. Thijssen, K. B. Tharaldsen, L. Vestad, S. K. Ertesvåg, T. Ogden, and P. N. Stallard (2025). The impact of teaching coping skills in schools on youth mental health and academic achievement: Evidence from a field experiment. Technical report, CESifo Working Paper.

Rosenberg, M. (1965). *Society and the adolescent self-image*, Volume 11. Princeton, NJ: Princeton University Press.

Salmivalli, C., C. F. Garandeau, and R. Veenstra (2012). KiVa anti-bullying program: Implications for school adjustment. In *Peer Relationships and Adjustment at School*, Adolescence and Education, pp. 279–305. Charlotte, NC, US: IAP Information Age Publishing.

Salmivalli, C., A. Kärnä, and E. Poskiparta (2009). From peer putdowns to peer support: A theoretical model and how it translated into a national anti-bullying program. In *Handbook of Bullying in Schools*, pp. 441–453. Routledge.

Salmivalli, C., A. Kärnä, and E. Poskiparta (2011). Counteracting bullying in Finland: The Kiva program and its effects on different forms of being bullied. *International Journal of Behavioral Development 35*(5), 405–411.

Salmivalli, C., T. Ojanen, J. Haanpää, and K. Peets (2005, mar). "I'm OK but you're not" and other peer-relational schemas: Explaining individual differences in children's social goals. *Developmental Psychology 41*(2), 363–375.

Salmivalli, C. and K. Peets (2018). Bullying and victimization. In W. Bukowski, B. Laursen, and K. H. Rubin (Eds.), *Handbook of peer interactions, relationships, and groups, 2nd ed.*, pp. 302–321. New York, NY, US: The Guilford Press.

Salmivalli, C. and E. Poskiparta (2012). KiVa antibullying program: Overview of evaluation studies based on a randomized controlled trial and national rollout in Finland. *International Journal of Conflict and Violence (IJCV) 6*(2), 293–301.

Sarzosa, M. (2024). Victimization and skill accumulation: The case of school bullying. *Journal of Human Resources 59*(1), 242–279.

Sorrenti, G., U. Zölitz, D. Ribeaud, and M. Eisner (2025). The Causal Impact of Socio-Emotional Skills Training on Educational Success. *The Review of Economic Studies 92*(1), 506–552.

UNESCO (2018). School violence and bullying: Global status and trends, drivers and consequences. Technical report, Paris.

Wolke, D. and S. T. Lereya (2015). Long-term effects of bullying. *Archives of disease in childhood 100*(9), 879–885.

