EFFECTIVENESS OF

THE KIVA ANTIBULLYING PROGRAM

by Antti Kärnä

From the Department of Psychology

University of Turku, Finland

Supervised by

Christina Salmivalli, PhD
Professor
Department of Psychology
University of Turku
Finland

Elisa Poskiparta, PhD
Senior researcher
Department of Psychology
University of Turku
Finland

Opponent

David P. Farrington, PhD
Professor
Institute of Criminology
Cambridge University
UK

Reviewed by

Marijtje A. J. van Duijn, PhD
Associate Professor
Department of Sociology
University of Groningen
The Netherlands

David P. Farrington, PhD
Professor
Institute of Criminology
Cambridge University
UK

Nancy Guerra, PhD
Professor
College of Arts and Sciences
University of Delaware
USA

Effectiveness of
the KiVa Antibullying Program

Antti Kärnä
Department of Psychology
University of Turku
Finland

ABSTRACT

According to the participant role approach (Salmivalli, Lagerspetz, Björkqvist, Österman, & Kaukiainen, 1996), bullying is a group phenomenon that is largely enabled and maintained by the classmates taking on different participant roles (e.g., reinforcers or assistants of the bully). There is, however, very little evidence on whether the bystander behaviors actually have an effect on the risk for victimization. Furthermore, the participant role approach implies that the bystanders should be used in putting an end to bullying. This view has been put into practice in the KiVa antibullying program, but it has not yet been investigated whether the program is effective. Four studies were conducted to investigate, (a) whether the behaviors of bystanders have an effect on the risk for victimization (Study I) and (b) whether the KiVa program reduces bullying and victimization and has other beneficial effects as well (Studies II–IV). The participants included large samples of elementary and lower secondary school students (Grades 1–9) from Finland. The assessments were done with web-based questionnaires including questions about bullying and victimization (both self- and peer reports), and about several bullying-related constructs. The results of this thesis suggest that bystander behaviors in bullying situations may influence the risk for victimization of vulnerable students. Moreover, the results indicate that the KiVa antibullying program is effective in reducing victimization and bullying. The program effects are larger in elementary schools than in lower secondary schools, whereas in Grades 8 and 9, they are larger for boys than girls for some peer-reported outcomes. The magnitude of the overall effects can be considered practically significant when obtained in a large-scale dissemination of the program.

CONTENTS

ACKNOWLEDGEMENTS

I am indebted to Professors David P. Farrington, Marijtje A. J. van Duijn, and Nancy Guerra who reviewed an earlier draft of this work and gave useful comments on it. In addition, Professor Farrington has agreed to serve as my opponent, which I greatly appreciate.

My parents Helmi and Pentti as well as my parents-in-law Vuokko and Aulis have provided important support to our family during the years in which I have focused on these studies. Finally, I wish to thank my daughter Heli and wife Eveliina for bringing joy and meaning to my life. I feel very lucky in having you by my side.

Turku, May, 2012

Antti Kärnä

LIST OF ORIGINAL PUBLICATIONS

I. **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.

II. **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.

III. **Kärnä, A.**, Voeten, M., Little, T. D., Alanen, E., Poskiparta, E., & Salmivalli, C. (2012). Effectiveness of the KiVa antibullying program: Grades 1–3 and 7–9. Manuscript submitted for publication.

IV. **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.

1. THEORETICAL BACKGROUND

1.1. SCHOOL BULLYING

Bullying and victimization are common problems in schools all over the world (e.g., Cook, Williams, Guerra, & Kim, 2010; Craig & Harel, 2004). For instance, in Finnish schools 1 in 6 school-aged children and adolescents are victimized and 1 in 10 bullies others (Study IV). This is a topic of public concern, because bullying is a particularly condemnable behavior distinguished by repeated acts of aggression against weaker victims who cannot easily defend themselves (Olweus, 1994). Victims of bullying suffer concurrently from loneliness and several kinds of internalizing symptoms (e.g., depression, anxiety, and low self-worth; Hawker & Boulton, 2000). Longitudinally, victimization has a substantial contribution to later risk for depression (Ttofi, Farrington, Lösel, & Loeber, 2011a) and anxiety disorder (Sourander et al., 2007), whereas bullying is a major risk factor for later criminal offending (Ttofi, Farrington, Lösel, & Loeber, 2011b) and antisocial personality disorder (Sourander et al., 2007). It is therefore imperative (a) to formulate and test theories of bullying and (b) to develop and evaluate theory-based intervention programs. In this thesis I examined the participant role approach to bullying (Salmivalli, Lagerspetz, Björkqvist, Österman, & Kaukiainen, 1996) and the effectiveness of the new KiVa antibullying program based on this view.[1]

