Effectiveness of the KiVa antibullying program: Grades 1–3 and 7–9

ARTICLE *in* JOURNAL OF EDUCATIONAL PSYCHOLOGY · JANUARY 2013

Impact Factor: 3.52 · DOI: 10.1037/a0031120

Journal of Educational Psychology
2013, Vol. 105, No. 2, 535–551

0022-0663/13/$12.00 DOI: 10.1037/a0030417

Effectiveness of the KiVa Antibullying Program: Grades 1–3 and 7–9

Antti Kärnä
University of Turku

Marinus Voeten
Radboud University

Todd D. Little
University of Kansas

Erkki Alanen, Elisa Poskiparta, and Christina Salmivalli
University of Turku

This study investigated the effectiveness of the KiVa Antibullying Program in two samples of students, one from Grades 1–3 (7–9 years old, $N = 6{,}927$) and the other from Grades 7–9 (13–15 years old, $N = 16, 503$). The Grades 1–3 students were located in 74 schools and Grades 7–9 students in 73 schools that were randomly assigned to intervention and control conditions. Multilevel regression analyses revealed that after 9 months of implementation, the intervention had beneficial effects in Grades 1–3 on self-reported victimization and bullying (odds ratios ≈ 1.5), with some differential effects by gender. In Grades 7–9, statistically significant positive results were obtained on 5 of 7 criterion variables, but results often depended on gender and sometimes age. The effects were largest for boys' peer reports: bullying, assisting the bully, and reinforcing the bully (Cohen's *d*s 0.11–0.19). Overall, the findings from the present study and from a previous study for Grades 4–6 (Kärnä, Voeten, Little, Poskiparta, Kaljonen, et al., 2011) indicate that the KiVa program is effective in reducing bullying and victimization in Grades 1–6, but the results are more mixed in Grades 7–9.

Keywords: bullying, victimization, prevention, intervention, evaluation

Supplemental materials: http://dx.doi.org/10.1037/a0030417.supp

Can bullying among children and youth be reduced by school-based interventions? Despite some previous, somewhat pessimistic views (e.g., Ferguson, San Miguel, Kilburn, & Sanchez, 2007; Merrell, Gueldner, Ross, & Isava, 2008; J. D. Smith, Schneider, Smith, & Ananiadou, 2004), a recent meta-analysis by Farrington and Ttofi (2010) concluded that the intervention programs are effective in reducing bullying and victimization, with an average decrease of about 20% in the prevalence of these problems. There was, however, considerable variability in results across studies, suggesting that the effectiveness of the programs may depend on the research methods, the nature of the intervention, and the target populations. The largest effects were obtained for intensive, long-duration programs with parent meetings and clear guidelines for tackling individual cases of bullying. It was also found that the effectiveness of programs increased steadily as the students got older (from 6 years to 14 years of age).

The results of Farrington and Ttofi (2010) concerning the influence of age are surprising and somewhat controversial. Specifically, several studies comparing the effects of one and the same program across age groups have shown that the programs actually work better for young rather than older students (Menesini, Codecasa, Benelli, & Cowie, 2003; Salmivalli, Kaukiainen, & Voeten, 2005) and better in primary than in secondary schools (Hanewinkel, 2004; Pitts & Smith, 1995; P. K. Smith & Sharp, 1994; Stevens, De Bourdeaudhuij, & Van Oost, 2000; see also Olweus, 2005, p. 4). All these studies were reviewed by P. K. Smith (2010, pp. 138–139), who concluded that antibullying programs often have less success in secondary than in primary schools. His main explanations were (a) developmental changes due to puberty and adolescence (e.g., in attitudes to victims) and (b) organizational changes resulting from larger and more complex structure of secondary schools. These organizational factors may make it more difficult to implement the intervention well.

As stated previously, the studies included in the meta-analysis by Farrington and Ttofi (2010) varied not only in the age of the target population but in other aspects as well. It is therefore important to continue studying the moderating effect of age while

This article was published Online First October 22, 2012.

Antti Kärnä, Department of Psychology, University of Turku, Turku, Finland; Marinus Voeten, Behavioral Science Institute, Radboud University, Nijmegen, the Netherlands; Todd D. Little, Department of Psychology and Schiefelbusch Institute for Life Span Studies, University of Kansas; Erkki Alanen, Elisa Poskiparta, and Christina Salmivalli, Department of Psychology, University of Turku.

The development of the KiVa program was financed by the Finnish Ministry of Education and Culture. The writing of this study was supported by the Academy of Finland Grant 134843 to the last author. We thank the whole KiVa project team for their contribution in the data collection process.

The University of Turku has the rights to disseminate the KiVa Antibullying Program. Because Christina Salmivalli and Elisa Poskiparta are among the authors of the KiVa material who stand to gain from a favorable report, they have distanced themselves from critical research activities such as primary data handling and analysis.

Correspondence concerning this article should be addressed to Antti Kärnä, Department of Psychology, University of Turku, Assistentinkatu 7, First floor, Turku 20500, Finland. E-mail: ankarna@utu.fi

using similar research methods and the same program across age groups. Furthermore, gender is another potentially important moderator of intervention effects, but the evidence on its role is scarce. The intervention effects on victimization have sometimes been larger for boys (Eslea & Smith, 1998; Olweus, 2004), whereas Olweus (2004) found larger reductions in girls' reports on bullying. In addition, age and gender are not only characteristics of individual students but also of classrooms. Students in a classroom form a social unit with a certain average age and a gender composition. It is quite possible that the social context in the form of these classroom-level characteristics strengthens or weakens the intervention effects on individual students. To our knowledge, however, in previous studies on antibullying programs, contextual effects have not been studied as moderators of intervention effects. Finally, in addition to positive results, there have been several interventions with statistically nonsignificant effects and even one with negative effects (Farrington & Ttofi, 2010). Despite the optimistic overall results of antibullying programs, it is therefore necessary to investigate the effectiveness of any new program when it is applied to its target population.

In the present article, we report the effects of the recently developed KiVa Antibullying Program on bullying, victimization, and other central outcome variables for Grades 1–3 and 7–9, thus extending the previous evaluation study for Grades 4–6 (Kärnä, Voeten, Little, Poskiparta, Kaljonen, et al., 2011). We also compare the effects of KiVa for children in Grades 1–3, 4–6, and 7–9. The results of the present study provide new knowledge both on the effectiveness of the KiVa program and more generally on the effectiveness of antibullying programs on students, with age and gender taken into account both at the student and the classroom levels.

The KiVa Antibullying Program

The Finnish Ministry of Education and Culture funded the development and initial evaluation of a new antibullying program named KiVa (an acronym for Kiusaamista Vastaan [Against Bullying]). The program was developed at the University of Turku, in collaboration between the Department of Psychology and the Centre for Learning Research. The program was meant for elementary and lower secondary schools, and it was introduced in the intervention schools during two school years: first for Grades 4–6 (2007–2008) and 1 year later (2008–2009) for Grades 1–3 and 7–9, with another group of schools in the intervention condition.

Theoretical Background of the KiVa Program

KiVa is a theory-based intervention program with a background in a particular view of bullying and social behavior (Kärnä, Voeten, Little, Poskiparta, Kaljonen, et al., 2011; Salmivalli, Kärnä, & Poskiparta, 2010a, 2010b). The program is based on (a) studies on the social standing of aggressive children in general (e.g., Cillessen & Mayeux, 2004; Rodkin, Farmer, Pearl, & Van Acker, 2000) and bullies in particular (Juvonen, Graham, & Schuster, 2003) and (b) research on participant roles in bullying (Salmivalli, Lagerspetz, Björkqvist, Österman, & Kaukiainen, 1996). At a more general level, social–cognitive theory (Bandura, 1989) is used as a framework for understanding the processes of social behavior.

Previous research suggests that bullying behavior is at least partly motivated by a pursuit of high status and a powerful position in the peer group (e.g., Juvonen & Galván, 2008; Salmivalli & Peets, 2008). Bullying can, in addition, be considered a group phenomenon, in which bystanders' behaviors have an effect on the maintenance of bullying and on the adjustment of the victims (Salmivalli, 2010; Salmivalli et al., 1996). Specifically, bystanders can contribute to the maintenance of bullying by assisting and reinforcing the bully, which provides bullies with the position of power; defending the victim, on the contrary, may make bullying an unsuccessful strategy for attaining and demonstrating high status (Salmivalli, Voeten, & Poskiparta, 2011). KiVa is predicated on the idea that a positive change in the bystanders' behaviors will reduce the rewards gained by bullies and consequently their motivation to bully in the first place. KiVa strongly emphasizes enhancing the empathy, self-efficacy, and antibullying attitudes of onlookers, who are neither bullies nor victims. This strategy is based on research relating these characteristics to defending and supporting victimized peers (Caravita, DiBlasio, & Salmivalli, 2009; Pöyhönen, Juvonen, & Salmivalli, 2010; Pöyhönen & Salmivalli, 2008; Salmivalli & Voeten, 2004). An important aim of KiVa is to make bystanders show that they are against bullying and to make them support the victim, instead of encouraging the bully. As another equally essential component, the KiVa program includes procedures for handling the acute bullying cases that come to the attention of the school personnel (for the program manuals, see Sainio et al., 2009, and Salmivalli, Poskiparta, Tikka, & Pöyhönen, 2009).

Universal interventions. The universal interventions of KiVa consist of three different age-appropriate versions that are, in the final version of KiVa, now widely implemented in Finnish schools, targeted at Grades 1, 4, and 7. During the evaluation study, these versions were introduced in Grades 1–3, 4–6, and 7–9, respectively. The universal interventions evaluated in Grades 1–3 were 10 double lessons for students (2 × 45 min each) given by classroom teachers during a school year. The lesson titles are "Let's Get to Know Each Other," "Emotions," "Our Class—Everyone Is Included!" "Difference Is Richness," "There Is No Bullying in KiVa School," "We Won't Join in Bullying!" "The Victim Needs Your Support," "I Will Not Be Bullied!" "Literature Lesson," and "KiVa Contract." The lesson goals are (a) to raise awareness of the role that the group plays in maintaining bullying, (b) to increase empathy toward victims, and (c) to promote children's strategies of supporting the victim and thus their self-efficacy to do so. The detailed lesson plans involve discussion, group work, role-play exercises, and short films about bullying. As the lessons proceed, class rules based on the central themes of the lessons are successively adopted one at a time. In the version now evaluated in Grades 7–9, four themes are described in the teachers' manual that can be introduced to students as series of lessons, whole theme days, or otherwise. The themes are "Group Interaction," "Me and the Others," "Forms of Bullying," and "The Consequences and Counterforces of Bullying." The recommended time to spend on the kick-off session, the four themes, and the concluding session compose 13–23 lessons altogether.

A unique feature of KiVa are the virtual learning environments involved. For primary school students (Grades 1–3 in the present study), there is an antibullying computer game that can be played during and between the student lessons. The game involves five

levels, and the topics and tasks in each level are closely connected to matters presented in the corresponding student lessons. By playing the game, students acquire new information and test their existing knowledge about bullying and learn new skills to act in constructive ways in bullying situations. Furthermore, they are encouraged to make use of these skills in real-life situations with their peers. For secondary school students, there is a different virtual learning environment called KiVa Street. It is an Internet forum where the students sign in and navigate to visit different places. For instance, they can go to a library and find information about bullying, or they can enter a movie theatre and watch short films about bullying. Similarly to the computer game, the KiVa Street aims to provide knowledge, skills, and motivation to change one's own behavior related to bullying. Both of these features thus form a component of the intervention in addition to those listed previously.