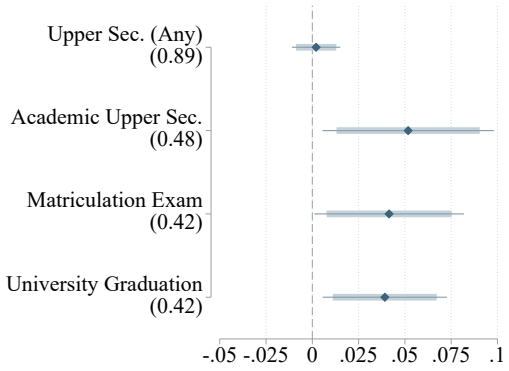
Tables and Figures

Table 1: Summary Statistics and Balance

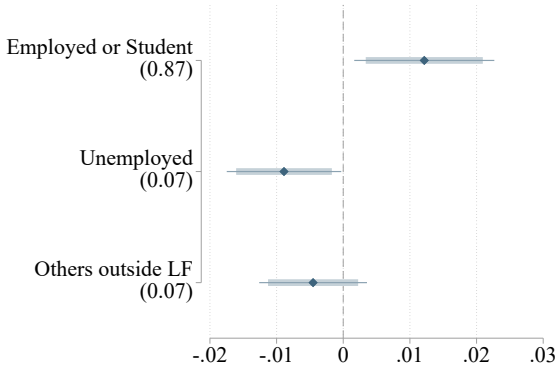
	Admin. Sample			Mechanism Sample		
	Control (1)	Treatment (2)	T - C (3)	Control (4)	Treatment (5)	T - C (6)
Panel A: Admin Controls						
Female	0.53	0.52	−0.00	0.53	0.52	−0.01
			(0.01)			(0.01)
Immigrant Background	0.01	0.02	0.01	0.01	0.02	0.01
			(0.01)			(0.01)
Lives in Region of Birth	0.89	0.89	0.00	0.90	0.89	−0.01
			(0.01)			(0.02)
Lives in Urban Area	0.73	0.76	0.04	0.73	0.77	0.04
			(0.05)			(0.05)
Young for Grade	0.01	0.01	−0.00	0.01	0.01	−0.00
			(0.00)			(0.00)
Old for Grade	0.05	0.05	0.00	0.04	0.04	−0.00
			(0.01)			(0.01)
Number Siblings	1.54	1.58	0.03	1.50	1.55	0.04
			(0.07)			(0.07)
Age Youngest Child in Family	10.11	10.14	0.03	10.50	10.55	0.06
			(0.13)			(0.15)
Single Parent Household	0.17	0.17	−0.00	0.17	0.17	−0.00
			(0.01)			(0.01)
Below Median Disposable Income	0.34	0.31	−0.02	0.33	0.31	−0.03
			(0.02)			(0.02)
Test for joint Orthogonality						
F-Stat			1.17			1.56
p-value			0.48			0.29
Panel B: KiVa Controls						
Bully (Wave 1)				0.15	0.15	−0.00
						(0.01)
Victim (Wave 1)				0.22	0.20	−0.01
						(0.02)
Bullying Index (Wave 1)				−0.02	−0.03	0.01
						(0.04)
Victimization Index (Wave 1)				−0.01	−0.03	−0.01
						(0.04)
Class Size (Wave 1)				19.51	19.81	0.12
						(0.51)
Students Observed in Class				17.03	17.26	0.14
						(0.51)
Test for joint Orthogonality						
F-Stat						1.30
p-value						0.54
N Observations			15,088			8,154
N Schools			73			68

This table shows summary statistics and balance for our main sample (Columns 1–3) and for the mechanism sample, consisting of students in grades 8 and 9 who participate in the KiVa baseline and endline survey (Columns 4–6). For each sample, the first and second column show the means of each variable for the control and treatment group, respectively. The third column shows the estimated treatment coefficient and standard error from a regression of each variable on treatment group assignment, controlling for strata fixed effects and clustering standard errors at the school level. Test for joint Orthogonality: F-Statistic and the p-value from a test of the joint significance of all covariates. P-value obtained via randomization inference. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

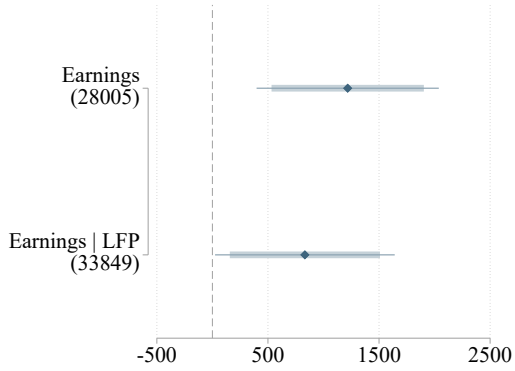
Figure 1: Academic and Labor Market Outcomes