1.2. PARTICIPANT ROLE APPROACH TO BULLYING

When a student is bullied at school, there are peer witnesses present in most of the incidents (Hawkins, Pepler & Craig, 2001; O'Connell, Pepler, & Craig, 1999; Salmivalli et al., 1996). The presence of bystanders has led researchers to investigate their reactions to the events and how they might contribute to the continuation or discontinuation of bullying. Salmivalli and her colleagues (1996) introduced a *participant role approach* in which bullying is viewed as a group phenomenon. This approach is based on the assumption that the bystanders adopt different participant roles in bullying situations: Assistants help the ringleader bullies; reinforcers provide positive feedback to the bullies (e.g., by cheering); outsiders withdraw from the situation; defenders try to comfort, support, and defend the victims. Accordingly, bullying can be viewed as a strategy for gaining and demonstrating a powerful position in the peer group (e.g., Juvonen & Galván, 2008; Salmivalli & Peets, 2008). Bystanders have a significant role in the process, because they in part maintain bullying behavior by assisting and reinforcing the

---

[1] As the original studies were conducted with the help of co-authors, the pronoun "we" is used instead of "I" in most of the text.

bully and thereby providing bullies the sought-after position of power. In contrast, defending the victim turns bullying into an unsuccessful strategy for attaining and demonstrating high status.

After the first Finnish studies (e.g., Salmivalli et. al., 1996; Salmivalli, Huttunen, & Lagerspetz, 1997; Salmivalli, Lappalainen, & Lagerspetz, 1998) several researchers in various countries (e.g., Andreou & Metallidou, 2004; Camodeca & Goossens, 2005, Goossens, Olthof, & Dekker, 2006; Menesini, Codecasa, Benelli, & Cowie, 2003; Schäfer & Korn, 2004; Sutton & Smith, 1999) have adopted a similar conceptualization of bullying. The studies on the participant roles have mainly (a) investigated the psychometric properties of the measures (Goossens, Olthof, & Dekker, 2006; Sutton & Smith, 1999), (b) described the students acting in various roles (Salmivalli et al., 1996), or (c) tried to explain, what makes students behave in different ways in bullying situations (Salmivalli & Voeten, 2004). Much less is known, however, about the potential effects of bystander behaviors on bullying or victimization. Prior to Study I, the only empirical studies on the effects of bystander behaviors were small-scale observational studies (Hawkins et al., 2001; O'Connell et al., 1999), in which the researchers observed students' behavior in the school yard. The evidence from these studies is limited, however, by the fact that the data were based on only brief episodes. This makes it impossible to determine, whether the interaction patterns actually were repeated, as required by the definition of bullying (Olweus, 1994).

Study I provides a preliminary test of some of the ideas behind the participant role approach. If the approach is valid, it is possible that the effects of risk factors of victimization (e.g., anxiety or peer rejection) depend on bystander behaviors. Specifically, it may be the case that reinforcing encourages bullying vulnerable peers, whereas defending serves to discourage bullying. These hypotheses are based on the view that knowledge of the likely positive or negative social consequences of actions (i.e., bullying) influences future student behavior (Bandura, 1977).

1.3. IMPLICATIONS OF THE PARTICIPANT ROLE APPROACH

The participant role approach (Salmivalli et al., 1996) has clear implications for designing antibullying intervention programs. The interventions should be targeted to changing bystander behaviors, because this will reduce the rewards gained by bullies and consequently their motivation for bullying. Research has found that empathy, self-efficacy, and antibullying attitudes predict defending and supporting the victimized peers (Caravita, DiBlasio, & Salmivalli, 2009; Pöyhönen, Juvonen, & Salmivalli, 2010; Salmivalli & Voeten, 2004), and therefore it may be possible to bring about positive change in the

bystander behaviors by increasing these characteristics in students. Furthermore, there is a discrepancy between students' generally negative attitudes towards bullying (Boulton, Bucci, & Hawker, 1999; Menesini et al., 1997; Rigby & Slee, 1991) and their actual behavior in bullying situations (e.g., Salmivalli et al., 1996), and this may be an important factor contributing to the persistence of the problem. By targeting the group norms about bullying we may be able to help students to act in accordance with their antibullying attitudes.

