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*

March 26, 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 attainment, labor market attachment, and criminal activity in adulthood. Treated students experience gains in human capital 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. These gains accrue to all groups of pupils, irrespective of gender or social role at baseline. We show that the likely mechanism is a reduction in bullying in the classroom, particularly among boys, which leads to a more positive learning environment for all students. A reduction in crime in adulthood among boys suggests that the program successfully mitigated harmful behavior beyond the intervention window.

JEL Codes: I20, I26, J24

*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, Tufts University and University of Zurich for valuable feedback. Elisa Alonso, Ralf Bloechinger 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. At the same time, bullying at school typically occurs within the social context of the classroom, thus raising the possibility that harmful behavior may not just deteriorate outcomes for those directly targeted but disrupt the learning environment more generally (Lazear, 2001).

In this paper, we establish a causal link between bullying and long-term educational and labor market outcomes, examining the consequences of harmful peer behavior not just on victims, but for all types of pupils in the classroom. To do so, we leverage the randomized rollout of KiVa, a school-based anti-bullying intervention launched in Finnish compulsory schools in the late 2000s. Linking rich survey data for 15,000 pupils from the original RCT to comprehensive administrative records, we track educational trajectories, 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 has higher earnings by their late twenties. We document that the likely mechanism behind these results is a reduction in bullying in the classroom, with decreases in harmful behavior concentrated primarily among boys. This translates into a better learning environment, greater academic motivation, and socio-emotional gains for all groups of students. Correspondingly, the returns of the program in terms of educational attainment and labor force attachment accrue to pupils independent of gender or social role at baseline. Finally, we document that the reduction in harmful behavior persists into adulthood: treated boys are significantly less likely to engage in criminal activity.

These findings establish that bullying that remains unaddressed has detrimental long-term consequences that extend beyond those directly affected. They also deliver important, policy-relevant insights: harmful behavior towards peers is malleable and a scalable intervention can deliver meaningful change. The reduction in adult criminal activity among boys—the group for whom the program reduced harmful behavior in the classroom—suggests that such programs may more permanently reduce toxic behavior in society.

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 random-

ization 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, aged 13-15), 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 variables from the register and survey data.

We first examine descriptive patterns of bullying and victimization. We utilize rich survey data from the RCT that elicited peer nominations for social roles based on classroom rosters, as well as self-reports. Based on self-reports on recurrent victimization, the most frequent categories of being bullied at baseline are being bullied sexually, being called nasty names, being the object of lies or gossip and physical harassment. Examining peer reports, boys are substantially more likely to be nominated as bullies by their peers, while gender differences in victimization are less pronounced. To differentiate between social roles, we classify pupils as bullies or victims if they score at or above the 75th percentile on an index across different categories in which peers could nominate for each social role. While gender is the strongest predictor of being a bully, and of being a victim, bullies tend to be from households that are slightly worse off economically, while victims are more likely to be a single child. Examining correlates of social roles and later life outcomes using the control group only, we observe—similar to prior work—a negative association between being a victim in middle school and academic achievement, educational attainment and adult earnings. The relationship between being a bully and educational outcomes is even more negative. In contrast, being a bully in middle school is correlated with higher future earnings and a higher probability of committing a crime in adulthood.

We then turn to examine the long-term causal impact of reducing bullying via the KiVa program. 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 (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. These increases in participation are accompanied by higher earnings: treated students earn about 4.3% more relative to the control group (unconditionally), and 2.5% conditional on labor market participation.

Turning to heterogeneity, we explore how these returns vary by gender and social role at baseline. We find no systematic differences in returns to KiVa by pupil gender, nor do we observe differential gains based on whether pupils 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 examine the direct impacts of the program: treated students are 3.8 ppt more likely to report a perceived reduction in bullying. When examining peer-reported social roles, we observe a decline in nominations as a bully that is primarily concentrated among boys. We then shed light on how the program achieved these reductions: teachers in treated schools are more likely to take action against bullying. For pupils, KiVa's design aimed to strengthen three key outcomes: anti-bullying attitudes, pupils' self-efficacy to intervene in bullying situations, and empathy towards victims. While we observe no meaningful treatment effects on attitudes, boys experience gains in self-efficacy (.098 SD) and empathy (.065 SD).

These direct impacts of the program, concentrated among boys—a group more likely to engage in bullying at baseline and exhibiting lower levels of empathy and self-efficacy—translate into a better learning environment and socio-emotional well-being for all groups of pupils. We show that all types of students report a better school and classroom climate and a higher academic self-concept, capturing dimensions such as the joy in and curiosity of learning. Treatment effects on academic grades that qualify for upper-secondary entry and are measured in register data are moderately sized and positive (.05 SD), but noisily estimated. Regarding well-being, treated students score higher on a self-esteem index (.04 SD) and report lower levels on the Beck et al. (1996) depression scale (-.06 SD). In contrast, the KiVa program is unlikely to deliver long-run returns via meaningfully changing the social network within the classroom: we observe no treatment impacts on how pupils perceive their peers and friendships within the class. Rather, these patterns are consistent with all pupils having a more positive and joyful learning experience that fosters a continued progression along the academic track as KiVa and classroom teachers were successful in shutting down toxic peer interactions.

As a last step, we examine whether these reductions in harmful behavior persist. Using records on criminal behavior in adulthood, we document that boys in the treatment group are significantly less likely to commit a crime. These patterns point towards KiVa being successful in reducing negative social externalities beyond the immediate time window of the intervention.

We contribute to three main strands of literature. First, while research on the consequences of bullying is abundant in disciplines spanning sociology, psychology, and education sciences, to date we lack evidence on the impacts of bullying that is causal and based on objective, long-term outcome measures. Our results clearly showcase that leaving bullying unaddressed has detrimental effects on educational attainment, labor market attachment, and earnings. 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 student- as well as parent-based approaches reduce self-reported bullying (Cunha et al., 2023; Huitsing et al., 2020; Olweus and Limber, 2010; Jiménez-Barbero et al., 2016), as do school closures during the Covid-19 pandemic (Bacher-Hicks et al., 2022; Werner and Woessmann, 2023).

Second, prior literature has examined the causal impact of being surrounded by disruptive peers at school 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; Eriksen et al., 2014; 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 does not allow for policy recommendations, 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.

Third, we highlight that fostering non-cognitive components in social interactions and curtailing negative behavior among older children delivers long-term labor market returns, as well as reductions in crime among those at elevated risk. 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 teaching skills such as patience, perspective taking, self-control, or coping strategies deliver improvements in classroom climate, violence at school, and short-term academic outcomes (Alan and Ertac, 2018; Alan et al., 2019, 2021; Rege et al., 2025), with some evidence that teaching such skills also generates long-run returns (Algan et al., 2022; Sorrenti et al., 2025). Work that employs non-cognitive approaches in reducing crime and recidivism among youths has shown that such interventions can be successful, especially among high-risk populations (Heller et al., 2017; Shem-Tov et al., 2024; Adukia et al., 2025).

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 offers descriptive statistics on bullying and victimization. Section 5 describes the experimental design, and Section 6 presents the main results. Section 7 discusses mechanisms, Section 8 presents a cost-benefit analysis of the intervention, 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 a team of 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 aim to foster three specific skills among pupils: anti-bullying attitudes, their own efficacy to stop bullying, and empathy towards victims (Kärnä et al., 2011). 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.

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. 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.

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

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

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 The Finnish Education 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.

3 Data

3.1 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 (Statistics Finland, 2020b), 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).[2] 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 (Statistics Finland, 2020a). This register includes students' final (teacher-assigned) grades in grade 9, the final year of compulsory education. We focus on students' average grade across compulsory subjects and on their performance in mathematics and language courses (including their native tongue, the other official national language, and English), 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.

3.2 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 as grade 7 students were still in primary 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 their teachers, who were supplied with detailed instructions. The teachers were told to act in such a way that the confidentiality of responses 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. To study mechanisms, we use data on three main domains: harmful behavior, skills and concepts directly targeted by the KiVa program, and measures of the learning environment and socio-emotional well-being. These domains are described below, and Table F1 provides further details on the

[2] 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.

questions capturing each dimension. To reduce the number of statistical tests, we create separate variance-weighted indices for each outcome family and wave, following Anderson (2008).

Harmful behavior — The survey collected measures of pupils' perceived changes in bullying in the classroom, and peer-reported bullying and victimization measures. For the peer-reported measures, 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, "Is being pushed around and hit", "Is called names and mocked", "Nasty rumors are spread about him/her". For each question, we observe the share of peers that nominate a child in a role among all responding classmates. We construct a variance-weighted index (Anderson, 2008) for bullying and for victimization by aggregating across the respective shares of nominating peers. We further construct indicators for being a (peer-reported) bully or victim if a pupil scores at or above the 75th percentile of the index of the control group at baseline.[3]

Concepts and skills directly targeted by the program — To measure the implementation of the program, we rely on four questions that ask students about their perception of the teachers' role around bullying. We further examine the three skills directly targeted by KiVa among pupils: disapproval of bullying, empathy, and efficacy. We construct a variance-weighted index across the question batteries used to elicit each of these concepts. Disapproval of bullying captures students' anti-bullying attitudes and support for defending victims, based on statements such as "it is a wrong thing to join in bullying". Empathy captures students' ability to understand and share the feelings of bullied peers, as reflected in statements such as "when the bullied pupil is sad, I also feel sad". Efficacy measures students' perceived ability to act in bullying situations and their expectations about the consequences of intervening, for example whether "trying to get the others to stop bullying" would be easy for them.

Learning environment and socio-emotional well-being — We use information on two dimensions of the learning environment and construct an index for each: school climate and academic self-concept. School climate captures students' perceptions of the overall atmosphere in their school and class, based on statements such as "there is a good atmosphere in my class" and "I feel safe at school." Academic self-concept elicits students' attitudes toward learning and their perceived performance at school, based on statements such as "learning brings me joy" and "In my opinion, I am doing fine at school."

To get at emotional well-being, we construct separate indices for depression, anxiety, and self-esteem. The depression battery is derived from the Beck Depression Inventory, a clinically validated instrument for assessing depressive symptoms (Beck et al., 1996). Social anxiety is captured using statements such as "I'm worried about what the others think of me" or "I feel quite shy even among those mates I know well." Self-esteem is measured via the Rosenberg Self-Esteem Scale, adapted to elicit how children feel about themselves among peers (Rosenberg, 1965; Salmivalli et al., 2005).

[3]The survey also collected self-reported measures of bullying and victimization following Olweus' Bully/Victim Questionnaire (Olweus, 1996). Figure B3 shows that self- and peer-reported measures are correlated. We discuss the comparison of these two measures in Section 4.

Finally, we use information on two dimensions of pupils' social network and construct an index for each: friendships and pupils' general perceptions of their peers. The Friendship index measures students' relations with their classmates, based on statements such as "I have good friends in my classroom." The question battery on the perception of peers asks students to report their view of their peers more broadly, with statements such as "[my] peers can really be relied on."

4 Descriptive Patterns on Bullying

What types of bullying are most prevalent, and which students are most affected? Figure B1, Panel (a) shows the share of pupils by gender who report having been victimized at least twice per month during the preceding two months, a relatively frequent and persistent notion of victimization (Kärnä et al., 2011). Around 10% of pupils were victimized overall, with boys reporting being victimized somewhat more (11% vs 8% of girls). Boys are more likely to be victimized across almost all the different victimization categories: they are more likely to be bullied sexually (boys: 10%, girls: 5%), called nasty names (boys: 10%, girls: 7%), and substantially more likely to be harassed physically (boys: 4%, girls: 1%). Girls are somewhat more likely to experience social exclusion (boys: 4%, girls: 5%), and report similar rates of being the object of lies or gossip (boys: 5%, girls: 4%).

Next to self-reports, the baseline survey also elicited peer-reported social roles. In particular, children received a roster of their classmates in which they could indicate which students engaged in bullying or were victimized based on three more general categories for each role (see Section 3.2). Figure B2 shows the distribution of the average share of nominating peers across all three categories by gender. We observe marked gender differences in nominations as a bully: while more than 60% of girls do not receive any nominations, more than 60% of boys receive at least one. Boys are also much more likely to be nominated as a bully by a high share of their classroom peers. As displayed in Panel (a), the distributions of peer nominations as a victim are more similar by gender, with about 70% of each gender receiving at least one nomination and girls only being somewhat less likely to be nominated as a victim by a high share of their peers. Relative to the self-reported variables, which capture high-frequency bullying and victimization, more children receive peer nominations, which are elicited without mentioning a specific time frame or frequency in the survey. In Appendix Figure B3 we examine the correlation between peer nominations and these high-frequency self-reports. For victims, the probability of self-identifying as being bullied frequently is close to zero when receiving no peer reports and increases markedly with the share of peer-reports. For bullies, the relationship between self- and peer-reported bullying behavior is relatively more attenuated, suggesting that bullies might be less willing to self-report frequent engagement in harmful behavior towards their peers. Throughout the paper, we will therefore classify students into social roles relying primarily on peer-reports.

What traits are associated with being nominated as a bully or a victim? To more easily examine such patterns, we construct a variance-weighted index across all three categories of peer nominations for bullies and victims (Anderson, 2008). We define an indicator variable for

each social role equal to one if a pupil receives peer nominations that are at or above the 75th percentile of the baseline distribution of this index in the control group. By definition, 25% of students in the control are thus categorized as a bully and a victim, respectively. Bullies and victims have a median share of nominating peers of 12.5% for bullies and of 15.4% for victims.[4]

Figure B5, Panel (a) shows correlates of social role by regressing each standardized baseline characteristic on these indicators for being a bully or a victim. Gender sticks out as the most striking correlate, with those being nominated as bullies being substantially less likely to be female (-.74 SD) and victims being somewhat less likely to be female (-.18 SD). Both victims and bullies are relatively older compared to their peers, while victims are more likely to be a single child (.13 SD), and bullies are somewhat more likely to come from a single parent household (.07 SD). For the control group, we can also examine how being nominated as a bully or a victim correlates with later life outcomes. Figure B5, Panel (b) shows that bullies on average have worse grades at the end of middle school (-.66 SD) and are subsequently less likely to pursue an academic track by attending academic high school, taking the university entrance exam, or graduating from university (-.43 – -.33 SD) . These patterns similarly present for victims, but the magnitudes are much smaller. Interestingly, bullies but not victims obtain higher earnings in the labor market. We also observe a relatively strong association between committing a crime as an adult and being classified as a bully in middle school (+.54 SD), suggesting that harmful behavior may extend beyond the particular social context and time period of growing up.

How fluid vs. sticky are these social roles over the course of the school year? We examine persistence of social roles in the control group between baseline and endline. Appendix Figure B6 shows students at baseline by their social roles and examines what share in each group was nominated for a particular endline role. For bystanders, i.e. students who are neither categorized as a victim nor a bully, more than 80% are still in that role at endline. For bullies and victims, roles are more fluid. About 50% of bullies and about 40% of victims are still in their same respective social role at the end of the school year. About 40% in each group switch to be a bystander, with the rest switching from being a bully to a victim and the other way around. These patterns suggest that the roles of bullies and victims are somewhat more fluid over the school year, relative to pupils who are not involved in either category.

5 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.[5]

[4]Figure B4 shows social roles by gender based on this dichotomized definition of social role. About 40% of boys and 10% of girls are classified as a bully, and 30% of boys and 20% of girls are classified as victims. 8.6% of pupils are both a bully and a victim. In later analyses we also allow for a continuous concept of social roles, which renders similar conclusions to using a dichotomized version.

[5]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 roll-out. However, since in the scaled-up program the KiVa lessons are only provided in grades 1, 4, and 7, these students were never

For 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 consists of 38 intervention and 35 control schools.[6]

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 for student i, and Treat_s is a binary variable equal to one if a student's school s was randomized into the treatment group. θ_s are school-level strata fixed effects 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.[7] Standard errors are clustered at the school level.

When studying outcomes that we observe at different ages for the different cohorts of pupils, such as earnings, we include (school) cohort fixed effects based on students' grade during KiVa. Where available (i.e., when studying mechanisms with survey data), we also control for the baseline measure of the outcome variable.

To account for multiple hypothesis testing, we report sharpened q-values (Anderson, 2008) across families of outcomes. To analyze the RCT's survey data on potential mechanisms, we construct variance-weighted indices across survey batteries for each group of outcomes (Anderson, 2008).

5.1 Sample and Balance

For our main analysis, our sample consists of all students who can be linked to administrative register data from Statistics Finland. We can match 90% of students that participated in KiVa to administrative records, and the match rate is similar for the treatment and the control group (Column 1 in Table 2). Our final sample consists of 15,088 students. The administrative records are largely complete.[8]

Columns 2–4 in Table 1 show that the control and treatment groups are well balanced across baseline characteristics measured in register data and we fail to reject the null in a test of joint orthogonality. The share of students with an immigrant background is quite low (2%). The majority of students live in the same region in which they were born (89%) and in areas that are predominantly urban (75%). While very few students are young for their grade, about 5% are

exposed to these more intensive elements of the intervention.

