Non-Toxic Classrooms: 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*

July 1, 2026

Abstract

We study the long-run impacts of the KiVa program, a randomized anti-bullying intervention in Finnish schools, linking RCT survey data for 15,000 pupils attending grades 7-9 to comprehensive administrative records. Treated students have better long-run outcomes: they are more likely to attend academic high school and graduate from university, and they earn higher wages by ages 27–29. These gains accrue to all groups of pupils, irrespective of gender or social role as a bully, victim, or bystander. We trace these effects to a reduction in bullying in the classroom, particularly among boys, which leads to a more positive learning environment for all students. The reduction in harmful behavior persists beyond the intervention window: treated boys are significantly less likely to engage in criminal activity in adulthood.

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 (INVEST flagship), email: christina.salmivalli@utu.fi. We would like to thank Amanda Agan, Lorenzo Casaburi, David Figlio, Hannes Schwandt, Elizabeth Setren, Roberto Weber, Ulf Zölitz, and Josef Zweimüller for helpful comments and suggestions. We thank Tiina Turunen for her support in linking the KiVa data with administrative records. We thank participants at NBER Children and Families Spring Meeting, UAM, ifo Institute for Education, JFest, CSIC-IAE, Uppsala University, Helsinki GSE, Tufts University, University of Zurich, UAB, CESifo Labor, CESifo Venice Summer Institute, and Essen Health Conference 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. The study was registered at the AEA RCT registry, RCT ID 0010879.

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; 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 outcomes spanning education, the labor market, and criminal activity, and examine the consequences of harmful peer behavior not just for 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 pupils' trajectories 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 mechanism behind these results is a reduction in bullying in the classroom. Decreases in harmful behavior are concentrated among boys, who at baseline are far more likely than girls to bully. 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 regardless of gender or social role as a bully, victim, or bystander. 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 speak to a broader question of considerable social importance: whether adult criminality can be traced to factors that are malleable during schooling. Our results identify bullying as one such factor and show that engaging in harmful behavior is not a fixed trait. Taken together, our findings point to addressing bullying in schools as a scalable lever for reducing the long-run social costs of toxic behavior.

The KiVa anti-bullying program was developed as a school-based intervention by a team of psychologists in Finland. It is based on the premise that bullying takes place in a group context and that successful mitigation requires active involvement of bystanders (Salmivalli, Kärnä, and Poskiparta, 2009). The program consists of two main components: about 20 age-appropriate classroom lessons delivered 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 between 2007 and 2009 as a large-scale RCT with a school-

level randomization design. The program was subsequently rolled out across all compulsory schools in Finland and has been adopted in 28 countries around the world.

In this paper, we leverage KiVa's original RCT design to study its long-term impact until early adulthood. Our main analysis sample comprises around 15,000 pupils in grades 7-9 (aged 13-15) 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 start by examining descriptive patterns of bullying and victimization at baseline, drawing on rich survey data from the RCT that collected both self-reports and peer nominations for social roles based on classroom rosters. The most frequent self-reported categories of bullying are being called nasty names or being bullied sexually (with sexual names, signs, or acts), being the object of lies or gossip, and being physically harassed. To examine the correlates of different social roles, we classify pupils as bullies or victims if they score at or above the 75th percentile on an index of peer nominations for each social role. Gender is the strongest predictor of being a bully, with boys substantially more likely to be nominated as bullies. Boys are also more likely to be a victim, but gender differences in victimization are much less pronounced. While bullies tend to be from households that are worse off economically, victims are more likely to be a single child. Turning to later-life outcomes using only the control group, 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 is correlated with higher future earnings, but also with a higher probability of being charged with a crime in adulthood.

We then turn to examine the long-term causal impact of reducing bullying via the KiVa program. We start with pupils' educational trajectory at age 16, when compulsory schooling ends. In Finland, this is the first consequential educational juncture and a strong determinant of subsequent academic and labor market trajectories (Silliman and Virtanen, 2022): pupils choose whether to apply to an academic high school or a vocational track for upper secondary school, with admission decisions based on middle-school GPA. While we do not find treatment effects on being enrolled in upper secondary education at the extensive margin, treated students are 5 percentage points (11% over the control mean) more likely to attend an academic high school track rather than vocational training. Subsequently, treated students are 4 ppt (9%) 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 upper-secondary level. These gains in academic attainment persist: by ages 27-29, treated students are 4 ppt (9%) more likely to hold a university degree.

Do these gains in educational attainment translate into returns in the labor market? We measure participation and earnings outcomes in 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 (1.4%) 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 employment.

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 these long-term gains, we leverage rich survey data from the original RCT for grades 8-9, together with administrative records on applications to upper secondary schools and GPA for the full sample. We first document the program's direct impacts on bullying: treated students are 3.8 ppt more likely to report a perceived reduction in bullying, and peer-reported nominations as a bully decline—with the latter decrease concentrated among boys. We then turn to how the program achieved these reductions. On the teacher side, staff in treated schools are more likely to take action against bullying. On the pupil side, KiVa is designed to strengthen three main levers: anti-bullying attitudes, self-efficacy to intervene in bullying situations, and empathy towards victims. While we observe no meaningful treatment effects on attitudes, boys experience gains in both self-efficacy (.10 SD) and empathy (.07 SD), mirroring the gendered pattern in the behavior response.

Although the program's direct impacts are concentrated among boys—a group more likely to engage in bullying and with lower baseline empathy and self-efficacy—these changes translate into improvements in the learning environment and higher well-being for all groups of pupils. All types of students report a better school and classroom climate and higher academic motivation, capturing dimensions such as the joy and curiosity of learning. Treated students score higher on a self-esteem index (.04 SD) and lower on the Beck, Steer, and Brown (1996) depression scale (-.06 SD). Using register data on applications for upper-secondary schools, we show that treated students also become more ambitious in their application behavior: they are more likely to apply to and obtain a slot (6.1 ppt) in the academic track. While we observe moderately sized (.05 SD) but noisily estimated increases in academic grades, a back-of-the-envelope calculation suggests that the increased academic-track enrollment is not driven by grade improvements alone, consistent with a general shift in pupils' academic interest and ambition. In contrast, the intervention does not appear to operate through changes in the classroom social network: we observe no treatment impacts on how pupils perceive their peers and friendships within the class. Rather, these patterns are consistent with all types of pupils having a more positive learning experience as the program successfully reduced toxic peer interactions, fostering a continued ambition to proceed along the academic track.

As a last step, we examine whether short-term reductions in harmful behavior persist over the long run. Using records on the universe of crimes charged in district courts, we document that boys in the treatment group are significantly less likely to be charged with a crime in adulthood. These patterns indicate that KiVa reduced negative social externalities beyond the immediate time window of the intervention, suggesting that adult criminality can in part be

traced to factors that are malleable during schooling.

We contribute to three main strands of literature. To the best of our knowledge, we provide the first causal evidence on the long-run consequences of bullying. While research in this area is abundant in sociology, psychology, and education sciences, evidence that is causal and based on objective, long-term outcomes remains scarce. Prior evaluations of KiVa (Kärnä et al., 2011*a*; Kärnä et al., 2013; Huitsing et al., 2020),[1] other anti-bullying programs (Cunha et al., 2023; Olweus and Limber, 2010; Jiménez-Barbero et al., 2016), and pandemic school closures (Bacher-Hicks et al., 2022; Werner and Woessmann, 2023) document short-term reductions in self-reported bullying and improvements in well-being. Linking the KiVa experiment to administrative records, we show that leaving bullying unaddressed has adverse consequences well into adulthood for educational attainment, earnings, and criminal activity.

Second, a large literature uses cross-cohort variation in classroom composition to identify the causal effects of disruptive peers on their non-disruptive classmates, documenting negative effects both in the short run (Lavy and Yancu, 2025; Kristoffersen et al., 2015; Eriksen, Nielsen, and Simonsen, 2014; Lavy, Paserman, and Schlosser, 2012; Lavy and Schlosser, 2011; Carrell and Hoekstra, 2010; Gould, Lavy, and Paserman, 2009; Figlio, 2007) and over the long run (Carrell, Hoekstra, and Kuka, 2018). A separate question—and the one we address—is whether harmful behavior itself can be changed, and who benefits when it is. Leveraging random assignment of the KiVa intervention, we provide direct evidence that the program improves long-run outcomes for bullies, victims, and bystanders alike, and identify the mechanisms through which it operates.

Third, we contribute to a growing literature evaluating programs that target non-cognitive skills and behavior in school-age children and youth. Recent work documents that such interventions can improve short-term academic outcomes and may reduce violent behavior and social isolation (Alan and Ertac, 2018; Alan, Boneva, and Ertac, 2019; Alan et al., 2021; Rege et al., 2025; Alan, Carlana, and Leone, forthcoming), with some evidence of long-run returns from interventions that teach general socioemotional skills (Algan et al., 2022; Sorrenti et al., 2025). Work that applies non-cognitive tools to reduce crime and recidivism among youth shows meaningful impacts for high-risk populations (Heller et al., 2017; Shem-Tov, Raphael, and Skog, 2024; Adukia, Feigenberg, and Momeni, 2025). We complement this evidence by evaluating a program with a sharply defined target—bullying—that has been implemented at scale in 28 countries, using a sample an order of magnitude larger than prior experimental studies. Our findings show that reducing this type of behavior among middle school boys lowers harmful acts in adulthood while delivering gains for the entire class.

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 and Section 4 presents the experimental design. Section 5 offers descriptive statistics on bullying and victimization and Section 6 presents the main results. Section 7 discusses mechanisms, Section 8 shows results on

[1] We provide an overview of previous research on the impacts of KiVa in Section 2. This work has exclusively focused on short-term, survey-based outcomes. In contrast, this paper links participants to administrative data and provides the first evidence on the long-run effects of KiVa on education and labor market trajectories as well as criminal behavior. Regarding short-term outcomes, we complement prior studies by pursuing an econometrically rigorous and more comprehensive approach that includes administrative data.

criminal behavior, Section 9 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: Kiusaamista Vastaan - Against Bullying) was launched as an RCT in 2006 by the Finnish Ministry of Education and Culture and a team of psychologists at the University of Turku to tackle bullying in Finnish schools through a common curriculum (Salmivalli, Kärnä, and Poskiparta, 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, Kärnä, and Poskiparta, 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 "Indicated Actions" target children involved in bullying incidents.

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*b*). 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.

"Indicated 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*b*; Kärnä et al., 2013).

2.2 Roll-out of KiVa and Prior Research

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 (primary school, ages 7–9) and grades 7–9 (secondary

school, ages 13–15). Between 2009 and 2011, KiVa was rolled out in all Finnish comprehensive schools, with the KiVa curriculum targeted to grades 1, 4, and 7 (Kärnä et al., 2011*b*). By 2011, 90 percent of all comprehensive schools were registered as KiVa users (Salmivalli and Poskiparta, 2012). Due to the fast national scale-up exposing almost all Finnish students to the program, in this paper we focus on the largest part of the RCT that was implemented among middle school students (grades 7–9), as its control group provides a clean counterfactual that was not directly exposed to the treatment (see Section 4 for additional details).

Previous evaluations of the KiVa RCT have examined its impact on survey measures collected immediately at the end of the academic year. These studies report reductions in bullying and victimization, with the strongest effects observed among fourth graders and the weakest among students in grades 7–9 (Kärnä et al., 2011*a*; Kärnä et al., 2013; Salmivalli and Poskiparta, 2012).[2] For grades 4–6, the program also improved subjective well-being, including a more positive attitude towards school (Kärnä et al., 2011*b*; Salmivalli, Garandeau, and Veenstra, 2012).

KiVa has been implemented across a total of 28 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

Appendix Figure A2 summarizes the Finnish education system. During the period of our study, compulsory education in Finland starts at age 7 and spans grades 1 through 9 in comprehensive school. Pupils start in primary school for the first 6 years, and subsequently attend lower-secondary education (hereafter: middle school) for 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 ages 18 or 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. Admission to both upper-secondary tracks depends on grade point averages (GPA) from the comprehensive-school leaving certificate, issued at the end of grade 9. While all subjects count towards admission to a vocational program, admission to academic high schools is usually only based on the grades from theoretical subjects such as languages and natural sciences. Students submit ranked applications through a centralized system during grade 9, and are placed in the highest-preferred school by GPA via deferred acceptance (Huttunen et al., 2023).

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

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 demographic information is the Finnish Population Register (Statistics Finland, 2023*a*), 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 KiVa implementation).[3] Regarding outcomes, 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 educational 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 university 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 middle school and applications to upper-secondary schools, we use data from the National Register of Applications to Upper Secondary Education (Statistics Finland, 2022). 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 second 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. We also study changes in pupils' application behavior: the number of submitted applications, the share of applications to academic programs, the probability of not applying

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

to any academic option, the highest-ranked offer the student receives, and the probability of being offered a slot in an academic high school. Finally, we use the register on Prosecutions, Sentences, and Punishments (Statistics Finland, 2023*b*) that records crimes charged in Finnish district courts (courts of first instance) up to the year 2023. We construct an indicator for being charged with any crime in adulthood, from age 18 as the age of reaching adult criminal liability 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 (end of academic year before the intervention), December 2008–February 2009, and May 2009 (endline survey). Baseline data were 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 Appendix Table G1 provides further details on the 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 collects measures of pupils' perceived changes in bullying in the classroom, and peer-reported bullying and victimization measures. For the peer-reported measures, students are asked to nominate classmates that bully others or are victimized. Students see a list of their classmates' names and mark any peer whose behavior matches the different questions. For bullies, the questions ask if a peer: "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.[4]

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:

[4]The survey also collected self-reported measures of bullying and victimization following Olweus' Bully/Victim Questionnaire (Olweus, 1996). Appendix Figure B1 shows that self- and peer-reported measures are correlated. We discuss the comparison of these two measures in Section 5.

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 motivation. 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 motivation 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, Steer, and Brown, 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).

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 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. Since the program was scaled up nationally starting in the fall of 2009, this ensures a clean control group that was not exposed or only minimally exposed to the program.[5] Due to the national roll-out, the control group for lower grades was eventually treated.