In all grade levels, KiVa provides prominent symbols such as bright vests for the recess supervisors to enhance their visibility and signal that bullying is taken seriously in the school and posters to remind students and school personnel about the KiVa program. Schools get presentation graphics they can use to introduce the program for the whole personnel and for parents. Parents also receive a guide that includes information about bullying and advice about what parents can do to prevent and reduce the problem.

Indicated interventions. In each school implementing the KiVa program, a team of three teachers or other school personnel, along with the classroom teacher, addresses each case of bullying that comes to their attention (Sainio et al., 2009; Figure 1). First, the team examines whether the reported case actually is an instance of bullying or not (e.g., a quarrel). The school team deals with bullying cases only; other conflicts are delegated to the classroom teacher. Second, individual discussions are organized with the victim. The victim gets a chance to relate his or her experiences, and the school team members communicate that they are on the victim's side and intend to put an end to bullying. This enhances the feeling of security for the bullied student. Third, each bully is taken, without prior notice, individually to discuss the bullying case. In this way, they do not have a chance to prepare themselves or to agree on a common story about the incidents. During the program evaluation phase, for research purposes, the school teams were randomized to implement one of two discussion methods: (a) a confronting approach, where the bullies are openly told that their behavior must stop immediately, and (b) a nonconfronting approach (cf. Pikas, 1989; Robinson & Maines, 1997), where the adult shares his or her concern about the victim and invites the bully to provide suggestions on what could improve the situation. Fourth, the school team meets with the bullies as a group to further confirm the agreements made individually. Fifth, there is a follow-up meeting with the victim to ascertain that bullying has stopped. An improvement in the situation is a requirement for the sixth phase, in which again a meeting is held with the bullies as a group. Also the victims may be included, if they want to be present. The goal of the meeting is to make sure that the bullying has stopped permanently.

In addition to the discussions with the involved students, the classroom teacher meets with between two and four prosocial and high-status classmates and encourages them to support the victimized child. For instance, this support may be shown by inviting the victim in different activities, by treating the victim in an egalitarian and friendly way, or by trying to make others stop bullying.

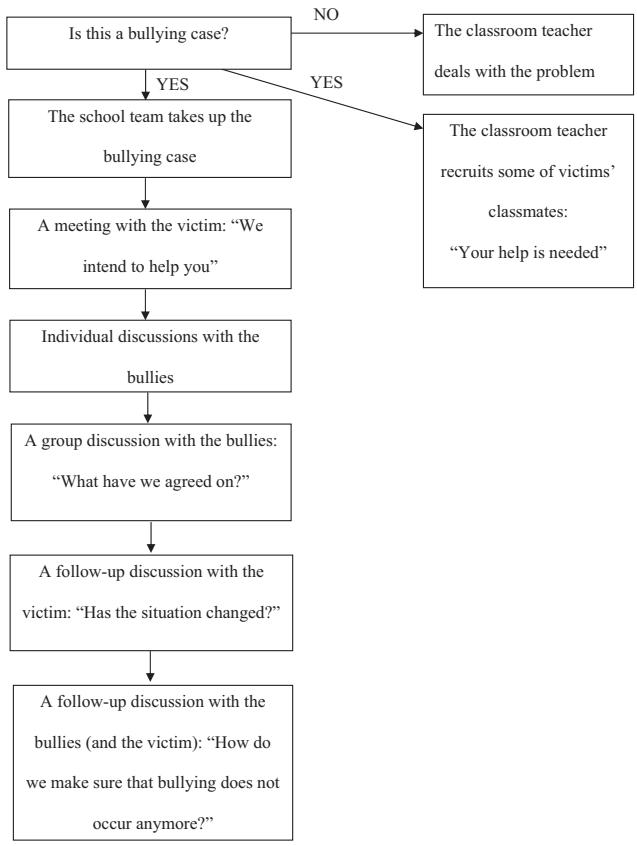

Figure 1. A flow chart for the individual and group discussions included in the indicated intervention.

Support in implementation. During the evaluation study, support was provided to teachers and schools to implement the program with fidelity. In addition to two full days of face-to-face training, networks of school teams were created, consisting of three school teams each. The network members met three times during the school year, with one person from the KiVa project guiding the network. The goal of the network meetings was to motivate the network members to implement the program and to help them overcome any possible obstacles in the process.

It is clear from previous description that KiVa is a whole-school antibullying program: Bullying is viewed as a systemic problem that has multiple causes at the student, classroom, and school levels (J. D. Smith et al., 2004). Like other whole-school programs (e.g., Olweus Bullying Prevention Program), the KiVa program includes components targeting individual students (e.g., discussion methods), classrooms (e.g., antibullying rules), and schools (e.g., a whole-school antibullying policy). There are, however, at least three features that taken together set KiVa apart from other antibullying programs (Kärnä, Voeten, Little, Poskiparta, Kaljonen, et al., 2011). First, KiVa offers a comprehensive collection of concrete and professionally designed materials to be used in antibullying activities, not just abstract principles and guidelines. Second, the program makes use of modern technology

such as computer games and an Internet forum, which are engaging learning environments for the students. Third, while several antibullying programs emphasize the bystanders' role, the KiVa program provides concrete, research-based methods for enhancing the bystanders' empathy, self-efficacy, and efforts to defend the victimized peers.

Previous Studies on the Effects of KiVa

So far the effectiveness studies of the KiVa Antibullying Program have focused on Grades 4–6. The main findings can be summarized as follows: KiVa has been found effective in reducing bullying and victimization, and it has also reduced witnesses' negative behaviors (assisting and reinforcing the bully) and increased their self-efficacy to support and defend the victimized peers (Kärnä, Voeten, Little, Poskiparta, Kaljonen, et al., 2011). The program effects have been found to generalize to multiple forms of victimization (Salmivalli, Kärnä, & Poskiparta, 2011). Furthermore, reductions in victimization in KiVa schools have been reported to predict decreases in depression, anxiety, and negative peer perceptions (Williford et al., 2012). Finally, the KiVa program has increased school liking, academic motivation, and self-reported academic achievement (Salmivalli, Garandeau, & Veenstra, 2012). All these effects were obtained during one school year (August to May) of implementation.

The Present Study

The effects of the KiVa program reported so far are promising. No study, however, has so far tested the effectiveness of KiVa on the main outcome variables among younger (Grades 1–3) or older (Grades 7–9) students in basic education. The present study offers new and important knowledge by examining the effectiveness of KiVa in new target populations. Furthermore, the study may shed light on the differential impact of antibullying programs on younger and older students, on boys versus girls, and on students in classrooms varying by average age and gender composition.

Similar to the researchers who conducted the previous study for Grades 4–6 (Kärnä, Voeten, Little, Poskiparta, Kaljonen, et al., 2011), we examined the program effects by comparing intervention-school students with control-school students at two time points: in the middle and at the end of the school year (i.e., 4 and 9 months after the beginning of program implementation; 7 and 12 months after the pretest measures). The baseline assessment took place at the end of the school year preceding the start of the intervention, because the intervention started right away at the beginning of the school year, and therefore, from a practical point of view, it was the latest possible date for a baseline assessment. The second measurement took place approximately halfway during the intervention year. This enabled us to investigate whether some intervention effects emerge already during the fall term.

We used self-reported bullying and victimization as the main outcomes; we expected that implementing the KiVa program would reduce these problems. For Grades 7–9, we also measured peer-reported bullying and victimization; reduction was expected for these outcomes in the intervention condition. For Grades 7–9, it was further hypothesized that the intervention would bring about beneficial changes in other outcomes. We expected (a) a decrease in assisting and reinforcing the bullies and (b) an increase in defending the victims.

Method

Design and procedure. To recruit schools, we sent letters in the fall of 2006 to all 3,418 schools providing basic education in mainland Finland. These included both Finnish-language and Swedish-language schools, because the basic education in Finland is given in both official languages. The letter included information about the goals and content of the KiVa Antibullying Program and an enrollment form. The 275 volunteering schools were stratified by province and language, and 125 of them (excluding special-education-only schools) were randomly assigned to the intervention (47 schools) or the control condition (78 schools). Furthermore, 31 schools that had previously been randomized into the control condition for Grades 4–6 now participated in the intervention condition. This procedure resulted in a sample of 156 schools: 79 schools (40 control and 39 intervention) for Grades 1–3 and 78 schools (39 control and 39 intervention) for Grades 7–9. One control school participated both with Grades 1–3 and 7–9, but otherwise there was no overlap in the two samples. The Swedish-language schools were oversampled slightly (13% in the sample; 9% in the population). Because the participating schools were quite diverse and located throughout the country, they can be considered representative of those Finnish elementary and lower secondary schools that have an active interest in implementing the KiVa program.

The school year in Finland ranges from mid-August to the end of May. Data were collected in three waves: May 2008, December 2008–February 2009, and May 2009. Students filled out Internet-based questionnaires in the schools' computer labs during regular school hours. The process was administered by the teachers, who were supplied with detailed instructions about 2 weeks prior to data collection. The teachers were told to act in such ways that the confidentiality of the response was secured to a maximum extent, and also both younger and older students were assured that their answers would not be revealed to teachers or parents. In addition, teachers were offered support through phone or e-mail prior to and during data collection.

Teachers distributed individual passwords to the students, who used them to log into the questionnaire. At the beginning of the session, the term *bullying* was defined for the students in the way formulated in the Olweus' Bully/Victim Questionnaire (Olweus, 1996), which emphasizes the repetitive nature of bullying and the power imbalance between the bully and the victim. Compared with the original definition used in Grades 7–9, the definition for students in Grades 1–3 was shortened and simplified to facilitate understanding of the concept measured (see online supplementary appendix). Additionally, to remind the students of the meaning of bullying, a short version of the definition appeared on the upper part of the computer screen when the students responded to a bullying-related question. For Grades 1–3, the teacher read out loud the questions and the answering options in order to facilitate answering, whereas students in Grades 7–9 answered at their own pace. For the older students, the order of questions, items, and scales was extensively randomized to alleviate any systematic order effect. The sessions took on average 26 min in Grades 1–3 and 21 min in Grades 7–9 (5% trimmed means in May 2009).

The program implementation fidelity was measured in both of the samples. For Grades 1–3, the classroom teachers were asked to fill out a questionnaire immediately after each of the 10 KiVa lessons, whereas for Grades 7–9, the schools reported via a web-based questionnaire (in May 2009) about the activities during the intervention year. In this study, the implementation fidelity was represented as school-level averages of the number of given lessons and themes. Further details on this implementation process will be reported in upcoming publications.

Sample. The target sample for Grades 1–3 consisted of 79 schools (40 in intervention and 39 in control conditions). Two intervention and three control schools dropped out without providing any data at all, and therefore we ended up with a data set of 74 schools (38 intervention and 36 control; for details on drop-out schools in Grades 1–3 and 7–9, see online supplementary appendix). In these 74 schools, there were 7,739 students, of whom 7,231 (i.e., 93.4%) received active parental consent to participate in the study; 508 students were excluded from the analyses because of lack of parental consent. Another 304 students left the sample schools after the first wave of measurement, and they were excluded from the analyses, because they were not in the schools at the time of the intervention. This exclusion left us with a sample of 6,927 students in 397 classrooms in 74 schools to be included into the analyses. The number of students per classroom ranged from 1 to 30 ($M = 17.45$, $SD = 5.37$).