[6] Originally, there were 39 schools in each treatment arm, but four control schools dropped out without providing any information, and one treatment school participated only in the baseline wave of data collection. Table 1 shows that students included in the final sample are similar in terms of observable characteristics.

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

[8] 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.

old relative to their peers.[9] 14% of students are a single child, and 17% live in a single-parent household. About a third of students lives in families with below median disposable income.

To study mechanisms, we use both administrative registers and survey data collected during the KiVa RCT. While we can observe outcomes in the administrative data for all pupils with a register link, not all pupils participated in all survey waves. In particular, the baseline survey was only fielded for students who were in grade 8 and 9 during the year in which KiVa was implemented.[10] As self-reported outcomes regarding personal well-being might arguably be more subjective and thus prone to noise, we restrict the sample to students from grades 8 and 9 in order to control for the baseline value of outcomes in the survey data. This results in a "Survey Sample" of 8,861 students.[11]

We show in columns 5–7 of Table 1 that the Survey Sample is balanced on baseline characteristics and we fail to reject the null of joint orthogonality. However, Table 2 shows that pupils did not participate in the endline survey at similar rates across treatment arms: The treatment group is about 9 ppt more likely to fill out the endline survey, and treated students who live in urban areas (and are thus from families with higher disposable income) are somewhat more likely to participate relative to their counterparts in the control group. We address concerns about selective attrition for survey-based variables in the mechanism section by showing that complementary results based on administrative data are not sensitive to this sample restriction, and by presenting additional estimates for survey-based variables with inverse probability weighting.

6 Long-Term Impacts of KiVa

6.1 Educational Attainment

We examine impacts on students' educational attainment by tracing pupils' trajectories through the key educational stages of the Finnish school system. After finishing middle school, schooling in Finland at the time of our study is no longer compulsory. However, almost all Finns continue in post-compulsory education by applying to enter either an academic or a vocational track at age 16 in the nationally organized allocation of study slots.

We report treatment impacts on educational attainment in Figure 1, Panel (a). At 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 the type of path that students are able to enter: among treated students, the share enrolling in the academic high school track (instead of vocational training) increases by 5.1 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, after three years of upper-secondary education, most students in the academic

[9] In Finland, children start school in they year which they turn 7 years old. We define pupils as young or old for their grade if they deviate from the typical age of their cohort.

[10] The baseline survey was fielded at the end of the academic year prior to the KiVa implementation, i.e., while grade 7 students were still in primary school.

[11] We impute baseline values and add a missing flag for pupils in this sample who have missing baseline data.

track take the matriculation or university-entrance exam. We see that treated students are 3.8 percentage points more likely to pass this exam, which also serves as a certificate of graduation from academic high school. This effect size corresponds to an 8.9% increase over the control group mean. Finally, in the last year we observe our cohorts in the administrative records, the treatment group is 3.9 percentage points or 9.3% 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.

6.2 Labor Market Outcomes

Do these gains in human capital attainment translate into returns in the labor market? To examine labor market outcomes, we use the last available year of data for each cohort when students are ages 27–29 and add grade fixed effects to our main specification.

In order to study attachment to the labor force, we classify young adults into three mutually exclusive categories: being "active", defined as being either in employment or a registered student, being unemployed, or being "inactive", i.e. out of the labor force without pursuing further education or training. Panel (b) in Figure 1 reports the impacts of the KiVa program. Treated students are slightly more likely to be active in the labor market (1.2 percentage points or 1.4% over the control mean) and 0.9 percentage points (13%) less likely to be unemployed.

Panel (c) shows effects on earnings. KiVa participation raises students' annual earnings by around 1,200 EUR, 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 EUR or 2.5% more relative to 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.

6.3 Heterogeneous Treatment Effects

Does KiVa affect children differentially depending on their gender or social role at baseline? As boys are more likely to engage in bullying behavior (see Section 4), they may also have been more likely to experience social disapproval over the course of the program, or they may have benefited more from the behavioral changes it induced. Appendix C reports tables displaying heterogeneity for all main outcomes. To estimate treatment effect heterogeneity, we interact the treatment indicator in Equation 1 with group status and add group fixed effects. Average treatment effects are displayed in Panel A, with Panel B reporting heterogeneity by gender. Impacts both on education and labor market outcomes are similar in magnitude for males and females, and we cannot reject the null that they are the same for any outcome.

Panel C estimates treatment effects by social role at baseline. We restrict the sample to pupils who were in grades 8 and 9 during the KiVa implementation, as grade 7 pupils did not participate in the baseline survey. We define a student as a victim or a bully if they received a share of peer nominations at or above the 75th percentile of the baseline distribution of an index for each role in the control group (see Section 4). We do not detect systematic or meaningful

heterogeneity by social role at baseline. Coefficients are somewhat more precisely estimated for the relatively larger bystander group, but we cannot reject the null of differential treatment effects for victims or bullies for the majority of outcomes.[12] These conclusions remain robust to using continuous measures to define social roles or to separately estimating impacts on pupils who are classified as both victims and bullies (Appendix Table E1). As documented in Section 4, social roles are fluid with students cycling in and out of different roles over the course of the school year. It is therefore possible that the signal for children whose role would remain "sticky" may be harder to pick up due to the heterogeneous composition of these groups. However, these average group effects are consistent with all groups of pupils benefiting precisely because any student may find themselves in a different role within the social structure of the classroom. We will further explore this notion in the next section, which documents the program's mechanisms by examining heterogeneous effects in reductions in bullying behavior.

7 Mechanisms

The results presented in Section 6 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 these trajectories, in addition to administrative records, we exploit the rich survey data collected during the RCT. In order to address concerns with selective attrition when using the Survey Sample based on grades 8–9 (see details in Section 5.1), we report additional specifications with inverse probability weighting for survey-based outcomes and show that results based on data from administrative registers are not sensitive to sample restrictions.[13]

7.1 Changes in Bullying

We start by examining if the program was successful in achieving its primary objective: reducing bullying. To this end, we use students' perceptions about the overall change in bullying and peer nominations for social roles. Panel (a) in Figure 2 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.

Panel (b) in Figure 2 documents treatment impacts on the indices for peer-reported bullying and victimization.[14] While, on average, point estimates are negative but noisy, we observe a significant decrease both in peer-reported bullying and victimization for boys. The reduction in

[12] The two outcomes for which we observe statistically significant differences are for being out of the labor force for reasons other than education as the main labor market activity, and for earnings unconditional on labor force participation. For both of these, victims are worse off relative to bystanders. Based on this gap closing once we condition on employment, and based on Appendix Table C3 showing university enrollment is higher among victims in their late twenties, we think that these differences are likely explained by longer engagement with education among victims in the treatment group.

[13] Table E2 further shows similar treatment effects on our main outcomes for the Survey Sample.

[14] We observe peer nominations for all students on the classroom roster, not just for those who participated in the survey, resulting in a slightly larger sample size.

bullying for boys amounts to 8.9 percent of a standard deviation, which accounts for closing about 20% of the baseline gender gap in bullying.[15] Appendix Tables E4 and E5 report robustness for these estimates when including self-reports and shows the dichotomized version of this variable: boys are 6.1 ppt less likely to be categorized as a bully, a 18% decrease over the control mean. This is accompanied by a decrease in victimization among boys of 8.8% of a standard deviation on the index or 4.4 ppt (18% over the mean) on the dichotomized variable. For girls, while point estimate across victimization and bullying measures are negative, they are not significantly different from zero. These results suggest that KiVa was effective in addressing problematic behavior of the group that was more frequently identified as harassing other students at baseline.

7.2 Direct Impacts of the Program

Next, we provide evidence on the mechanisms through which KiVa achieved these reductions in harmful behavior.

Implementation of the program by classroom teachers — We first explore the role of teachers using pupils' perceptions as reported in the RCT's endline survey. In Figure D1, we see that treated pupils are substantially more likely to report that their classroom teacher has become active regarding bullying: we observe a 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. Treated students further perceive their teachers as somewhat more capable to do something against bullying, with a 10% increase in perceiving that their teacher has at least some influence to reduce bullying and a 11% 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.

Skills taught by KiVa — As discussed in Section 2, KiVa's approach is based on the idea that changing bystanders' attitudes is an important component in reducing bullying. In particular, the program targets three specific levers (see Kärnä et al., 2011): i) fostering students' anti-bullying attitudes, ii) promoting empathy towards victims, and iii) enhancing their beliefs about their own effectiveness to intervene against bullying (self-efficacy).

Figure 3 shows treatment impacts on each of these three targeted dimensions. The top panel of the figure shows effects on an index of anti-bullying attitudes. Reported disapproval of bullying is already high at baseline, with e.g. 77% of students saying that it is not OK to call kids nasty names or nearly 70% saying that it is wrong to participate in bullying. We find small and insignificant treatment effects on this index, both on average and when studying separate impacts for male and female students. The second panel shows treatment impacts on an index of empathy towards the victim. Treated boys score .065 of a SD higher on this index, but we do not see an increase in empathy for female students (who have higher levels of empathy to begin with). Finally, the bottom panel presents effects on an index measuring students' perceived efficacy in fighting bullying. We again see a significant increase in perceived efficacy among boys

[15] The gender gap in the bullying index in the control group is .43 SD; see Appendix Table E4.

(by .098 of a SD), who are more likely to report that their individual actions would be effective in fighting bullying, but no impact on female students.

These results suggest that, while the program did not meaningfully shift attitudes towards bullying—already predominantly negative at baseline—it was effective in fostering boys' empathy towards victims and in making them more aware of their own ability to fight bullying. Given male students' higher propensity to be involved in bullying situations, these changes might have been fundamental in shaping group dynamics and achieving reductions in bullying. The changes in these skills and beliefs could also partly account for the beneficial effects of the program on boys' long-term outcomes. However, as discussed in Section 6.3, education and labor market gains are visible also for females and for students who did not participate in bullying at baseline. In the next section we thus investigate how the reductions in harmful behavior translate into academic gains for the average student in the classroom.

7.3 Indirect Impacts of the Program

Learning environment and academic achievement — We start by studying the impact of the program on students' learning environment. 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, an effect mostly driven by students' increased enjoyment of school. Consistent with students' improved perception of the school environment, we observe positive treatment impacts of 5 percent of a standard deviation on an index of academic self-concept that elicits students' joy and curiosity for learning. We document in Appendix Table D4 that these gains materialize for all groups of pupils, irrespective of their gender or social role.

Do these impacts translate into higher academic achievement? The bottom bars of Figure 4 shows treatment effects on teacher-assigned, standardized grades in grade 9 (the last year of compulsory school), which are the main criterion by which students' applications to post-compulsory education institutions are evaluated. We report the average grade across compulsory subjects and for languages and mathematics as core subjects. All point estimates for treated students are positive and of moderate size (average grade: .05 SD, math: .05 SD, languages: .08 SD); however, none of the coefficients are statistically different from zero. If we were to take those estimates at face value, could higher grades explain why treated students are more likely to enter the academic track of upper secondary education? In Appendix Table D5, we estimate the association between grades and the probability of attending the academic track in the control group. A 1 SD increase in the average grade increases the likelihood of attending the academic track by 31.5 ppt. An increase in average grades of .05 of an SD for our treated students could thus account for about a third of the observed treatment impact on attending the academic track.

Overall, these results suggest that the intervention created a more positive classroom and school environment in which students are able to pursue their academic curiosity and experience learning as a joyful activity. Taken together with moderately sized (but noisily estimated)

increases in academic grades, these changes offer a plausible explanation for why treated pupils embark and continue on a more academically-oriented track.

Socio-emotional wellbeing — Improvements in the learning environment are directly tied to improvements in students' own well-being. We present these results in Figure D2. While we do not observe statistically significant reductions in a (social) anxiety index, we see a significant decrease of .06 SD in depressive symptoms, as measured by Beck et al. (1996)'s depression scale. This increase in students' well-being is also visible in their self-esteem in relation to others: the last bar of Figure D2 shows that treated students score higher on an index measuring their self-concept among their peers (Rosenberg, 1965).

Social network — While treated students perceive a better atmosphere in their class, they do not seem to have more friends or have an improved perception of their peers. Appendix Figure D3 shows no treatment impacts on an index that combines three questions on having good friends and relationships with classmates, and on an index aggregating students' perceptions of the qualities of their peers.

7.4 Discussion and Long-Run Effects on Crime

These different pieces of evidence suggest that KiVa was effective in reducing harmful behavior in the classroom, primarily among boys. The benefits from reducing negative peer interactions extend to everyone in class: we see an improved perception of the classroom environment, more curiosity and joy in learning, and increases in self-reported well-being while in school. Appendix Tables D4 and D6 show that these gains are visible for all groups of students, irrespective of gender or social role at baseline.[16] 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.

How persistent are the behavioral changes induced by the program? While we document education and labor market benefits of KiVa up to thirteen years after the program ended, these in themselves do not tell us whether the reductions in harmful behavior were limited to the period of the intervention or sustained over time. To answer this question, we turn to administrative data from crime registers and study treatment impacts on the probability of committing a crime as an adult up to the last available year in our data, when individuals in our sample are aged 28–30. The results are presented in Figure 5 and mimic our findings on bullying behavior in the short term. We do not find significant changes in crime rates on average or for females. However, males in the treatment arm are 2.7 percentage points less likely to commit a crime. This represents a 17.5% decrease compared to the average crime rate of males in the control group (15%). These results suggest that KiVa was successful in achieving persistent reductions

[16] Appendix Tables D2-D6 show that these results are robust to inverse-probability weighting, suggesting that these findings are not driven by selective responding in the endline survey. This conclusion is further supported by Appendix Table E3, which compares the estimated treatment impacts on administrative outcomes—average grades and crime—in the full sample of KiVa participants linked to administrative data from grades 8 and 9 and when restricting to those who also filled in the endline survey. For both outcomes, we cannot reject the null hypothesis of equal effects across the two samples.

in harmful behavior, thus decreasing negative social externalities beyond the time window of the intervention.

8 Cost-Benefit Analysis

We compare the costs and benefits of the program using the marginal value of public funds (MVPF) framework from Hendren and Sprung-Keyser (2020). The MVPF is defined as the ratio of the aggregate willingness to pay (WTP) of the policy's beneficiaries to the net cost of the policy to the government:

$$\text{MVPF} = \frac{\text{WTP}}{G} \tag{2}$$

where G denotes the net present value of all fiscal costs net of fiscal offsets. We compute WTP and G using estimates from our randomized design and convert them to real terms in 2008 EUR. See Online Appendix G for details on the calculations.

8.1 Government Costs

Figure G1, Panel (a) documents the net costs of KiVa for the government. The initial program costs EUR 97.22 per student, which includes the direct fees for materials, as well as the costs of teachers' training (direct costs and opportunity costs) and time in class delivering KiVa lessons.

Because the KiVa intervention increases the probability that treated students will attend university, we include the additional public expenditure on higher education as a government cost. We use the average public expenditure per tertiary student in Finland (EUR 77,912, see OECD, 2024) and apply our estimated increase in the probability of university attendance (3.9 ppt, see Appendix Table C1). This gives an expected additional government expenditure of EUR 3,039 per treated student.

The government recoups its upfront costs through increased income tax revenue generated by the intervention's positive effect on earnings. Beyond age 29, we take a conservative approach and assume that the earnings gap between the treatment and control group remains constant through retirement age. Assuming an average tax rate of 30% (OECD, 2025), the present value of additional tax revenue is EUR 4,694 per student.

Combining these different components leads to a negative G: net costs to the government equal EUR $-1{,}558$, which implies that the program has an infinite MVPF as long as $WTP > 0$.

8.2 Willingness to Pay

Figure G1 Panel (b) shows our estimates of WTP for KiVa participants. We restrict the WTP calculation to the effect of the intervention on post-tax labor income, following the envelope-theorem argument in Hendren and Sprung-Keyser (2020): under the assumption that income gains arise from a genuine increase in human capital, rather than from costly additional effort, the present value of after-tax income changes provides a valid first-order estimate of willingness to pay.

For ages 16–29, we use empirical estimates of Δy_t drawn directly from our estimates of the program's effect on earnings at each age. Between ages 16 and 23, WTP is only EUR 156, as university attendance delays labor market entry. From ages 24 to 29, earnings gains for the treatment group rise to EUR 1,549. Projected earnings gains from ages 30 to 65 accumulate to EUR 9,247. In total, this yields an estimate of WTP equal to EUR 10,953 per treated student.