For recruitment, the KiVa team sent letters to all comprehensive schools in mainland Finland in the fall of 2006. These included both Finnish-language and Swedish-language schools. The

[5]Students in the control group during grade 9 were never exposed to the intervention. Some of the students in grades 7 and 8 in the control group could later have been exposed to some of the light-touch elements of the intervention, such as the school posters, once their school adopted the program as part of the national roll-out. Since in the scaled-up program, KiVa lessons are only provided in grades 1, 4, and 7, these students were never exposed to these more intensive elements of the intervention.

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 treatment and 35 control schools.[6]

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

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

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_j are strata fixed effects for the province and the language of instruction of the school. 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, Chernozhukov, and Hansen, 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 examining outcomes that we observe at different ages for the different cohorts (such as earnings) we include cohort fixed effects based on students' grade level during KiVa. Where available (i.e., when examining 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).[8]

4.1 Sample and Balance

For our main analysis, our sample consists of all students in the RCT for grades 7–9 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.[9]

[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. We present difference-in-differences estimates across cohorts within school in Section 6.3 to address potential concerns with sample selection of schools and students.

[7] Full list of variables in Panel A in Table 1. The lasso-chosen controls for each model are shown in Appendix Table G2.

[8] Before obtaining the administrative data, we pre-registered our main specification and primary outcomes in a pre-analysis plan at the AEA RCT registry, RCT ID 0010879. The pre-analysis plan is accessible here: https://anacostaramon.github.io/kiva/KiVa_PAP.pdf. Due to privacy concerns, we have been unable to get access to: i) health data that we were planning to use for studying both the contemporaneous and long-term health impacts of the program, ii) family identifiers to directly link participants to their family members and their characteristics. The main results in this paper therefore include pre-registered education trajectories, labor market outcomes, and criminal activity for grades 7–9 for whom we have a clean counterfactual for long-term outcomes.

[9] 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, labor market outcomes, and criminal activity for over

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 pupils with an immigrant background is quite low (2%). The majority live in the same region in which they were born (89%) and in areas that are predominantly urban (75%). While very few are young for their grade, about 5% are old relative to their peers.[10] 14% of pupils are a single child, and 17% live in a single-parent household. About a third of pupils 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 everyone 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.[11] 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 measured with survey data. This results in a "Survey Sample" of 8,861 students from grades 8 and 9 who participated in the endline.[12]

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.

5 Descriptive Patterns on Bullying

To shed light on the types and prevalence of bullying, we use the RCT survey data at baseline (available for grades 8 and 9). Appendix Figure B2, 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*b*). Around 10% of pupils are 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 with sexual names, signs or acts (boys: 10%, girls: 5%), called nasty names (boys: 10%, girls: 7%), and substantially more likely to be harassed physically (boys: 4%, girls: 1%). Girls report roughly similar rates in experiencing

97% of the sample.

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

[11] 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.

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

social exclusion (boys: 4%, girls: 5%) and 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 for details). Appendix Figure B3 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 in Panel (b): 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 periodicity in the survey. Appendix Figure B1 examines 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). For each social role, we define an indicator variable 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. The median share of nominating peers is 12.5% for bullies and 15.4% for victims.[13]

Appendix Figure B5, Panel (a) shows correlates of social role by regressing each standardized baseline characteristic on these indicators. 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. Appendix 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

[13] Appendix 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.

by attending academic high school, taking the university entrance exam, or graduating from university (-.43 – -.33 SD). These patterns are 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 being charged with a crime in adulthood 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 is nominated for a particular role at endline. 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.

6 Long-Term Impacts of KiVa

6.1 Educational Attainment

We examine KiVa's 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 students 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) and Appendix Table C1. 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 increase of 11% relative to the control mean.

Track choices at age 16 in Finland send students onto different—and, in practice, relatively sticky—educational paths (Silliman and Virtanen, 2022; Schaede and Mankki, 2026). Consistent with this, 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 track take the matriculation exam that qualifies for university entry. 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, i.e. at least a Bachelor's degree. These findings suggest that exposure to the KiVa program effectively shifts students' educational trajectories towards the academic track after finishing compulsory education, 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 when students are ages 27–29 and add cohort 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. Figure 1 Panel (b) and Appendix Table C2 report 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.

Figure 1 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 Robustness Main Results

While our main results are based on a large number of students whose baseline characteristics are balanced across treatment and control, there are two sources of drop-out that may raise some concern about selection bias in our setting. First, five schools whose identity we do not observe dropped out between randomization and data collection. Second, our main analysis sample contains only those students for whom we are able to establish a link from the RCT rosters to the administrative data.

In order to address such concerns, we implement a difference-in-differences design that compares the outcomes of all students that are in a KiVa cohort to their counterparts in the same school who graduated up to four years earlier across treatment and control schools. For this robustness check, we use data for all students in a cohort based on register data (Statistics Finland, 2022), irrespective of whether they enter our main analysis sample. Due to data privacy restrictions, we are, however, only able to use these records collapsed at the school-by-cohort level and focus on two main outcomes: whether pupils attend the academic track at age 16, the decisive step that sets students onto different educational trajectories, and earnings measured at age 29. The value of this exercise lies in the fact that it rests on a different identifying assumption: rather than relying on randomization holding across the surviving schools and students observed in the analysis sample, it requires that – absent KiVa – treated-school cohorts' outcomes would have evolved in parallel to those of the control schools.

We estimate event studies of the following form:

$$\bar{Y}_{sc} = \alpha_s + \delta_c + \sum_{c \neq 2008} \beta_c \left(\text{Treat}_s \times \mathbf{1}\{\text{cohort} = c\} \right) + \nu_{sc}, \tag{2}$$

with $\bar{Y}_{sc}$ the average outcome across students at school s in cohort c, where c indicates the year of graduation from middle school. Treat_s is an indicator equal to one if a school received the KiVa treatment, which we interact with cohort indicators. Each β_c measures the treatment–control difference in cohort c relative to the cohort graduating in 2008 (the last graduation cohort before the KiVa RCT). α_s and δ_c are school and graduation cohort fixed effects, respectively. Standard errors are clustered at the school level and observations are weighted by the number of students in each school-by-cohort cell.

Panels (a) and (c) in Appendix Figure F1 show the raw data at the cohort level for the four cohorts that graduated from the middle schools in our sample prior to the KiVa RCT (2005-2008) and for the KiVa cohorts, who graduated starting in the year 2009 and thereafter.[14] The upper panel shows the fraction of pupils attending the academic track at age 16, while the bottom panel shows earnings measured at age 29. Outcomes for the cohorts that graduated before KiVa are similar in levels and move in parallel across treatment and control schools, while a visible gap opens up for cohorts graduating after the intervention was implemented in treated schools.

Event study estimates are displayed in Panels (b) and (d) in Appendix Figure F1. We observe no meaningful pre-treatment differences in trends. For the cohorts that received the KiVa program, we observe increases in pupils' likelihood of attending the academic track, as well as higher earnings. While some individual point estimates may lack precision, a pooled difference-in-differences estimate across all treated cohorts displayed in the right corner clearly rejects the null hypothesis that the treatment had no impact. Moreover, the pooled treatment effects are consistent in sign and magnitude with our RCT estimates based on individual level data: treated cohorts are 4 ppt more likely to attend academic high school, and earn 1,479 EUR more by age 29. Taken together, these results should alleviate concerns that our main results are driven by selection of schools or individuals into our main analysis sample.

6.4 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 5), 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.

[14] Students who received the KiVa RCT in grade 9 graduate in 2009, those in grade 8 in 2010 and those in grade 7 in 2011. For this exercise, we measure earnings at age 29 to get a more stable snapshot of income when students are established in the labor market, and we thus observe this outcome only for two cohorts.

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 5). 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.[15] 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 F1). As documented in Section 5, 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.

7 Mechanisms

The results presented in Section 6 show that the KiVa program has a long-lasting impact on students' outcomes, increasing educational attainment and earnings in early adulthood. In order to understand the mechanisms through which the intervention may have shifted these trajectories, we exploit the rich survey data collected during the RCT in addition to administrative records. In order to address concerns with selective attrition when using the Survey Sample based on grades 8–9 (see details in Section 4.1), 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.[16]

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 measure 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 control mean) less likely to agree that

[15] 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 treated victims in their late twenties, we think that these differences are likely explained by longer engagement with education among victims in the treatment group.

[16] Appendix Table F2 further shows similar treatment effects on our main outcomes for the Survey Sample.

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 whether pupils were nominated as a bully or victim by their peers, using the indices for peer-reported bullying and victimization as an outcome.[17] While, on average, point estimates are negative but noisy, we observe a significant decrease both in peer-reported bullying and victimization for boys. Appendix Tables F3 and F4 report point estimates and treatment effects for the dichotomized version of this variable (columns 1 and 2), and robustness when including self-reports (columns 3 and 4). The reduction in bullying for boys is .09 SD (Appendix Table F3, column 1, Panel B) which accounts for closing about 20% of the baseline gender gap in bullying (.43 SD, see Appendix Table F3). Boys are 6.1 ppt less likely to be nominated as a bully, a 18% decrease over the control mean (Appendix Table F3, column 2, Panel B). This is accompanied by a decrease in nominations as a victim among boys of .09 SD on the index or 4.4 ppt (18% over the mean) on the dichotomized version (Appendix Table F4, columns 1 and 2, Panel B). For girls, while point estimates across bullying and victimization measures are negative, they are not significantly different from zero. Taken together, these patterns suggest that KiVa was effective in addressing problematic interactions by reducing harmful behavior among boys—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 how 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 Appendix 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 (115% over control group mean) increase in the teacher having discussed bullying at least twice during the academic year. Treated students are also about 19 ppt (46%) more likely to report that their teacher has taken actions in order to decrease bullying ("Fought Bullying"). Treated students further perceive their teachers as somewhat more capable to do something against bullying, with a 7 ppt (11%) increase in perceiving that their teacher opposes bullying, and a 5 ppt (10%) increase in perceiving that their teacher has at least some influence to reduce 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 taking 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*b*): i) fostering students' anti-bullying attitudes, ii) promoting empathy towards victims, and iii) enhancing beliefs about their own effectiveness to intervene against bullying (self-efficacy).

[17] 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.

Appendix Figure D2 shows treatment impacts on each of these three targeted dimensions. The top bar of the figure shows effects on an index of anti-bullying attitudes. Reported disapproval of bullying is already high at baseline, with 77% of students saying that it is not OK to call kids nasty names and 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 bar shows treatment impacts on an index of empathy towards the victim. Treated boys score .07 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 bar presents effects on an index measuring students' perceived self-efficacy in fighting bullying. We again see a significant increase in perceived efficacy among boys (.10 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. These changes in skills and beliefs could also partly account for the beneficial effects of the program on boys' long-term outcomes. However, as discussed in Section 6.4, 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 all groups of students in the classroom.

7.3 Indirect Impacts of the Program

Learning environment and socio-emotional wellbeing — We start by studying the impact of the program on students' perceptions of their learning environment and well-being in Appendix Figure D3 and Appendix Table D4. The first bar in Appendix Figure D3 and column 1 in Appendix Table D4 show 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 .06 SD 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 .05 SD on an index of academic motivation that elicits students' joy and curiosity for learning (second bar in the figure; column 3 in the table). Improvements in the learning environment go hand in hand with improvements in students' own well-being. While we do not observe statistically significant reductions in a (social) anxiety index (third bar; column 5), we see a significant decrease of .06 SD in depressive symptoms, as measured by the Beck, Steer, and Brown (1996)'s depression scale (fourth bar; column 7). This increase in students' well-being is also visible in their self-esteem in relation to others: the last bar of Appendix Figure D3 and column 9 in Appendix Table D4 show that treated students score higher on an index measuring their self-concept among their peers (Rosenberg, 1965). These gains materialize for all groups

of pupils, irrespective of their gender or social role at baseline (Appendix Table D4, Panels B and C).

Academic achievement and ambition — Do these improvements in school climate translate into higher academic achievement and ambition? To examine this, we leverage register data on grades and application choices for pupils' applications to upper-secondary education options after middle school. Appendix Table D5, columns 1–3, shows treatment effects on teacher-assigned, standardized grades in grade 9 (the last year of middle 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, languages: .08 SD, math: .05 SD); however, none of the coefficients are statistically different from zero.[18] Columns 4–8 report results on students application behavior in this centralized application system, in which pupils rank up to five upper secondary education choices. While treated students on average submit the same number of applications (column 4), they become academically more ambitious: a higher share of their applications goes to the academic track (+ 6.2 ppt, column 5), and they are less likely to submit a ranked list that contains no academic option among their choices (column 6). In terms of getting an offer, column 7 shows that treated students do not obtain a higher ranked choice relative to the control group, but are 6.1 ppt more likely to be offered a slot for an academic track (column 8). While moderately higher grades may be able to account for about a third of the impact on attending the academic track documented in Section 6.1, the results on pupils' ranked choice lists suggest that the decision to apply to academic high school in the first place is an important component in explaining these results.

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 Table D7 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 of Mechanism

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 and well-being, and more curiosity and joy in learning. This translates into treated students being more likely to apply to and receive a slot for the academic track after middle school. Appendix Tables D4 and D5 show that these gains are visible for all groups of students, irrespective of

[18] Taking those point estimates at face value, higher grades can explain roughly one third of the observed treatment effect on attending academic high school: in Appendix Table D6, 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 SD for our treated students could thus account for about a third of the observed treatment impact on attending the academic track.

gender or social role at baseline.[19] Taken together, these patterns are consistent with a Lazear (2001) model of education, where reducing harmful behavior by a few students mitigates negative externalities for the whole class.

8 Impacts on Criminal Behavior in Adulthood

How persistent are the behavioral changes induced by the program? While the education and labor market gains from KiVa accrue up to thirteen years after the program ended, they do not tell us whether the underlying reduction in harmful behavior is itself sustained over time. This question carries particular weight as bullying during middle school is strongly correlated with later criminal conduct (see Appendix Figure B5)—an observable form of harmful adult behavior with large social costs. We turn to administrative registers that record criminal charges in Finnish district courts to estimate treatment impacts on the probability of being charged with a crime as an adult.