(a) Educational Attainment



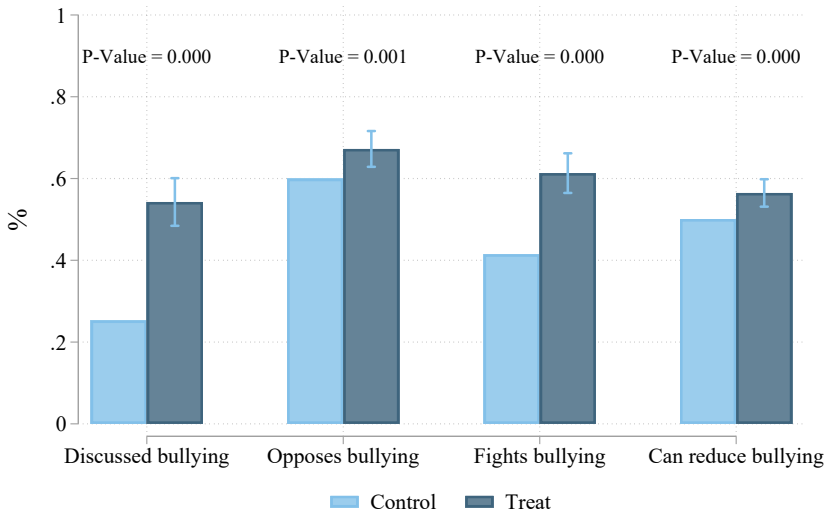
(b) Labor Market Attachment



(c) Earnings

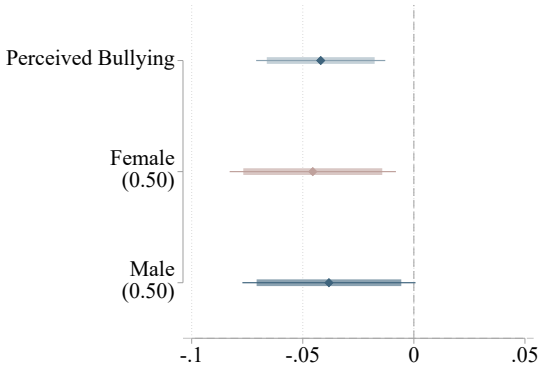
Notes: This figure shows treatment effects (Equation 1) on academic attainment in panel (a), on labor market status in panel (b), and on earnings in panel (c). All regressions include strata fixed effects and PDS lasso selected controls. The regression for university graduation in panel (a) and all regressions in panels (b) and (c) also control for school cohort fixed effects. Standard errors are clustered at the school level. Control group mean reported in brackets below variable labels. Bars indicate 90% (thick) and 95% (thin) confidence intervals.

Figure 2: Implementation by Teachers

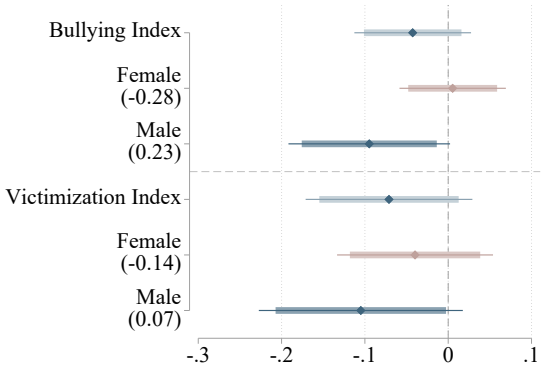


Notes: This figure shows students' perceived engagement of their classroom teacher against bullying by treatment status. Control group mean depicted in light blue, treatment mean in navy. 95% confidence interval and p-value reported from a regression of each outcome on an indicator for treatment with strata fixed effects. Standard errors are clustered at the school level.

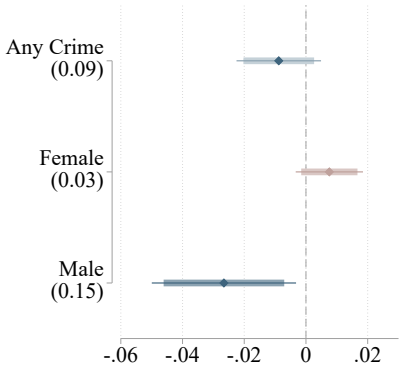
Figure 3: Harmful Behavior



(a) Perceived Change in Bullying



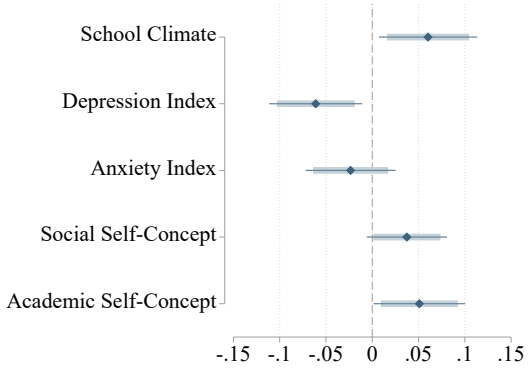
(b) Peer-reported Bullying and Victimization



(c) Crime

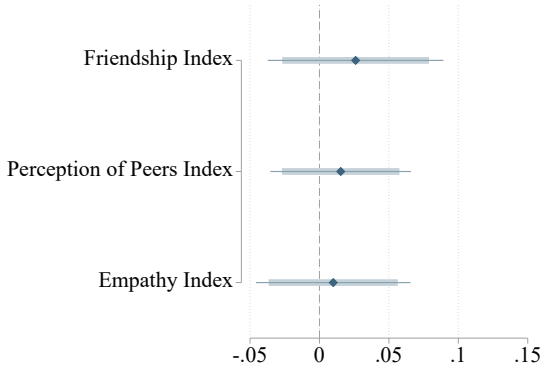
Notes: This figure shows treatment effects (Equation 1) for the full sample and separately by gender on perceived changes in bullying in panel (a), on bullying and victimization indices in panel (b), and on crime rates in panel (c). All regressions include strata fixed effects and PDS lasso selected controls. Regressions in panels (a) and (b) control for the baseline value of the outcome variable and regressions in panel (c) for school cohort fixed effects. Standard errors are clustered at the school level. Control group mean reported in brackets below variable labels. Bars indicate 90% (thick) and 95% (thin) confidence intervals.