Parallel to Grades 1–3, the target sample for Grades 7–9 consisted of 78 schools (39 intervention and 39 control). Four control schools dropped out without providing any data, and one intervention school participated only in the first wave of data collection. After excluding these five schools from the analyses, we were left with 38 intervention schools and 35 control schools, in which there were 19,191 students. Of these students, 16,764 (i.e., 87.4%) gained active parental consent to participate. Altogether, 261 students left the sample after Wave 1 and were excluded. The final analysis sample consisted of 16,503 students in 1,000 classrooms in 73 schools. The number of students per classroom ranged from 1 to 26 ($M = 16.50$, $SD = 4.65$).

The youngest cohorts in our two subsamples were students in Grade 1 and Grade 7 during the intervention year. For these cohorts, we only had the posttest measurements, for two reasons. First, at the time of the pretest, these students were mostly not yet in the schools participating in the present study, and therefore it was impossible to collect pretest measurements from them. Second, we wanted to reduce the burden of data collection for schools participating in the study. Not all Grade 1 students would have had the required cognitive skills to respond to the questionnaire by Wave 2 (in December–February).

Due to the missing pretest measures for Grades 1 and 7, we fitted separate models for students at these grade levels. In the following, the reporting is focused on results for Grades 2–3 and 8–9; the posttest-only sample description and results for Grades 1 and 7 are summarized in the online supplemental material. In the final sample for Grades 2–3, there were 4,704 students in 273 classrooms in 74 schools, whereas in Grades 8–9, there were 11,070 students in 686 classrooms in 73 schools.

Classroom changes were not taken into account for either of the subsamples (Grades 2–3 or 8–9); in the fitted models, students were assigned at pretest to the classrooms they were in during the posttest measurements. This was done to keep the modeling task tractable, and it was justified by the fact that the classroom student composition remained rather similar during the study. Comparing pre- and posttest measurements reveals that 80% of the classrooms remained at least 85% similar in Grades 2 and 3, whereas 90% of the classrooms remained at least 90% similar in Grades 8 and 9.

Missing data. There were several patterns of attrition in the data. For the independent variables, some values were missing, but they could be imputed on the basis of school records. The rates of unintentional missing data for dependent variables varied by variable and wave, but in general, the missingness was higher for self-reports (8.2%–18.4%) than for peer reports (3.2%–7.7%; for details, see Table A1 in the online supplemental material). For Grades 2 and 3, attrition was highest at posttest in the intervention schools, whereas for Grades 8 and 9, attrition was highest at posttest in the control schools.

There were students who responded at Wave 1, but whose answers were missing at Wave 3. Compared with Wave-3 responders, the largest Wave-1 differences were the following (Table A2 in the online supplementary material): The Wave-3 nonresponders had a higher level of some peer-reported behaviors–victimization (Cohen's $d = 0.11$), defending ($d = 0.08$), bullying ($d = 0.07$), and assisting the bully ($d = 0.06$)–and they had a higher level of self-reported bullying in Grades 2–3 ($d = 0.10$) as well as in Grades 8–9 ($d = 0.05$). For about half of the outcomes (five of nine), the drop-outs were at a disadvantage, but the differences were generally small. Additionally, in order to investigate differential attrition, we compared the intervention and control group differences at Wave 1 between the Wave-3 responders and nonresponders. The results indicated some potential for positive bias (i.e., inflation of intervention effects) in self-reported victimization in Grades 2 and 3, and in self-reported bullying and peer-reported defending in Grades 8 and 9. For other outcomes, the biasing effects of attrition were either small or negative. These mean comparisons do not reveal conclusively the mechanism of missingness (Enders, 2010), but they nevertheless suggest that missing data must be specifically taken into account in the models, and the results must be viewed with some caution.

Students could have one, two, or three measurements for a dependent variable. We treated these measurements as nested within students. The measurements were defined as Level 1 of the multilevel models. In this way, we could allow missing values at Level 1 and use all the available information, including the responses of students with partly missing data. The parameters of the models were estimated by full information maximum likelihood (FIML). Schafer and Graham (2002) considered this method as a state-of-the-art missing data technique (see also Jeličić, Phelps, & Lerner, 2009). This FIML approach works well (i.e., gives unbiased estimates) when the missing data can be assumed missing at random (MAR) and when the distributional assumptions for the residuals of the model are met (see, for instance, Enders, 2010, Chapter 4). Our analyses of the missing data showed differential attrition to some extent, and it is therefore unlikely that the missing data are missing completely at random. But MAR is a much less stringent assumption: MAR means that the probability of a missing value does not depend on the missing value itself but that this probability depends on the observed data used in the analysis model. This implies that all variables related to missingness need to be in the model. We can never know whether that is true, but it

is believed that FIML is rather robust to violations of the assumption (Collins, Schafer, & Kam, 2001).

Variables and instrumentation.

Self-reported bullying and self-reported victimization. The questionnaire started with demographic questions (e.g., gender and age) followed by questions about bullying and victimization. To measure bullying and victimization, we used the global items from the revised Olweus' Bully/Victim Questionnaire (Olweus, 1996): "How often have you been bullied at school in the last couple of months?" and "How often have you bullied others at school in the last couple of months?" Students answered with one of five frequency categories (*0 = not at all, 1 = only once or twice, 2 = two or three times a month, 3 = about once a week,* and *4 = several times a week*). For the younger subsample, the five answering options were provided with different colors. That is, in addition to having the response options written in the web-based questionnaire, the teacher who was giving the instructions could also refer to the five colors when helping students pick the right alternative.

Students who reported they had been bullied two or three times a month, every week, or several times a week (Response Categories 2–4) during the past couple of months were categorized as victims, whereas those reporting they had bullied others at the same frequency were categorized as bullies. The cutoff point chosen agrees with the repetitive nature of bullying. With this cutoff point, victims and bullies differ markedly from noninvolved students in conceptually related variables (Solberg & Olweus, 2003). Furthermore, using this criterion (i.e., more than once or twice) facilitates comparisons between the present study and the previous studies, for instance those reviewed by Farrington and Ttofi (2010) in their meta-analysis. Dichotomization was also a practical way to deal with the extremely skewed distributions of the variables (Table 1). In addition, the scores on the self-reported bullying and victimization variables are only categorical or ordinal at most and clearly cannot satisfy the required distributional assumptions (e.g., normality).

To investigate the validity of the global bullying and victimization questions, we calculated school-level correlations between (a) the dichotomized global bullying and victimization items (used in the study), (b) averages for self-reported forms of bullying and victimization (e.g., name calling), and (c) peer reports of bullying and victimization (Kärnä, Voeten, Little, Poskiparta, Alanen, et al., 2011). The results indicated that the associations between the global questions and the questions concerning the respective forms of bullying or victimization were substantial and fairly similar in all grade levels (Grades 1–9) of the KiVa data (rs = .65–.87, $p < .001$). Also the correlations between the global bullying and victimization items were of similar magnitude across all grades (rs = .59–.65, $p < .001$). In addition, for Grade Levels 4–9, school-level correlations between the global self-report items and peer-reported bullying and victimization were .46–.75 ($p < 001$). Taken together, these results provide clear evidence for the construct validity of the global bullying and victimization items. We chose to use only the global items because they provide unambiguous estimates for the prevalence of bullying and victimization. Averaging over ordinal responses concerning the various forms of bullying or victimization would have produced aggregate scores with no clear prevalence interpretation.

Participant roles in bullying situations and peer-reported victimization. When answering the Participant Role Questionnaire (Salmivalli & Voeten, 2004), students were instructed to think of situations in which someone was bullied. They were presented with items describing different ways to behave in such situations, and they were asked to nominate, from a list of classmates presented on the computer screen, an unlimited number of classmates who usually behave in the way described in each item. They were allowed also to choose "no one." The 12 items used in this study form four scales reflecting different participant roles: bullying ("Starts bullying," "Makes the others join in the bullying," and "Always finds new ways of harassing the victim"), assisting the bully ("Joins in the bullying, when someone else has started it," "Assists the bully," and "Helps the bully, maybe by catching the victim"), reinforcing the bully ("Comes around to watch the situation," "Laughs," and "Incites the bully by shouting or saying, 'Show him/her!'"), and defending the victim ("Comforts the victim or encourages him/her to tell the teacher about the bullying," "Tells the others to stop bullying," and "Tries to make the others stop bullying"). In order to measure peer-reported victimization, students nominated classmates treated in the following ways: "He/she is being pushed around and hit," "He/she is called names and mocked," and "Nasty rumors are spread about him/her" (Kärnä, Voeten, Poskiparta, & Salmivalli, 2010).

Table 1
Frequencies of Responses in the Five Categories of the Self-Reported Bullying and Victimization Variables at Wave 3

	Grades 1–3				Grades 7–9			
	Victimization		Bullying		Victimization		Bullying	
Variable	Frequency	%	Frequency	%	Frequency	%	Frequency	%
Occurrence								
Not at all	3,203	53.6	4,296	72.0	10,660	77.4	10,880	79.5
Only once or twice	1,745	29.2	1,333	22.3	2,031	14.7	1,987	14.5
2 or 3 times a month	446	7.5	197	3.3	402	2.9	344	2.5
About once a week	297	5.0	90	1.5	312	2.3	196	1.4
Several times a week	281	4.7	49	0.8	375	2.7	279	2.0
Participants								
Respondents *n*	5,972	100.0	5,965	100.0	13,780	100.0	13,686	100.0
Missing *n*	955		962		2,723		2,817	
Total *N*	6,927		6,927		16,503		16,503	

Peer nominations received were totaled and divided by the number of classmates responding, resulting in a score ranging from 0.00 to 1.00 for each student on each item. The proportion scores were averaged across the three items for each scale. The participant role scales have shown good internal consistencies (e.g., Salmivalli & Voeten, 2004), and in the present sample, Cronbach's alpha coefficients were .90 for the Bully scale, .86 for the Assistant scale, .83 for the Reinforcer scale, .88 for the Defender scale, and .73 for the Victim scale. The relatively low value of the latter is an exception, also compared with results from another sample (Kärnä, Voeten, Little, Poskiparta, Kaljonen, et al., 2011). Finally, we examined the associations among these behaviors prior to intervention in Grades 7 and 8 (n = 10,589). There were strong positive correlations (rs = .76–.85, $p < .001$) among the probullying behaviors (bullying, assisting, and reinforcing). The probullying behaviors had weak negative associations with defending (rs ranging from −.20 to −.26, $p < .001$) but positive associations with victimization (rs = .12–.15, $p < .001$). Defending correlated with victimization only weakly (r = −.08, $p < .001$).

Results

Descriptive Statistics for Outcome Variables

Before fitting the models, we examined the means and standard deviations, without statistical tests yet, for all dependent variables separately for the intervention and control groups at the three waves (Table 2). The comparisons revealed that, in general, there was an overall decrease in the mean levels and standard deviations of all dependent variables. These results suggest (a) that probullying and antibullying (i.e., defending) behaviors as well as victimization decreased over time and (b) that students became more similar over time. To some extent, the decreasing trends may also be a result of attrition (with more problematic students dropping out).