Figure 3 and Appendix Table E1, column 1 present results for an indicator equal to one if a pupil is charged with any crime as an adult. Overall, 9.2% of pupils in the control group are charged, with 15.7% of young men and 3.5% of young women engaging in any criminal activity. Point estimates in Appendix Table E1 mimic our findings on bullying behavior during school years: while the reduction in the full sample is noisily estimated (Panel A), we observe an economically meaningful and statistically significant reduction in being charged with a crime among young men of 2.6 ppt or a decrease of 16.6% compared to the average crime rate of males in the control group (Panel B). As depicted in Panel C, these reductions are also concentrated among bullies, the group with the highest crime prevalence in the control group (21.0%). Appendix Table E1 columns 2–6 show results disaggregated by type of crime: for boys, we see similar-sized reductions in criminal activity across categories of traffic, violent, property, and drug crimes. Taken together, these results suggest that KiVa was successful in achieving persistent reductions in harmful behavior, thus decreasing negative social externalities beyond the time window of the intervention.

9 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:

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

[19] Appendix Tables D1-D4 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 F5, which compares the estimated treatment impacts on administrative outcomes—average grades, share of applications to the academic track, 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 all three outcomes, we cannot reject the null of equal effects across the two samples.

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 H for details on the calculations.

9.1 Government Costs

Appendix Figure H1, Panel (a) documents the net costs of KiVa for the government. We consider two scenarios for the cost of the program. The first takes an upper-bound view based on the total costs incurred by the Finnish government at the time of the intervention, covering program development, licenses, researcher involvement, and teacher training. These costs amounted to EUR 2.6 million. We distribute them across the 8,548 treated students in grades 7–9, yielding a per-student cost of EUR 304.16. This is a conservative scenario since the RCT was also implemented separately in the lower comprehensive-school grades (1–6); spreading the same fixed costs across all treated students would lower the per-student figure.

The second scenario instead uses the current marginal cost of delivering KiVa per student, as estimated in the broader literature on its implementation. This cost is EUR 97.12 per student and covers direct fees for materials, teacher training (direct and opportunity costs), and class time spent delivering KiVa lessons.[20] This figure is arguably the more policy-relevant input for any government considering whether to adopt the program today, since it reflects the marginal cost a replicating implementation would actually face.

Because the KiVa intervention increases the probability that treated students will graduate from 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, derived from OECD, 2024) and apply our estimated increase in the probability of university graduation (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 a tax rate of 30% (OECD, 2025), the present value of additional tax revenue is EUR 4,694 per student.

Combining these components yields a negative net cost to the government, G, under both scenarios: EUR $-1{,}352$ in the first scenario and EUR $-1{,}559$ in the second. Because $G < 0$, the program has an infinite MVPF as long as $WTP > 0$.

9.2 Willingness to Pay

Appendix Figure H1 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

[20] Information on direct costs is taken from Persson et al. (2018), except for the costs of teacher manuals, borrowed from Bowes et al. (2024).

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 potential Pareto improvement.

10 Conclusion

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. These reductions create a more positive learning environment and improve socio-emotional well-being for all groups of pupils, generating similar educational and labor market gains across students regardless of their direct involvement in bullying dynamics. The behavioral changes also persist well past the schooling years: treated boys are significantly less likely to engage in criminal activity in adulthood. Together, these results suggest that bullying in the classroom imposes costs on all students irrespective of their social role, and that adult criminality can, at least in part, be traced to factors that are malleable during schooling.

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 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's 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 childhood education programs such as Head Start or for expansions of Medicaid coverage to children (Hendren and Sprung-Keyser, 2020). The MVPF is based on partial equilibrium estimates; in general equilibrium, displacement from 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.

How large are the effects of KiVa relative to other childhood interventions? Focusing on the small set of studies that estimate the long-run effects of programs targeting non-cognitive skills, KiVa's effects on academic high school enrollment are about half the effect size of PATHS in Zurich, Switzerland (Sorrenti et al., 2025), and about one fourth of the effect size on obtaining a secondary schooling diploma for 250 high-risk boys in the Montreal Longitudinal Experimental Study (MLES) (Algan et al., 2022). For earnings, the estimated effects of KiVa amount to about one third of those documented in Algan et al. (2022). These results are remarkable because KiVa operates through a relatively indirect channel: rather than fostering individual skills such as self-control or patience that map more directly onto academic outcomes, it collectively targets the disruptive behaviors and group dynamics that undermine the classroom environment. Finally, our effect sizes are very similar in magnitude, though opposite in sign, to the long-run impact of exposure to one additional disruptive peer in class (Carrell, Hoekstra, and Kuka, 2018). The KiVa intervention thus provides a policy-actionable and scalable solution to mitigate the negative long-run impacts of disruptive peers.

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 long-run gains for all groups of students in the classroom.

References

Adukia, Anjali, Benjamin Feigenberg, and Fatemeh Momeni. 2025. "From Retributive to Restorative: An Alternative Approach to Justice in Schools." *American Economic Review*, 115(8): 2722–2754.

Alan, Sule, and Seda Ertac. 2018. "Fostering Patience in the Classroom: Results from Randomized Educational Intervention." *Journal of Political Economy*, 126(5): 1865–1911.

Alan, Sule, Ceren Baysan, Mert Gumren, and Elif Kubilay. 2021. "Building Social Cohesion in Ethnically Mixed Schools: An Intervention on Perspective Taking." *Quarterly Journal of Economics*, 134(4): 2147–2194.

Alan, Sule, Michela Carlana, and Marinella Leone. forthcoming. "Inclusive Teaching: Spotting Social Isolation in the Classroom." *American Economic Journal: Economic Policy.*

Alan, Sule, Teodora Boneva, and Seda 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.

Algan, Yann, Elizabeth Beasley, Sylvana Côté, Jungwee Park, Richard E. Tremblay, and Frank Vitaro. 2022. "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, Michael 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, Andrew, Joshua Goodman, Jennifer Greif Green, and Melissa K. Holt. 2022. "The COVID-19 Pandemic Disrupted Both School Bullying and Cyberbullying." *American Economic Review: Insights*, 4(3): 353–70.

Beck, Aaron T, Robert A Steer, and Gregory Brown. 1996. *Beck Depression Inventory–II.* APA PsycTests.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014. "Inference on Treatment Effects after Selection among High-Dimensional Controls." *The Review of Economic Studies*, 81(2): 608–650.

Bowes, Lucy, Malavika Babu, Julia R Badger, Matthew R Broome, Rebecca Cannings-John, Suzy Clarkson, Elinor Coulman, Rhiannon Tudor Edwards, Tamsin Ford, Richard 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, Antonio, Kevin Runions, Rosario Ortega-Ruiz, and Eva M. Romera. 2023. "Bullying and Cyberbullying Perpetration and Victimization: Prospective Within-Person Associations." *Journal of Youth and Adolescence*, 52(2): 406–418.

Carrell, Scott E., and Mark L. Hoekstra. 2010. "Externalities in the Classroom: How Children Exposed to Domestic Violence Affect Everyone's Kids." *American Economic Journal: Applied Economics*, 2(1): 211–28.

Carrell, Scott E., Mark Hoekstra, and Elira Kuka. 2018. "The Long-Run Effects of Disruptive Peers." *American Economic Review*, 108(11): 3377–3415.

Cunha, Flavio, Qinyou Hu, Yiming Xia, and Naibao Zhao. 2023. "Reducing Bullying: Evidence from a Parental Involvement Program on Empathy Education." *NBER Working Paper*, 30827.

Eriksen, Tine Louise Mundbjerg, Helena Skyt Nielsen, and Marianne Simonsen. 2014. "Bullying in Elementary School." *Journal of Human Resources*, 49(4): 839–871.

Figlio, David N. 2007. "Boys Named Sue: Disruptive Children and Their Peers." *Education Finance and Policy*, 2(4): 376–394.

Gorman, Emma, Colm Harmon, Silvia Mendolia, Anita Staneva, and Ian Walker. 2021. "Adolescent School Bullying Victimization and Later Life Outcomes." *Oxford Bulletin of Economics and Statistics*, 83(4): 1048–1076.

Gould, Eric D, Victor 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, Sara B, Anuj K Shah, Jonathan Guryan, Jens Ludwig, Sendhil Mullainathan, and Harold 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, Nathaniel, and Ben Sprung-Keyser. 2020. "A Unified Welfare Analysis of Government Policies." *The Quarterly Journal of Economics*, 135(3): 1209–1318.

Huitsing, Gijs, Gerine M.A. Lodder, Wiliam J. Browne, Beau Oldenburg, Rozemarijn Van der Ploeg, and René Veenstra. 2020. "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, Kristiina, Tuomas Pekkarinen, Roope Uusitalo, and Hanna Virtanen. 2023. "Lost Boys? Secondary Education and Crime." *Journal of Public Economics*, 218: 104804.

Irwin, Véronique, Ke Wang, Jiashan Cui, and Alexandra Thompson. 2024. "Report on Indicators of School Crime and Safety: 2023. NCES 2024-145/NCJ 309126." *National Center for Education Statistics*.

Jiménez-Barbero, José Antonio, José Antonio Ruiz-Hernández, Laura Llor-Zaragoza, María Pérez-García, and Bartolomé Llor-Esteban. 2016. "Effectiveness of Anti-Bullying School Programs: A Meta-Analysis." *Children and Youth Services Review*, 61: 165–175.

Kärnä, Antti, Marinus Voeten, Todd D. Little, Elisa Poskiparta, Anne Kaljonen, and Christina Salmivalli. 2011*a*. "A Large-Scale Evaluation of the KiVa Antibullying Program: Grades 4-6." *Child Development*, 82(1): 311–330.

Kärnä, Antti, Marinus Voeten, Todd D. Little, Elisa Poskiparta, Eija Alanen, and Christina Salmivalli. 2011*b*. "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, Antti, Marinus Voeten, Todd D. Little, Erkki Alanen, Elisa Poskiparta, and Christina Salmivalli. 2013. "Effectiveness of the Kiva Antibullying Program: Grades 1-3 and 7-9." *Journal of Educational Psychology*, 105(2): 535–551.

Kristoffersen, Jannie Helene Grøne, Morten Visby Krægpøth, Helena Skyt Nielsen, and Marianne Simonsen. 2015. "Disruptive School Peers and Student Outcomes." *Economics of Education Review*, 45: 1–13.

Lavy, Victor, and Analia Schlosser. 2011. "Mechanisms and Impacts of Gender Peer Effects at School." *American Economic Journal: Applied Economics*, 3(2): 1–33.

Lavy, Victor, and Assaf Yancu. 2025. "Violent Peers at School: Impacts and Mechanisms." National Bureau of Economic Research Working Paper 34482.

Lavy, Victor, M. Daniele Paserman, and Analia 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.

Lazear, Edward 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, Dan. 1996. *The Revised Olweus Bully/Victim Questionnaire.* Bergen:University of Bergen, Research Center for Health Promotion.

Olweus, Dan, and Susan P. Limber. 2010. "Bullying in School: Evaluation and Dissemination of the Olweus Bullying Prevention Program." *American Journal of Orthopsychiatry*, 80(1): 124–134.

Persson, Mattias, Linn Wennberg, Linda Beckman, Christina Salmivalli, and Mikael Svensson. 2018. "The Cost-Effectiveness of the Kiva Antibullying Program: Results from a Decision-Analytic Model." *Prevention Science*, 19(6): 728–737.

Ponzo, Michela. 2013. "Does Bullying Reduce Educational Achievement? An Evaluation Using Matching Estimators." *Journal of Policy Modeling*, 35(6): 1057–1078.

Rege, Mari, Edvin Bru, Ingeborg F Solli, Maximiliaan WPT Thijssen, Kjersti B Tharaldsen, Lene Vestad, Sigrun K Ertesvåg, Terje Ogden, and Paul N Stallard. 2025. "The Impact of Teaching Coping Skills in Schools on Youth Mental Health and Academic Achievement: Evidence from a Field Experiment." CESifo Working Paper.

Rosenberg, Morris. 1965. *Society and the Adolescent Self-Image.* Vol. 11, Princeton, NJ:Princeton University Press.

Salmivalli, Christina, and Elisa 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.

Salmivalli, Christina, and Kätlin Peets. 2018. "Bullying and Victimization." In *Handbook of peer interactions, relationships, and groups, 2nd ed.*. , ed. W.M. Bukowski, B. Laursen and K. H. Rubin, 302–321. New York, NY, US:The Guilford Press.

Salmivalli, Christina, Antti Kärnä, and Elisa 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.* 441–453. Routledge.

Salmivalli, Christina, Antti Kärnä, and Elisa 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, Christina, Claire F. Garandeau, and René Veenstra. 2012. "KiVa Anti-Bullying Program: Implications for School Adjustment." In *Peer Relationships and Adjustment at School. Adolescence and Education*, 279–305. Charlotte, NC, US:IAP Information Age Publishing.

Salmivalli, Christina, Tuna Ojanen, Jemina Haanpää, and Kätlin 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.

Schaede, Ursina, and Ville Mankki. 2026. "Quota versus Quality? Long-Term Gains from an Unusual Gender Quota." *American Economic Review*, 116(6): 1996–2037.

Shem-Tov, Yotam, Steven Raphael, and Alissa Skog. 2024. "Can Restorative Justice Conferencing Reduce Recidivism? Evidence from the Make-It-Right Program." *Econometrica*, 92(1): 61–78.

Silliman, Mikko, and Hanna Virtanen. 2022. "Labor Market Returns to Vocational Secondary Education." *American Economic Journal: Applied Economics*, 14(1): 197–224.

Sorrenti, Giuseppe, Ulf Zölitz, Denis Ribeaud, and Manuel Eisner. 2025. "The Causal Impact of Socio-Emotional Skills Training on Educational Success." *The Review of Economic Studies*, 92(1): 506–552.

Statistics Finland. 2022. "EDUC TYHR [Toisen asteen yhteishaku - moduuli]." Accessed 13-Sep-2024.

Statistics Finland. 2023*a*. "FOLK basic [Perustieto]." Accessed 13-Sep-2024.