This positive WTP estimate implies an infinite MVPF of the program, given the negative net costs of the intervention for the government. In other words, the program more than pays for itself, and implementing KiVa generates a Pareto improvement.

9 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 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. These results suggest that bullying in the classroom imposes costs on all students, irrespective of their social role.

While the landscape of peer interaction has changed considerably since our intervention, the behavioral tendencies underlying bullying appear to be persistent (Camacho et al., 2023). At the time of the intervention, online bullying was already present in Finnish schools: roughly 7% of students reported cyber-victimization at least once per month. Yet the salience and reach of peer harassment have likely increased with the expansion of smartphones and social media. Our analysis provides proof of concept that reducing harmful behavior in adolescence can generate long-run benefits. This logic may be even more relevant today, when bullying can extend beyond the classroom and into online spaces.

A cost-benefit analysis of the intervention following the MVPF framework suggests that the KiVa program pays for itself from the government point of view: even when abstracting from potential reductions in fiscal costs from decreases in crime, the direct and indirect costs of the program are more than offset by the additional income tax revenue generated by the intervention. The estimated MVPF for KiVa is thus infinite, similar to that found for some early education programs such as Head Start or for expansions of Medicaid coverage to children (Hendren and Sprung-Keyser, 2020). This estimate is based on partial equilibrium estimates; in general equilibrium, displacement in university slots could attenuate the labor market returns of the intervention. However, our finding of reductions in criminal behavior among male students is not subject to the same concerns: reducing harmful behavior by one student does not come

at the expense of another, and likely generates positive externalities of its own.

Overall, our findings show that programs aimed at stopping negative peer dynamics, even among older children, can improve socio-emotional well-being and yield meaningful labor market returns.

References

Adukia, A., B. Feigenberg, and F. Momeni (2025). From retributive to restorative: An alternative approach to justice in schools. *American Economic Review 115*(8), 2722–2754.

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.

Bacher-Hicks, A., J. Goodman, J. G. Green, and M. K. Holt (2022, September). The covid-19 pandemic disrupted both school bullying and cyberbullying. *American Economic Review: Insights 4*(3), 353–70.

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.

Bowes, L., M. Babu, J. R. Badger, M. R. Broome, R. Cannings-John, S. Clarkson, E. Coulman, R. T. Edwards, T. Ford, R. P. Hastings, et al. (2024). The effects and costs of an anti-bullying program (KiVa) in UK primary schools: a multicenter cluster randomized controlled trial. *Psychological Medicine 54*(15), 4362–4373.

Camacho, A., K. Runions, R. Ortega-Ruiz, and E. M. Romera (2023). Bullying and Cyberbullying Perpetration and Victimization: Prospective Within-Person Associations. *Journal of Youth and Adolescence 52*(2), 406–418.

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.

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.

Heller, S. B., A. K. Shah, J. Guryan, J. Ludwig, S. Mullainathan, and H. A. Pollack (2017). Thinking, fast and slow? Some field experiments to reduce crime and dropout in Chicago. *The Quarterly Journal of Economics 132*(1), 1–54.

Hendren, N. and B. Sprung-Keyser (2020). A unified welfare analysis of government policies. *The Quarterly Journal of Economics 135*(3), 1209–1318.

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.

OECD (2010). *Education at a Glance 2010: OECD Indicators*. Paris: OECD Publishing. Table D3.1: Teachers' salaries (2008). Retrieved: March 2026.

OECD (2024). Expenditure per student by educational level [dataset: DSD_EAG_UOE_FIN@DF_UOE_INDIC_FIN_PERSTUD, v3.1]. In *Education at a Glance* (ISSN 1999-1487). OECD Data Explorer. Retrieved March 2026.

OECD (2025). *Taxing Wages 2025: Decomposition of Personal Income Taxes and the Role of Tax Reliefs*. Paris: OECD Publishing.

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.

Persson, M., L. Wennberg, L. Beckman, C. Salmivalli, and M. Svensson (2018, aug). The Cost-Effectiveness of the Kiva Antibullying Program: Results from a Decision-Analytic Model. *Prevention Science 19*(6), 728–737.

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). "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.

Shem-Tov, Y., S. Raphael, and A. Skog (2024). Can restorative justice conferencing reduce recidivism? Evidence from the make-it-right program. *Econometrica 92*(1), 61–78.

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.

Statistics Finland (2020a). EDUC TYHR [Toisen asteen yhteishaku - moduuli]. Accessed 12-Dec-2025.

Statistics Finland (2020b). FOLK basic [Perustieto]. Accessed 12-Dec-2025.

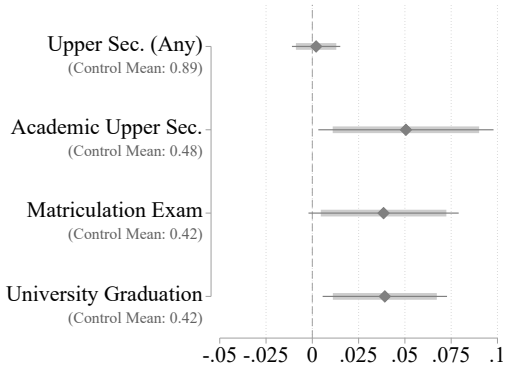
UNESCO (2018). School violence and bullying: Global status and trends, drivers and consequences. Technical report, Paris.

Werner, K. and L. Woessmann (2023). The legacy of covid-19 in education. *Economic Policy 38*(115), 609–668.

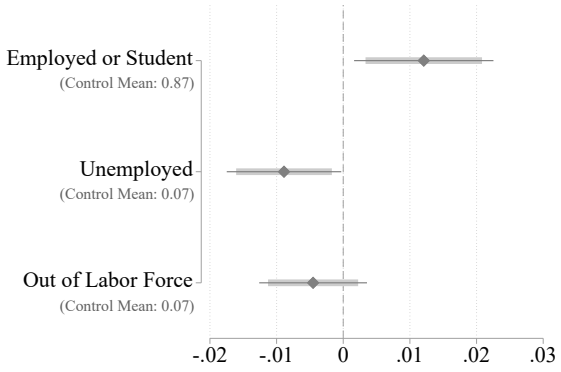
Wolke, D. and S. T. Lereya (2015). Long-term effects of bullying. *Archives of disease in childhood 100*(9), 879–885.

Tables and Figures

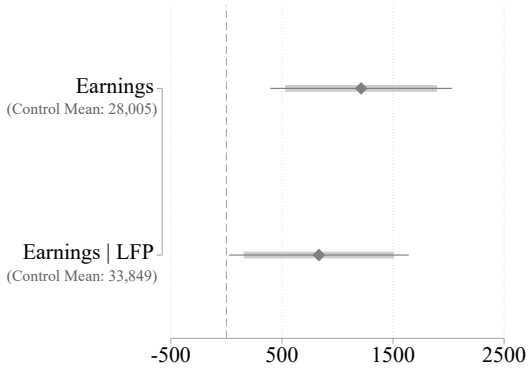
Figure 1: Long-Term Academic and Labor Market Outcomes



(a) Educational Attainment



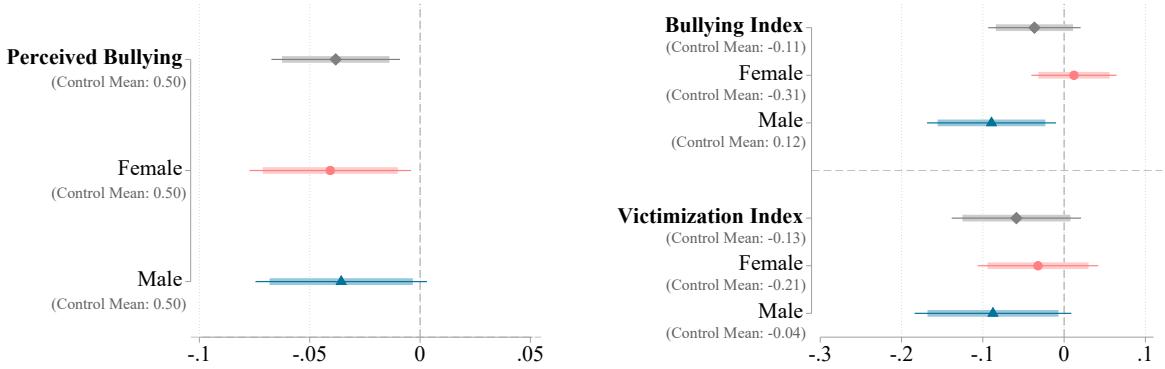
(b) Labor Market Attachment



(c) Earnings in Euro

Notes: This figure shows treatment effects (Equation 1) on educational attainment in panel (a), on labor market status in panel (b), and on earnings in panel (c). Outcomes Panel (a) from top to bottom, Indicators for: Enrollment in upper secondary school at age 16 (Upper Sec. (Any)), Enrollment in academic track of upper secondary school at age 16 (Academic Upper Sec.), Passing the matriculation exam by age 19 (Matriculation Exam), Graduating from university by 2022 (University Graduation). Outcomes Panel (b) from top to bottom, measured in 2022. Indicators for: Being employed or student (Employed or Student), Being unemployed (Unemployed), Being outside the labor force and not in education (Out of Labor Force). Outcomes Panel (c) from top to bottom, measured in 2022: Annual earnings in EUR (Earnings), Annual earnings conditional on being employed in EUR (Earnings | LFP). All regressions include strata fixed effects and controls selected with PDS lasso. The regression for university graduation in panel (a) and all regressions in panels (b) and (c) also control for grade 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: Harmful Behavior at School

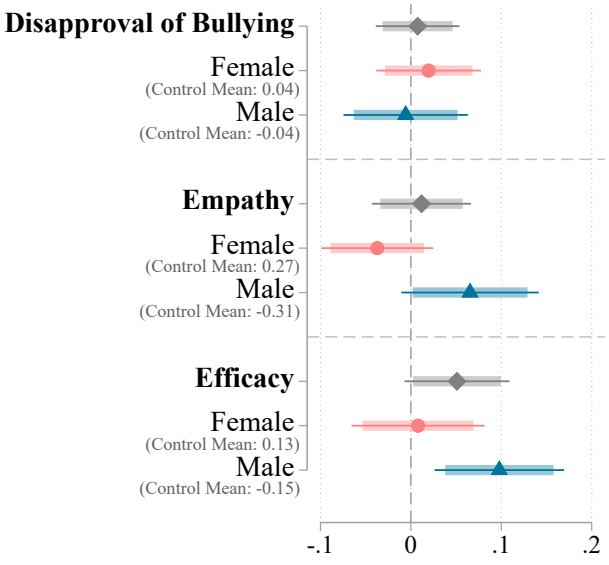


(a) Perceived Change in Bullying

(b) Peer-reported Bullying and Victimization

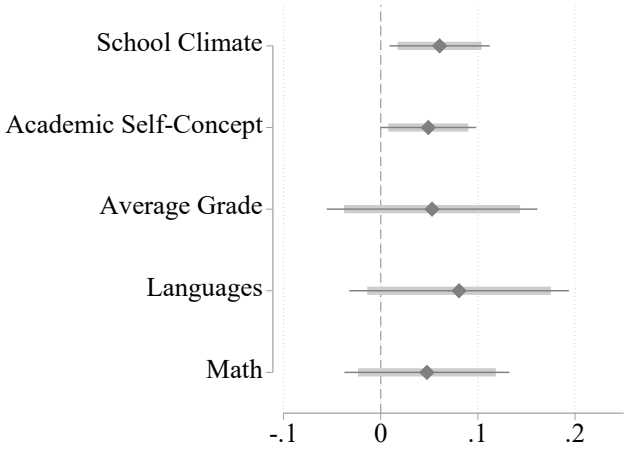
Notes: This figure shows treatment effects (Equation 1) for the full sample and by pupils' gender on an indicator for perceived bullying in panel (a) and indices for peer-reported bullying and victimization in panel (b). Outcome Panel a: Indicator equal to one if a student indicates that bullying stayed constant or increased over the past academic year (Perceived Bullying). Outcome Panel b: Variance-weighted index of the share of nominating peers for each role (bullym, victim) across three categories of bullying and victimization (Bullying and Victimization Index, see Section 3.2 for details). All regressions include strata fixed effects and the baseline value of the outcome variable, additional controls selected with PDS lasso. Standard errors are clustered at the school level. Bars indicate 90% (thick) and 95% (thin) confidence intervals.

Figure 3: Skills Taught by KiVa Intervention



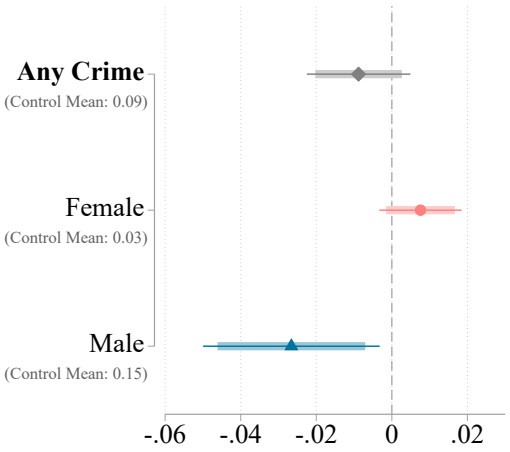
Notes: This figure shows treatment effects (Equation 1) on indices for the skills taught by the KiVa program. Outcomes from top to bottom are indices for: Disapproval of bullying, Empathy towards the victim, and Self-Efficacy to intervene in a bullying incident (see Appendix Table F1). All regressions include strata fixed effects and the baseline value of the outcome variable, additional controls are selected with PDS lasso. 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: Learning Environment: Climate and Grades



Notes: This figure shows treatment effects (Equation 1) on measures of school climate based on survey data and standardized grades obtained in grade 9 from register data. Outcomes from top to bottom: Index for school climate, Index for academic self-concept, Average grade across all subjects that qualify for admission to upper secondary school, Grade for languages (mother tongue, second national language and foreign language), Grade for mathematics. See Appendix Table F1 for details on survey-based outcomes. All regressions include strata fixed effects and the baseline value of the outcome variable for survey-based outcomes. Additional controls selected with PDS lasso. Standard errors are clustered at the school level. Bars indicate 90% (thick) and 95% (thin) confidence intervals.

Figure 5: Long-term Harmful Behavior: Crime in Adulthood



Notes: This figure shows treatment effects (Equation 1) on an indicator for having committed any crime in adulthood. All regressions include strata fixed effects, controls selected with PDS lasso, and grade 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.