In pretest measures for Grades 2–3 and 8–9, the differences in averages between control and intervention groups were small (ranging from 0.00 to 0.01). For Grades 2 and 3, there was a clear positive change from Wave 1 to Wave 3 in the means of self-

Table 2
Descriptive Statistics for the Dependent Variables in Grades 2–3 and 8–9: Means and Standard Deviations

Variable	Control			Intervention		
	Wave 1	Wave 2	Wave 3	Wave 1	Wave 2	Wave 3
Grades 2–3						
Self-reported victimization						
M	0.23	0.16	0.17	0.22	0.13	0.13
SD	0.42	0.37	0.38	0.42	0.34	0.33
N	1,987	2,086	2,018	2,030	2,230	2,020
Self-reported bullying						
M	0.07	0.05	0.06	0.07	0.04	0.04
SD	0.25	0.23	0.23	0.26	0.20	0.20
N	1,966	2,083	2,018	2,027	2,224	2,019
Grades 8–9						
Self-reported victimization						
M	0.10	0.08	0.07	0.09	0.06	0.07
SD	0.30	0.27	0.26	0.29	0.24	0.25
N	4,333	4,360	3,847	5,694	5,535	5,252
Self-reported bullying						
M	0.08	0.06	0.07	0.07	0.06	0.05
SD	0.26	0.23	0.25	0.25	0.23	0.23
N	4,327	4,358	3,816	5,690	5,530	5,216
Peer-reported victimization						
M	0.07	0.06	0.05	0.06	0.06	0.05
SD	0.10	0.09	0.07	0.09	0.08	0.07
N	4,633	4,779	4,488	5,951	5,940	5,894
Peer-reported bullying						
M	0.05	0.05	0.04	0.05	0.05	0.04
SD	0.10	0.09	0.07	0.10	0.09	0.07
N	4,633	4,779	4,488	5,951	5,939	5,885
Peer-reported assisting						
M	0.07	0.06	0.05	0.07	0.07	0.05
SD	0.11	0.10	0.08	0.11	0.10	0.07
N	4,633	4,779	4,488	5,951	5,939	5,885
Peer-reported reinforcing						
M	0.11	0.11	0.08	0.12	0.10	0.07
SD	0.12	0.11	0.09	0.12	0.11	0.09
N	4,633	4,779	4,488	5,951	5,939	5,885
Peer-reported defending						
M	0.09	0.08	0.07	0.10	0.08	0.06
SD	0.10	0.10	0.09	0.10	0.10	0.08
N	4,633	4,779	4,488	5,951	5,939	5,885

reported victimization and bullying, whereas for Grades 8 and 9, the intervention/control differences in the mean-level changes were more mixed. They were in the positive direction for self-reported bullying and peer-reported reinforcing but in the negative direction (i.e., the change being more positive in the control group) for self- and peer-reported victimization and defending the victims.

Implementation Fidelity

We also examined whether the schools had actually used the KiVa program. To this end, we calculated school-level means and standard deviations for the implemented lessons and themes. In Grades 1–3, the teachers had given on average nine of the 10 prescribed lessons ($M = 9.1$, $SD = 1.1$; $N = 36$). The lower secondary schools were instructed to implement four themes plus introductory and concluding sessions, which amounts to six components. On average, the teachers had implemented five of the six prescribed components ($M = 5.1$, $SD = 0.9$; $N = 32$). It can be noted that the number of implemented lessons and themes corresponds well with the recommendations. This is, as one could expect, because the schools were participating on a voluntary basis.

Variances and Intraclass Correlations

We estimated the variance during pretest for Grades 2–3 and 8–9 (i.e., in the spring of Grades 1–2 and 7–8) for each dependent variable at three levels: students, classrooms, and schools (Table 3). There was nonzero variance for all variables at each level. We also calculated intraclass correlations (ICCs), which provide estimates of the proportion of variance due to differences among students, classrooms, and schools (for notation and formulas, see the note for Table 3). Classroom-level ICCs were higher for peer-reported (ICCs = .15–.25) than for self-reported (ICCs = .07–.15) data. It should be kept in mind, however, that ICCs at the classroom level include both classroom- and school-level variance (classrooms are nested within schools). The highest proportions of variance associated with classroom or school factors were obtained for peer reports of defending (ICC = .25), reinforcing (ICC = .24), and victimization (ICC = .22). For all outcomes, the classroom-level variance was clearly higher than the school-level variance. Between-school variance was highest for peer-reported defending (ICC = .05), and, in Grades 2–3, for self-reported victimization (ICC = .05). Overall, these ICCs show that students sharing the same social environment were more alike than students from different classrooms or schools.

Table 3
Variance Estimates and Intraclass Correlations for Dependent Variables: Student (U), Classroom (V), and School (F) Levels

Variable	Variance			Intraclass correlations	
	$\hat{\sigma}_u^2$	$\hat{\sigma}_v^2$	$\hat{\sigma}_f^2$	ICC_1	ICC_2
Grades 2–3					
Self-reported victimization	1.304	0.156	0.074	.15	.05
Self-reported bullying	2.616	0.100	0.085	.07	.03
Grades 8–9					
Self-reported victimization	2.647	0.157	0.090	.09	.03
Self-reported bullying	3.812	0.241	0.071	.08	.02
Peer-reported victimization	0.734	0.186	0.025	.22	.03
Peer-reported bullying	0.667	0.099	0.015	.15	.02
Peer-reported assisting	0.745	0.127	0.020	.16	.02
Peer-reported reinforcing	0.803	0.205	0.044	.24	.04
Peer-reported defending	0.736	0.186	0.053	.25	.05

Note. All variances were statistically significant (one-tailed p-values at least $< .05$, but mostly $p < .001$), except for between-classroom variance for self-reported bullying in Grades 2 and 3 ($p = .059$). $\hat{\sigma}_u^2$ = variance between students; $\hat{\sigma}_v^2$ = variance between classrooms; $\hat{\sigma}_f^2$ = variance between schools; ICC = intraclass correlation. ICC_1 = proportion of total variance at the classroom level and the school level: $ICC_1 = (\hat{\sigma}_v^2 + \hat{\sigma}_f^2) / (\hat{\sigma}_u^2 + \hat{\sigma}_v^2 + \hat{\sigma}_f^2)$. ICC_2 = proportion of total variance at the school level: $ICC_2 = (\hat{\sigma}_f^2) / (\hat{\sigma}_u^2 + \hat{\sigma}_v^2 + \hat{\sigma}_f^2)$.

Multilevel Models

We used multilevel modeling with MLwiN Version 2.22 (Rasbash, Charlton, Browne, Healy, & Cameron, 2009) to estimate the intervention effects in the presence of the nested data structures. In a nested data structure, the observations are nonindependent: Children from the same classroom or school are more likely to be similar in their responses than children from a different social context. If not modeled, this nonindependence may produce inaccurate standard errors (Raudenbush & Bryk, 2002). Multilevel regression models are therefore preferable to traditional regression models because of their ability to accurately estimate the standard errors by decomposing the total variance into the various hierarchical levels of the data (Snijders & Bosker, 1999).

The data sets for Grades 2–3 and 8–9 contained three measurements; these were treated as a separate level, time points within students. We fitted four-level models to represent change over time, individual student differences, differences between classrooms, and between-school differences. We examined the gains for KiVa schools compared with control schools after controlling for baseline levels of the variable of interest, gender, age, and language of instruction at school (Finnish or Swedish). The models bear a resemblance to the models in the previous study on KiVa program effects in Grades 4–6 (Kärnä, Voeten, Little, Poskiparta, Kaljonen, et al., 2011); the model specification is described in detail in the online supplemental appendix. For the dichotomous self-reports of victimization and bullying, logistic regression analysis was used.

Because gender and age are important predictors of bullying and victimization (see, e.g., Salmivalli & Voeten, 2004; Whitney & Smith, 1993), we included them as covariates in all models. It is known that not only means but also variances of bullying-related variables may differ between boys and girls (Salmivalli & Voeten, 2004). Therefore, student-level variance was specified as a function of gender. In addition, gender and age were also entered as predictors at the classroom level (i.e., proportion of boys and average age of children in the classroom). The inclusion of these covariates enabled us to control for their effects and to investigate their potential interactions with the intervention effects. We added also the language of instruction into our models, because earlier analyses have shown that Swedish-speaking minority students may report lower levels of bullying and victimization than Finnish-speaking students (Kärnä, Voeten, Little, Poskiparta, Kaljonen, et al., 2011). Furthermore, it has been previously argued that in Finland, the Swedish-speaking minority deviates positively

in some respects from the Finnish-speaking majority, for instance, in juvenile delinquency (Obstbaum, 2006).

There were several dummy-coded variables in the models. To test intervention effects at Waves 2 and 3 separately, we coded the three waves of data collection with two dummy variables: Time 2, or *T2* (Wave 2 = 1, other waves = 0), and *T3* (Wave 3 = 1, other waves = 0). In addition, gender (girls = 0, boys = 1), language of instruction (Finnish = 0, Swedish = 1), and condition (control school = 0, intervention school = 1) were entered into the models with dummy coding. Student age was centered around the average age of the students in Grade 2 for the Grades 2–3 data or in Grade 8 for the Grades 8–9 data (grand-mean centering, see Enders & Tofigi, 2007). This implies that the intercepts and the variance components estimated refer to these average ages rather than to the age of zero, which would not make sense. Furthermore, at the classroom level, we included average age of the students in the classroom (*AgeCl*), which is highly correlated with grade level, and proportion of boys in the classroom (*BoyCl*) to test the difference between within-classroom and between-classroom regressions (see, e.g., Snijders & Bosker, 1999, pp. 27–29, and especially pp. 52–56) and to take the influence of the classroom context on bullying and victimization into account. Given our grand-mean centering and inclusion of classroom means in the models, the student-level coefficients of age and gender in Tables 4 and 5 measure the effects of age and gender within classrooms; the classroom-level coefficients for age and gender measure the extent to which these effects at the classroom level differ from those at the student level. The classroom-level coefficients therefore indicate whether the classroom average age or the gender composition

Table 4
Multilevel Modeling Results: Intervention Effects for Self-Reported Victimization and Bullying in Grades 2–3 and 8–9

Variable	Grades 2–3		Grades 8–9	
	Victimization	Bullying	Victimization	Bullying
Baseline				
Intercept	−1.38***	−3.31***	−2.37***	−3.00***
Student level				
Boy	0.48***	1.22***	0.53***	0.80***
Age	—	—	0.09	0.12*
Classroom level				
BoyCl	−0.06	—	0.50	—
AgeCl	—	—	−0.29**	—
School level				
Swedish	−0.40**	−0.84***	—	—
Intervention	−0.02	0.04	0.03	−0.13
Intervention × Boy	−0.10	—	−0.16	—
Intervention × BoyCl	0.91	—	−1.00*	—
Change by Wave 2				
T2	−0.44***	−0.24*	−0.40***	−0.26***
Student level				
Boy × T2	—	—	0.18*	—
Age × T2	—	—	0.18**	—
School level				
Intervention × T2	−0.21*	−0.34*	−0.19*	0.08
Change by Wave 3				
T3	−0.22*	−0.15	−0.54***	−0.15
Student level				
Boy × T3	−0.29*	—	0.36***	—
Classroom level				
BoyCl × T3	0.75	—	—	—
School level				
Intervention × T3	−0.49**	−0.36*	−0.04	−0.08
Intervention × Boy × T3	0.44*	—	—	—
Intervention × BoyCl × T3	−1.67*	—	—	—
Variance components				
Student level				
Baseline for girls	1.965***	6.879***	3.871***	6.414***
Baseline for boys	1.626***	2.591***	2.742***	3.015***
Classroom level				
Intercept	0.167***	0.575**	0.382***	0.432***
Boy	—	1.053**	0.601***	0.530**
School level				
Intercept	0.048*	0.014	0.088**	0.065**

Note. Estimates of covariances omitted. An em dash indicates that the estimate was not included in the model. Cl = classroom level; T2 = Time 2; T3 = Time 3.
* $p < .05$. ** $p < .01$. *** $p < .001$ (one-tailed tests for variances).