Statistics Finland. 2023*b*. "SYYTTR [Syytetyt, tuomitut ja rangaistukset]." Accessed 13-Sep-2024.

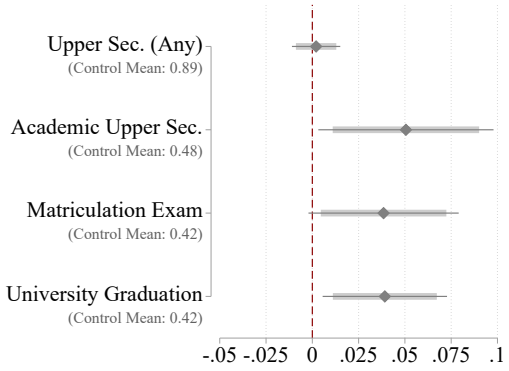
UNESCO. 2018. "School Violence and Bullying: Global Status and Trends, Drivers and Consequences." Paris.

Werner, Katharina, and Ludger Woessmann. 2023. "The Legacy of COVID-19 in Education." *Economic Policy*, 38(115): 609–668.

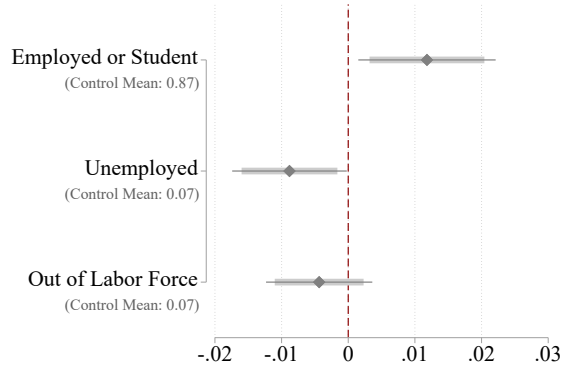
Wolke, Dieter, and Suzet Tanya Lereya. 2015. "Long-Term Effects of Bullying." *Archives of Disease in Childhood*, 100(9): 879–885.

Tables and Figures

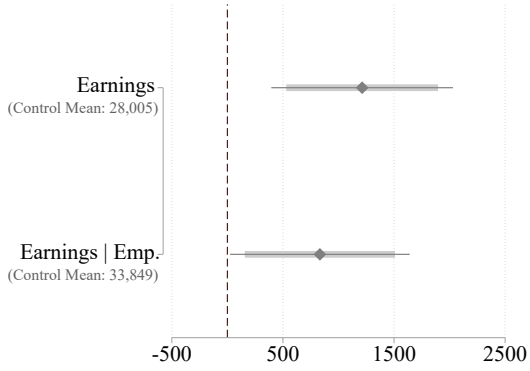
Figure 1: Long-Term Academic and Labor Market Outcomes



(a) Educational Attainment



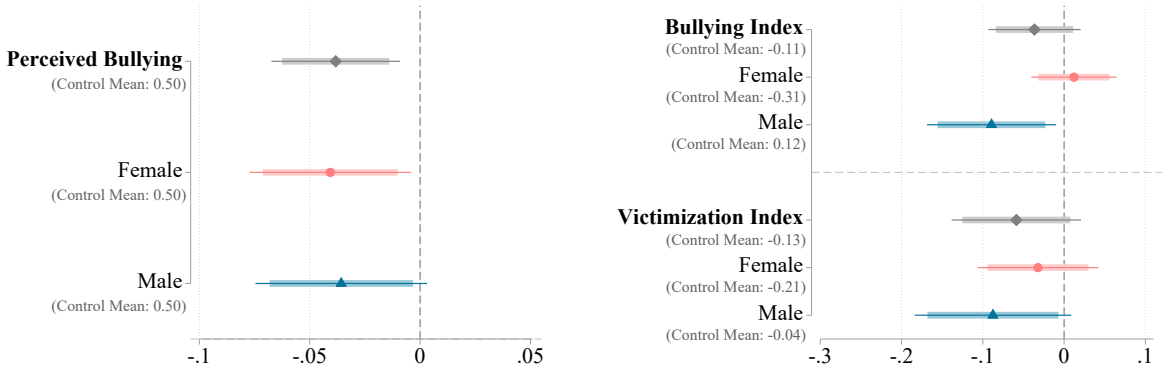
(b) Labor Market Attachment



(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 | Emp.). 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

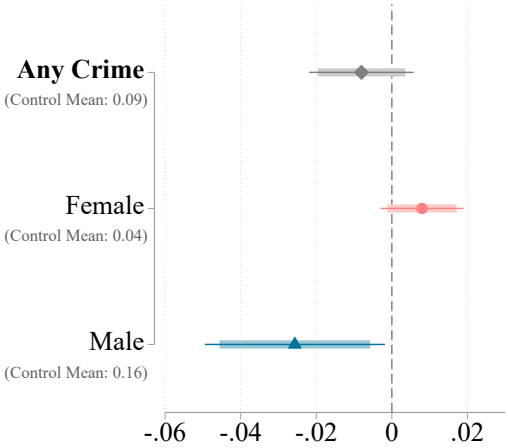


(a) Perceived Change in Bullying

(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 (bully, victim) across three categories of bullying and victimization (Bullying and Victimization Index, see Section 3.2 for details). All regressions include strata fixed effects, the baseline value of the outcome variable, and 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: Long-term Harmful Behavior: Crime in Adulthood



Notes: This figure shows treatment effects (Equation 1) on the full sample and by pupils' gender on an indicator for being charged with at least one 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

Notes: 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

Notes: 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

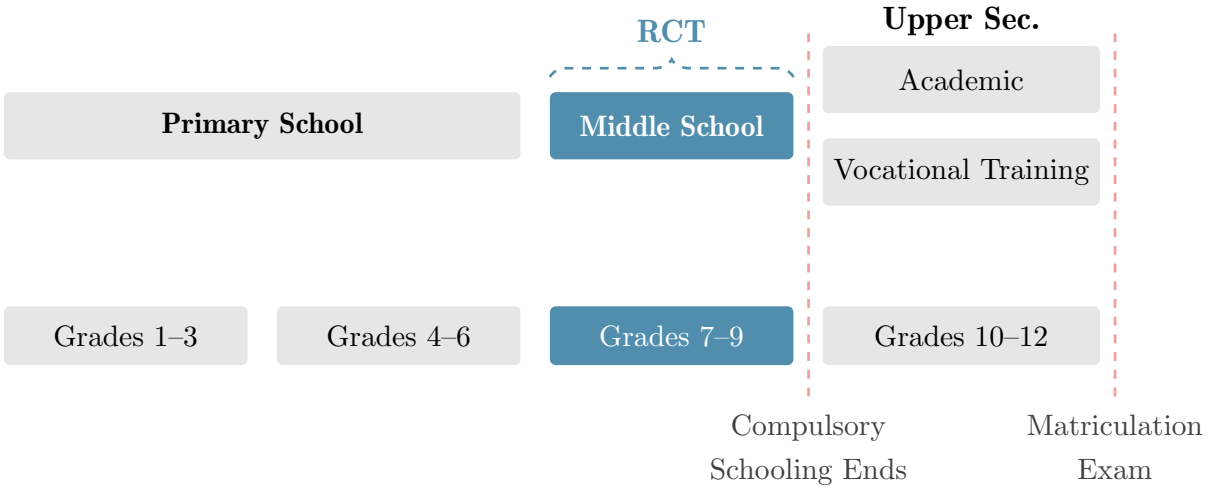
Contents

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.

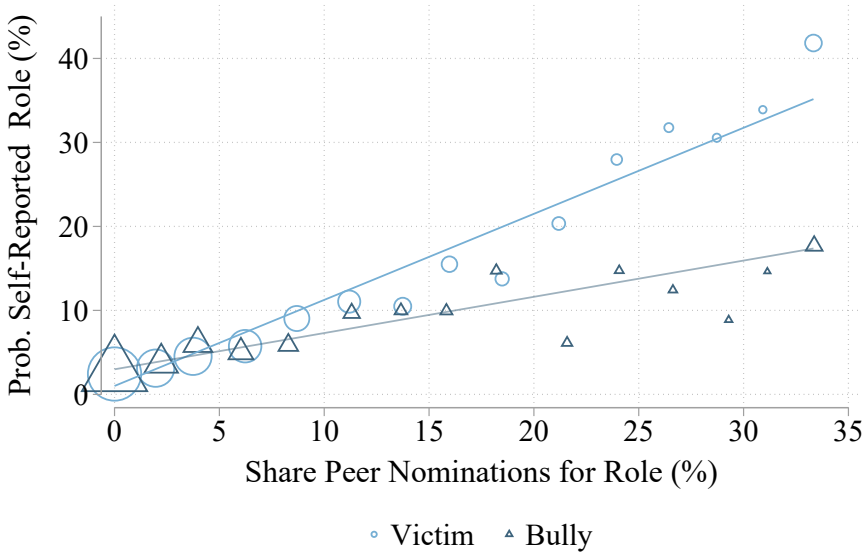
Figure A2: The Finnish Education System



Notes: The figure summarizes the structure of the Finnish education system. The KiVa anti-bullying RCT we study was implemented in lower-secondary education (grades 7–9). Compulsory schooling ends after grade 9; the matriculation examination concludes the academic upper-secondary track.

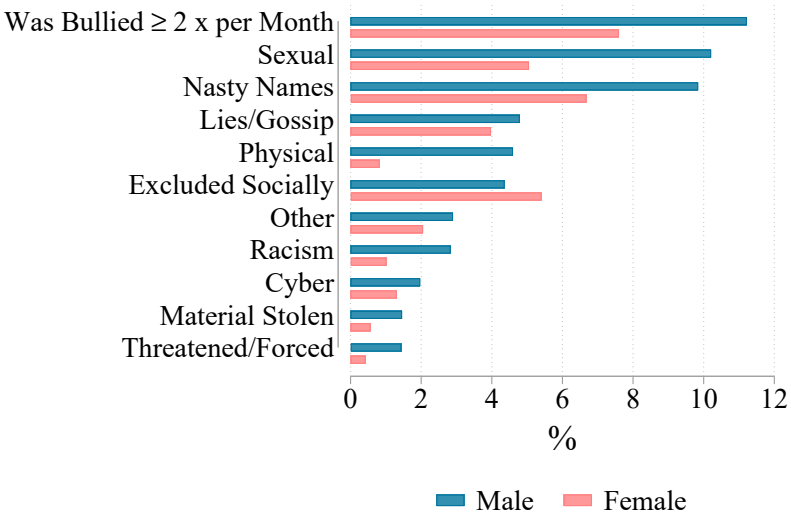
B Descriptives: Bullying and Victimization

Figure B1: Correlation between Self- and Peer-Reported Social Roles

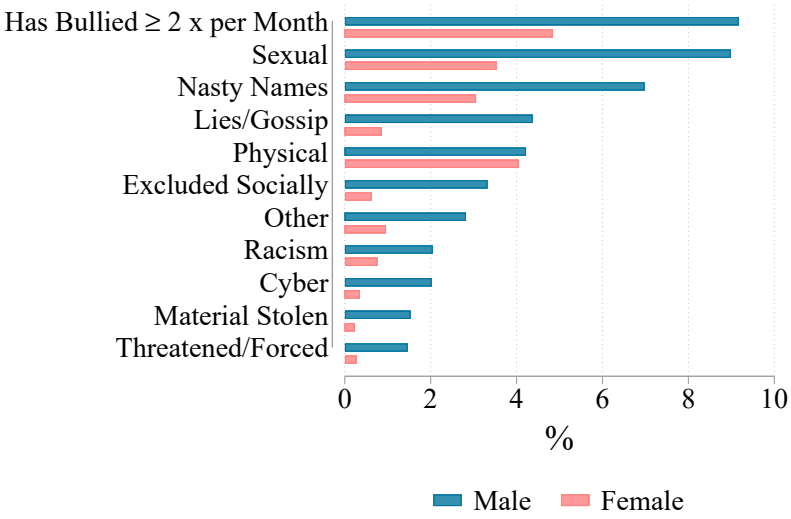


Notes: This figure shows the correlation between the probability of self-nominating as bully or victim and the share of peer nomination received. The share of peer nominations are averaged across the three questions for each role and binned. Observations with more than 35% of nominations are included 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*a*). 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 B2: Categories of Bullying and Victimization



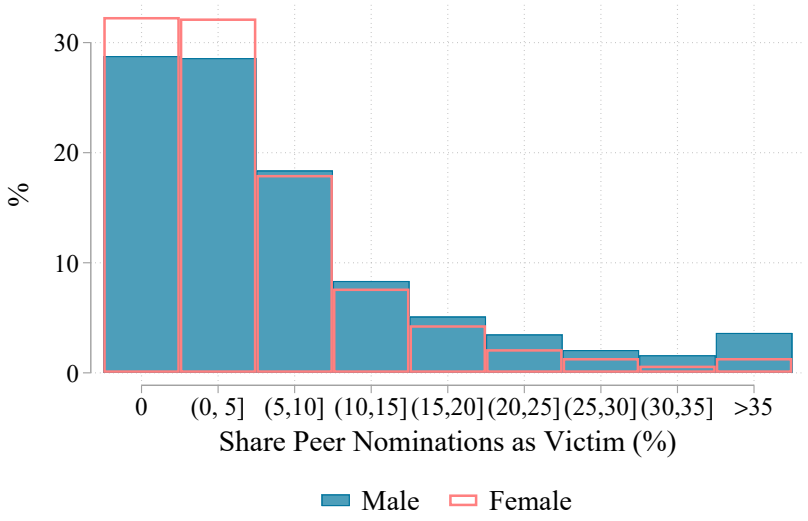
(a) Victimization



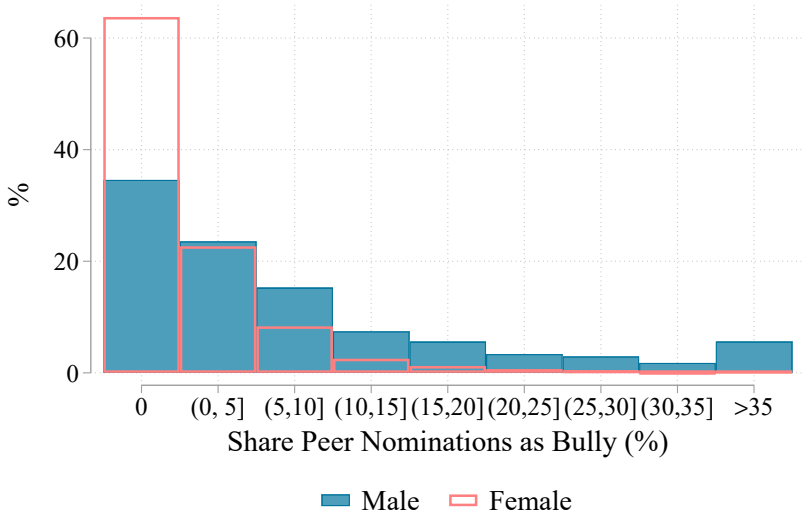
(b) Bullying

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*a*).