Figure 4: Socio-Emotional Well-being



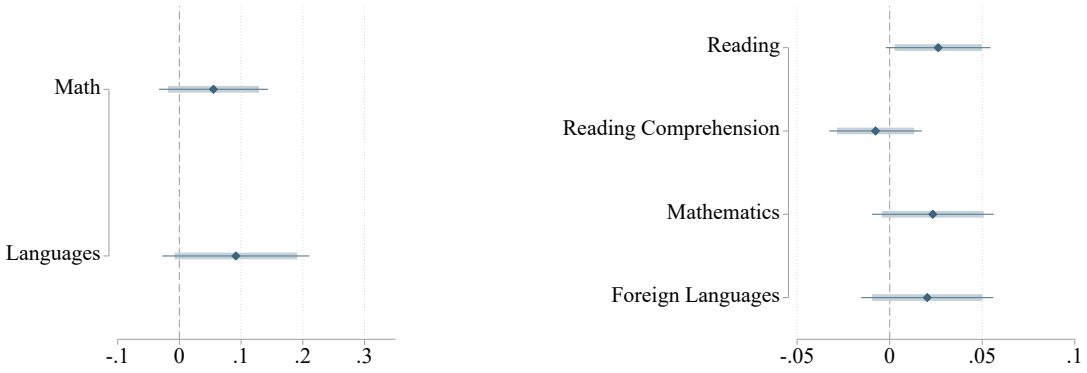
Notes: This figure shows treatment effects (Equation 1) on the indices for school climate, depression, anxiety, social- and academic self-concept. All regressions include strata fixed effects, PDS lasso selected controls, and the baseline value of the outcome variable. Standard errors are clustered at the school level. Bars indicate 90% (thick) and 95% (thin) confidence intervals.

Figure 5: Friendship and Social Connections



Notes: This figure shows treatment effects (Equation 1) on indices of friendship, general perception of peers, and empathy towards victims. All regressions include strata fixed effects, PDS lasso selected controls, and the baseline value of the outcome variable. Standard errors are clustered at the school level. Bars indicate 90% (thick) and 95% (thin) confidence intervals.

Figure 6: Academic Performance



(a) Math and Language Grades

(b) Self-reported Academic Difficulties

Notes: This figure shows average treatment effects (Equation 1) on math and language grades in grade 9, standardized to have a mean zero and standard deviation of one in the control group in panel (a), and on indicators for self-reported academic difficulties in different subjects in panel (b). All regressions include strata fixed effects and PDS lasso selected controls. Standard errors are clustered at the school level. Bars indicate 90% (thick) and 95% (thin) confidence intervals.

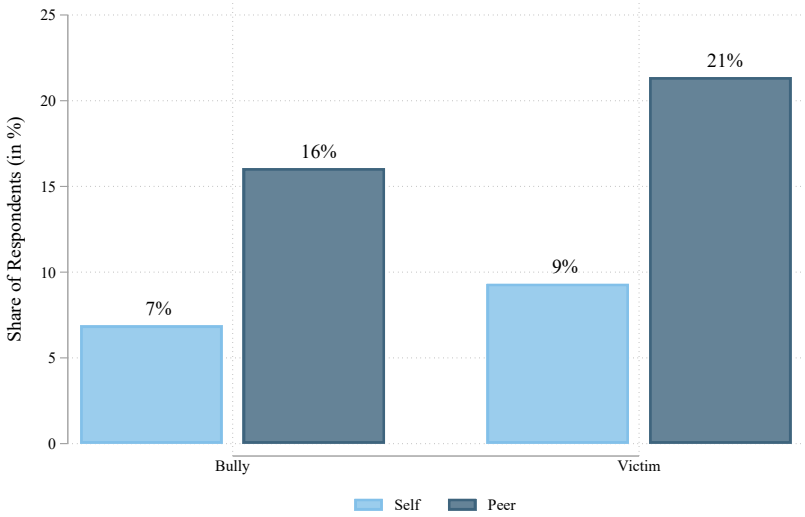
Appendix

Figure A1: Examples of KiVa Classroom Sessions



Notes: This figure shows scenes from the universal components of the KiVa program, including teacher-led classroom lessons, group discussions, and role-playing activities. These sessions aim to foster empathy, build students' confidence to intervene as bystanders, and promote anti-bullying norms.