Table 1: Summary Statistics and Balance

		Main Sample (Grades 7–9)			Survey Sample (Grades 8–9)		
	Mean	Control	Treatment	T - C	Control	Treatment	T - C
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Administrative Data							
Female	0.52	0.53	0.52	−0.00	0.53	0.51	−0.01
	(0.50)			(0.01)			(0.01)
Immigrant Background	0.02	0.01	0.02	0.01	0.01	0.02	0.01
	(0.14)			(0.01)			(0.01)
Lives in Region of Birth	0.89	0.89	0.89	0.00	0.90	0.89	−0.00
	(0.31)			(0.01)			(0.02)
Lives in Urban Area	0.75	0.73	0.76	0.04	0.73	0.77	0.04
	(0.43)			(0.05)			(0.05)
Young for Grade	0.01	0.01	0.01	−0.00	0.01	0.01	−0.00
	(0.08)			(0.00)			(0.00)
Old for Grade	0.05	0.05	0.05	0.00	0.05	0.05	−0.00
	(0.21)			(0.01)			(0.01)
Single Child	0.14	0.14	0.14	−0.00	0.15	0.15	−0.00
	(0.35)			(0.01)			(0.01)
Age Youngest Child in Family	10.13	10.11	10.14	0.03	10.50	10.54	0.06
	(3.74)			(0.13)			(0.15)
Single Parent Household	0.17	0.17	0.17	−0.00	0.17	0.17	−0.01
	(0.38)			(0.01)			(0.01)
Below Median Disposable Income	0.32	0.34	0.31	−0.02	0.34	0.31	−0.04*
	(0.47)			(0.02)			(0.02)
Test for joint Orthogonality							
F-Stat				1.00			1.09
p-value				0.61			0.55
Panel B: KiVa Survey Data							
Bully Indicator	0.25				0.25	0.24	−0.01
	(0.43)						(0.02)
Victim Indicator	0.24				0.25	0.24	−0.00
	(0.43)						(0.02)
Bullying Index	-0.02				−0.02	−0.02	0.01
	(0.95)						(0.04)
Victimization Index	-0.02				−0.01	−0.03	−0.01
	(0.97)						(0.04)
Students Observed in Class	17.59				17.53	17.64	0.10
	(3.05)						(0.46)
Test for joint Orthogonality							
F-Stat							1.17
p-value							0.61
N Observations				15,088			8,861
N Schools				73			70

This table shows summary statistics (Column 1) and balance for our main sample (Columns 2–4) and for the survey sample, consisting of students in grades 8 and 9 who participate in the KiVa endline survey (Columns 5–7). 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 coefficient on treatment from a regression of each variable on treatment group assignment, controlling for strata fixed effects and clustering standard errors at the school level. Standard errors are displayed in parentheses. Test for joint Orthogonality: F-Statistic and the p-value from a test of the joint significance of all covariates in Panel A (column 4), and all covariates in Panel A and B (column 7). P-value obtained via randomization inference. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table 2: Attrition in Administrative Data and Endline Survey

	Main Sample (Grades 7–9) (1)	Survey Sample (Grades 8–9) (2)
Panel A: Selective Attrition		
Treat	0.032	0.090**
	(0.021)	(0.038)
Panel B: Determinants of Attrition		
Treat × Female		−0.004
		(0.018)
Treat × Immigrant Background		0.073
		(0.052)
Treat × Lives in Region of Birth		−0.013
		(0.025)
Treat × Lives in Urban Area		0.087*
		(0.046)
Treat × Young for Grade		0.094
		(0.079)
Treat × Old for Grade		−0.083*
		(0.043)
Treat × Single Child		0.005
		(0.030)
Treat × Age Youngest Child in Family		−0.002
		(0.002)
Treat × Single Parent Household		0.003
		(0.024)
Treat × Below Median Disposable Income		−0.048**
		(0.019)
KiVa Survey Data		
Treat × Bully Indicator		−0.030
		(0.031)
Treat × Victim Indicator		0.015
		(0.027)
Treat × Students Observed in Class		−0.002
		(0.009)
Mean of Dependent Variable	0.90	0.83
N Observations	16,736	10,634
N Schools	73	73

This table reports attrition in the administrative data sample (Column 1) and the survey sample (Column 2) in Panel A. Differential attrition by baseline covariates in Panel B is estimated in a single regression that includes all interaction terms simultaneously. The dependent variable is an indicator equal to one if a student remains in the sample. The survey sample consists of students in grades 8 and 9 who participated in the KiVa endline survey, the sample in column 2 is therefore restricted to all students in grades 8 and 9. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

For Online Publication: Appendix Tables and Figures

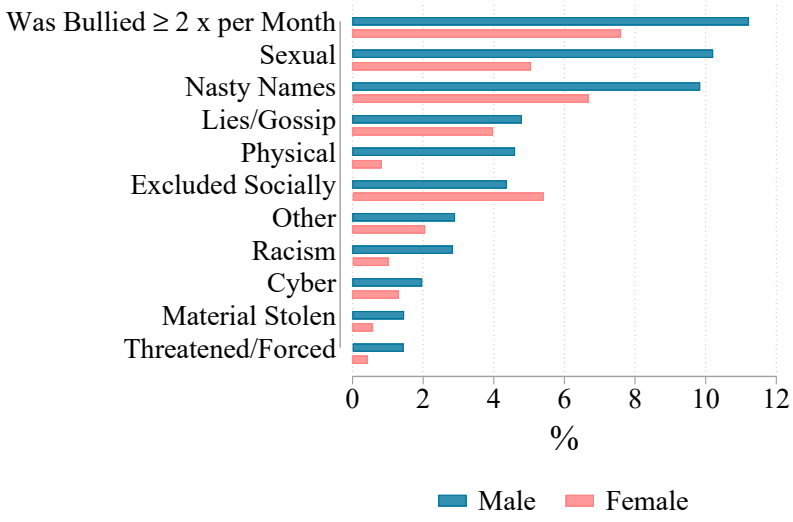
A Background

Figure A1: Examples of KiVa Classroom Sessions

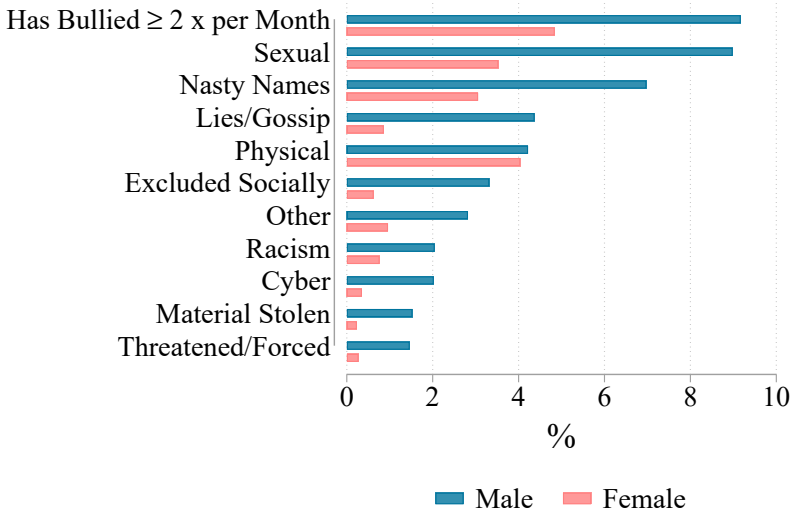
Notes: This figure shows scenes from the universal components of the KiVa program, including teacher-led classroom lessons, group discussions, and role-playing activities. These sessions aim to foster empathy, build students' confidence to intervene as bystanders, and promote anti-bullying attitudes.

B Descriptives: Bullying and Victimization

Figure B1: Categories of Bullying and Victimization



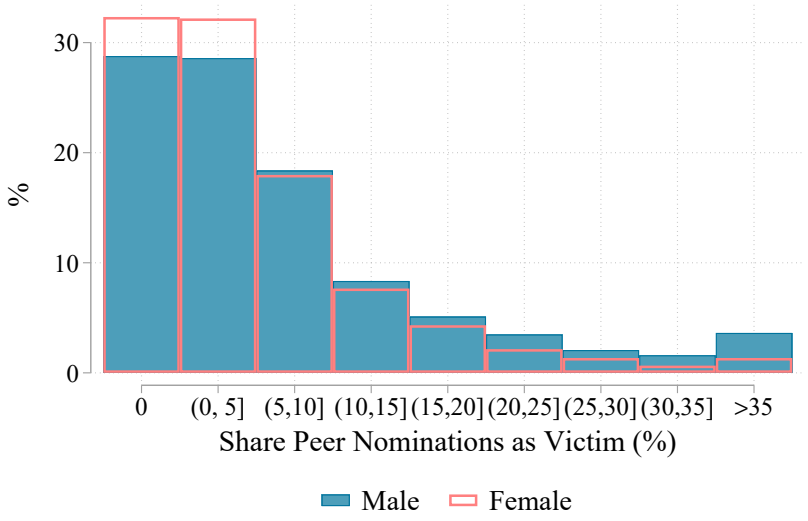
(a) Victimization



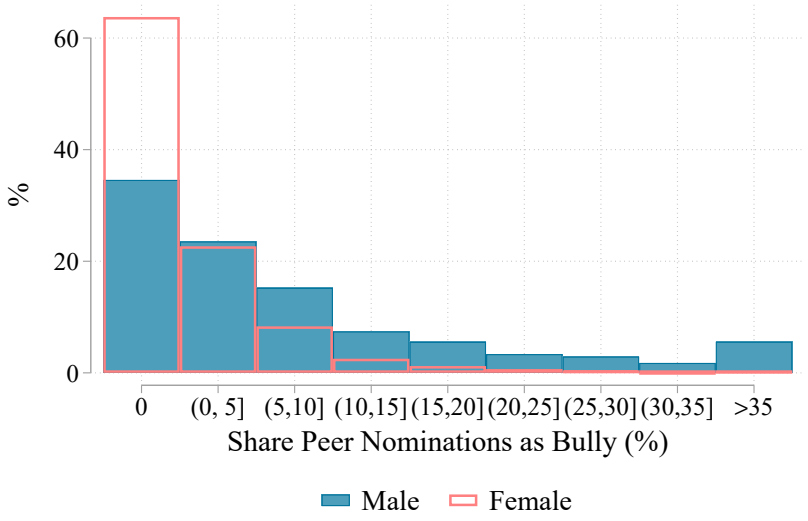
(b) Bullying

Notes: This figure shows the share of pupils by gender who self-report being affected by different categories of victimization (Panel a), or committing different categories of bullying (Panel b). We classify a student as belonging to a category if they report having been victimized or having bullied in a specific category at least twice a month over the past two months, following Kärnä et al. (2011).

Figure B2: Peer-Nomination Shares by Gender



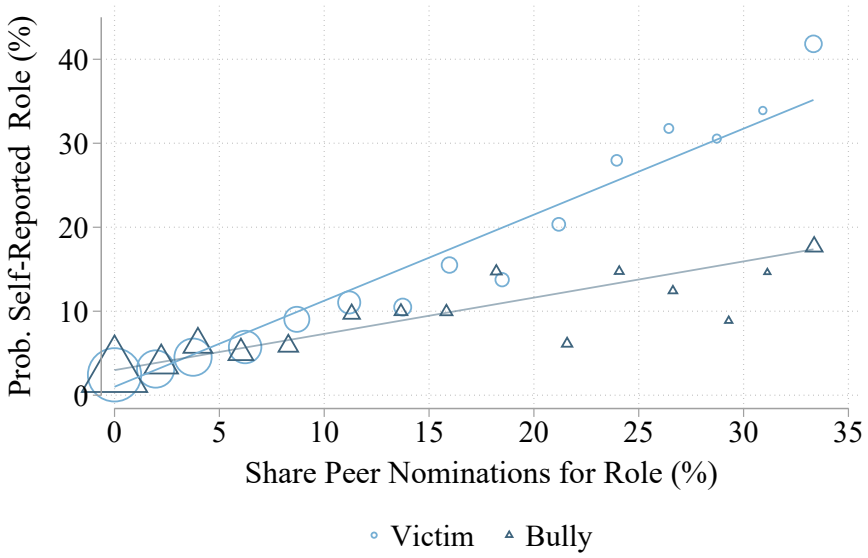
(a) Victimization



(b) Bullying

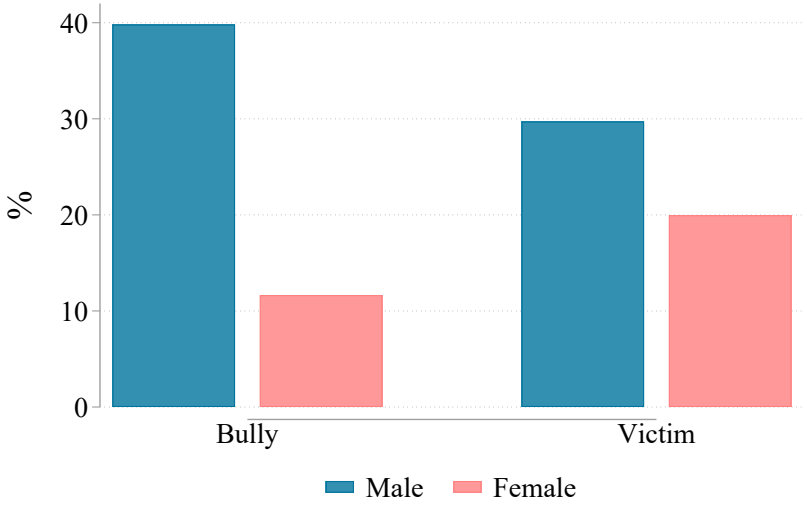
Notes: This figure shows the distribution of the share of peer nominations received by gender for victimization (Panel a) and bullying (Panel b). The share of peer nominations is calculated as the average share of classmates nominating a student as a victim or bully, respectively, across the three questions for each role (see Appendix Table F1 for details).

Figure B3: Correlation between Self- and Peer-Reported Social Roles



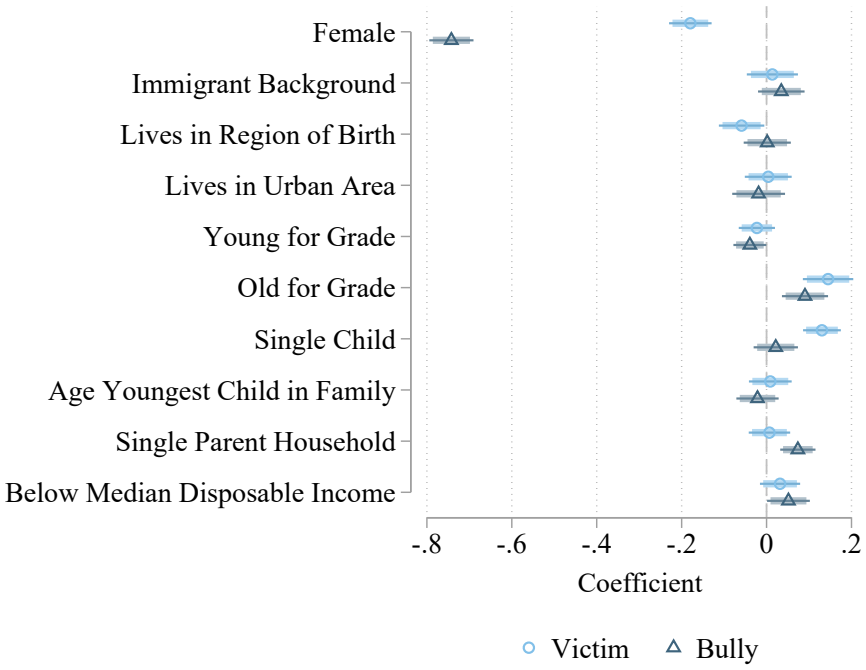
Notes: This figure shows the correlation between the probability of self-nominating as bully or victim and the share of peer nomination received. The share of peer nominations are averaged across the three questions for each role and binned. Observations with more than 35% of nominations are included in the last bin. We classify a student as identifying for a social role via self-reports if they report having been victimized or having bullied at least twice a month over the past two months, following Kärnä et al. (2011). Lines show a regression fit based on binned data, weighted by observations per bin. Marker size is proportional to the number of observations in each bin.

Figure B4: Social Roles: Prevalence by Gender

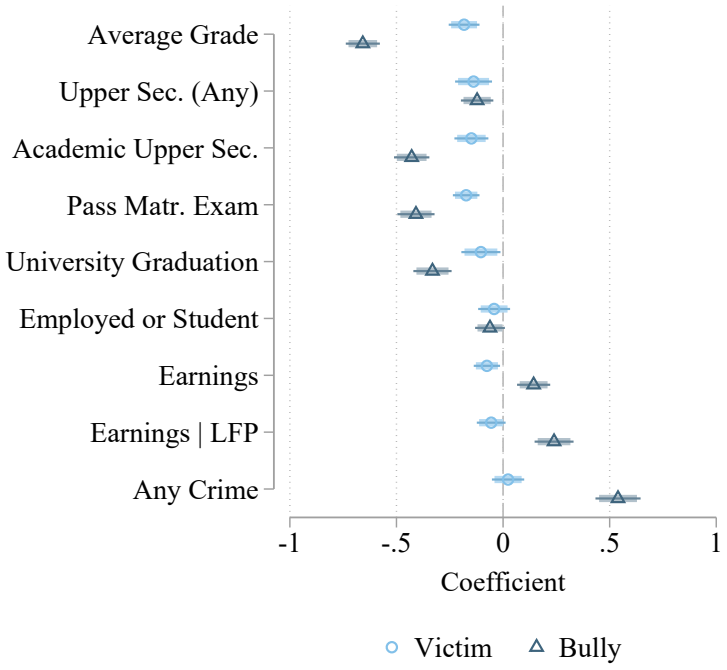


Notes: This figure shows the prevalence of bullies and victims by gender at baseline based on the peer-reported role definitions. Bully and Victim are dummies equal to one if the corresponding variance-weighted index of the share of nominating peers is at or above the 75th percentile of the distribution of the control group at baseline (see Data Section 3.2 for details). Social roles are non-exclusive categories.