Table 5
Multilevel Modeling Results: Intervention Effects for Peer-Reported Victimization, Bullying, and Bystander Behaviors in Grades 8–9

Variable	Victimization	Bullying	Bystander behavior		
			Assisting	Reinforcing	Defending
Baseline					
Intercept	−0.09	−0.25***	−0.34***	−0.40***	0.40***
Student level					
Boy	0.19***	0.61***	0.80***	0.92***	−0.71***
Age	0.07*	0.05**	—	0.04	0.03
Classroom level					
BoyCl	−0.39**	−0.10	−0.45**	−0.30	0.26
AgeCl	0.03	0.01	—	−0.07	−0.12
School level					
Intervention	−0.03	−0.01	0.03	−0.09	0.02
Intervention × Boy	—	0.03	0.00	0.03	0.03
Intervention × Age	−0.09*	—	—	0.02	−0.01
Intervention × BoyCl	—	0.27	0.33	0.34	−0.40
Intervention × AgeCl	0.13	—	—	0.17*	0.05
Change by Wave 2					
T2	0.10***	0.03	0.05**	0.05*	−0.01
Student level					
Boy × T2	−0.04*	−0.02	−0.05*	−0.05*	0.08***
Age × T2	0.02	−0.02	—	—	—
Classroom level					
BoyCl × T2	0.32***	0.11	0.30***	0.07	0.26**
AgeCl × T2	−0.21***	—	—	—	—
School level					
Intervention × T2	−0.06**	0.02	−0.01	−0.02	−0.05*
Intervention × Boy × T2	—	−0.04	−0.03	−0.03	−0.03
Intervention × Age × T2	0.01	—	—	—	—
Intervention × BoyCl × T2	—	−0.31	−0.59***	−0.21	0.25*
Intervention × AgeCl × T2	0.13**	—	—	—	—
Change by Wave 3					
T3	0.14***	0.08***	0.09***	0.12***	−0.14***
Student level					
Boy × T3	−0.11***	−0.11***	−0.12***	−0.15***	0.18***
Age × T3	−0.02	−0.06***	—	−0.03	−0.03
Classroom level					
BoyCl × T3	—	−0.01	—	—	0.35***
AgeCl × T3	−0.15***	—	—	−0.08***	0.10*
School level					
Intervention × T3	−0.10***	0.00	−0.06*	−0.02	−0.10***
Intervention × Boy × T3		−0.09**	−0.10*	−0.16***	—
Intervention × Age × T3	0.08**	—	—	—	0.05
Intervention × BoyCl × T3	—	−0.36**	—	—	—
Intervention × AgeCl × T3	0.02	—	—	—	−0.11*
Variance components					
Student level					
Baseline for girls	0.694***	0.366***	0.382***	0.426***	0.696***
Baseline for boys	0.793***	0.768***	0.810***	0.702***	0.501***
T2	0.656***	0.509***	0.522***	0.568***	0.665***
T3	0.801***	0.651***	0.670***	0.725***	0.834***
Classroom level					
Intercept	0.183***	0.101***	0.141***	0.226***	0.200***
School level					
Intercept	0.024*	0.011**	0.014**	0.023**	0.033***

Note. Estimates of covariances omitted. An em dash indicates that the estimate was not included in the model. Cl = classroom level; T2 = Time 2; T3 = Time 3.
* $p < .05$. ** $p < .01$. *** $p < .001$ (one-tailed tests for variances).

has effects on the outcome over and above the individual student's age or gender.

The models contained several interaction terms. The intervention effects were represented by the coefficients for the interaction terms Intervention × T2 and Intervention × T3, and they can be interpreted as intervention-control differences in the average change scores by Waves 2 and 3. More specifically, the model was defined as follows: $\hat{Y}_{tijk}$ is used to indicate time points, i is used for individual students, j is used to denote classrooms, and k to denote schools:

$$\begin{aligned}\hat{Y}_{tijk} = {} & b_{0ijk} + b_{1i}T2_{tijk} + b_{2i}T3_{tijk} + b_{3i}Boy_{ijk} + b_4Age_{ijk} \\ & + b_5BoyCl_{jk} + b_6AgeCl_{jk} + b_7BoyT2_{tijk} + b_8BoyT3_{tijk} \\ & + b_9AgeT2_{tijk} + b_{10}AgeT3_{tijk} + b_{11}BoyClT2_{tijk} \\ & + b_{12}BoyClT3_{tijk} + b_{13}AgeClT2_{tijk} + b_{14}AgeClT3_{tijk} \\ & + b_{15}Swedish_k + b_{16}Intervention_k + b_{17}InterventionT2_{tijk} \\ & + b_{18}InterventionT3_{tijk}.\end{aligned}$$

Furthermore, we examined moderator effects of gender and age of the student, to see whether intervention effects were dependent on student or classroom characteristics. This was tested by including the terms Intervention × Age × T2, Intervention × Classroom Average Age × T2, Intervention × Age × T3, Intervention × Classroom Average Age × T3, Intervention × Boy × T2, Intervention × Classroom Proportion of Boys × T2, Intervention × Boy × T3, Intervention × Classroom Proportion of Boys × T3. We tested these interactions with multivariate Wald tests. Statistically nonsignificant interaction effects were removed from the models unless they were needed because of significant higher order interactions. To achieve model parsimony and convergence, we omitted some random effects, and we deleted all covariates with statistically nonsignificant effects from the models (for details of final models, see Tables 4 and 5). The random part of the models was kept as simple as possible and the same across dependent variables, but variance heterogeneity by gender was allowed at the student level. Random slopes of the time variables were introduced both at the classroom and school levels but were omitted in all final models because of estimation problems with some models (in these models, it appeared not to be possible to get appropriate parameter estimates when random slopes for the time variables were included). In those cases, estimates were obtained that implied correlations larger than 1.0 between slope and intercept. Parameter estimates remained practically the same whether the random slopes were in the model. So, conclusions were not affected by this simplification of the models. At the classroom level, random slopes for gender were allowed when they were statistically significant. But in the models for Table 5, these slopes for gender were removed, because in these variables there were only minor differences in classroom-level variances between boys and girls and because the removal did not affect the other results.

When statistically significant moderator effects were found, predictors were recentered to test the simple slopes (Aiken & West, 1991). *Age* was recentered from the average value at Grade 8 (or Grade 2) to the average value in Grade 9 (Grade 3) to test the simple slopes, both at the student and classroom levels. If a moderator effect of gender was found, *boy* was recoded at student level (boy = 0, girl = 1), and the *classroom proportion of boys* was recentered to various values ranging from low (35% boys; $M - 1\ SD$), to average (50% boys), to high (65% boys; $M + 1\ SD$).

Results From the Multilevel Models

We used in total seven dependent variables: self-reported bullying and victimization, peer-reported bullying and victimization, and three peer-reported bystanders' behaviors (assisting, reinforcing, defending); the peer reports were obtained only in Grades 7–9. Because of skewness in the distributions of the continuous dependent variables, we transformed them into normal scores (Blom, 1958). The method used creates *z* scores corresponding to the estimated cumulative proportions. After transformation, the new distribution resembled more closely a normal distribution than the raw-score distribution did (e.g., Crocker & Algina, 2008, pp. 442–444). To achieve simplicity and brevity of presentation, we explicated only the Wave-3 results for Grades 2–3 and 8–9 in the text. The interested reader may want to compare them with the Wave-2 results included in the tables. A summary of the Grade-1 and Grade-7 results (from the posttest-only design) is provided in the online supplementary appendix.

Statistical tests. Tables 4 and 5 present the parameter estimates for the final models for each criterion variable in Grades 2–3 and 8–9, including the results for Waves 2 and 3 (T2 and T3). Unstandardized regression coefficients and variance components are reported, and for the dichotomized self-reports, logistic regression coefficients are shown (Table 4). There were residual variances at Levels 2, 3, and 4 but not at Level 1, because two dummy variables were used to represent the three time points. The tables contain only the residual variances at the student, classroom, and school levels plus random slope variances at the student (Table 5) or classroom (Table 4) levels, omitting all covariances.

Baseline equivalency between control and intervention schools. In Tables 4 and 5, the coefficients for the *intervention* variable represent the differences between control and intervention schools at baseline. In examining the descriptive statistics, we already noted that these differences were small (Table 2), and here it can be seen that they are not statistically significant, except for a few interaction effects.

Intervention effects at Wave 3: Grades 2–3 and 8–9. Intervention results concerning self-reported bullying and victimization are reported in Table 4. Compared with the control school students, second and third grade boys and girls in KiVa schools bullied less ($b = -0.36$, $p = .036$). For victimization, however, the effect depended on gender at both the student ($b = 0.44$, $p = .017$) and the classroom ($b = -1.67$, $p = .029$) levels. In other words, there were two separate interaction effects: Intervention × Boy × T3 and Intervention × Classroom Proportion of Boys × T3. Together, these interactions imply that the significant reduction of victimization associated with the intervention, Intervention × T3, was restricted to girls in classrooms with an average proportion (50%) of boys ($b = -0.49$, $p = .001$). This reduction became even stronger when the proportion of boys increased (with 65% boys, $b = -0.74$, $p < .001$). The reduction of victimization, however, was not significant for girls in classrooms with a low proportion (35%) of boys ($b = -0.23$, $p = .179$). For boys, the reduction of victimization at Wave 3 approached statistical significance only in classrooms with 65% boys ($b = -0.30$, $p = .055$). For students in Grades 8–9, the intervention showed no statistically significant effects on self-reported bullying or victimization.

The intervention effects on peer-reported outcomes in Grades 8–9 are presented in Table 5. The intervention reduced peer-reported victimization ($b = -0.10$, $p < .001$). There was an interaction with age of student ($b = 0.08$, $p < .01$), however. Victimization decreased significantly for younger students (at or below the average for students in Grade 8), but for students who

were at the average age for Grade 9, there was hardly any effect ($b = -0.01$, $p = .670$).

The intervention effect on peer-reported bullying was statistically nonsignificant ($b = 0.00$, $p = .854$; Table 5). But because of interaction effects (i.e., Intervention × Boy × T3 and Intervention × Classroom Proportion of Boys × T3), this result applies only to girls in classrooms with an average proportion of boys. At the student level, there was a significant interaction with gender ($b = -0.09$, $p < .01$), and the interaction was significantly stronger at the classroom than at the individual level ($b = -0.36$, $p = .008$). By probing the interactions, we found that bullying was reduced for boys and the more so when the proportion of boys in the classroom was higher (35% boys, $b = -0.04$, $p = .237$; 50% boys, $b = -0.09$, $p < .001$; 65% boys, $b = -0.15$, $p < .001$). Bullying was not reduced for girls, but the effect approached statistical significance when a girl was in a classroom with a high proportion of boys (with 65% boys, $b = -0.07$, $p = .060$).