Figure B3: Peer-Nomination Shares by Gender



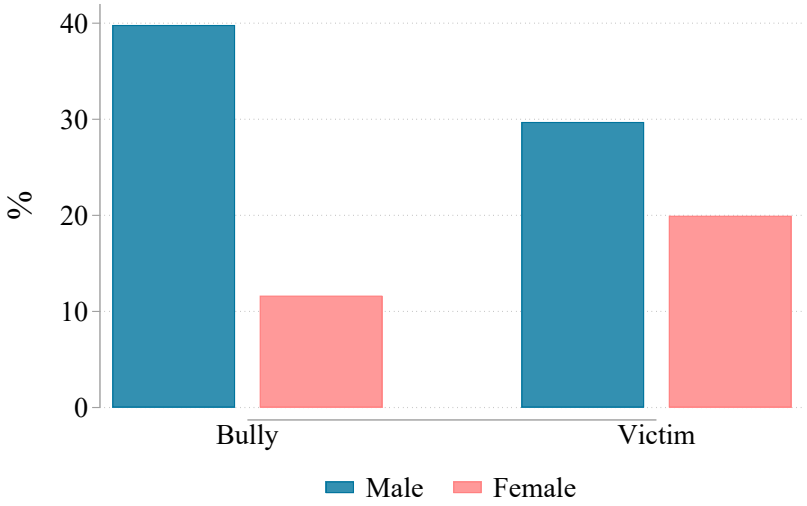
(a) Victimization



(b) Bullying

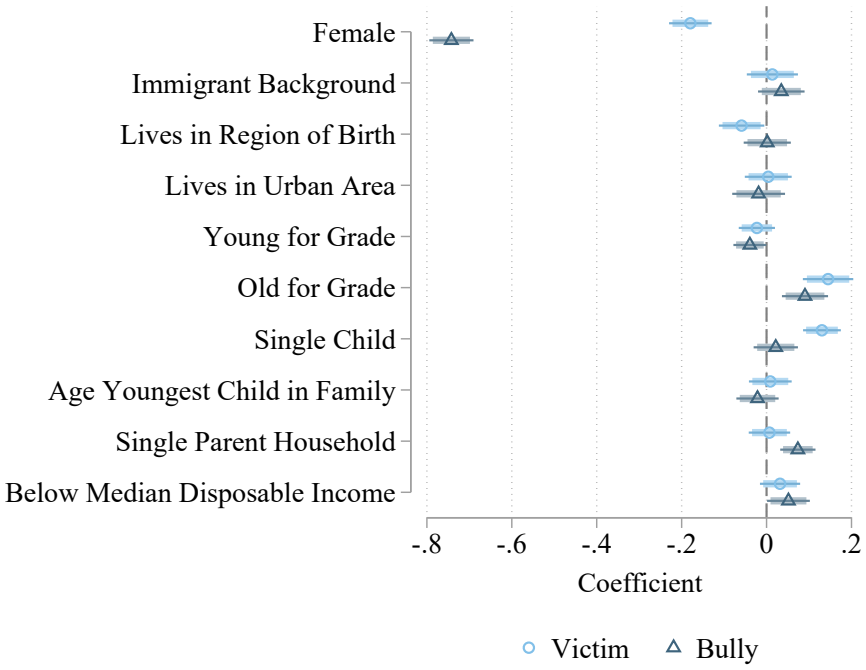
Notes: This figure shows the distribution of the share of peer nominations received by gender for victimization (Panel a) and bullying (Panel b). The share of peer nominations is calculated as the average share of classmates nominating a student as a victim or bully, respectively, across the three questions for each role (see Appendix Table G1 for details).

Figure B4: Social Roles: Prevalence by Gender

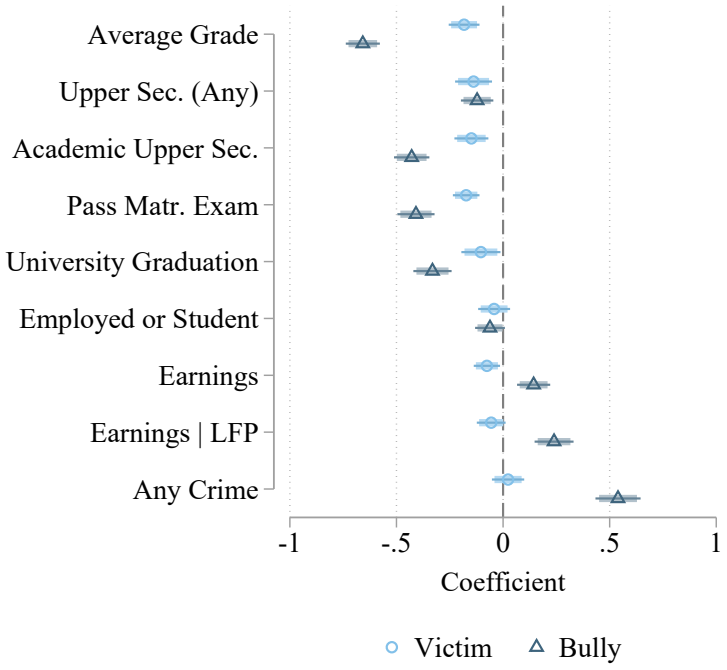


Notes: This figure shows the prevalence of bullies and victims by gender at baseline based on the peer-reported role definitions. Bully and Victim are dummies equal to one if the corresponding variance-weighted index of the share of nominating peers is at or above the 75th percentile of the distribution 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



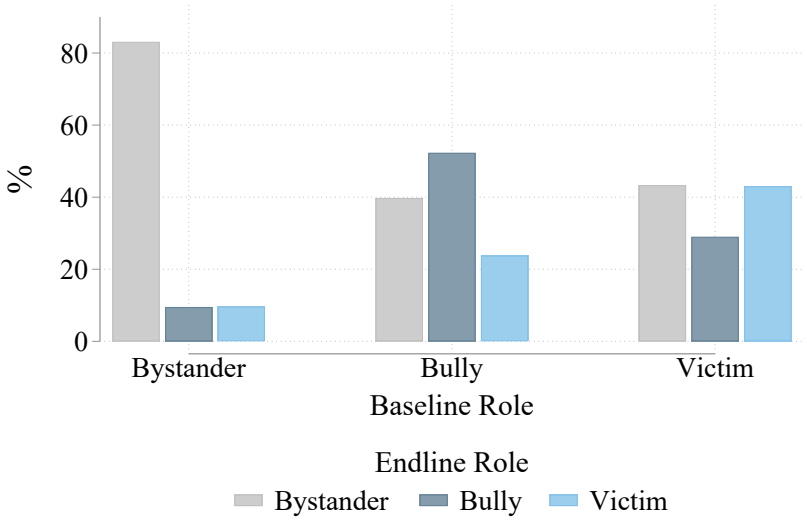
(a) Baseline Characteristics



(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: 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: Treatment Effects on Labor Market Outcomes

	Employed or Student (1)	Unemployed (2)	Out of Labor Force (3)	Earnings (4)	Earnings | Employed (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.

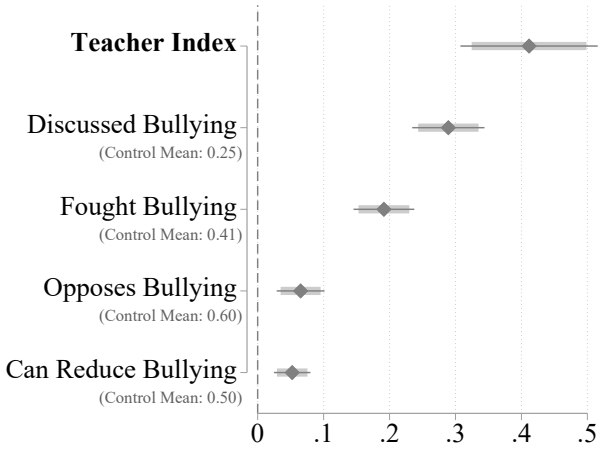
Table C3: Treatment Effects on University Enrollment by Age

	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.

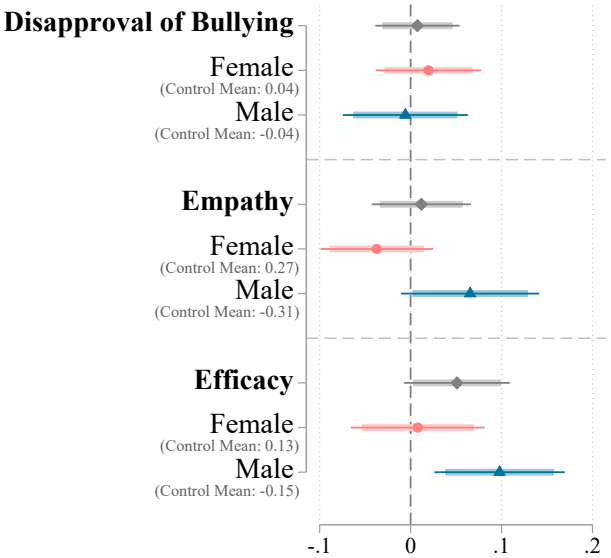
D Additional Results Mechanism

Figure D1: Perceptions about Teachers' Actions against Bullying



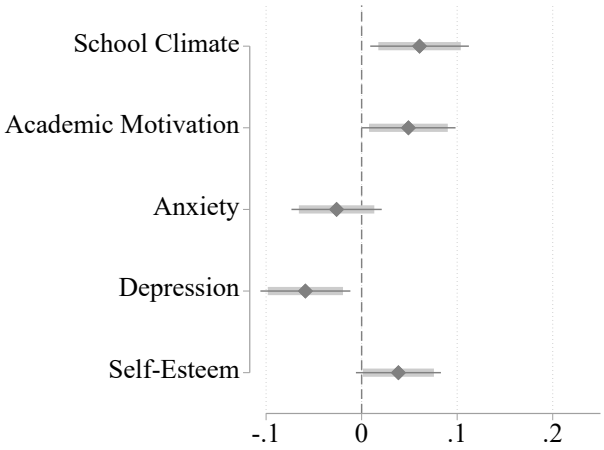
Notes: This figure shows treatment effects (Equation 1) of students' perceptions of teachers' actions and attitudes against bullying. Outcomes from top to bottom: Index across the subsequent four indicators. Indicators for: Teacher discussed bullying at least twice during the academic year (Discussed 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, the baseline value of the outcome variable, and 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: Skills Taught by KiVa Intervention



Notes: This figure shows treatment effects (Equation 1) for the full sample and by pupils' gender 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 G1). All regressions include strata fixed effects, the baseline value of the outcome variable, and 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 D3: School Environment and Socio-Emotional Wellbeing



Notes: This figure shows treatment effects (Equation 1) on indices for school environment and socio-emotional wellbeing. Outcomes from top to bottom: Index for school climate, Index for academic motivation, Index for anxiety, Index for depression based on Beck, Steer, and Brown (1996), Index for self-esteem (see Appendix Table G1). The indices are constructed over raw scales. All regressions include strata fixed effects, the baseline value of the outcome variable, and 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	-0.026	-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**	−0.040**	−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	0.62	0.71	0.62
Victim/ Bystander	0.57	0.39	0.12	0.69

Notes: This table shows treatment effects (Equation 1) on harmful behavior. Column 1 and 2: Indicator equal to one if a student indicates that bullying stayed constant or increased over the past academic year. 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 (1)	IPW (2)	Main (3)	IPW (4)	Main (5)	IPW (6)	Main (7)	IPW (8)	Main (9)	IPW (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	0.001	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***	0.162***	0.148***	0.147***	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)
	[0.001]		[0.001]		[0.001]		[0.115]		[0.043]	
Treat × Victim	0.365***	0.362***	0.232***	0.227***	0.133***	0.133***	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)
	[0.001]		[0.001]		[0.001]		[0.006]		[0.003]	
Treat × Bystander	0.386***	0.391***	0.289***	0.292***	0.168***	0.172***	0.073***	0.076***	0.034*	0.035*
	(0.056)	(0.056)	(0.029)	(0.029)	(0.026)	(0.026)	(0.019)	(0.020)	(0.018)	(0.019)
	[0.001]		[0.001]		[0.001]		[0.001]		[0.014]	
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	0.03	0.00	0.00	0.57	0.50	0.06	0.06	1.00	0.84
Victim/ Bystander	0.74	0.65	0.10	0.06	0.20	0.16	0.57	0.46	0.14	0.14

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)
	[1.000]		[1.000]		[1.000]	
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)
	[0.114]		[0.076]		[0.306]	
Treat × Bystander	0.015	0.017	−0.029	−0.030	0.052	0.054
	(0.033)	(0.033)	(0.030)	(0.030)	(0.035)	(0.035)
	[0.357]		[0.296]		[0.198]	
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	0.85	0.22	0.20	0.77	0.65
Victim/ Bystander	0.19	0.20	0.05	0.07	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 Socio-Emotional Wellbeing