Figure B5: Correlation of Social Roles with Baseline Characteristics and Long-Term Outcomes



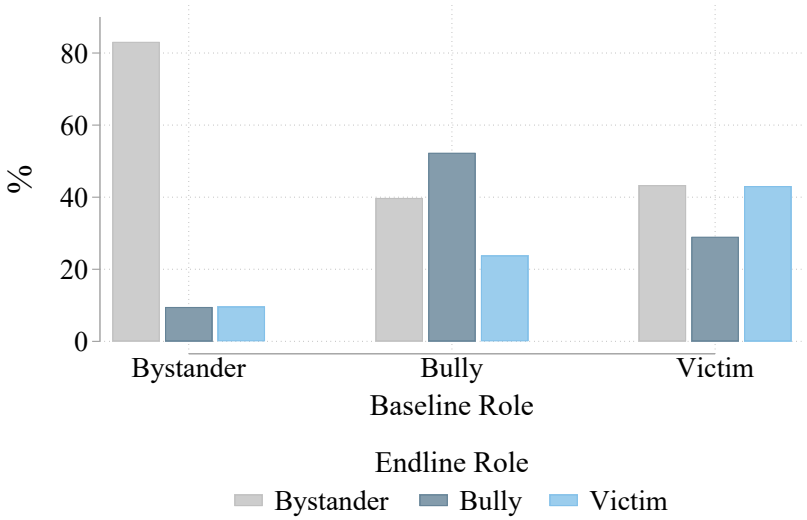
(a) Baseline Characteristics



(b) Long-Term Outcomes

Notes: This figure shows correlations between social role at baseline and baseline characteristics (Panel a) as well as long-term outcomes (Panel b). Panel (b) is restricted to the control group only. Coefficients from a regression of the standardized baseline characteristic or outcome on role dummies with strata fixed effects. A pupil is classified as belonging to a social role if the corresponding variance-weighted index of the share of nominating peers is at or above the 75th percentile of the distribution of the control group at baseline. Standard errors clustered at the school level. Bars indicate 90% (thick) and 95% (thin) confidence intervals.

Figure B6: Persistence of Social Roles



Notes: This figure shows the persistence of social roles based on peer reports. For each baseline role, the bars show the percentage of observations in each baseline role at the endline survey. A student is classified as belonging to a social role if the corresponding variance-weighted index of the share of nominating peers is at or above the 75th percentile of the distribution of the control group at baseline. Social roles are non-exclusive categories.

C Additional Main Results and Heterogeneity

Table C1: Heterogeneous Treatment Effects on Education Outcomes

	Attend			
	Upper Sec. (Any)	Academic Upper Sec.	Pass Matr. Exam	University Graduation
	(1)	(2)	(3)	(4)
A. Main Estimates				
Treat	0.002	0.051**	0.038*	0.039**
	(0.007)	(0.024)	(0.020)	(0.017)
	[0.234]	[0.080]	[0.080]	[0.080]
Grade	7-9	7-9	7-9	7-9
N	15,087	15,087	15,068	13,833
Control Mean	0.894	0.478	0.425	0.420
Adj. R^2	0.42	0.10	0.08	0.06
B. Heterogeneity: Gender				
Treat × Female	0.001	0.060**	0.037*	0.040*
	(0.007)	(0.025)	(0.022)	(0.020)
	[0.295]	[0.073]	[0.099]	[0.089]
Treat × Male	0.003	0.040	0.040*	0.038**
	(0.009)	(0.027)	(0.023)	(0.018)
	[0.230]	[0.163]	[0.163]	[0.163]
Grade	7-9	7-9	7-9	7-9
N	15,087	15,087	15,068	13,833
Control Mean				
Female	0.910	0.538	0.484	0.494
Male	0.877	0.411	0.358	0.335
P-value	0.81	0.31	0.87	0.94
C. Heterogeneity: Role				
Treat × Bully	-0.010	0.036	0.020	0.012
	(0.014)	(0.028)	(0.024)	(0.025)
	[1.000]	[1.000]	[1.000]	[1.000]
Treat × Victim	0.004	0.067**	0.057**	0.041
	(0.012)	(0.026)	(0.023)	(0.030)
	[0.294]	[0.031]	[0.031]	[0.128]
Treat × Bystander	0.007	0.057**	0.045**	0.044**
	(0.007)	(0.024)	(0.022)	(0.021)
	[0.095]	[0.066]	[0.066]	[0.066]
Grade	8-9	8-9	8-9	8-9
N	10,209	10,209	10,198	9,375
Control Mean				
Bully	0.863	0.313	0.270	0.301
Victim	0.861	0.407	0.348	0.371
Bystander	0.917	0.561	0.511	0.503
P-value				
Bully/ Bystander	0.24	0.45	0.38	0.28
Victim/ Bystander	0.82	0.71	0.59	0.93

Notes: This table shows treatment effects (Equation 1) on education outcomes. Column 1: Indicator for enrollment in upper secondary school at age 16. Column 2: Indicator for enrollment in academic upper secondary school at age 16. Column 3: Indicator for passing the matriculation exam by age 19. Column 4: Indicator for university graduation by 2022, the last year of observation. All regressions include strata fixed effects and controls selected with PDS lasso. Column 4 includes controls for grade fixed effects. Standard errors clustered at the school level in parentheses and sharpened q-values for each row reported in square brackets. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table C2: Heterogeneous Treatment Effects on Labor Outcomes

	Employed or Student (1)	Unemployed (2)	Out of Labor Force (3)	Earnings (4)	Earnings | LFP (5)
A. Main Estimates					
Treat	0.012**	−0.009**	−0.005	1,213***	833**
	(0.005)	(0.004)	(0.004)	(410)	(406)
	[0.047]	[0.047]	[0.059]	[0.021]	[0.047]
Grade	7-9	7-9	7-9	7-9	7-9
N	14,639	14,639	14,639	14,639	11,641
Control Mean	0.867	0.067	0.066	28,005	33,849
Adj. R^2	0.01	0.01	0.01	0.05	0.07
B. Heterogeneity: Gender					
Treat × Female	0.013	-0.008	-0.006	1,182**	990*
	(0.008)	(0.006)	(0.006)	(567)	(570)
	[0.215]	[0.215]	[0.215]	[0.215]	[0.215]
Treat × Male	0.011	−0.010*	−0.003	1,246**	664
	(0.007)	(0.006)	(0.005)	(591)	(541)
	[0.254]	[0.235]	[0.349]	[0.235]	[0.254]
Grade	7-9	7-9	7-9	7-9	7-9
N	14,639	14,639	14,639	14,639	11,641
Control Mean					
Female	0.868	0.055	0.077	24,982	30,062
Male	0.866	0.081	0.053	31,317	38,016
P-value	0.82	0.80	0.74	0.94	0.67
C. Heterogeneity: Role					
Treat × Bully	0.025**	-0.007	-0.019*	815	35
	(0.012)	(0.010)	(0.011)	(893)	(828)
	[0.224]	[0.570]	[0.224]	[0.570]	[1.000]
Treat × Victim	−0.013	−0.008	0.018	−578	42
	(0.016)	(0.011)	(0.011)	(671)	(670)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Treat × Bystander	0.017**	−0.010	−0.008	1,730***	1,474**
	(0.007)	(0.006)	(0.006)	(637)	(708)
	[0.044]	[0.074]	[0.075]	[0.042]	[0.056]
Grade	8-9	8-9	8-9	8-9	8-9
N	9,897	9,897	9,897	9,897	7,984
Control Mean					
Bully	0.855	0.077	0.067	30,814	37,382
Victim	0.856	0.081	0.063	27,779	34,095
Bystander	0.880	0.055	0.065	28,538	33,656
P-value					
Bully/ Bystander	0.59	0.81	0.33	0.42	0.18
Victim/ Bystander	0.11	0.89	0.04	0.01	0.13

Notes: This table shows treatment effects (Equation 1) on labor outcomes. Column 1: Indicator for being employed or student in 2022. Column 2: Indicator for being unemployed in 2022. Column 3: Indicator for being outside the labor force for other reasons in 2022. Column 4: Annual earnings in Euro in 2022. Column 5: Annual earnings in Euro in 2022, conditional on being employed. All regressions include strata fixed effects, controls selected with PDS lasso, and controls for grade fixed effects. Standard errors clustered at the school level in parentheses and sharpened q-values for each row reported in square brackets. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

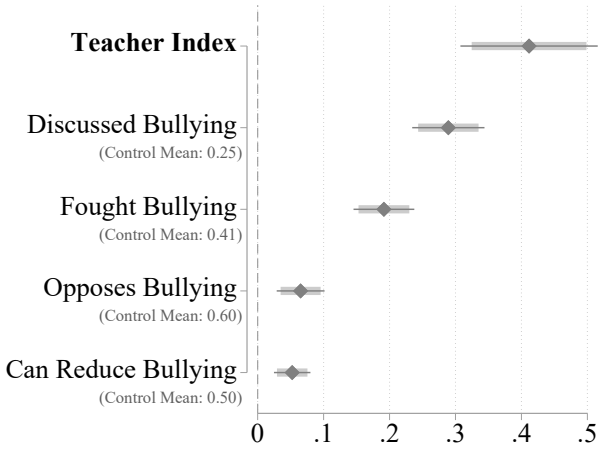
Table C3: Heterogeneous Treatment Effects on University Enrollment

	Age 19 (1)	Age 20 (2)	Age 21 (3)	Age 22 (4)	Age 23 (5)	Age 24 (6)	Age 25 (7)	Age 26 (8)	Age 27 (9)	Age 28 (10)	Age 29 (11)
A. Main Estimates											
Treat	0.024***	0.028**	0.030*	0.035*	0.032	0.033	0.039**	0.026	0.017	0.000	0.004
	(0.009)	(0.013)	(0.017)	(0.019)	(0.021)	(0.021)	(0.019)	(0.016)	(0.013)	(0.009)	(0.008)
	[0.084]	[0.154]	[0.172]	[0.172]	[0.172]	[0.172]	[0.154]	[0.172]	[0.209]	[0.311]	[0.237]
Grade	7-9	7-9	7-9	7-9	7-9	7-9	7-9	7-9	7-9	7-9	7-9
N	15,014	14,955	14,912	14,869	14,831	14,771	14,720	14,687	14,667	14,606	10,463
Control Mean	0.173	0.298	0.369	0.404	0.391	0.346	0.281	0.229	0.190	0.127	0.082
Adj. R^2	0.05	0.05	0.05	0.05	0.04	0.03	0.03	0.01	0.01	0.00	0.00
B. Heterogeneity: Gender											
Treat × Female	0.013	0.024	0.036**	0.036*	0.024	0.030	0.025	0.011	0.010	-0.011	0.000
	(0.012)	(0.016)	(0.018)	(0.021)	(0.023)	(0.023)	(0.020)	(0.016)	(0.014)	(0.009)	(0.011)
	[0.651]	[0.651]	[0.651]	[0.651]	[0.651]	[0.651]	[0.651]	[0.651]	[0.651]	[0.651]	[0.665]
Treat × Male	0.036***	0.031**	0.024	0.035*	0.040*	0.035	0.054**	0.041**	0.025	0.013	0.008
	(0.012)	(0.016)	(0.020)	(0.021)	(0.021)	(0.022)	(0.021)	(0.020)	(0.016)	(0.011)	(0.009)
	[0.035]	[0.119]	[0.186]	[0.149]	[0.119]	[0.149]	[0.076]	[0.119]	[0.149]	[0.196]	[0.208]
Grade	7-9	7-9	7-9	7-9	7-9	7-9	7-9	7-9	7-9	7-9	7-9
N	15,014	14,955	14,912	14,869	14,831	14,771	14,720	14,687	14,667	14,606	10,463
Control Mean											
Female	0.191	0.336	0.402	0.434	0.403	0.340	0.279	0.229	0.195	0.141	0.091
Male	0.154	0.255	0.333	0.370	0.379	0.352	0.284	0.228	0.183	0.112	0.073
P-value	0.14	0.68	0.50	0.98	0.32	0.77	0.09	0.05	0.28	0.03	0.56
C. Heterogeneity: Role											
Treat × Bully	0.007	0.013	-0.000	-0.008	0.015	0.011	0.027	0.004	0.003	-0.003	-0.003
	(0.014)	(0.018)	(0.021)	(0.021)	(0.022)	(0.021)	(0.022)	(0.021)	(0.018)	(0.018)	(0.015)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Treat × Victim	0.054***	0.036	0.033	0.043	0.027	0.038	0.033	0.054**	0.050**	0.042**	0.023**
	(0.019)	(0.023)	(0.028)	(0.031)	(0.030)	(0.030)	(0.026)	(0.023)	(0.020)	(0.019)	(0.012)
	[0.071]	[0.139]	[0.167]	[0.162]	[0.255]	[0.162]	[0.162]	[0.076]	[0.076]	[0.076]	[0.084]
Treat × Bystander	0.024**	0.028*	0.036**	0.046**	0.038*	0.041*	0.034*	0.015	0.003	0.003	−0.004
	(0.012)	(0.016)	(0.018)	(0.020)	(0.021)	(0.022)	(0.020)	(0.018)	(0.014)	(0.013)	(0.010)
	[0.163]	[0.163]	[0.163]	[0.163]	[0.163]	[0.163]	[0.163]	[0.254]	[0.447]	[0.447]	[0.447]
Grade	8-9	8-9	8-9	8-9	8-9	8-9	8-9	8-9	8-9	8-9	8-9
N	10,164	10,125	10,093	10,064	10,044	10,003	9,962	9,932	9,925	9,893	9,859
Control Mean											
Bully	0.105	0.178	0.248	0.291	0.288	0.264	0.227	0.193	0.161	0.141	0.076
Victim	0.152	0.257	0.328	0.360	0.354	0.309	0.272	0.209	0.176	0.163	0.087
Bystander	0.209	0.365	0.438	0.467	0.454	0.399	0.327	0.260	0.214	0.190	0.090
P-value											
Bully/ Bystander	0.30	0.45	0.15	0.05	0.36	0.26	0.80	0.67	0.98	0.79	0.98
Victim/ Bystander	0.15	0.73	0.89	0.90	0.69	0.90	0.96	0.12	0.03	0.05	0.06

Notes: This table shows treatment effects (Equation 1) on university enrollment. The outcome is an indicator for being enrolled at university at a certain age. All regressions include strata fixed effects and controls selected with PDS lasso. Standard errors clustered at the school level in parentheses and sharpened q-values for each row reported in square brackets. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

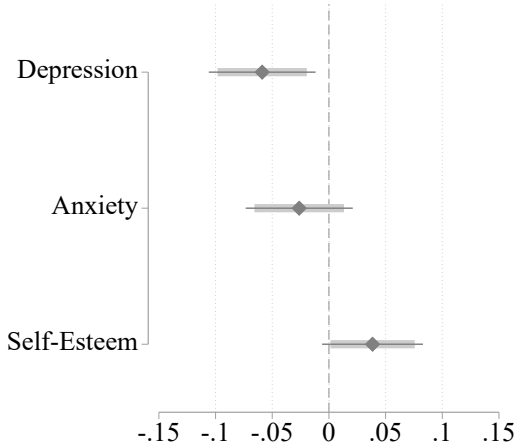
D Additional Results Mechanism

Figure D1: Perceptions about Teachers' Actions against Bullying



Notes: This figure shows treatment effects (Equation 1) of students' perceptions of teachers' actions and attitudes against bullying. Outcomes from top to bottom: Index across the subsequent four indicators. Indicators for: Teacher discussed bullying at least twice during the academic year (Discussed Bullying), Teacher actively intervened in bullying (Fought Bullying), Teacher is opposed to bullying (Opposes Bullying), Teacher can decrease bullying (Can Reduce Bullying). All regressions include strata fixed effects and the baseline value of the outcome variable, additional controls selected with PDS lasso. 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 D2: Socio-Emotional Wellbeing



Notes: This figure shows treatment effects (Equation 1) on indices for socio-emotional wellbeing. Outcomes from top to bottom: Index for depression based on Beck et al. (1996), Index for anxiety, Index for self-esteem (see Appendix Table F1). The indices are constructed over raw scales. All regressions include strata fixed effects and the baseline value of the outcome variable, additional controls selected with PDS lasso. Standard errors are clustered at the school level. Bars indicate 90% (thick) and 95% (thin) confidence intervals.