Figure A2: Self- vs Peer-reported Measures



Notes: This figure shows the prevalence of bullies and victims by whether the definition comes from self- or peer-reported survey measures. The peer-reported indicators are dummies equal to one if the student was nominated as a victim or bully, respectively, by at least 10% of their peers. The self-reported indicators are equal to one if the student reported being bullied or victimized at least twice a month, following Kärnä et al. (2011).

Table A1: Heterogeneous Treatment Effects on Education by Gender

	Attend			
	Upper Sec. (Any)	Academic Upper Sec.	Pass Matr. Exam	University Graduation
	(1)	(2)	(3)	(4)
A. Main Estimates				
Treat	0.002	0.052**	0.041**	0.037**
	(0.007)	(0.023)	(0.020)	(0.017)
	[0.234]	[0.063]	[0.063]	[0.063]
B. Heterogeneity				
Treat × Female	0.001	0.061**	0.040*	0.037*
	(0.007)	(0.024)	(0.021)	(0.021)
	[0.295]	[0.064]	[0.082]	[0.082]
Treat × Male	0.003	0.042	0.043*	0.037**
	(0.009)	(0.026)	(0.023)	(0.018)
	[0.219]	[0.145]	[0.145]	[0.145]
Grade	7-9	7-9	7-9	7-9
Adj. R^2 (Panel A)	0.42	0.10	0.08	0.06
N (Panel A)	15,087	15,087	15,068	13,833
N (Panel B)	15,087	15,087	15,068	13,833
Control mean	0.894	0.478	0.425	0.420
Control mean (Female)	0.910	0.538	0.484	0.494
Control mean (Male)	0.877	0.411	0.358	0.335
P-value	0.81	0.34	0.83	0.98

This table shows average treatment effects on education outcomes in Panel A and heterogeneity by gender in Panel B. All regressions include strata fixed effects and PDS lasso selected controls and cluster standard errors at the school level. Column (4) also controls for school cohort fixed effects. P-value refers to the p-value from testing the null hypothesis of equal effects for males and females. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table A2: Heterogeneous Treatment Effects on Labor Outcomes by Gender

	Employed or Student	Unemployed	Others outside LF	Earnings	Earnings | LFP
	(1)	(2)	(3)	(4)	(5)
A. Main Estimates					
Treat	0.012**	−0.009**	−0.005	1,218***	833**
	(0.005)	(0.004)	(0.004)	(411)	(406)
	[0.047]	[0.047]	[0.059]	[0.021]	[0.047]
B. Heterogeneity					
Treat × Female	0.013	−0.008	−0.006	1,174**	990*
	(0.008)	(0.006)	(0.006)	(569)	(570)
	[0.225]	[0.225]	[0.225]	[0.225]	[0.225]
Treat × Male	0.011	−0.010*	−0.003	1,265**	664
	(0.007)	(0.006)	(0.005)	(591)	(541)
	[0.242]	[0.220]	[0.322]	[0.220]	[0.254]
Grade	7-9	7-9	7-9	7-9	7-9
Adj. R^2 (Panel A)	0.01	0.01	0.01	0.05	0.07
N (Panel A)	14,639	14,639	14,639	14,639	11,641
N (Panel B)	14,639	14,639	14,639	14,639	11,641
Control mean	0.867	0.067	0.066	28,005	33,849
Control mean (Female)	0.868	0.055	0.077	24,982	30,062
Control mean (Male)	0.866	0.081	0.053	31,317	38,016
P-value	0.85	0.80	0.74	0.91	0.67

This table shows average treatment effects on labor market outcomes at ages 27-29 in Panel A and heterogeneity by gender in Panel B. All regressions include strata fixed effects, PDS lasso selected controls, and school cohort fixed effects, and cluster standard errors at the school level. P-value refers to the p-value from testing the null hypothesis of equal effects for males and females. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table A3: Heterogeneous Treatment Effects on Education by Baseline Role