According to the Table 5 results, the intervention reduced assisting ($b = -0.06$, $p = .010$). This intervention effect applied to girls; for boys, the intervention effect was stronger, shown by the significant interaction with gender at the student level ($b = -0.10$, $p = .001$). More specifically, a significant reduction of assisting was seen for both girls ($b = -0.06$, $p = .010$) and boys ($b = -0.16$, $p = .023$).

For peer-reported reinforcing, the intervention effect was statistically nonsignificant ($b = -0.02$, $p = .473$; Table 5). Due to an interaction, this result, however, applies only to girls: There was an interaction effect with gender at the student level ($b = -0.16$, $p < .001$). Although for girls, the intervention effect at Wave 3 was not statistically significant; for boys, a larger and statistically significant reduction of reinforcing was observed ($b = -0.18$, $p < .001$).

The intervention effect on defending was statistically significant, but it was not in the expected direction ($b = -0.10$, $p < .001$; Table 5). That is, defending the victims decreased in the intervention condition. The intervention effect appeared the same for boys and girls but differed by age. The negative effect on defending did not depend on the age of the student ($b = 0.05$, $p = .292$) but increased with the average age in the classroom ($b = -0.11$, $p = .041$). There was no significant effect for older students in a classroom with primarily younger classmates (more specifically, for students whose age was equal to the Grade-9 average but who were in classrooms with average age equal to the Grade-8 average [$b = -0.06$, $p = .219$]). For students in Grade-9 classrooms, the negative effect on defending became larger ($b = -0.16$, $p < .001$ and $b = -0.21$, $p < .001$, for average student ages at Grades 8 and 9, respectively).

Effect sizes. We calculated effect sizes for the intervention at Waves 2 and 3: model-based odds ratios (*OR*s) for the dichotomous self-reports and Cohen's *d*s for the continuous variables (Tables 6 and 7). The calculations were done assuming an average proportion of boys in classroom (about 50%). The odds ratios were converted to represent the odds of being a bully or a victim in a control school compared with the respective odds in an intervention school. For Cohen's *d*s, an effect size with a positive sign stands here for a positive (i.e., desired) intervention effect.

The Wave-3 results show that in Grades 2–3, the odds of being a victim or a bully were approximately 1.5 times larger in the control schools than in the intervention schools. The only exception was for boys in Grades 2–3, for whom there was no intervention effect on victimization (*OR* = 1.04). For Grades 8–9, the results were not statistically significant.

Table 6
Odds Ratios and 95% Confidence Intervals for the Intervention Effects in Grades 2–3 and 8–9

Variable	Wave 2 odds ratio [95% CI]	Wave 3 odds ratio [95% CI]
	Grades 2–3	
Victimization	1.23 [1.04, 1.42]	—
Girls	—	1.63 [1.34, 1.91]
Boys	—	1.04 [0.79, 1.30]
Bullying	1.41 [1.07, 1.75]	1.43 [1.10, 1.77]
	Grades 8–9	
Victimization	1.21 [1.04, 1.38]	1.04 [0.86, 1.22]
Bullying	0.93 [0.73, 1.12]	1.08 [0.88, 1.28]

Note. A confidence interval that does not include 1.00 implies that $p < .05$. An em dash indicates that the effect size was not calculated. CI = confidence interval.

In Grades 8–9, the largest effects on peer reports were obtained at Wave 3 for boys' reinforcing ($d = 0.19$), assisting ($d = 0.18$), and bullying ($d = 0.11$), and for victimization among both boys and girls in Grade 8 ($d = 0.10$). Many effect sizes for the peer-reported variables in Grades 8 and 9 depended on the proportion of boys in the classroom. Typically, positive intervention effects increased when the proportion of boys was higher. All other effect sizes were small. and several of them rather close to zero.

To make the results of the KiVa program comparable with previous studies (Farrington & Ttofi, 2010), we estimated how much the entire KiVa program reduced the odds for bullying and victimization and the prevalence of these problems. To this end, we used the whole KiVa project sample comprising of Grades 1–9 students from the two phases of the randomized controlled trial evaluation of the program. We included into the calculations those students who had gained parental permission to participate and who had responded both at pretest and posttest (except for Grades 1 and 7 for which only posttest was included). The total sample size was 24,138 for victimization and 24,002 for bullying (with response rates of 70% and 69%). We categorized the students into victims, bullies, and noninvolved children by dichotomizing the self-reported bullying and victimization, as described previously. We calculated odds ratios for victimization and bullying, controlling for the pretest differences (except for Grades 1 and 7, where it was not possible). Next, average weighted means were calculated across the grade levels, and the standard errors were corrected for clustering at the school level by multiplying them with the design effect (based on the ICCs and the average school sizes in our sample; for the formulas, see Farrington & Ttofi, 2010). For victimization, the odds ratios were 1.33, 1.53, and 1.13, and for bullying 1.50, 1.41, and 1.21 in Grades 1–3, 4–6, and 7–9, respectively. The effect sizes were larger in primary school (Grades 1–6) than in secondary school (Grades 7–9). The average weighted odds ratios across all grade levels were for victimization 1.28 with 95% confidence interval (CI) [1.17, 1.40] and for bullying 1.30 with 95% CI [1.15, 1.48]. Therefore, the odds of being a victim or being a bully were about 1.3 times higher for a control-school student than for a student in an intervention school. This corresponds to a reduction of about 20% in the prevalence of bullying and victimization.

Table 7
Cohen's ds for the Intervention Effects in Grades 8–9

	Wave 2	Wave 3
Peer-reported victimization		
Grade 8	0.06	0.10
Grade 9	−0.08	0.01
Peer-reported bullying		
Boys	0.02	0.11
Girls	−0.03	0.00
Peer-reported assisting		
Boys	0.04	0.18
Girls	0.01	0.06
Peer-reported reinforcing		
Boys	0.05	0.19
Girls	0.02	0.02
Peer-reported defending		
Boys	−0.08	—
Girls	−0.05	—
Grade 8	—	−0.11
Grade 9	—	−0.17

Note. Effect sizes were computed as gain for intervention group minus gain for control group. Cohen's d was calculated as the adjusted group mean difference divided by unadjusted pooled within-group standard deviation:

$$d = \frac{b}{\sqrt{\frac{(n_1 - 1)SD_1^2 + (n_2 - 1)SD_2^2}{n_1 + n_2}}}$$

where b is the coefficient for the intervention's effect, which represents the group mean difference adjusted for student- and school-level covariates (Intervention × T2 or Intervention × T3); n_1 and n_2 are the student-level sample sizes; and SD_1 and SD_2 are the student-level unadjusted pretest standard deviations for the intervention group and the control group, respectively. The sign of b was determined such that a positive d always signifies a positive (i.e., desired) intervention effect. All the results are provided in the table regardless of their statistical significance. An em dash indicates that the effect size was not calculated.

Discussion

The present study examined the effectiveness of the KiVa program for Grades 1–3 and 7–9, and it thereby complements the previous findings for Grades 4–6 (Kärnä, Voeten, Little, Poskiparta, Kaljonen, et al., 2011). The results indicate that KiVa is effective in reducing bullying and victimization not only in Grades 4–6 but in Grades 1–3 as well. In Grades 8–9, there were significant positive effects on four of the five peer-reported outcomes, but these effects depended on student and classroom characteristics. In contrast to Grades 4–6, no significant positive effects were found on any of the other outcomes.[1] As a whole, the intervention effects on bullying and victimization appeared larger and more consistent in elementary than in lower secondary schools. The current study thus makes a unique contribution to the literature (a) by providing new knowledge about the effectiveness of the KiVa antibullying intervention program and (b) by supplying evidence about the effectiveness of the program for students and classrooms varying in age and gender.

A comparison of the effect sizes across the present and the previous study (Kärnä, Voeten, Little, Poskiparta, Kaljonen, et al., 2011), provides a somewhat complicated pattern of results. Nevertheless, on the basis of odds ratios calculated in the same way for the whole KiVa sample, it seems that the intervention effects on self-reported victimization and bullying are generally larger in Grades 1–6 than in Grades 7–9. Compared with previous studies, the overall effect sizes in the present study for Grades 1–9 (*OR*s 1.28 and 1.30 for victimization and bullying, respectively) correspond well to the results of a recent meta-analysis (Farrington & Ttofi, 2010), which showed that the average odds ratio for antibullying intervention programs (excluding the KiVa program) was 1.29 (95% CI [1.17, 1.41]) for victimization and 1.36 (95% CI [1.26, 1.47]) for bullying (David P. Farrington, personal communication, March 8, 2010). Furthermore, the KiVa program effects are larger than the average effects for studies with randomized design for victimization ($OR = 1.17$, 95% CI [1.00, 1.37]) and for bullying ($OR = 1.10$, 95% CI [0.97, 1.26]; Farrington, 2010).

With regard to other outcomes measured only in Grades 4–9, the intervention effects were larger and more consistent in Grades 4–6 than in Grades 7–9. In both samples, the intervention effects on peer-reported bullying, victimization, assisting, and reinforcing were at least equally large compared with the effects on self-reported bullying and victimization.

In Grades 8 and 9, positive and significant effects were found on four of the five peer-reported role scales (victimization, bullying, assisting, and reinforcing). But the size of the intervention effects depended on gender and sometimes also on age. In several instances, there were stronger effects for boys and in classrooms with a high proportion of boys. No statistically significant effects were found in Grade 7 for the peer-reported outcomes, and most of the effect sizes were close to zero. These results are less dependable because of the lacking pretest, however. We had no specific prior hypotheses about the moderating effects of gender or age, which makes the analyses somewhat explorative, and the results should be replicated in further studies. The stronger positive effects for boys may be a consequence of boys' high scores that make them suitable targets for the intervention to reduce bullying, assisting, and reinforcing. It is an interesting result that the effects on individual students were largest in classrooms with a high proportion of boys. Perhaps in such classrooms there is the largest potential for improvement because a large concentration of boys may lead to an increase of problematic behaviors; this increase may be counteracted by the intervention.

It has been proposed that as children turn into adolescents, their social intelligence increases, and this may cause (a) a decrease in physical and verbal aggression and (b) an increase in indirect aggression (e.g., Björkqvist, Lagerspetz, & Kaukiainen, 1992; Björkqvist, Österman, & Kaukiainen, 1992). The KiVa program can be less effective in reducing the indirect forms of bullying and victimization, which may partially account for the weaker effects in the lower secondary schools compared with the primary schools. It is also possible that as students age, it becomes increasingly difficult to influence the bullying-related classroom norms (i.e., that a mediation effect of norms is moderated by age). Further studies are needed to investigate these possibilities empirically.

[1] We examined the intervention effects also on antibullying attitudes, empathy toward victims, self-efficacy for defending and wellbeing at school in Grades 7–9 (cf. Kärnä, Voeten, Little, Poskiparta, Kaljonen, et al., 2011, for a definition of these variables). The effect sizes for the intervention effects on these outcomes were in Grades 7–9 practically zero; to conserve space, these analyses were not reported in the present article.