Table D1: Treatment Effects on Bullying and Victimization

	Perceived Bullying		Bullying Index	Victimization Index
	Main	IPW		
	(1)	(2)	(3)	(4)
A. Main Estimates				
Treat	−0.038**	−0.039***	−0.037	−0.059
	(0.015)	(0.014)	(0.028)	(0.040)
	[0.035]		[0.158]	[0.158]
Grade	8-9	8-9	8-9	8-9
N	8,576	8,575	10,010	10,019
Control Mean	0.497	0.496	−0.108	−0.127
Adj. R^2	0.04	0.04	0.40	0.30
B. Heterogeneity: Gender				
Treat × Female	−0.041**	−0.043**	0.012	−0.032
	(0.018)	(0.018)	(0.026)	(0.037)
	[0.099]		[0.757]	[0.643]
Treat × Male	−0.036*	−0.036*	−0.089**	−0.088*
	(0.019)	(0.020)	(0.040)	(0.048)
	[0.080]		[0.080]	[0.080]
Grade	8-9	8-9	8-9	8-9
N	8,576	8,575	10,010	10,019
Control Mean				
Female	0.496	0.494	−0.311	−0.205
Male	0.499	0.498	0.119	−0.039
P-value	0.84	0.76	0.01	0.08
C. Heterogeneity: Role				
Treat × Bully	-0.026	-0.025	-0.048	-0.066
	(0.022)	(0.022)	(0.071)	(0.045)
	[0.585]		[0.585]	[0.585]
Treat × Victim	−0.022	−0.017	−0.071**	−0.073
	(0.025)	(0.026)	(0.033)	(0.072)
	[0.353]		[0.114]	[0.353]
Treat × Bystander	−0.036**	−0.038**	−0.023	−0.046
	(0.017)	(0.017)	(0.021)	(0.031)
	[0.107]		[0.227]	[0.166]
Grade	8-9	8-9	8-9	8-9
N	8,232	8,231	9,634	9,643
Control Mean				
Bully	0.476	0.474	0.523	0.019
Victim	0.467	0.465	0.052	0.418
Bystander	0.527	0.526	−0.340	−0.321
P-value				
Bully/ Bystander	0.70	0.67	0.71	0.62
Victim/ Bystander	0.57	0.44	0.12	0.69

Notes: This table shows treatment effects (Equation 1) on harmful behavior. Column 1 and 2: Indicator for a perceived increase in bullying. Column 3: Index for being peer-reported as a bully. Column 4: Index for being peer-reported as a victim. The indices are constructed over raw scales. All regressions include strata fixed effects, controls selected with PDS lasso, and the baseline control of the outcome variable. In Column 2, observations are weighted by the inverse estimated probability of being observed in the survey sample, conditional on baseline characteristics. Regressions in Columns 3 and 4 are estimated using all observations in grade 8 and 9 for which peer reports are available. Standard errors clustered at the school level in parentheses and sharpened q-values for each row reported in square brackets. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table D2: Treatment Effects on Students' Perception of Teachers

	Teacher Index		Discussed Bullying		Fought Bullying		Opposes Bullying		Can Reduce Bullying	
	Main	IPW	Main	IPW	Main	IPW	Main	IPW	Main	IPW
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A. Main Estimates										
Treat	0.412***	0.413***	0.289***	0.289***	0.191***	0.193***	0.065***	0.067***	0.052***	0.054***
	(0.052)	(0.052)	(0.027)	(0.027)	(0.023)	(0.023)	(0.018)	(0.019)	(0.014)	(0.014)
	[0.001]		[0.001]		[0.001]		[0.001]		[0.001]	
Grade	8-9	8-9	8-9	8-9	8-9	8-9	8-9	8-9	8-9	8-9
N	8,549	8,548	8,549	8,548	8,540	8,539	8,545	8,544	8,535	8,534
Control Mean	0.000	0.002	0.252	0.251	0.413	0.412	0.597	0.597	0.501	0.501
Adj. R^2	0.13	0.13	0.11	0.11	0.09	0.09	0.09	0.09	0.07	0.07
B. Heterogeneity: Gender										
Treat × Female	0.454***	0.458***	0.335***	0.335***	0.198***	0.202***	0.083***	0.085***	0.039**	0.040**
	(0.057)	(0.057)	(0.031)	(0.031)	(0.026)	(0.026)	(0.021)	(0.021)	(0.018)	(0.018)
	[0.001]		[0.001]		[0.001]		[0.001]		[0.007]	
Treat × Male	0.365***	0.364***	0.240***	0.239***	0.185***	0.185***	0.046*	0.047*	0.067***	0.069***
	(0.060)	(0.060)	(0.028)	(0.028)	(0.026)	(0.026)	(0.026)	(0.026)	(0.017)	(0.017)
	[0.001]		[0.001]		[0.001]		[0.016]		[0.001]	
Grade	8-9	8-9	8-9	8-9	8-9	8-9	8-9	8-9	8-9	8-9
N	8,549	8,548	8,549	8,548	8,540	8,539	8,545	8,544	8,535	8,534
Control Mean										
Female	0.026	0.028	0.245	0.243	0.416	0.414	0.621	0.622	0.499	0.501
Male	−0.030	−0.029	0.260	0.261	0.411	0.410	0.571	0.570	0.503	0.501
P-value	0.09	0.08	0.00	0.00	0.59	0.49	0.21	0.20	0.19	0.17
C. Heterogeneity: Role										
Treat × Bully	0.244***	0.247***	0.163***	0.163***	0.148***	0.148***	0.018	0.021	0.034*	0.040**
	(0.054)	(0.054)	(0.027)	(0.027)	(0.025)	(0.026)	(0.026)	(0.026)	(0.020)	(0.020)
	[0.197]		[0.322]		[0.180]		[0.001]		[0.001]	
Treat × Victim	0.365***	0.362***	0.232***	0.227***	0.133***	0.134***	0.059**	0.058**	0.078***	0.080***
	(0.072)	(0.072)	(0.035)	(0.034)	(0.033)	(0.033)	(0.026)	(0.027)	(0.026)	(0.026)
	[0.109]		[0.024]		[0.099]		[0.001]		[0.001]	
Treat × Bystander	0.386***	0.395***	0.289***	0.294***	0.168***	0.173***	0.073***	0.077***	0.034*	0.037**
	(0.056)	(0.055)	(0.029)	(0.028)	(0.026)	(0.026)	(0.019)	(0.019)	(0.018)	(0.018)
	[0.027]		[0.097]		[0.065]		[0.001]		[0.001]	
Grade	8-9	8-9	8-9	8-9	8-9	8-9	8-9	8-9	8-9	8-9
N	8,205	8,204	8,205	8,204	8,196	8,195	8,201	8,200	8,192	8,191
Control Mean										
Bully	−0.110	−0.112	0.240	0.239	0.394	0.393	0.520	0.520	0.480	0.474
Victim	−0.082	−0.080	0.250	0.250	0.409	0.408	0.548	0.551	0.475	0.472
Bystander	0.060	0.063	0.256	0.256	0.424	0.424	0.636	0.635	0.514	0.517
P-value										
Bully/ Bystander	0.03	0.02	0.00	0.00	0.57	0.48	0.06	0.06	1.00	0.91
Victim/ Bystander	0.74	0.61	0.10	0.05	0.20	0.15	0.57	0.44	0.14	0.15

Notes: This table shows treatment effects (Equation 1) on student's perceptions about teachers. Column 1 and 2: Index of the following four indicators. Column 3 and 4: Indicator for whether the teacher has discussed bullying at least twice since last autumn. Column 5 and 6: Indicator that the teacher has fought bullying. Column 7 and 8: Indicator that the teacher opposes bullying. Column 9 and 10: Indicator that the teacher can decrease bullying. All regressions include strata fixed effects, controls selected with PDS lasso, and the baseline control of the outcome variable. In the second column for each outcome, observations are weighted by the inverse estimated probability of being observed in the survey sample, conditional on baseline characteristics. Standard errors clustered at the school level in parentheses and sharpened q-values for each row reported in square brackets. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table D3: Treatment Effects on Skills Taught by KiVa

	Disapproval of Bullying		Empathy		Efficacy	
	Main (1)	IPW (2)	Main (3)	IPW (4)	Main (5)	IPW (6)
A. Main Estimates						
Treat	0.007	0.012	0.012	0.009	0.051*	0.050*
	(0.023)	(0.023)	(0.027)	(0.027)	(0.029)	(0.030)
	[0.427]		[0.427]		[0.092]	
Grade	8-9	8-9	8-9	8-9	8-9	8-9
N	8,632	8,631	8,548	8,547	8,637	8,636
Control Mean	0.000	−0.004	0.000	0.003	0.000	0.001
Adj. R^2	0.01	0.01	0.23	0.23	0.18	0.18
B. Heterogeneity: Gender						
Treat × Female	0.020	0.023	−0.037	−0.038	0.008	0.010
	(0.029)	(0.029)	(0.031)	(0.031)	(0.037)	(0.038)
	[0.602]		[0.308]		[0.644]	
Treat × Male	−0.006	−0.001	0.065*	0.061	0.098***	0.094**
	(0.034)	(0.035)	(0.038)	(0.038)	(0.036)	(0.037)
	[0.211]		[0.073]		[0.017]	
Grade	8-9	8-9	8-9	8-9	8-9	8-9
N	8,632	8,631	8,548	8,547	8,637	8,636
Control Mean						
Female	0.038	0.033	0.275	0.278	0.134	0.136
Male	−0.042	−0.045	−0.313	−0.307	−0.152	−0.151
P-value	0.56	0.58	0.02	0.02	0.05	0.07
C. Heterogeneity: Role						
Treat × Bully	0.018	0.028	0.030	0.033	0.039	0.033
	(0.048)	(0.047)	(0.050)	(0.050)	(0.041)	(0.042)
	[0.512]		[0.576]		[0.314]	
Treat × Victim	−0.079	−0.074	0.089*	0.079	0.023	0.020
	(0.059)	(0.059)	(0.047)	(0.048)	(0.049)	(0.050)
	[0.202]		[0.039]		[0.171]	
Treat × Bystander	0.015	0.018	−0.029	−0.030	0.052	0.054
	(0.033)	(0.033)	(0.030)	(0.030)	(0.035)	(0.035)
	[0.042]		[0.161]		[0.120]	
Grade	8-9	8-9	8-9	8-9	8-9	8-9
N	8,286	8,285	8,203	8,202	8,291	8,290
Control Mean						
Bully	−0.118	−0.127	−0.308	−0.304	−0.094	−0.095
Victim	0.041	0.039	−0.061	−0.050	−0.161	−0.164
Bystander	0.003	0.002	0.104	0.107	0.067	0.068
P-value						
Bully/ Bystander	0.95	0.86	0.22	0.20	0.77	0.65
Victim/ Bystander	0.19	0.20	0.05	0.08	0.60	0.56

Notes: This table shows treatment effects (Equation 1) on skills taught by KiVa. Column 1 and 2: Index for disapproval of bullying. Column 3 and 4: Index for empathy. Column 5 and 6: Index for efficacy. The indices are constructed over raw scales. All regressions include strata fixed effects, controls selected with PDS lasso, and the baseline control of the outcome variable. In the second column for each outcome, observations are weighted by the inverse estimated probability of being observed in the survey sample, conditional on baseline characteristics. Standard errors clustered at the school level in parentheses and sharpened q-values for each row reported in square brackets. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table D4: Treatment Effects on Learning Environment and Academic Performance

	School Climate		Academic Self-Concept		Grades		
					Average	Languages	Math
	Main	IPW	Main	IPW			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Main Estimates							
Treat	0.061**	0.060**	0.049*	0.050**	0.053	0.081	0.048
	(0.026)	(0.027)	(0.025)	(0.025)	(0.054)	(0.057)	(0.043)
	[0.124]		[0.124]		[0.251]	[0.190]	[0.251]
Grade	8-9	8-9	8-9	8-9	7-9	7-9	7-9
N	8,839	8,838	8,345	8,344	15,033	15,033	15,032
Control Mean	0.000	−0.001	0.000	−0.003	−0.000	−0.000	−0.000
Adj. R^2	0.17	0.17	0.18	0.18	0.11	0.13	0.04
B. Heterogeneity: Gender							
Treat × Female	0.067**	0.068**	0.056*	0.059**	0.040	0.066	0.045
	(0.030)	(0.030)	(0.029)	(0.029)	(0.054)	(0.059)	(0.045)
	[0.167]		[0.167]		[0.385]	[0.314]	[0.314]
Treat × Male	0.054	0.052	0.041	0.039	0.067	0.097	0.051
	(0.036)	(0.036)	(0.034)	(0.033)	(0.060)	(0.062)	(0.049)
	[0.452]		[0.452]		[0.452]	[0.452]	[0.452]
Grade	8-9	8-9	8-9	8-9	7-9	7-9	7-9
N	8,839	8,838	8,345	8,344	15,033	15,033	15,032
Control Mean							
Female	0.102	0.103	0.102	0.102	0.241	0.280	0.104
Male	−0.114	−0.117	−0.116	−0.123	−0.268	−0.311	−0.115
P-value	0.76	0.67	0.69	0.61	0.47	0.45	0.89
C. Heterogeneity: Role							
Treat × Bully	-0.012	-0.011	0.042	0.041	0.002	0.031	0.022
	(0.047)	(0.047)	(0.045)	(0.045)	(0.055)	(0.060)	(0.051)
	[0.915]		[0.915]		[0.533]	[0.001]	[0.001]
Treat × Victim	0.081*	0.077	0.034	0.028	0.136**	0.110*	0.064
	(0.047)	(0.047)	(0.048)	(0.048)	(0.054)	(0.059)	(0.054)
	[0.211]		[0.049]		[0.175]	[0.001]	[0.001]
Treat × Bystander	0.063**	0.064**	0.040	0.046	0.052	0.069	0.043
	(0.029)	(0.030)	(0.028)	(0.028)	(0.056)	(0.062)	(0.044)
	[0.061]		[0.219]		[0.153]	[0.001]	[0.001]
Grade	8-9	8-9	8-9	8-9	8-9	8-9	8-9
N	8,488	8,487	8,002	8,001	10,192	10,192	10,192
Control Mean							
Bully	−0.217	−0.224	−0.310	−0.313	−0.499	−0.477	−0.377
Victim	−0.165	−0.167	−0.080	−0.078	−0.182	−0.175	−0.121
Bystander	0.117	0.117	0.106	0.102	0.227	0.224	0.165
P-value							
Bully/ Bystander	0.15	0.15	0.96	0.92	0.39	0.55	0.72
Victim/ Bystander	0.73	0.81	0.91	0.73	0.12	0.48	0.66

Notes: This table shows treatment effects (Equation 1) on the learning environment and academic performance. Column 1 and 2: Index for school climate. Column 3 and 4: Index for academic self-concept. Column 5: Average grade in compulsory subjects in grade 9 (standardized). Column 6: Grade in Mathematics in grade 9 (standardized). Column 7: Average language grade (native tongue, the other official language, and English) in grade 9 (standardized). The indices are constructed over raw scales. All regressions include strata fixed effects and controls selected with PDS lasso. Columns 1 to 4 also include the baseline control of the outcome variable. In Columns 2 and 4 observations are weighted by the inverse estimated probability of being observed in the survey sample, conditional on baseline characteristics. Standard errors clustered at the school level in parentheses and sharpened q-values for each row reported in square brackets. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table D5: Association between Average Grades and Attending Academic Track

	Academic Upper Sec.		
	(1)	(2)	(3)
Average Grade	0.323***	0.323***	0.315***
	(0.006)	(0.006)	(0.006)
% of Effect Explained	32.8	32.8	32.0
Adj. R^2	0.42	0.42	0.44
N	6,506	6,506	6,506
Strata FE		Yes	Yes
Controls			Yes

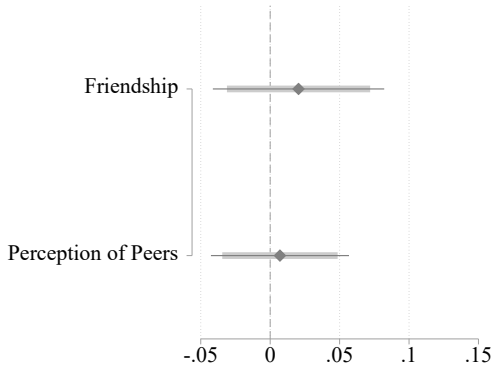
Notes: This table shows the association between the average grade obtained in grade 9 and academic upper secondary school attendance in the control group. Columns 2 and 3 include strata fixed effects. Column 3 includes controls selected with PDS lasso. Standard errors clustered at the school level in parentheses. The middle panel reports the share of the estimated treatment effect on attending the academic track of upper secondary school (in Table C1) that could be accounted for by treatment impacts on grades (Table D4) given the estimated association in each column. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table D6: Treatment Effects on Socio-Emotional Wellbeing

	Depression		Anxiety		Self-Esteem	
	Main	IPW	Main	IPW	Main	IPW
	(1)	(2)	(3)	(4)	(5)	(6)
A. Main Estimates						
Treat	−0.059**	−0.062**	−0.026	−0.030	0.038*	0.039*
	(0.024)	(0.024)	(0.024)	(0.023)	(0.022)	(0.022)
	[0.082]		[0.279]		[0.217]	
Grade	8-9	8-9	8-9	8-9	8-9	8-9
N	8,555	8,554	8,545	8,544	8,552	8,551
Control Mean	−0.000	0.004	−0.000	0.000	0.000	−0.003
Adj. R^2	0.19	0.20	0.13	0.13	0.20	0.21
B. Heterogeneity: Gender						
Treat × Female	−0.085***	−0.087***	−0.061**	−0.065**	0.052	0.053
	(0.025)	(0.026)	(0.025)	(0.025)	(0.035)	(0.035)
	[0.006]		[0.036]		[0.167]	
Treat × Male	−0.029	−0.033	0.012	0.008	0.023	0.024
	(0.040)	(0.041)	(0.038)	(0.038)	(0.029)	(0.028)
	[1.000]		[1.000]		[1.000]	
Grade	8-9	8-9	8-9	8-9	8-9	8-9
N	8,555	8,554	8,545	8,544	8,552	8,551
Control Mean						
Female	0.073	0.072	0.096	0.099	−0.068	−0.068
Male	−0.083	−0.074	−0.110	−0.111	0.077	0.072
P-value	0.22	0.25	0.10	0.10	0.54	0.53
C. Heterogeneity: Role						
Treat × Bully	-0.016	-0.021	0.029	0.014	-0.010	-0.003
	(0.061)	(0.061)	(0.048)	(0.048)	(0.035)	(0.035)
	[1.000]		[1.000]		[0.839]	
Treat × Victim	−0.067	−0.063	0.020	0.026	0.051	0.045
	(0.049)	(0.051)	(0.050)	(0.050)	(0.042)	(0.042)
	[0.272]		[0.066]		[0.220]	
Treat × Bystander	−0.050**	−0.053**	−0.019	−0.023	0.023	0.022
	(0.024)	(0.024)	(0.028)	(0.028)	(0.027)	(0.027)
	[0.082]		[0.280]		[0.180]	
Grade	8-9	8-9	8-9	8-9	8-9	8-9
N	8,211	8,210	8,201	8,200	8,209	8,208
Control Mean						
Bully	0.088	0.101	−0.125	−0.119	−0.060	−0.068
Victim	0.096	0.100	0.052	0.052	−0.012	−0.015
Bystander	−0.058	−0.058	0.019	0.018	0.039	0.039
P-value						
Bully/ Bystander	0.57	0.59	0.39	0.52	0.47	0.59
Victim/ Bystander	0.77	0.86	0.47	0.37	0.55	0.62

Notes: This table shows treatment effects (Equation 1) on socio-emotional wellbeing. Column 1 and 2: Index for depression. Column 3 and 4: Index for anxiety. Column 5 and 6: Index for self-esteem. The indices are constructed over raw scales. All regressions include strata fixed effects, controls selected with PDS lasso, and baseline control of the outcome variable. In the second column for each outcome, observations are weighted by the inverse estimated probability of being observed in the survey sample, conditional on baseline characteristics. Standard errors clustered at the school level in parentheses and sharpened q-values for each row reported in square brackets. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Figure D3: Social Network



Notes: This figure shows treatment effects (Equation 1) on indices for friendship and perception of peers. All regressions include strata fixed effects, controls selected with PDS lasso, and the baseline value of the outcome variable. 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.