	School Climate		Academic Motivation		Anxiety		Depression		Self-Esteem	
	Main (1)	IPW (2)	Main (3)	IPW (4)	Main (5)	IPW (6)	Main (7)	IPW (8)	Main (9)	IPW (10)
A. Main Estimates										
Treat	0.061**	0.060**	0.049*	0.050**	−0.026	−0.030	−0.059**	−0.062**	0.038*	0.039*
	(0.026)	(0.027)	(0.025)	(0.025)	(0.024)	(0.023)	(0.024)	(0.024)	(0.022)	(0.022)
	[0.059]		[0.059]		[0.123]		[0.059]		[0.072]	
Grade	8-9	8-9	8-9	8-9	8-9	8-9	8-9	8-9	8-9	8-9
N	8,839	8,838	8,345	8,344	8,545	8,544	8,555	8,554	8,552	8,551
Control Mean	0.000	−0.001	0.000	−0.003	−0.000	0.000	−0.000	0.004	0.000	−0.003
Adj. R^2	0.17	0.17	0.18	0.18	0.13	0.13	0.19	0.20	0.20	0.21
B. Heterogeneity: Gender										
Treat × Female	0.067**	0.068**	0.056*	0.059**	−0.061**	−0.065**	−0.085***	−0.087***	0.052	0.053
	(0.030)	(0.030)	(0.029)	(0.029)	(0.025)	(0.025)	(0.025)	(0.026)	(0.035)	(0.035)
	[0.042]		[0.045]		[0.036]		[0.006]		[0.061]	
Treat × Male	0.054	0.052	0.041	0.039	0.012	0.008	−0.029	−0.033	0.023	0.024
	(0.036)	(0.036)	(0.034)	(0.033)	(0.038)	(0.038)	(0.040)	(0.041)	(0.029)	(0.028)
	[1.000]		[1.000]		[1.000]		[1.000]		[1.000]	
Grade	8-9	8-9	8-9	8-9	8-9	8-9	8-9	8-9	8-9	8-9
N	8,839	8,838	8,345	8,344	8,545	8,544	8,555	8,554	8,552	8,551
Control Mean										
Female	0.102	0.103	0.102	0.102	0.096	0.099	0.073	0.072	−0.068	−0.068
Male	−0.114	−0.117	−0.116	−0.123	−0.110	−0.111	−0.083	−0.074	0.077	0.072
P-value	0.76	0.68	0.69	0.61	0.10	0.10	0.22	0.25	0.54	0.53
C. Heterogeneity: Role										
Treat × Bully	−0.012	−0.011	0.042	0.041	0.029	0.014	−0.016	−0.021	−0.010	−0.003
	(0.047)	(0.047)	(0.045)	(0.045)	(0.048)	(0.048)	(0.061)	(0.061)	(0.035)	(0.035)
	[1.000]		[1.000]		[1.000]		[1.000]		[1.000]	
Treat × Victim	0.081*	0.077	0.034	0.028	0.020	0.026	−0.067	−0.063	0.051	0.045
	(0.047)	(0.047)	(0.048)	(0.048)	(0.050)	(0.050)	(0.049)	(0.051)	(0.042)	(0.042)
	[0.613]		[0.613]		[0.613]		[0.613]		[0.613]	
Treat × Bystander	0.063**	0.064**	0.040	0.046	−0.019	−0.021	−0.050**	−0.053**	0.023	0.022
	(0.029)	(0.030)	(0.028)	(0.028)	(0.028)	(0.028)	(0.024)	(0.024)	(0.027)	(0.027)
	[0.109]		[0.192]		[0.367]		[0.109]		[0.367]	
Grade	8-9	8-9	8-9	8-9	8-9	8-9	8-9	8-9	8-9	8-9
N	8,488	8,487	8,002	8,001	8,201	8,200	8,211	8,210	8,209	8,208
Control Mean										
Bully	−0.217	−0.224	−0.310	−0.313	−0.125	−0.119	0.088	0.101	−0.060	−0.068
Victim	−0.165	−0.167	−0.080	−0.078	0.052	0.052	0.096	0.100	−0.012	−0.015
Bystander	0.117	0.117	0.106	0.102	0.019	0.018	−0.058	−0.058	0.039	0.039
P-value										
Bully/ Bystander	0.15	0.15	0.96	0.92	0.39	0.54	0.57	0.59	0.47	0.59
Victim/ Bystander	0.73	0.81	0.91	0.73	0.47	0.39	0.77	0.86	0.55	0.62

Notes: This table shows treatment effects (Equation 1) on the learning environment and socio-emotional wellbeing. Column 1 and 2: Index for school climate. Column 3 and 4: Index for academic motivation. Column 5 and 6: Index for anxiety. Column 7 and 8: Index for depression. Column 9 and 10: Index for self-esteem. 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 D5: Treatment Effects on Grades and Application Behavior

	Grades			Application Choices			Slot Offered	
	Average	Languages	Math	# Appl.	Share Acad. Upper Sec.	No Acad. Upper Sec.	Rank	Acad. Upper Sec.
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
A. Main Estimates								
Treat	0.053	0.081	0.048	0.078	0.062**	−0.052**	0.033	0.061**
	(0.054)	(0.057)	(0.043)	(0.175)	(0.027)	(0.021)	(0.027)	(0.026)
	[0.314]	[0.251]	[0.288]	[0.492]	[0.066]	[0.066]	[0.274]	[0.066]
Grade	7-9	7-9	7-9	7-9	7-9	7-9	7-9	7-9
N	15,033	15,033	15,032	15,033	14,950	14,950	14,448	14,448
Control Mean	−0.000	−0.000	−0.000	3.245	0.488	0.354	1.185	0.498
Adj. R^2	0.11	0.13	0.04	0.09	0.09	0.08	0.01	0.07
B. Heterogeneity: Gender								
Treat × Female	0.040	0.066	0.045	0.044	0.066**	-0.053***	0.032	0.068**
	(0.054)	(0.059)	(0.045)	(0.190)	(0.025)	(0.019)	(0.024)	(0.027)
	[0.495]	[0.361]	[0.361]	[0.693]	[0.039]	[0.039]	[0.303]	[0.039]
Treat × Male	0.067	0.097	0.051	0.114	0.058*	−0.050*	0.034	0.053*
	(0.060)	(0.062)	(0.049)	(0.167)	(0.031)	(0.028)	(0.034)	(0.030)
	[0.296]	[0.294]	[0.296]	[0.330]	[0.294]	[0.294]	[0.296]	[0.294]
Grade	7-9	7-9	7-9	7-9	7-9	7-9	7-9	7-9
N	15,033	15,033	15,032	15,033	14,950	14,950	14,448	14,448
Control Mean								
Female	0.241	0.280	0.104	3.183	0.562	0.281	1.172	0.561
Male	−0.268	−0.311	−0.115	3.313	0.406	0.435	1.199	0.429
P-value	0.47	0.46	0.89	0.36	0.66	0.86	0.94	0.51
C. Heterogeneity: Role								
Treat × Bully	0.002	0.031	0.023	0.071	0.045	-0.022	0.058	0.043
	(0.055)	(0.060)	(0.051)	(0.135)	(0.028)	(0.033)	(0.044)	(0.028)
	[1.000]	[0.995]	[0.995]	[0.995]	[0.995]	[0.995]	[0.995]	[0.995]
Treat × Victim	0.137**	0.111*	0.064	−0.028	0.066**	−0.070**	0.002	0.065**
	(0.054)	(0.059)	(0.054)	(0.154)	(0.026)	(0.029)	(0.031)	(0.028)
	[0.047]	[0.054]	[0.138]	[0.477]	[0.047]	[0.047]	[0.477]	[0.047]
Treat × Bystander	0.052	0.069	0.043	0.020	0.060**	−0.063***	0.016	0.065**
	(0.056)	(0.062)	(0.044)	(0.180)	(0.026)	(0.020)	(0.024)	(0.026)
	[0.417]	[0.417]	[0.417]	[0.890]	[0.055]	[0.017]	[0.590]	[0.048]
Grade	8-9	8-9	8-9	8-9	8-9	8-9	8-9	8-9
N	10,189	10,189	10,189	10,189	10,132	10,132	9,782	9,782
Control Mean								
Bully	−0.499	−0.477	−0.377	3.602	0.313	0.526	1.249	0.328
Victim	−0.182	−0.175	−0.121	3.431	0.424	0.419	1.235	0.440
Bystander	0.227	0.224	0.165	3.184	0.574	0.279	1.159	0.584
P-value								
Bully/ Bystander	0.39	0.55	0.73	0.70	0.57	0.17	0.25	0.44
Victim/ Bystander	0.11	0.46	0.66	0.66	0.83	0.82	0.68	0.98

Notes: This table shows treatment effects (Equation 1) on standardized grades and students' application behavior in the centralized application system for upper secondary education choices in grade 9. Column 1: Average grade in compulsory subjects (standardized). Column 2: Average language grade (native tongue, second national language, and English) (standardized). Column 3: Grade in Mathematics (standardized). Column 4: Total number of applications (ranked choices) to upper secondary education. Column 5: Share of applications to academic upper secondary. Column 6: Indicator for ranked choice list containing no academic upper secondary option. Column 7: Rank of obtained choice. Column 8: Indicator for getting an offer from an academic upper secondary school. 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.

Table D6: 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	33.4	33.4	32.6
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 D5) given the estimated association in each column. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table D7: Treatment Effects on Social Network

	Friendship		Perception of Peers	
	Main	IPW	Main	IPW
	(1)	(2)	(3)	(4)
A. Main Estimates				
Treat	0.020	0.024	0.007	0.004
	(0.031)	(0.031)	(0.025)	(0.024)
	[1.000]		[1.000]	
Grade	8-9	8-9	8-9	8-9
N	8,559	8,558	8,539	8,538
Control Mean	0.000	−0.004	−0.000	0.000
Adj. R^2	0.13	0.13	0.16	0.16
B. Heterogeneity: Gender				
Treat × Female	0.024	0.028	0.009	0.005
	(0.037)	(0.036)	(0.036)	(0.036)
	[0.525]		[0.683]	
Treat × Male	0.017	0.020	0.005	0.003
	(0.038)	(0.038)	(0.034)	(0.033)
	[1.000]		[1.000]	
Grade	8-9	8-9	8-9	8-9
N	8,559	8,558	8,539	8,538
Control Mean				
Female	0.150	0.151	0.122	0.128
Male	−0.171	−0.179	−0.138	−0.144
P-value	0.86	0.85	0.95	0.97
C. Heterogeneity: Role				
Treat × Bully	0.016	0.024	−0.062	−0.051
	(0.055)	(0.054)	(0.051)	(0.050)
	[1.000]		[1.000]	
Treat × Victim	−0.003	−0.013	0.089*	0.076*
	(0.056)	(0.057)	(0.045)	(0.044)
	[0.913]		[0.257]	
Treat × Bystander	0.027	0.033	−0.006	−0.011
	(0.030)	(0.030)	(0.027)	(0.027)
	[0.993]		[1.000]	
Grade	8-9	8-9	8-9	8-9
N	8,215	8,214	8,196	8,195
Control Mean				
Bully	−0.285	−0.291	−0.171	−0.175
Victim	−0.211	−0.213	−0.212	−0.210
Bystander	0.139	0.136	0.129	0.130
P-value				
Bully/ Bystander	0.85	0.88	0.29	0.44
Victim/ Bystander	0.61	0.44	0.09	0.11

Notes: This table shows treatment effects (Equation 1) on students' social network. Column 1 and 2: Index for friendship. Column 3 and 4: Index for positive perception of peers. 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.

E Additional Results Criminal Behavior

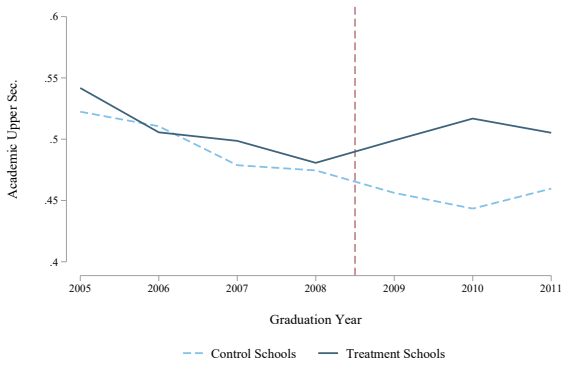
Table E1: Treatment Effects on Criminality in Adulthood

		Type of Crime				
	Any Crime	Traffic	Violent	Property	Drug	Other
	(1)	(2)	(3)	(4)	(5)	(6)
A. Main Estimates						
Treat	−0.008	−0.005	−0.002	−0.006**	−0.006*	−0.002
	(0.007)	(0.005)	(0.003)	(0.002)	(0.003)	(0.004)
	[0.504]	[0.621]	[0.767]	[0.169]	[0.312]	[0.767]
Grade	7-9	7-9	7-9	7-9	7-9	7-9
N	14,869	14,865	14,865	14,865	14,865	14,865
Control Mean	0.092	0.055	0.021	0.019	0.024	0.037
Adj. R^2	0.05	0.03	0.01	0.01	0.01	0.03
B. Heterogeneity: Gender						
Treat × Female	0.008	0.003	0.004*	-0.000	-0.002	0.002
	(0.005)	(0.004)	(0.003)	(0.002)	(0.003)	(0.003)
	[0.819]	[0.819]	[0.819]	[1.000]	[0.819]	[0.819]
Treat × Male	−0.026**	−0.013	−0.009	−0.012**	−0.009	−0.007
	(0.012)	(0.010)	(0.006)	(0.005)	(0.006)	(0.007)
	[0.118]	[0.193]	[0.193]	[0.118]	[0.193]	[0.315]
Grade	7-9	7-9	7-9	7-9	7-9	7-9
N	14,869	14,865	14,865	14,865	14,865	14,865
Control Mean						
Female	0.035	0.018	0.007	0.009	0.010	0.014
Male	0.157	0.095	0.037	0.029	0.039	0.063
P-value	0.01	0.14	0.05	0.05	0.28	0.18
C. Heterogeneity: Role						
Treat × Bully	-0.040*	-0.027	-0.010	-0.009	-0.013	-0.006
	(0.021)	(0.019)	(0.011)	(0.009)	(0.011)	(0.012)
	[0.673]	[0.673]	[0.673]	[0.673]	[0.673]	[0.893]
Treat × Victim	0.017	0.013	0.003	0.005	0.001	0.011
	(0.014)	(0.011)	(0.006)	(0.006)	(0.006)	(0.010)
	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]	[1.000]
Treat × Bystander	−0.001	−0.005	0.006*	−0.004	−0.006**	0.002
	(0.007)	(0.005)	(0.003)	(0.003)	(0.003)	(0.004)
	[0.802]	[0.480]	[0.190]	[0.276]	[0.190]	[0.590]
Grade	8-9	8-9	8-9	8-9	8-9	8-9
N	10,055	10,053	10,053	10,053	10,053	10,053
Control Mean						
Bully	0.210	0.132	0.056	0.040	0.050	0.082
Victim	0.111	0.063	0.026	0.016	0.021	0.042
Bystander	0.052	0.031	0.007	0.011	0.012	0.019
P-value						
Bully/ Bystander	0.08	0.28	0.18	0.64	0.54	0.53
Victim/ Bystander	0.26	0.11	0.69	0.14	0.33	0.35