The effect sizes discussed previously can be considered small or moderate. The significant interactions imply, however, that the intervention effects may differ for different types of students. If that is true, then average effect sizes for victimization and bullying (whether self-reports or peer reports) and for reinforcing and assisting may look somewhat modest because of that. This is apparent in the interaction effects with student and classroom characteristics found in the data from Grades 8–9. The magnitude of the effect sizes may also be related to the high proportion of (consistently) noninvolved students. Students who during the school year were never involved in bullying and who were never victimized cannot show an intervention effect. Only when the proportion of such students is lower in intervention schools than in control schools can we have an intervention effect. Intervention effects in the form of reducing negative behaviors can only be found for those students who were involved in bullying or victimization. This may also be related to the finding that intervention effects for some variables were stronger for boys than girls and stronger in classrooms with higher rather than lower proportions of boys. Furthermore, the difference in results for self-reported and peer-reported bullying and victimization may in part be explained by the fact that there was a decrease in the control condition for the self-reports but not for the peer reports. If the problematic behaviors decrease to some extent even without an intervention (or with "treatment as usual," as Finnish schools are obliged by law to counteract bullying), detecting the intervention effects becomes more difficult.

For defending the victims, a significant effect in the wrong direction was found: Compared with their peers in control schools, the students in intervention schools on average defended the victims less. This is a surprising finding, because it suggests that turning adolescents into defenders of victims is more difficult in Grades 7–9 than in Grades 4–6 (Kärnä, Voeten, Little, Poskiparta, Kaljonen, et al., 2011). These results are in contradiction with a recent meta-analysis by Polanin, Espelage, and Pigott (2012) in which the average effect size for the bystander intervention was larger for older (Grades 9–12; $g = .43$) than for younger (Grades 3–8; $g = .14$) students. If the KiVa results were indeed replicated, it would be an important task to investigate in detail the processes involved. A partial explanation may be that KiVa did not succeed in Grades 7–9 in increasing students' antibullying attitudes, empathy toward victims, or self-efficacy for defending: These are characteristics that have been found to predict defending behavior (Pöyhönen et al., 2010; Salmivalli & Voeten, 2004).

Limitations

The fact that there were only posttest data for students in Grades 1 and 7 makes it impossible to control for potential preexisting differences between intervention and control conditions, and it weakens the evidence for these grade levels. Furthermore, for students in Grades 1–3, the outcome variables included only self-reports of bullying and victimization. These young students' global self-reports correlated with respective questions about different forms of bullying and victimization equally well as for older students in Grades 4–9. This indicates that although only self-reports were gathered in Grades 1–3, they were as such valid measures of the phenomena under study. Another limitation is that we did not investigate the effectiveness of KiVa on different forms of victimization and bullying. It remains an important topic for future studies to find out whether the relative impact of the KiVa program varies by the specific form of victimization or bullying (e.g., physical, verbal, or indirect) measured at various grade levels. Yet another important limitation is that the results were assessed solely by questionnaire data. It is possible that the less favorable results in the higher grades may partially be a consequence of the measurement method: There were signs (i.e., implausible or impossible responses) suggesting that the students were not always answering the questions sincerely. Finally, student surveys were administered by teachers. When students answer questions about undesirable behaviors in the presence of their teacher, they may be influenced to answer the way they think the teacher would like. We sought to prevent this by giving teachers detailed instructions on how to act in the survey administration, and the students were told that their answers remain confidential.

In the field trials of intervention programs, several threats to the internal validity of conclusions must be dealt with (e.g., Shadish, Cook, & Campbell, 2002). Most of the alternative explanations for the results obtained in the present study can be considered rather implausible, because we used (a) random assignment of schools into intervention and control conditions and (b) FIML to deal with missing data. Although it was not possible to actually test the MAR assumption of FIML, methodologists have demonstrated that in many realistic cases, failing to take into account a cause or correlate of missingness has only a minor impact on estimates and standard errors (Collins et al., 2001).

Some of the intervention schools ($n = 31$) participating in the present study were randomized into a control condition for Grades 4–6, and during the school year 2007–2008, they were on a waiting list to receive the KiVa program. These schools were promised that they would be included in the intervention condition during the following school year, because otherwise some of them might have easily dropped out of the study. This waiting list may be considered a limitation of our present randomization procedure. But it should be kept in mind that (a) all the schools were originally randomized into intervention and control conditions, (b) they belonged to the same pool of volunteers, and (c) we might have ended up with the present samples even without the waiting list element. The waiting list procedure is unlikely to have had any noticeable effect on the results. The only obvious consequence for the measurements is that a subgroup of Grade 7 students had previously answered the study questionnaires as Grade 6 students, but these students were not included in the main analyses, which involved only Grades 8–9.

With regard to external validity, it can be noted that we had a diverse sample of schools from all over the mainland Finland, including both Finnish- and Swedish-language schools. On the other hand, all our schools were volunteering to take part into the study, and during the study, there was some attrition, with a larger proportion of more problematic students dropping out. Such attrition limits the generalizability of our results to some extent, and further studies are needed to investigate the effectiveness of the KiVa program when it is disseminated widely to a larger sample of schools. Actually, one such study has already been published (Kärnä, Voeten, Little, Poskiparta, Alanen, et al., 2011). There we found on the basis of a sample consisting of 888 schools that the KiVa program also was effective when disseminated broadly throughout the country. The effects were somewhat smaller (the

*OR*s for self-reported bullying and victimization equaled about 1.2), but the pattern of effects was similar to the present study: larger effects in elementary schools compared with lower secondary schools. Another possible limitation is that some cultural specificities of the Finnish context or school system have contributed to the differential effectiveness of the KiVa program. There are research projects in progress in both the Netherlands and the United States to investigate the effectiveness of the KiVa program; the results from these evaluations will provide some idea about the relevance of the school system for the effectiveness of the KiVa program.

Implications

Considered as a whole, the results from the present and the previous study (Kärnä, Voeten, Little, Poskiparta, Kaljonen, et al., 2011) support the view that after 9 months of implementation, the KiVa program is effective in primary school (Grades 1–6), whereas the positive effects in secondary school (Grades 7–9) are more modest and depend especially on the gender of the student. These findings are in contrast with the conclusion of Farrington and Ttofi (2010) that the effects of antibullying programs increase steadily as a function of age. They even suggested that "[antibullying] programs should be targeted on children aged 11 or older rather than on younger children" (Farrington & Ttofi, 2010, p. 72). According to our findings, even students in Grades 1–3 can benefit significantly from antibullying programs, whereas it may be much more difficult to reduce bullying and victimization among secondary school students. Actually, the results from the present study are in accordance with views of P. K. Smith (2010), who argued that the interventions are less effective in secondary schools than in primary schools. He proposed several explanations for the modest results, such as developmental changes related to adolescence (e.g., changes in peer relations), difficulty of change in large organizations such as secondary schools, and differences in teacher roles between primary and secondary schools. These are all possible explanations for the differences in effectiveness of the primary- and secondary-school versions of the KiVa program.

Future Directions

The ultimate aim of the Finnish Ministry of Education and Culture was to develop a research-based antibullying program that could be used in all Finnish elementary and lower secondary schools. The dissemination of KiVa to Finnish schools started in 2009, and after the first 3 years, 90% of all schools in the country have joined in. As the students in the participating schools will answer questionnaires every spring, this will create a unique opportunity to investigate the long-term effects of KiVa on bullying and victimization.

After the present and previous studies (e.g., Kärnä, Voeten, Little, Poskiparta, Alanen, et al., 2011; Kärnä, Voeten, Little, Poskiparta, Kaljonen, et al., 2011) on the main effects of KiVa, several important questions remain for future research. For instance, it is important to try to understand why the effects of the KiVa program seem to be larger for primary schools than for secondary schools. This requires investigation of mediators and moderators of program effects in the different age groups at the multiple systemic levels of student, classroom, and school. The degree of fidelity of program implementation may, to some extent, explain variation in the intervention outcomes (e.g., Olweus & Alsaker, 1991; Salmivalli et al., 2005; Whitney, Rivers, Smith, & Sharp, 1994). Research on the association between implementation and intervention results will give some idea of how much the intervention results can be improved by providing support for schools in program implementation. Finally, investigating the predictors of implementation (Kallestad & Olweus, 2003) will provide information about what kind of schools need additional resources for high-quality implementation of the KiVa program.

References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: Sage.

Bandura, A. (1989). Social cognitive theory. In R. Vasta (Ed.), *Annals of child development: Six theories of child development* (pp. 1–60). Greenwich, CT: JAI Press.

Björkqvist, K., Lagerspetz, K. M. J., & Kaukiainen, A. (1992). Do girls manipulate and boys fight? *Aggressive Behavior, 18,* 117–127.

Björkqvist, K., Österman, K., & Kaukiainen, A. (1992). The development of direct and indirect aggressive strategies in males and females. In K. Björkqvist & P. Niemelä (Eds.), *Of mice and women: Aspects of female aggression* (pp. 51–64). San Diego, CA: Academic Press.

Blom, G. (1958). *Statistical estimates and transformed beta variables*. New York, NY: Wiley.

Caravita, S., DiBlasio, P., & Salmivalli, C. (2009). Unique and interactive effects of empathy and social status on involvement in bullying. *Social Development, 18,* 140–163. doi:10.1111/j.1467-9507.2008.00465.x

Cillessen, A. H. N., & Mayeux, L. (2004). From censure to reinforcement: Developmental changes in the association between aggression and social status. *Child Development, 75,* 147–163. doi:10.1111/j.1467-8624.2004.00660.x

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing-data procedures. *Psychological Methods, 6,* 330–351. doi:10.1037/1082-989X.6.4.330

Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Mason, OH: Cengage.

Enders, C. K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.

Enders, C. K., & Tofighi, D. (2007). Centering predictor variables in cross-sectional multilevel models: A new look at an old issue. *Psychological Methods, 12,* 121–138. doi:10.1037/1082-989X.12.2.121

Eslea, M., & Smith, P. K. (1998). The long-term effectiveness of antibullying work in primary schools. *Educational Research, 40,* 203–218. doi:10.1080/0013188980400208

Farrington, D. P., & Ttofi, M. M. (2010). School-based programs to reduce bullying and victimization. *Campbell Collaboration Library of Systematic Reviews, 6*. Retrieved from http://www.campbellcollaboration.org/library.pl

Ferguson, C., San Miguel, C., Kilburn, J., & Sanchez, P. (2007). The effectiveness of school-based anti-bullying programs: A meta-analytic review. *Criminal Justice Review, 32,* 401–414. doi:10.1177/0734016807311712

Hanewinkel, R. (2004). Prevention of bullying in German schools: An evaluation of an anti-bullying approach. In P. K. Smith, D. Pepler, & K. Rigby (Eds.), *Bullying in schools: How successful can interventions be?* (pp. 81–98). Cambridge, England: Cambridge University Press. doi:10.1017/CBO9780511584466.006

Jeličić, H., Phelps, E., & Lerner, R. M. (2009). Use of missing data methods in longitudinal studies: The persistence of bad practices in developmental psychology. *Developmental Psychology, 45,* 1195–1199. doi:10.1037/a0015665

Juvonen, J., & Galván, A. (2008). Peer influence in involuntary social

groups: Lessons from research on bullying. In M. J. Prinstein & K. A. Dodge (Eds.), *Understanding peer influence in children and adolescents* (pp. 225–244). New York, NY: Guilford Press.