E Robustness

Table E1: Robustness: Definition of Social Roles

	Academic Upper Sec. (1)	Pass Matr. Exam (2)	University Graduation (3)	Employed or Student (4)	Earnings (5)	Earnings | LFP (6)
A. Main Estimates						
Treat	0.051**	0.038*	0.039**	0.012**	1,213***	833**
	(0.024)	(0.020)	(0.017)	(0.005)	(410)	(406)
	[0.047]	[0.051]	[0.042]	[0.042]	[0.025]	[0.047]
Grade	7-9	7-9	7-9	7-9	7-9	7-9
N	15,087	15,068	13,833	14,639	14,639	11,641
Control Mean	0.478	0.425	0.420	0.867	28,005	33,849
Adj. R^2	0.10	0.08	0.06	0.01	0.05	0.07
B. Heterogeneity: Peer-reported (Anderson Index)						
Treat	0.060**	0.043**	0.043**	0.013**	1,044**	780
	(0.024)	(0.020)	(0.018)	(0.006)	(498)	(520)
	[0.051]	[0.051]	[0.051]	[0.051]	[0.051]	[0.051]
Treat × Bully Index	−0.003	0.003	−0.013	0.002	−655	−638
	(0.011)	(0.011)	(0.009)	(0.006)	(497)	(483)
	[0.666]	[0.666]	[0.624]	[0.666]	[0.624]	[0.624]
Treat × Victim Index	0.016	0.007	−0.002	−0.003	−453	−110
	(0.011)	(0.010)	(0.011)	(0.008)	(370)	(354)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Treat × Victim Index × Bully Index	−0.000	0.001	−0.003	−0.008	154	507
	(0.005)	(0.005)	(0.006)	(0.005)	(318)	(351)
	[1.000]	[1.000]	[1.000]	[0.859]	[1.000]	[0.859]
Grade	8-9	8-9	8-9	8-9	8-9	8-9
N	10,209	10,198	9,375	9,897	9,897	7,984
C. Heterogeneity: Self-reported (Indicators)						
Treat	0.056**	0.042*	0.048**	0.009	1,236**	1,144**
	(0.025)	(0.022)	(0.018)	(0.007)	(525)	(570)
	[0.062]	[0.062]	[0.062]	[0.073]	[0.062]	[0.062]
Treat × Bully	0.013	0.031	-0.005	0.021	-1,269	-2,792*
	(0.039)	(0.036)	(0.041)	(0.032)	(1,752)	(1,519)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[0.725]
Treat × Victim	−0.015	−0.046	−0.035	0.006	−667	−446
	(0.041)	(0.034)	(0.035)	(0.028)	(1,272)	(1,384)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Treat × Bully & Victim	0.141**	0.048	0.002	0.010	−2,720	−2,406
	(0.070)	(0.058)	(0.058)	(0.050)	(2,728)	(2,489)
	[0.417]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Grade	8-9	8-9	8-9	8-9	8-9	8-9
N	9,739	9,730	8,960	9,448	9,448	7,629
Control Mean						
Bully	0.335	0.281	0.348	0.846	30,557	37,111
Victim	0.452	0.408	0.399	0.844	25,513	31,898
Bully & Victim	0.230	0.207	0.253	0.847	27,066	33,183
Bystander	0.504	0.454	0.449	0.878	29,072	34,454

Notes: This table shows treatment effects (Equation 1) on main outcomes in Panel A and heterogeneity by social roles in Panels B and C. Panel B regresses the outcome on the treatment indicator, the continuous bullying and victimization Anderson indices, their interaction, and their interactions with the treatment indicator. Panel C regresses the outcome on indicators for self-reported bully, self-reported victim, and self-reported bully & victim, along with each indicator's interaction with treatment. 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). All regressions include strata fixed effects and PDS lasso selected controls. Columns 3–7 also control for grade fixed effects. Standard errors are clustered at the school level. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table E2: Robustness: Treatment Effects on Education and Labor Outcomes for Different Samples

	Academic Upper Sec.			University Graduation			Earnings		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A. Main Estimates									
Treat	0.051**	0.057**	0.058**	0.039**	0.045**	0.059***	1,213***	1,408***	1,646***
	(0.024)	(0.025)	(0.026)	(0.017)	(0.018)	(0.018)	(410)	(513)	(571)
N	15,087	10,633	8,860	13,833	9,745	8,169	14,639	10,307	8,596
Control Mean	0.478	0.480	0.478	0.420	0.431	0.421	28,005	28,511	28,234
Adj. R^2	0.10	0.10	0.10	0.06	0.06	0.06	0.05	0.05	0.05
P-value (Column 2/3)			0.98			0.58			0.76
B. Heterogeneity: Gender									
Treat × Female	0.060**	0.056**	0.055*	0.040*	0.031	0.047**	1,182**	1,380*	1,499*
	(0.025)	(0.027)	(0.029)	(0.020)	(0.022)	(0.023)	(567)	(720)	(792)
Treat × Male	0.040	0.059**	0.060**	0.038**	0.061***	0.072***	1,246**	1,438**	1,803**
	(0.027)	(0.027)	(0.027)	(0.018)	(0.019)	(0.018)	(591)	(682)	(763)
N	15,087	10,633	8,860	13,833	9,745	8,169	14,639	10,307	8,596
Control Mean									
Female	0.538	0.547	0.545	0.494	0.512	0.503	24,982	25,150	25,022
Male	0.411	0.405	0.403	0.335	0.336	0.326	31,317	32,269	31,808
P-value	0.31	0.89	0.84	0.94	0.14	0.29	0.94	0.95	0.77
C. Heterogeneity: Role									
Treat × Bully		0.036	0.032		0.012	0.007		867	1,244
		(0.028)	(0.030)		(0.025)	(0.027)		(886)	(936)
Treat × Victim		0.066**	0.069**		0.041	0.048		−562	−563
		(0.026)	(0.032)		(0.030)	(0.031)		(677)	(819)
Treat × Bystander		0.057**	0.057**		0.044**	0.061***		1,718***	1,749**
		(0.024)	(0.025)		(0.021)	(0.021)		(640)	(701)
N		10,209	8,508		9,375	7,852		9,897	8,252
Control Mean									
Bully		0.313	0.316		0.301	0.302		30,814	30,462
Victim		0.406	0.406		0.371	0.370		27,805	27,717
Bystander		0.561	0.555		0.503	0.486		28,538	28,318
P-value									
Bully/ Bystander		0.45	0.38		0.28	0.08		0.46	0.68
Victim/ Bystander		0.71	0.73		0.93	0.71		0.01	0.02
Grade	7-9	8-9	8-9	7-9	8-9	8-9	7-9	8-9	8-9
Conditional			Endline			Endline			Endline

Notes: This table shows treatment effects (Equation 1) on education and labor outcomes for different samples. Column 1–3: Indicator for enrollment in upper secondary school at age 16. Column 4–6: Indicator for university graduation by 2022, the last year of observation. Column 7–9: Annual earnings in Euro in 2022. All regressions include strata fixed effects and controls selected with PDS lasso. Columns 4–9 include controls for grade fixed effects. Regressions are estimated for three samples: grades 7–9, grades 8–9, and the survey sample (students in grades 8–9 who participated in the endline survey). Standard errors clustered at the school level in parentheses. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table E3: Robustness: Treatment Effects on Grades and Crime for Different Samples

	Average Grade			Any Crime		
	(1)	(2)	(3)	(4)	(5)	(6)
A. Main Estimates						
Treat	0.053	0.062	0.083	−0.009	−0.009	−0.011
	(0.054)	(0.056)	(0.057)	(0.007)	(0.008)	(0.009)
N	15,033	10,598	8,838	15,088	10,634	8,861
Control Mean	−0.000	0.005	−0.007	0.091	0.093	0.092
Adj. R^2	0.11	0.11	0.11	0.05	0.05	0.05
P-value (Column 2/3)			0.79			0.87
B. Heterogeneity: Gender						
Treat × Female	0.040	0.027	0.044	0.008	0.011*	0.012*
	(0.054)	(0.056)	(0.058)	(0.005)	(0.007)	(0.006)
Treat × Male	0.067	0.100	0.125*	−0.027**	−0.032**	−0.036**
	(0.060)	(0.063)	(0.064)	(0.012)	(0.014)	(0.015)
N	15,033	10,598	8,838	15,088	10,634	8,861
Control Mean						
Female	0.241	0.259	0.251	0.035	0.034	0.031
Male	−0.268	−0.281	−0.294	0.154	0.161	0.159
P-value	0.47	0.07	0.07	0.00	0.00	0.00
C. Heterogeneity: Role						
Treat × Bully		0.002	-0.004		-0.042*	-0.036
		(0.055)	(0.059)		(0.021)	(0.022)
Treat × Victim		0.137**	0.174***		0.016	0.010
		(0.054)	(0.058)		(0.014)	(0.016)
Treat × Bystander		0.052	0.073		−0.001	−0.002
		(0.056)	(0.055)		(0.007)	(0.008)
N		10,189	8,499		10,210	8,509
Control Mean						
Bully		−0.499	−0.505		0.206	0.202
Victim		−0.183	−0.193		0.110	0.108
Bystander		0.227	0.211		0.051	0.050
P-value						
Bully/ Bystander		0.39	0.18		0.06	0.14
Victim/ Bystander		0.11	0.09		0.26	0.48
Grade	7-9	8-9	8-9	7-9	8-9	8-9
Conditional			Endline			Endline

Notes: This table shows treatment effects (Equation 1) on the average grade and crime for different samples. Column 1–3: Average grade in compulsory subjects in grade 9 (standardized). Column 4–6: Indicator whether a crime was committed in adulthood. Regressions are estimated for three samples: grades 7–9, grades 8–9, and the survey sample (students in grades 8–9 who participated in the endline survey). All regressions include strata fixed effects and controls selected with PDS lasso. Columns 4–6 includes controls for grade fixed effects. Standard errors clustered at the school level in parentheses. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table E4: Robustness: Definition of Bullying

	Peer-reported		Peer- and Self-reported	
	Index	Indicator	Index	Indicator
	(1)	(2)	(3)	(4)
A. Main Estimates				
Treat	−0.037	−0.035*	−0.047	−0.044**
	(0.028)	(0.020)	(0.036)	(0.021)
Grade	8-9	8-9	8-9	8-9
N	10,010	10,010	10,017	10,017
Control mean	−0.108	0.233	−0.000	0.287
Adj. R^2	0.40	0.20	0.24	0.18
B. Heterogeneity: Gender				
Treat × Female	0.012	-0.011	0.008	-0.028
	(0.026)	(0.021)	(0.027)	(0.021)
Treat × Male	−0.089**	−0.061**	−0.107*	−0.062**
	(0.040)	(0.024)	(0.054)	(0.025)
Grade	8-9	8-9	8-9	8-9
N	10,010	10,010	10,017	10,017
Control Mean				
Female	−0.311	0.134	−0.219	0.189
Male	0.119	0.343	0.246	0.398
P-value	0.01	0.01	0.02	0.11
C. Heterogeneity: Role				
Treat × Bully	-0.026	-0.016	-0.022	-0.019
	(0.064)	(0.035)	(0.074)	(0.035)
Treat × Victim	−0.073**	−0.048	−0.071**	−0.053
	(0.034)	(0.032)	(0.034)	(0.032)
Treat × Bystander	−0.023	−0.033*	−0.031	−0.040**
	(0.022)	(0.017)	(0.024)	(0.017)
Grade	8-9	8-9	8-9	8-9
N	8,497	8,497	8,503	8,503
Control Mean				
Bully	0.523	0.548	0.684	0.598
Victim	0.052	0.321	0.171	0.380
Bystander	−0.340	0.113	−0.254	0.166
P-value				
Bully/ Bystander	0.97	0.62	0.91	0.51
Victim/ Bystander	0.11	0.57	0.23	0.64

Notes: This table shows treatment effects (Equation 1) on different specifications of the bullying variable. Column 1: Index for being peer-reported as a bully. Column 2: Indicator for being at or above the 75th percentile of the baseline distribution of the peer-reported bullying index in the control group. Column 3: Index for being self- or peer-reported as a bully. Column 4: Indicator for being at or above the 75th percentile of the baseline distribution of the self- and peer-reported bullying index in the control group. The indices are constructed over raw scales. All regressions include strata fixed effects, controls selected with PDS lasso, and the baseline control of the outcome variable. Standard errors clustered at the school level in parentheses. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table E5: Robustness: Definition of Victimization

	Peer-reported		Peer- and Self-reported	
	Index	Indicator	Index	Indicator
	(1)	(2)	(3)	(4)
A. Main Estimates				
Treat	−0.059	−0.030	−0.075	−0.043
	(0.040)	(0.023)	(0.047)	(0.026)
Grade	8-9	8-9	8-9	8-9
N	10,019	10,019	10,021	10,021
Control mean	−0.127	0.205	−0.000	0.278
Adj. R^2	0.30	0.13	0.20	0.10
B. Heterogeneity: Gender				
Treat × Female	-0.032	-0.017	-0.035	-0.034
	(0.037)	(0.022)	(0.045)	(0.027)
Treat × Male	−0.088*	−0.044*	−0.119**	−0.053*
	(0.048)	(0.025)	(0.058)	(0.030)
Grade	8-9	8-9	8-9	8-9
N	10,019	10,019	10,021	10,021
Control Mean				
Female	−0.205	0.176	−0.118	0.245
Male	−0.039	0.238	0.132	0.315
P-value	0.08	0.07	0.05	0.34
C. Heterogeneity: Role				
Treat × Bully	-0.081	-0.045	-0.118**	-0.056*
	(0.051)	(0.028)	(0.055)	(0.030)
Treat × Victim	−0.071	−0.039	−0.050	−0.011
	(0.072)	(0.037)	(0.087)	(0.039)
Treat × Bystander	−0.047	−0.020	−0.056	−0.045*
	(0.033)	(0.021)	(0.036)	(0.025)
Grade	8-9	8-9	8-9	8-9
N	8,506	8,506	8,507	8,507
Control Mean				
Bully	0.019	0.270	0.184	0.348
Victim	0.418	0.461	0.624	0.532
Bystander	−0.321	0.110	−0.233	0.183
P-value				
Bully/ Bystander	0.42	0.30	0.18	0.65
Victim/ Bystander	0.71	0.59	0.94	0.32