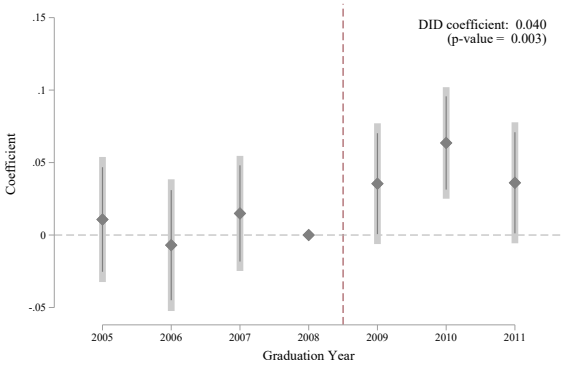
Notes: This table shows treatment effects (Equation 1) on criminal activity in adulthood. Column 1: Indicator for being charged with at least one crime. Columns 2 to 6: Indicator for being charged with at least one crime in the indicated category. Crimes are categorized into types following Huttunen et al. (2023). All regressions include grade and 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.

F Robustness

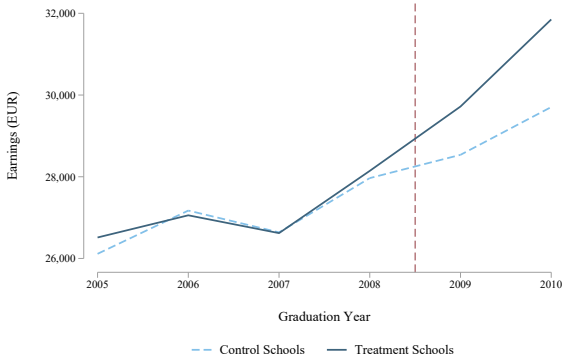
Figure F1: Robustness Main Results: Cohort-Level Analysis



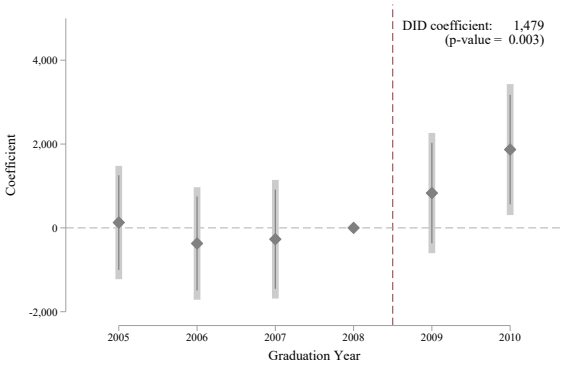
(a) Academic Upper Sec: Raw Trends



(b) Academic Upper Sec.: Event Study



(c) Earnings at Age 29: Raw Trends



(d) Earnings at Age 29: Event Study

Notes: This figure presents the cohort-level difference-in-differences analysis described in Section 6.3. The sample is based on register data for all students in a cohort, irrespective of whether they enter the main analysis sample, collapsed at the school-by-cohort level. Panels (a) and (c) show raw weighted average outcomes by graduation cohort for treated and control schools, for the four cohorts that graduated before the KiVa RCT (2005–2008) and the cohorts that graduated from 2009 onward. Panels (b) and (d) plot the event-study estimates of β_c from Equation 2, which measure the treatment–control difference in each cohort relative to the 2008 cohort. Outcomes are an indicator for enrollment in academic upper secondary school at age 16 (Panels a–b) and annual earnings in EUR at age 29 (Panels c–d). All specifications include school and cohort fixed effects, and observations are weighted by the number of students in each school-by-cohort cell. Standard errors are clustered at the school level; thick and thin bars indicate 90% and 95% confidence intervals, respectively. The pooled difference-in-differences estimate and its p-value are reported in the top-right corner of each event-study panel.

Table F1: 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*a*). All regressions include strata fixed effects and PDS lasso selected controls. Columns 3–6 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 F2: 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. For each outcome, the first column is based on the sample for grades 7–9, the second column on the sample for grades 8–9, and the third column on the survey sample (students in grades 8–9 who participated in the endline survey). Outcomes are: Column 1–3: Indicator for enrollment in academic track of 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. Standard errors clustered at the school level in parentheses. Significance levels: * p < 0.10, ** p < 0.05, *** p < 0.01.

Table F3: 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 F4: 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 victim. 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.

Table F5: Robustness: Treatment Effects on Grades, Applications and Crime for Different Samples

	Average Grade			Share Applications Acad. Upper Sec.			Any Crime		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
A. Main Estimates									
Treat	0.053	0.062	0.083	0.062**	0.065**	0.071**	−0.008	−0.009	−0.011
	(0.054)	(0.056)	(0.057)	(0.027)	(0.028)	(0.028)	(0.007)	(0.009)	(0.009)
N	15,033	10,598	8,838	14,950	10,539	8,796	14,869	10,471	8,729
Control Mean	−0.000	0.005	−0.007	0.488	0.492	0.484	0.092	0.094	0.093
Adj. R^2	0.11	0.11	0.11	0.09	0.10	0.09	0.05	0.05	0.05
P-value (Column 2/3)			0.79			0.88			0.88
B. Heterogeneity: Gender									
Treat × Female	0.040	0.027	0.044	0.066**	0.062**	0.063**	0.008	0.012*	0.012*
	(0.054)	(0.056)	(0.058)	(0.025)	(0.027)	(0.028)	(0.005)	(0.007)	(0.006)
Treat × Male	0.067	0.100	0.125*	0.058*	0.067**	0.081**	−0.026**	−0.032**	−0.036**
	(0.060)	(0.063)	(0.064)	(0.031)	(0.032)	(0.031)	(0.012)	(0.015)	(0.016)
N	15,033	10,598	8,838	14,950	10,539	8,796	14,869	10,471	8,729
Control Mean									
Female	0.241	0.259	0.251	0.562	0.570	0.566	0.035	0.034	0.031
Male	−0.268	−0.281	−0.294	0.406	0.403	0.393	0.157	0.163	0.162
P-value	0.47	0.07	0.07	0.66	0.81	0.40	0.01	0.00	0.00
C. Heterogeneity: Role									
Treat × Bully		0.002	-0.004		0.045	0.050		-0.040*	-0.034
		(0.055)	(0.059)		(0.028)	(0.030)		(0.021)	(0.023)
Treat × Victim		0.137**	0.174***		0.066**	0.066**		0.017	0.011
		(0.054)	(0.058)		(0.026)	(0.030)		(0.014)	(0.016)
Treat × Bystander		0.052	0.073		0.060**	0.067**		−0.001	−0.001
		(0.056)	(0.055)		(0.026)	(0.026)		(0.007)	(0.008)
N		10,189	8,499		10,132	8,458		10,055	8,382
Control Mean									
Bully		−0.499	−0.505		0.313	0.309		0.210	0.207
Victim		−0.183	−0.193		0.423	0.424		0.111	0.109
Bystander		0.227	0.211		0.574	0.564		0.052	0.051
P-value									
Bully/ Bystander		0.39	0.18		0.57	0.55		0.08	0.17
Victim/ Bystander		0.11	0.09		0.83	0.98		0.27	0.48
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 the average grade, applications to academic upper secondary, and crime for different samples. For each outcome, the first column is based on the sample for grades 7–9, the second column on the sample for grades 8–9, and the third column on the survey sample (students in grades 8–9 who participated in the endline survey). Outcomes are: Column 1–3: Average grade in compulsory subjects in grade 9 (standardized). Column 4–6: Share of applications to academic upper secondary. Column 7–9: Indicator for being charged with a crime in adulthood. All regressions include strata fixed effects and controls selected with PDS lasso. Columns 7–9 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.

G Documentation

Table G1: 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

Notes: This table provides an overview of the survey outcome measures (continues on the next page). All standardized indices flip variables in the same direction and are normed by the control group.

Table G1: 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 Motivation	"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		
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
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
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

Notes: This table provides an overview of the survey outcome measures (continued from previous page). All standardized indices flip variables in the same direction and are normed by the control group.

Table G2: LASSO Control Variables Used (Part I)

Outcome Variable	LASSO Controls Used
Treatment Effects on Educational Outcomes	
Upper Sec. (Any)	Young for Grade, Old for Grade, Below Median Disposable Income
Academic Upper Sec.	Female, Lives in Urban Area, Young for Grade, Old for Grade, Single Parent Household, Below Median Disposable Income
Pass Matr. Exam	Female, Lives in Urban Area, Young for Grade, Old for Grade, Single Child (Missing Flag), Age Youngest Child in Family, Age Youngest Child in Family (Missing Flag), Single Parent Household, Below Median Disposable Income
University Graduation	Female, Old for Grade, Single Parent Household, Below Median Disposable Income
Treatment Effects on Labor Outcomes	
Employed or Student	Lives in Urban Area (Missing Flag), Old for Grade, Single Child, Single Child (Missing Flag), Single Parent Household, Single Parent Household (Missing Flag), Below Median Disposable Income
Unemployed	Female, Old for Grade, Single Parent Household, Below Median Disposable Income
Out of Labor Force	Female, Old for Grade, Below Median Disposable Income
Earnings	Female, Old for Grade, Single Child, Single Child (Missing Flag), Single Parent Household, Below Median Disposable Income
Earnings | LFP	Female, Old for Grade, Single Parent Household, Below Median Disposable Income, Below Median Disposable Income (Missing Flag)
Treatment Effects on University Enrollment	
Age 19	Young for Grade, Old for Grade, Single Parent Household, Below Median Disposable Income
Age 20	Female, Immigrant Background, Young for Grade, Old for Grade, Single Child (Missing Flag), Single Parent Household, Single Parent Household (Missing Flag), Below Median Disposable Income
Age 21	Female, Lives in Urban Area (Missing Flag), Old for Grade, Single Parent Household, Below Median Disposable Income
Age 22	Female, Lives in Urban Area (Missing Flag), Old for Grade, Single Parent Household, Below Median Disposable Income
Age 23	Lives in Urban Area, Lives in Urban Area (Missing Flag), Old for Grade, Single Parent Household, Below Median Disposable Income
Age 24	Lives in Urban Area, Lives in Urban Area (Missing Flag), Old for Grade, Single Child, Below Median Disposable Income
Age 25	Lives in Urban Area, Old for Grade, Single Child, Below Median Disposable Income
Age 26	Old for Grade, Below Median Disposable Income
Age 27	Old for Grade, Below Median Disposable Income
Age 28	Age Youngest Child in Family, Below Median Disposable Income
Age 29	Young for Grade, Below Median Disposable Income
Treatment Effects on Bullying and Victimization	
Perceived Bullying	None
Perceived Bullying (IPW)	Below Median Disposable Income (Missing Flag)
Bullying Index	Female, Below Median Disposable Income (Missing Flag)
Victimization Index	Below Median Disposable Income (Missing Flag)
Treatment Effects on Students' Perception of Teachers	
Teacher Index	Female
Teacher Index (IPW)	Female, Below Median Disposable Income (Missing Flag)
Discussed Bullying	Immigrant Background (Missing Flag)
Discussed Bullying (IPW)	Below Median Disposable Income (Missing Flag)
Fought Bullying	None
Fought Bullying (IPW)	Below Median Disposable Income (Missing Flag)
Opposes Bullying	Female
Opposes Bullying (IPW)	Female, Below Median Disposable Income (Missing Flag)
Can Reduce Bullying	None
Can Reduce Bullying (IPW)	Below Median Disposable Income (Missing Flag)

Notes: This table shows the control variables used with each outcome or specification (continues on the next page). The controls are selected from the full set of potential controls listed through post-double-selection lasso as outlined in Belloni, Chernozhukov, and Hansen (2014). Full set of possible controls: Female, Immigrant Background, Lives in Region of Birth, Lives in Urban Area, Young for Grade, Old for Grade, Single Child, Age Youngest Child in Family, Single Parent Household, Below Median Disposable Income. Flags for missing values are included for all controls in the set of potential control variables and denoted by the suffix 'Missing Flag'.