Juvonen, J., Graham, S., & Schuster, M. A. (2003). Bullying among young adolescents: The strong, the weak, and the troubled. *Pediatrics, 112,* 1231–1237. doi:10.1542/peds.112.6.1231

Kallestad, J. H., & Olweus, D. (2003). Predicting teachers' and schools' implementation of the Olweus Bullying Prevention Program: A multilevel study. *Prevention & Treatment, 6,* 3–21. doi:10.1037/1522-3736.6.1.621a

Kärnä, A., Voeten, M., Little, T. D., Poskiparta, E., Alanen, E., & Salmivalli, C. (2011). Going to scale: A nonrandomized nationwide trial of the KiVa antibullying program for Grades 1–9. *Journal of Consulting and Clinical Psychology, 79,* 796–805. doi:10.1037/a0025740

Kärnä, A., Voeten, M., Little, T. D., Poskiparta, E., Kaljonen, A., & Salmivalli, C. (2011). A large-scale evaluation of the KiVa antibullying program: Grades 4–6. *Child Development, 82,* 311–330. doi:10.1111/j.1467-8624.2010.01557.x

Kärnä, A., Voeten, M., Poskiparta, E., & Salmivalli, C. (2010). Vulnerable children in varying classroom contexts: Bystanders' behaviors moderate the effects of risk factors on victimization. *Merrill–Palmer Quarterly, 56,* 261–282. doi:10.1353/mpq.0.0052

Menesini, E., Codecasa, E., Benelli, B., & Cowie, H. (2003). Enhancing children's responsibility to take action against bullying: Evaluation of a befriending intervention in Italian middle schools. *Aggressive Behavior, 29,* 1–14. doi:10.1002/ab.80012

Merrell, K., Gueldner, B., Ross, S., & Isava, D. (2008). How effective are school bullying intervention programs? A meta-analysis of intervention research. *School Psychology Quarterly, 23,* 26–42. doi:10.1037/1045-3830.23.1.26

Obstbaum, Y. (2006). *Brottslighet bland finskspråkiga och svenskspråkiga ungdomar* [Crime among Finnish-speaking and Swedish-speaking young people] (Oikeuspoliittisen tutkimuslaitoksen tutkimustiedonantoja 69) [National Research Institute of Legal Policy Research Communications 69]. Helsinki, Finland: National Research Institute of Legal Policy.

Olweus, D. (1996). *The Revised Olweus Bully/Victim Questionnaire*. Bergen, Norway: University of Bergen, Research Center for Health Promotion (HEMIL Center).

Olweus, D. (2004). The Olweus Bullying Prevention Programme: Design and implementation issues and a new national initiative in Norway. In P. K. Smith, D. Pepler, & K. Rigby (Eds.), *Bullying in schools: How successful can interventions be?* (pp. 13–36). Cambridge, England: Cambridge University Press. doi:10.1017/CBO9780511584466.003

Olweus, D. (2005). *New positive results with the Olweus Bullying Prevention Program in 37 Oslo schools*. Unpublished report, University of Bergen, Research Center for Health Promotion (HEMIL Center), Bergen, Norway.

Olweus, D., & Alsaker, F. (1991). Assessing change in a cohort-longitudinal study with hierarchical data. In D. Magnusson, L. Bergman, G. Rudinger, & B. Törestad (Eds.), *Problems and methods in longitudinal research: Stability and change* (pp. 107–132). New York, NY: Cambridge University Press. doi:10.1017/CBO9780511663260.008

Pikas, A. (1989). The common concern method for the treatment of mobbing. In E. Roland & E. Munthe (Eds.), *Bullying: An International Perspective* (pp. 91–104). London, England: David Fulton.

Pitts, J., & Smith, P. K. (1995). *Preventing school bullying*. London, England: Home Office.

Polanin, J. R., Espelage, D. L., & Pigott, T. D. (2012). A meta-analysis of school-based bullying prevention programs' effects on bystander intervention behavior. *School Psychology Review, 41,* 47–65.

Pöyhönen, V., Juvonen, J., & Salmivalli, C. (2010). What does it take to stand up for the victim of bullying? The interplay between personal and social factors. *Merrill–Palmer Quarterly, 56,* 143–163. doi:10.1353/mpq.0.0046

Pöyhönen, V., & Salmivalli, C. (2008). New directions in research and practice addressing bullying: Focus on defending behavior. In D. Pepler & W. Craig (Eds.), *An international perspective on understanding and addressing bullying* (PREVNet Publication Series, 1, pp. 26–43). Bloomington, IN: AuthorHouse.

Rasbash, J., Charlton, C., Browne, W. J., Healy, M., & Cameron, B. (2009). *MLwiN (Version 2.11) [Computer software]*. Bristol, England: University of Bristol, Centre for Multilevel Modelling.

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Robinson, G., & Maines, B. (1997). *Crying for help: The no blame approach to bullying*. Bristol, England: Lucky Duck.

Rodkin, P. C., Farmer, T. W., Pearl, R., & Van Acker, R. (2000). Heterogeneity of popular boys: Antisocial and prosocial configurations. *Developmental Psychology, 36,* 14–24. doi:10.1037/0012-1649.36.1.14

Sainio, M., Kaukiainen, A., Willför-Nyman, U., Annevirta, T., Pöyhönen, V., & Salmivalli, C. (2009). *KiVa: Teacher's guide, Unit 3 (Research into Practice Publication Series, No. 4)*. Turku, Finland: University of Turku, Psychology Department.

Salmivalli, C. (2010). Bullying and the peer group: A review. *Aggression and Violent Behavior, 15,* 112–120. doi:10.1016/j.avb.2009.08.007

Salmivalli, C., Garandeau, C., & Veenstra, R. (2012). KiVa anti-bullying program: Implications for school adjustment. In A. M. Ryan & G. W. Ladd (Eds.), *Peer relationships and adjustment at school* (pp. 279–305). Charlotte, NC: Information Age.

Salmivalli, C., Kärnä, A., & Poskiparta, E. (2010a). Development, evaluation, and diffusion of a national anti-bullying program KiVa. In B. Doll, W. Pfohl, & J. Yoon (Eds.), *Handbook of youth prevention science* (pp. 238–252). New York, NY: Routledge.

Salmivalli, C., Kärnä, A., & Poskiparta, E. (2010b). From peer putdowns to peer support: A theoretical model and how it translated into a national anti-bullying program. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *Handbook of bullying in schools: An international perspective* (pp. 441–454). New York, NY: Routledge.

Salmivalli, C., Kärnä, A., & Poskiparta, E. (2011). Counteracting bullying in Finland: The KiVa program and its effects on different forms of being bullied. *International Journal of Behavioral Development, 35,* 405–411. doi:10.1177/0165025411407457

Salmivalli, C., Kaukiainen, A., & Voeten, M. (2005). Anti-bullying intervention: Implementation and outcome. *British Journal of Educational Psychology, 75,* 465–487. doi:10.1348/000709905X26011

Salmivalli, C., Lagerspetz, K., Björkqvist, K., Österman, K., & Kaukiainen, A. (1996). Bullying as a group process: Participant roles and their relations to social status within the group. *Aggressive Behavior, 22,* 1–15. doi:10.1002/(SICI)1098-2337(1996)22:1<1::AID-AB1>3.0.CO;2-T

Salmivalli, C., & Peets, K. (2008). Bullies, victims, and bully–victim relationships. In K. Rubin, W. Bukowski, & B. Laursen (Eds.), *Handbook of peer interactions, relationships, and groups* (pp. 322–340). New York, NY: Guilford Press.

Salmivalli, C., Poskiparta, E., Tikka, A., & Pöyhönen, V. (2009). *KiVa: Teacher's guide, Unit 1 (Research into Practice Publication Series, No. 2)*. Turku, Finland: University of Turku, Department of Psychology.

Salmivalli, C., & Voeten, M. (2004). Connections between attitudes, group norms, and behaviors associated with bullying in schools. *International Journal of Behavioral Development, 28,* 246–258. doi:10.1080/01650250344000488

Salmivalli, C., Voeten, M., & Poskiparta, E. (2011). Bystanders matter: Associations between defending, reinforcing, and the frequency of bullying in classrooms. *Journal of Clinical Child and Adolescent Psychology, 40,* 668–676. doi:10.1080/15374416.2011.597090

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7,* 147–177. doi:10.1037/1082-989X.7.2.147

Shadish, W., Cook, T., & Campbell, D. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston, MA: Houghton–Mifflin.

Smith, J. D., Schneider, B. H., Smith, P. K., & Ananiadou, K. (2004). The effectiveness of whole-school anti-bullying programs: A synthesis of evaluation research. *School Psychology Review, 33,* 547–560.

Smith, P. K. (2010). Bullying in primary and secondary schools: Psychological and organizational comparisons. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *Handbook of bullying in schools: An international perspective* (pp. 137–150). New York, NY: Guilford Press.

Smith, P. K., & Sharp, S. (Eds.). (1994). *School bullying: Insights and perspectives*. New York, NY: Routledge. doi: 10.4324/9780203425497

Snijders, T. A. B., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. London, England: Sage.

Solberg, M. E., & Olweus, D. (2003). Prevalence estimation of school bullying with the Olweus Bully/Victim Questionnaire. *Aggressive Behavior, 29,* 239–268. doi:10.1002/ab.10047

Stevens, V., De Bourdeaudhuij, I., & Van Oost, P. (2000). Bullying in Flemish schools: An evaluation of anti-bullying intervention in primary and secondary schools. *British Journal of Educational Psychology, 70,* 195–210. doi:10.1348/000709900158056

Whitney, I., Rivers, I., Smith, P. K., & Sharp, S. (1994). The Sheffield Project: Methodology and findings. In P. K. Smith & S. Sharp (Eds.), *School bullying: Insights and perspectives* (pp. 20–56). London, England: Routledge.

Whitney, I., & Smith, P. K. (1993). A survey of the nature and extent of bullying in junior/middle and secondary schools, *Educational Research, 35*. 3–25. doi:10.1080/0013188930350101

Williford, A., Boulton, A., Noland, B., Kärnä, A., Little, T. D., & Salmivalli, C. (2012). Effects of the KiVa Anti-Bullying Program on adolescents' perception of peers, depression, and anxiety. *Journal of Abnormal Child Psychology, 40,* 289–300. doi:10.1007/s10802-011-9551-1

Received November 22, 2011
Revision received August 9, 2012
Accepted August 28, 2012 ■

Correction to Kärnä et al. (2012)

The article "Effectiveness of the KiVa Antibullying Program: Grades 1–3 and 7–9," by Antti Kärnä, Marinus Voeten, Todd D. Little, Erkki Alanen, Elisa Poskiparta, and Christina Salmivalli (*Journal of Educational Psychology*, Advance online publication. October 22, 2012. doi: 10.1037/a0030417) omitted some wording in the text. The sentence in the first paragraph below Table 5 beginning with, "More specifically, the model . . . ." should have read "More specifically, the model was defined as follows: $\hat{Y}_{tijk}$ is the predicted value, t is used to indicate time points, i is used for individual students, j is used to denote classrooms, and k to denote schools:"

DOI: 10.1037/a0031120