Notes: This table shows treatment effects (Equation 1) on different specifications of the victimization variable. Column 1: Index for being peer-reported as a bully. Column 2: Indicator for being at or above the 75th percentile of the baseline distribution of the peer-reported victimization index in the control group. Column 3: Index for being self- or peer-reported as a victim. Column 4: Indicator for being at or above the 75th percentile of the baseline distribution of the self- and peer-reported victimization index in the control group. The indices are constructed over raw scales. All regressions include strata fixed effects, controls selected with PDS lasso, and the baseline control of the outcome variable. Standard errors clustered at the school level in parentheses. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

F Documentation

Table F1: Survey Outcome Variables (Part I)

Variable	Survey Question	Coding
Harmful Behavior		
Perceived Bullying	"Has the amount of bullying changed in your class from the situation last autumn?" (Likert scale: 0 = Increased a lot, 4 = Decreased a lot)	$\mathbf{1}(\text{Value} < 3)$
Bullying Index	*Who in your class acts like this in bullying situations?* "Starts bullying" "Makes others join in the bullying" "Always finds new ways of harassing the victim." (Share of nominating peers)	Standardized index over raw values using GLS weighting
Victimization Index	*Who in your class acts like this in bullying situations?* "S/He is pushed and hit?" "S/He is called with nasty names or made fun of?" "S/he is usually talked about with bad tone?" (Share of nominating peers)	Standardized index over raw values using GLS weighting
Perceptions about Teachers		
Teacher Index	Index of Discussed Bullying, Fought Bullying, Opposes Bullying, Can Reduce Bullying (see below)	Standardized index over dummies using GLS weighting
Discussed Bullying	"Has the teacher touched the issue of bullying during any lesson since last autumn?" (Scale: 0 (0), 1 (1), 2 (2–5), 3 (5–8), 4 (8+ times))	$\mathbf{1}(\text{Value} > 1)$
Fought Bullying	"How much has the teacher done in order to decrease bullying since last autumn?" (Likert scale: 0 = Nothing, 4 = Very much)	$\mathbf{1}(\text{Value} > 1)$
Opposes Bullying	"How does your teacher think of bullying?" (Likert scale: 0 = Good thing, 4 = Absolutely wrong)	$\mathbf{1}(\text{Value} > 2)$
Can Reduce Bullying	"How much can the teacher do in order to decrease bullying?" (Likert scale: 0 = Nothing, 4 = Very much)	$\mathbf{1}(\text{Value} > 1)$
Skills Taught by KiVa		
Disapproval of Bullying	"It's okay to call some kids nasty names" "It is funny to see kids get upset when they are teased" "A bully is really a coward" "Kids who get picked on a lot usually deserve it" "I feel bad seeing a child bullied" "It is a wrong thing to join in bullying" "Kids who are weak are just asking for trouble" "Soft kids make me sick" "Nobody likes a wimp" "I like it when someone stands up for kids who are being bullied" "It is a good thing to help children who can't defend themselves" "It irritates me when nobody defends a bullied child" (Likert scale: 0 = Completely disagree, 4 = Completely agree). "The ending of bullying is to me ..." "The bullied person not being sad is to me ..." "Me being thought highly of is to me ..." "Nobody being bullied in my class is to me ..." "The bullied still enjoys staying in our class is to me ..." "The decrease of bullying is to me ..." "Me being known as a person who helps others is to me ..." "The bullied person feeling better is to me ..." "Me being liked by classmates is to me ..." (Likert scale: 0 = Not at all important, 3 = Very important)	Standardized index over raw values using GLS weighting
Empathy	"When the bullied pupil is sad, I also feel sad" "When the bullied pupil feels sad, I want to comfort him/her" "When the bullied pupil starts to cry, I also feel bad" "When someone is bullied, I start to get angry on his/her behalf" "I can understand how the bullied pupil must feel" "I can see how the bullied pupil is feeling bad" "I can imagine how the bullied pupil must feel, even if he/she would not tell" (Likert scale: 0 = Never, 3 = Always)	Standardized index using over raw values using GLS weighting
Efficacy	*How easy or difficult it would be for you to act following ways?:* "Trying to get the others stop bullying would be for me ..." "Comforting the bullied person or encouraging him/her to report about the bullying to the teacher would be for me ..." "Asking others to stop bullying or saying that bullying is stupid would be for me ..." (Likert scale: 0 = Very easy, 3 = Very difficult). *How likely do you consider the consequences of the following?* For each of the actions: i) If you tried to stop bullying, ii) If you comforted the bullied person or told him/her to report the bullying to the teacher, iii) If you asked others to stop bullying or said bullying is stupid... "It would end or decrease bullying", "It would increase bullying", "It would make the bullied person feel better", "It would make the bullied person feel worse", "It would make the others think highly of you", "It would make you unpopular and you would be bullied" (Likert scale: 0 = Not at all likely, 3 = Very likely).	Standardized index over raw values using GLS weighting

This table provides an overview of the survey outcome measures (continues on the next page).

Table F1: Survey Outcome Variables (Part II)

Variable	Survey Question	Coding
Learning Environment		
School Climate	"There is a good atmosphere in my class" "Helping others is common in our class" "I am happy to be in my class" "I feel safe at school" "I am satisfied with the atmosphere of my school" "Also those pupils who are different from the others are accepted at school" "I am thought highly of at school" "I feel being accepted as I am at school" "My schooldays are usually nice" "I like going to school" "I am happy with going to school in general" (Likert scale: 0 = Totally disagree, 4 = Totally agree)	Standardized index over raw values using GLS weighting
Academic Self-Concept	"Learning brings me joy" "I want to know and learn many different things" "I am doing fine at school (in my opinion)" (Likert scale: 0 = Totally disagree, 4 = Totally agree)	Standardized index over raw values using GLS weighting
Socio-emotional Wellbeing		
Depression	"How was your mood?" "How do you feel about the future?" "How do you feel about your life?" "How satisfied or dissatisfied do you feel about yourself?" "How do you see yourself?" "Do you feel senses of disappointment?" "How do feel about your being and appearance?" (Likert scale: 0 = Sunny & good, 4 = So depressed and melancholic that I cannot stand)	Standardized index over raw values using GLS weighting
Anxiety	"I'm worried about what the others think of me" "I'm afraid the others won't like me" "I'm worried about what the others talk about me" "I'm worried that the others don't like me" "If I have to argue about something, I'm afraid that the other won't like me" "I stay quiet when I'm in a group of people" "I'm afraid of asking others to do things with me as they might turn me down" "I feel quite shy even among those mates I know well" "It's difficult for me to ask others to do things with me" (Likert scale: 0 = Not at all, 4 = All the time)	Standardized index over raw values using GLS weighting
Self-Esteem	*How do you feel about yourself among peers? When I am with them ...* "I am more or less satisfied with myself" "I feel I am not good enough for anything" "I feel that I have a number of good qualities" "I feel I do things as well as the others" "I don't feel I have much to be proud of" "I sometimes feel really useless" "I feel I am as valuable (as a person) as the others" "I hope I could respect myself more" "I consider myself a failure" "I have positive thoughts of myself" (Likert scale: 0 = Not true at all, 4 = Exactly true)	Standardized index over raw values using GLS weighting
Social Network		
Friendship	"I have good friends in my classroom" "I have mates/friends in my own class" "I feel it easy to get along with my classmates" (Likert scale: 0 = Totally disagree, 4 = Totally agree)	Standardized index over raw values using GLS weighting
Perception of Peers	*How do you consider your mates of the same age? When responding don't think of your best friends only, but tell us your impression in general. They...* "Can really be relied on" "Really care about what happens to me" "Are there for me whenever I need help" "Shouldn't be trusted too much" "Don't really care about me" "Only think about their own interest" "Betray one's trust whenever they get the chance" "Want to hurt me" "Can be confided in" "Are honest with me" "Think bad things about me" "Usually have good intentions" "Are hostile" (Likert scale: 0 = Not true at all, 4 = Exactly true)	Standardized index over raw values using GLS weighting

This table provides an overview of the survey outcome measures (continued from previous page).

G Marginal Value of Public Funds Calculation

G.1 Overview of the MVPF Framework

We compute the Marginal Value of Public Funds (MVPF) of the KiVa anti-bullying intervention, following the framework of Hendren and Sprung-Keyser (2020). The MVPF is defined as the ratio of the aggregate willingness to pay (WTP) of the policy's beneficiaries to the net cost of the policy to the government:

$$\text{MVPF} = \frac{\text{WTP}}{G}, \tag{3}$$

where G denotes the net present value of all fiscal costs net of fiscal offsets. A policy with a negative net cost and a positive WTP has MVPF = ∞, meaning it more than pays for itself from the government's perspective while generating positive value for recipients.

We restrict the WTP calculation to the effect of the intervention on post-tax labor income, following the envelope-theorem argument in Hendren and Sprung-Keyser (2020): under the assumption that income gains arise from a genuine increase in human capital (rather than costly additional effort), the present value of after-tax income changes constitutes a valid first-order estimate of willingness to pay. All monetary values are expressed in 2008 EUR, the price level at the time of the intervention. We apply a discount rate of $r = 3\%$, following Hendren and Sprung-Keyser (2020), and a constant marginal tax rate of $\tau = 30\%$ (the average tax rate in 2024 for an average-wage worker without children in Finland, see OECD, 2025). Present values are discounted to age 13, the earliest age at which the program we study was delivered (grades 7–9).

G.2 Willingness to Pay

The WTP is the present value of the stream of after-tax income gains induced by the intervention:

$$\text{WTP} = \sum_{t=16}^{65} \frac{(1-\tau)\,\Delta y_t}{(1+r)^{t-13}}, \tag{4}$$

where Δy_t is the estimated income gain at age t (in 2008 EUR).

For ages 16–29 we use empirical estimates of Δy_t drawn directly from our estimates of the program's effect on earnings at each age. Note that these reflect slightly lower earnings for students in their early 20s, as university attendance on average delays labor market entry. From age 30 onward, we assume a constant annual income gain equal to the average estimated effect at ages 27-29. The no-effect assumption prior to age 16 reflects the expectation that the labor market returns to the intervention accrue only once participants enter employment.

We do not include utility gains from reductions in bullying or crime, or from improvements in mental health, even though these are likely an important channel through which the program operates. Including them would increase the WTP estimate. We also abstract from any general equilibrium effects on wages.

G.3 Direct Program Costs

Upfront program costs are computed on a per-student basis using information on the costs of implementing KiVa reported by Persson et al. (2018) and Bowes et al. (2024). All cost components are expressed per exposed student, assuming 200 students per school and 18 students per class.

The costs of the program include a one-off training fee for school staff (EUR 2,943), a license fee (EUR 150), and materials consisting of teacher manuals, parental guides, and posters and vests (EUR 2,084 in total per school of 200 students, equivalent to EUR 25.89 per student).

Moreover, we use information on indirect costs of the program. The opportunity cost of classroom time devoted to the program is proxied by the teacher's salary for those hours. We assume 23 lesson hours per class per year and an hourly teacher wage of EUR 47, sourced from OECD (2010, Table D3.1). With 18 students per class, the resulting per-student cost is EUR 60.06. We also take into account the time cost of training for KiVa teachers: a two-day (16 hours) training course attended by three KiVa-designated teachers per school. At the same hourly wage, the total training cost per school is EUR 2,256, or EUR 11.28 per student.

G.4 Other Government Costs: University Attendance

Because the KiVa intervention increases the probability that treated students go on to attend university, we include the additional public expenditure on higher education as a government cost. We use annual public expenditure per tertiary student in Finland of EUR 18,435 (in 2020 EUR, including R&D), drawn from OECD (2024). We assume that university attendance corresponds to a five-year degree (bachelor's plus master's), which is the standard trajectory in Finland and yields a more conservative estimate. We apply an estimated increase in the probability of university attendance of 3.9 ppt attributable to the intervention (see Table C1). This gives an expected additional government expenditure of EUR 3,039 per treated student (in 2008 EUR).

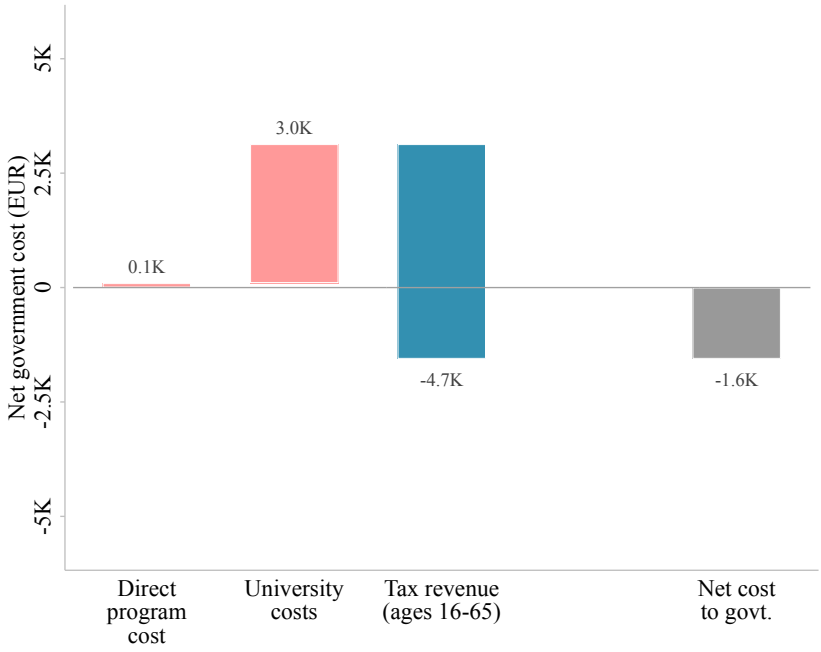
G.5 Tax Revenue Offset

The government recoups its upfront costs through increased income tax revenue generated by the intervention's positive effect on earnings. We apply the same age-income profile described in Subsection G.2 and a flat marginal tax rate of $\tau = 30\%$. The present value of additional tax revenue is EUR 4,694 per student (in 2008 EUR).

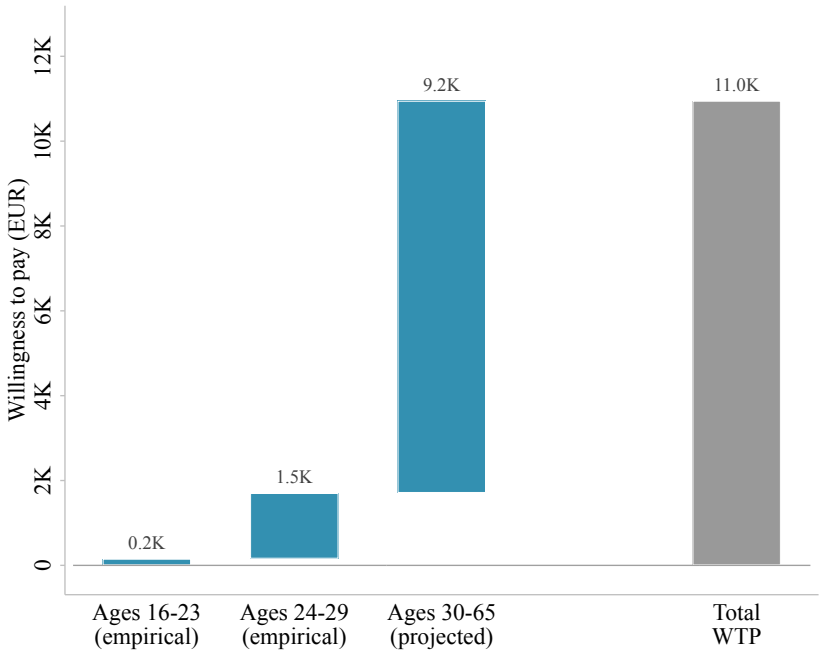
G.6 Summary Results

Figure G1 reports the present-value totals for each component of the MVPF calculation. The net government cost is negative (EUR −1,558), meaning that even before accounting for any additional fiscal benefits (such as reduced crime or lower healthcare costs), the intervention more than pays for itself through the increased income tax revenue it generates. The MVPF is therefore infinite. Importantly, this conclusion does not depend on the assumptions we make to calculate WTP, as they would hold with any positive WTP estimate.

Figure G1: KiVa: Marginal Value of Public Funds



(a) Government Cost Decomposition



(b) WTP Decomposition

Notes: This figure decomposes the Marginal Value of Public Funds (MVPF) for the KiVa program following Hendren and Sprung-Keyser (2020). Panel (a) shows the government cost decomposition: direct program costs, university costs induced by higher educational attainment, and tax revenue from increased earnings, yielding a negative net cost to government. Panel (b) shows the willingness to pay decomposition by age group, measured as the present value of post-tax income gains. All values are in 2008 EUR, discounted to age 13 at $r = 0.03$ with $\tau = 0.30$.