Table G2: LASSO Control Variables Used (Part II)

Outcome Variable	LASSO Controls Used
Treatment Effects on Skills Taught by KiVa	
Disapproval of Bullying	Female, Immigrant Background (Missing Flag)
Disapproval of Bullying (IPW)	Female, Below Median Disposable Income (Missing Flag)
Empathy	Female, Immigrant Background (Missing Flag)
Empathy (IPW)	Female, Below Median Disposable Income (Missing Flag)
Efficacy	Female, Immigrant Background (Missing Flag), Below Median Disposable Income (Missing Flag)
Efficacy (IPW)	Female, Below Median Disposable Income (Missing Flag)
Treatment Effects on Learning Environment and Socio-Emotional Wellbeing	
School Climate	Female, Below Median Disposable Income (Missing Flag)
School Climate (IPW)	Female, Below Median Disposable Income (Missing Flag)
Academic Motivation	Female
Academic Motivation (IPW)	Female, Below Median Disposable Income (Missing Flag)
Anxiety	Female
Anxiety (IPW)	Female, Below Median Disposable Income (Missing Flag)
Depression	Immigrant Background (Missing Flag), Below Median Disposable Income (Missing Flag)
Depression (IPW)	Below Median Disposable Income (Missing Flag)
Self-Esteem	Immigrant Background (Missing Flag), Single Parent Household, Below Median Disposable Income (Missing Flag)
Self-Esteem (IPW)	Single Parent Household, Below Median Disposable Income (Missing Flag)
Treatment Effects on Grades and Application Behavior	
Grades: Average	Female, Young for Grade, Old for Grade, Single Parent Household, Below Median Disposable Income
Grades: Languages	Female, Lives in Urban Area (Missing Flag), Young for Grade, Old for Grade, Age Youngest Child in Family, Single Parent Household, Below Median Disposable Income
Grades: Math	Female, Immigrant Background (Missing Flag), Young for Grade, Old for Grade, Single Child, Single Parent Household, Below Median Disposable Income
Application Choices: # Appl.	Female, Immigrant Background, Old for Grade, Single Parent Household
Application Choices: Share Acad. Upper Sec.	Female, Lives in Urban Area, Young for Grade, Old for Grade, Single Parent Household, Below Median Disposable Income
Application Choices: No Acad. Upper Sec.	Female, Lives in Urban Area, Young for Grade, Old for Grade, Single Parent Household, Below Median Disposable Income
Slot Offered: Rank	Old for Grade, Single Parent Household
Slot Offered: Acad. Upper Sec.	Female, Lives in Urban Area, Young for Grade, Old for Grade, Single Parent Household, Below Median Disposable Income
Association between Average Grades and Attending Academic Track	
Average Grade	Female, Lives in Region of Birth, Young for Grade, Old for Grade, Single Parent Household, Below Median Disposable Income
Treatment Effects on Social Network	
Friendship	Female, Immigrant Background (Missing Flag), Old for Grade
Friendship (IPW)	Female, Old for Grade, Below Median Disposable Income (Missing Flag)
Perception of Peers	Female, Old for Grade
Perception of Peers (IPW)	Female, Old for Grade, Below Median Disposable Income (Missing Flag)
Treatment Effects on Criminality in Adulthood	
Any Crime	Female, Old for Grade, Age Youngest Child in Family, Single Parent Household, Single Parent Household (Missing Flag), Below Median Disposable Income
Type of Crime: Traffic	Female, Old for Grade, Below Median Disposable Income
Type of Crime: Violent	Female, Age Youngest Child in Family, Below Median Disposable Income
Type of Crime: Property	Female, Young for Grade, Old for Grade, Single Parent Household, Below Median Disposable Income
Type of Crime: Drug	Female, Young for Grade, Single Parent Household, Below Median Disposable Income
Type of Crime: Other	Female, Old for Grade, Single Child (Missing Flag), Single Parent Household, Single Parent Household (Missing Flag), Below Median Disposable Income

Notes: This table shows the control variables used with each outcome or specification (continued from previous page and continues on following page). The controls are selected from the full set of potential controls through post-double-selection lasso as outlined in Belloni, Chernozhukov, and Hansen (2014). Full set of possible controls: Female, Immigrant Background, Lives in Region of Birth, Lives in Urban Area, Young for Grade, Old for Grade, Single Child, Age Youngest Child in Family, Single Parent Household, Below Median Disposable Income. Flags for missing values are included for all controls in the set of potential control variables and denoted by the suffix 'Missing Flag'.

Table G2: LASSO Control Variables Used (Part III)

Outcome Variable	LASSO Controls Used
Robustness: Definition of Social Roles	
Academic Upper Sec.	Female, Lives in Urban Area, Young for Grade, Old for Grade, Single Parent Household, Below Median Disposable Income
Pass Matr. Exam	Female, Lives in Urban Area, Young for Grade, Old for Grade, Single Child (Missing Flag), Age Youngest Child in Family, Age Youngest Child in Family (Missing Flag), Single Parent Household, Below Median Disposable Income
University Graduation	Female, Old for Grade, Single Parent Household, Below Median Disposable Income
Employed or Student	Lives in Urban Area (Missing Flag), Old for Grade, Single Child, Single Child (Missing Flag), Single Parent Household, Single Parent Household (Missing Flag), Below Median Disposable Income
Earnings	Female, Old for Grade, Single Child, Single Child (Missing Flag), Single Parent Household, Below Median Disposable Income
Earnings | LFP	Female, Old for Grade, Single Parent Household, Below Median Disposable Income, Below Median Disposable Income (Missing Flag)
Treatment Effects on Education and Labor Outcomes for Different Samples	
Academic Upper Sec. (Grade 7-9)	Female, Lives in Urban Area, Young for Grade, Old for Grade, Single Parent Household, Below Median Disposable Income
Academic Upper Sec. (Grade 8-9)	Female, Lives in Urban Area, Old for Grade, Single Parent Household, Below Median Disposable Income
Academic Upper Sec. (Grade 8-9, Conditional on Endline)	Female, Lives in Urban Area, Old for Grade, Single Parent Household, Below Median Disposable Income
University Graduation (Grade 7-9)	Female, Old for Grade, Single Parent Household, Below Median Disposable Income
University Graduation (Grade 8-9)	Female, Old for Grade, Single Parent Household, Below Median Disposable Income
University Graduation (Grade 8-9, Conditional on Endline)	Female, Old for Grade, Single Parent Household, Below Median Disposable Income
Earnings (Grade 7-9)	Female, Old for Grade, Single Child, Single Child (Missing Flag), Single Parent Household, Below Median Disposable Income
Earnings (Grade 8-9)	Female, Old for Grade, Single Parent Household, Below Median Disposable Income
Earnings (Grade 8-9, Conditional on Endline)	Female, Old for Grade, Single Parent Household, Below Median Disposable Income
Treatment Effects on Grades, Applications and Crime for Different Samples	
Average Grade (Grade 7-9)	Female, Young for Grade, Old for Grade, Single Parent Household, Below Median Disposable Income
Average Grade (Grade 8-9)	Female, Young for Grade, Old for Grade, Single Parent Household, Below Median Disposable Income
Average Grade (Grade 8-9, Conditional Endline)	Female, Young for Grade, Old for Grade, Single Parent Household, Below Median Disposable Income
Share Applications Acad. Upper Sec. (Grade 7-9)	Female, Lives in Urban Area, Young for Grade, Old for Grade, Single Parent Household, Below Median Disposable Income
Share Applications Acad. Upper Sec. (Grade 8-9)	Female, Lives in Urban Area, Old for Grade, Single Parent Household, Below Median Disposable Income
Share Applications Acad. Upper Sec. (Grade 8-9, Conditional Endline)	Female, Lives in Urban Area, Old for Grade, Single Parent Household, Below Median Disposable Income
Any Crime (Grade 7-9)	Female, Old for Grade, Age Youngest Child in Family, Single Parent Household, Single Parent Household (Missing Flag), Below Median Disposable Income
Any Crime (Grade 8-9)	Female, Old for Grade, Age Youngest Child in Family, Single Parent Household, Below Median Disposable Income
Any Crime (Grade 8-9, Conditional Endline)	Female, Old for Grade, Age Youngest Child in Family, Single Parent Household, Below Median Disposable Income
Definition of Bullying	
Peer-reported Index	Old for Grade, Below Median Disposable Income (Missing Flag)
Peer-reported Indicator	Female, Below Median Disposable Income (Missing Flag)
Peer- and Self-reported Index	Old for Grade, Below Median Disposable Income (Missing Flag)
Peer- and Self-reported Indicator	Female, Below Median Disposable Income (Missing Flag)
Definition of Victimization	
Peer-reported Index	Old for Grade, Below Median Disposable Income (Missing Flag)
Peer-reported Indicator	Female, Below Median Disposable Income (Missing Flag)
Peer- and Self-reported Index	Old for Grade, Below Median Disposable Income (Missing Flag)
Peer- and Self-reported Indicator	Female, Below Median Disposable Income (Missing Flag)

Notes: This table shows the control variables used with each outcome or specification (continued from previous page). The controls are selected from the full set of potential controls through post-double-selection lasso as outlined in Belloni, Chernozhukov, and Hansen (2014). Full set of possible controls: Female, Immigrant Background, Lives in Region of Birth, Lives in Urban Area, Young for Grade, Old for Grade, Single Child, Age Youngest Child in Family, Single Parent Household, Below Median Disposable Income. Flags for missing values are included for all controls in the set of potential control variables and denoted by the suffix 'Missing Flag'.

H Marginal Value of Public Funds Calculation

H.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:

$$\text{MVPF} = \frac{\text{WTP}}{G}, \tag{4}$$

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 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).

H.2 Willingness to Pay

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

$$\text{WTP} = \sum_{t=16}^{65} \frac{(1-\tau)\,\Delta y_t}{(1+r)^{t-13}}, \tag{5}$$

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.

H.3 Direct Program Costs

We consider two scenarios for the direct cost of the program.

Scenario 1: Total Finnish program costs. The first scenario takes an upper-bound view based on the total costs incurred by the Finnish government at the time of the intervention. These amounted to EUR 2.6 million and covered program development, licenses, researcher involvement, and teacher training. We distribute these total costs across the 8,548 treated students in grades 7–9, yielding a per-student cost of EUR 304.16. This is a conservative figure since the RCT was also implemented separately in the lower comprehensive-school grades (1–6); spreading the same fixed costs across all treated students would lower the per-student amount.

Scenario 2: Marginal cost of implementation. The second scenario uses the current marginal cost of delivering KiVa per student, as estimated in the broader literature on its implementation. This is arguably the more policy-relevant input for any government considering whether to adopt the program today, since it reflects the marginal cost a replicating implementation would actually face.

Per-student costs are computed using information on the cost of implementing KiVa reported by Persson et al. (2018) and Bowes et al. (2024), 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 130), 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 25.79 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. Summing these components yields a total direct cost of EUR 97.12 per student under this scenario.

H.4 Other Government Costs: University Graduation

Because the KiVa intervention increases the probability that treated students graduate from 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 graduation 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).

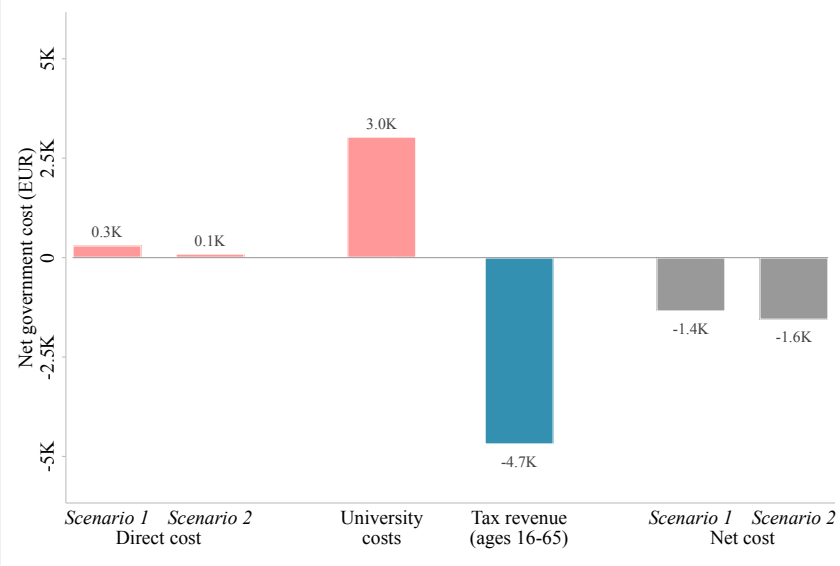
H.5 Tax Revenue Offset

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 H.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).

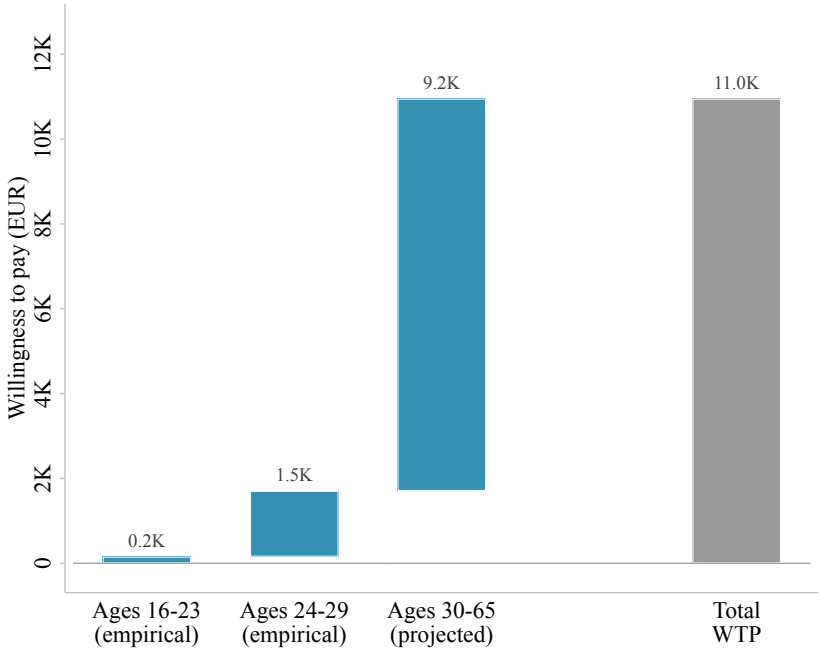
H.6 Summary Results

Figure H1 reports the present-value totals for each component of the MVPF calculation. The net government cost is negative under both scenarios: EUR −1,352 in Scenario 1 (total Finnish program costs) and EUR −1,559 in Scenario 2 (marginal cost of implementation). This means 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 under both scenarios. Importantly, this conclusion does not depend on the assumptions we make to calculate the WTP, as it would hold for any positive WTP estimate.

Figure H1: 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 under two scenarios (Scenario 1: total Finnish program costs allocated across the 8,548 treated students in grades 7–9; Scenario 2: marginal cost of implementation drawn from the literature), university costs induced by higher educational attainment, and tax revenue from increased earnings. The net cost to government is negative under both scenarios. 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 $\tau = 0.30$.