	Attend			
	Upper Sec. (Any)	Academic Upper Sec.	Pass Matr. Exam	University Graduation
	(1)	(2)	(3)	(4)
A. Main Estimates				
Treat	0.002	0.052**	0.041**	0.037**
	(0.007)	(0.023)	(0.020)	(0.017)
	[0.234]	[0.063]	[0.063]	[0.063]
B. Heterogeneity				
Treat × Bully	−0.005	0.064*	0.048*	−0.011
	(0.019)	(0.032)	(0.028)	(0.023)
	[0.659]	[0.211]	[0.211]	[0.659]
Treat × Victim	0.004	0.066**	0.049**	0.038
	(0.013)	(0.026)	(0.022)	(0.026)
	[0.237]	[0.055]	[0.055]	[0.104]
Treat × Bystander	0.005	0.055**	0.043*	0.049**
	(0.007)	(0.025)	(0.022)	(0.020)
	[0.153]	[0.069]	[0.069]	[0.069]
Grade	8-9	8-9	8-9	8-9
Adj. R^2 (Panel A)	0.42	0.10	0.08	0.06
N (Panel A)	15,087	15,087	15,068	13,833
N (Panel B)	10,213	10,213	10,202	9,378
Control mean	0.894	0.478	0.425	0.420
Control mean (Bully)	0.840	0.250	0.206	0.274
Control mean (Victim)	0.848	0.406	0.347	0.367
Control mean (Bystander)	0.920	0.547	0.498	0.486
P-value (Bully/ Victim)	0.71	0.96	0.97	0.09
P-value (Bully/ Bystander)	0.63	0.79	0.86	0.03
P-value (Victim/ Bystander)	0.96	0.67	0.77	0.71

This table shows average treatment effects on education outcomes in Panel A and heterogeneity by baseline role in relation to bullying in Panel B. Bully and Victim are indicators equal to 1 if a child is nominated as a bully or a victim, respectively, by at least 10% of their peers in the baseline survey. Bystander takes value 1 for those who are not bullies or victims. All regressions include strata fixed effects and PDS lasso selected controls and cluster standard errors at the school level. Column (4) also controls for school cohort fixed effects. P-values from testing the null hypothesis of equal effects for different groups are reported at the bottom of the table. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table A4: Heterogeneous Treatment Effects on Labor Outcomes by Baseline Role

	Employed or Student	Unemployed	Others outside LF	Earnings	Earnings | LFP
	(1)	(2)	(3)	(4)	(5)
A. Main Estimates					
Treat	0.012**	−0.009**	−0.005	1,218***	833**
	(0.005)	(0.004)	(0.004)	(411)	(406)
	[0.047]	[0.047]	[0.059]	[0.021]	[0.047]
B. Heterogeneity					
Treat × Bully	0.030*	−0.012	−0.019	1,019	39
	(0.015)	(0.014)	(0.012)	(1,181)	(1,116)
	[0.352]	[0.438]	[0.352]	[0.438]	[1.000]
Treat × Victim	−0.025	0.003	0.021*	−1,094	7
	(0.017)	(0.012)	(0.011)	(793)	(763)
	[0.402]	[0.659]	[0.402]	[0.402]	[0.659]
Treat × Bystander	0.017**	−0.009	−0.009	1,563***	1,154*
	(0.007)	(0.006)	(0.006)	(585)	(648)
	[0.034]	[0.086]	[0.086]	[0.034]	[0.086]
Grade	8-9	8-9	8-9	8-9	8-9
Adj. R^2 (Panel A)	0.01	0.01	0.01	0.05	0.07
N (Panel A)	14,639	14,639	14,639	14,639	11,641
N (Panel B)	9,901	9,901	9,901	9,901	7,984
Control mean	0.867	0.067	0.066	28,005	33,849
Control mean (Bully)	0.843	0.087	0.070	31,535	38,395
Control mean (Victim)	0.853	0.080	0.067	27,341	33,707
Control mean (Bystander)	0.881	0.056	0.063	28,833	34,058
P-value (Bully/ Victim)	0.04	0.51	0.04	0.14	0.98
P-value (Bully/ Bystander)	0.41	0.83	0.42	0.69	0.41
P-value (Victim/ Bystander)	0.03	0.40	0.01	0.00	0.21

This table shows average treatment effects on labor market outcomes at ages 27-29 in Panel A and heterogeneity by baseline role in relation to bullying in Panel B. Bully and Victim are indicators equal to 1 if a child is nominated as a bully or a victim, respectively, by at least 10% of their peers in the baseline survey. Bystander takes value 1 for those who are not bullies or victims. All regressions include strata fixed effects, PDS lasso selected controls, and school cohort fixed effects, and cluster standard errors at the school level. P-values from testing the null hypothesis of equal effects for different groups are reported at the bottom of the table. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.