These implications have been put into practice in the new KiVa antibullying program (Salmivalli, Kärnä, & Poskiparta, 2010a; Salmivalli, Kärnä, & Poskiparta, 2010b). The KiVa program attempts to turn the bystanders into supporters and defenders of victims. The program also includes procedures for handling acute bullying cases, and it therefore consists of both universal (targeted at all students) and indicated (targeted at students involved in bullying) interventions. Specifically, it is assumed that the program reduces bullying and victimization (a) by encouraging students to support and defend the victimized peers, (b) by communicating to children that bullying others cannot be tolerated, and (c) by providing adults (parents and school personnel) information about bullying, and by enhancing their efficacy to prevent it and to intervene in it.

1.4. EFFECTIVENESS OF THE ANTIBULLYING INTERVENTION PROGRAMS

During the last 20 years, there has been a marked increase of studies on the effectiveness of antibullying intervention programs (Farrington & Ttofi, 2009). The results of these research reports have been summarized in several reviews and meta-analyses (Baldry & Farrington, 2007; Farrington, & Ttofi, 2009; Ferguson, San Miguel, Kilburn, & Sanchez, 2007; Merrell, Gueldner, Ross, & Isava, 2008; Polanin, Espelage, & Pigott, 2012; J. D. Smith, Schneider, Smith, & Ananiadou, 2004; P. K. Smith, Ananiadou, & Cowie, 2003; Vreeman & Carroll, 2007). The authors of the most comprehensive and authoritative meta-analysis (Farrington & Ttofi, 2009) concluded that the intervention programs are, in general, effective in reducing the prevalence of bullying and victimization. The interventions reduced the prevalence of bullying and victimization by 20–23% and 17–20%, respectively. There was, however, considerable variability in results across studies: In addition to positive results, there were several interventions with statistically nonsignificant effects, and even one with negative effects. This variability may result partly from differences in the intervention components and in the target populations. It was found, for example, that formal engagement of peers in antibullying work was

associated with *increases* in victimization and bullying (the increase for bullying being, however, statistically nonsignificant).

Although these results do not demonstrate the causes of effectiveness, they nevertheless suggest that "work with peers should not be used" (Farrington & Ttofi, 2009). This is in contrast with the views described above (e.g., Salmivalli et al., 2010a, 2010b). Furthermore, Farrington and Ttofi (2009) found that the program effects are larger for older children and concluded that interventions should be targeted to students of aged 11 years or above. This view has been contested by P. K. Smith (2010) who argued, based on a different sample of studies, that the interventions are actually less effective for older students and for those attending secondary schools.

In addition to bullying and victimization, evaluation studies for antibullying programs have sometimes included other outcomes as well (Merrell et al., 2008; Polanin et al., 2012). Merrell and his colleagues (2008) calculated average effect sizes for a wide range of outcomes. For instance, the effect sizes for attitudes towards bullying (Cohen's $d = 0.15$; 4 studies) and for intervening to stop bullying ($d = 0.17$; 3 studies) were in the desired direction, but sympathy towards victims seemed to decrease after intervention ($d = -0.10$; 2 studies). Yet another recent meta-analysis summarized the effectiveness of intervention programs in terms of increasing both bystander intervention in bullying situations and empathy for the victim (Polanin et al., 2012).[2] They concluded that the programs are successful in enhancing intervening behavior (average Hedges' $g = 0.20$, 95% CI [0.11, 0.29]; 11 studies), but the program effects were smaller for younger (Grades 3–8; $g = 0.14$, 95% CI [0.11, 0.18]; 8 studies) than for older students (Grades 9–12; $g = 0.43$, 95% CI [0.33, 0.52]; 4 studies). The average effect size for empathy for the victim, however, was rather close to zero ($g = 0.05$, 95% CI [−0.07 to 0.17]; 8 studies). As noted by the authors, all these effect sizes must be interpreted with caution, because only a small number of studies was included in the calculations (Merrell et al., 2008; Polanin et al., 2012).

Overall, it seems clear that antibullying programs are on average effective for reducing bullying and victimization (Farrington & Ttofi, 2009). It is nevertheless important to continue investigating the effectiveness of these programs, for several reasons. First, the promising average results from the meta-analysis by Farrington and Ttofi (2009) do not inevitably generalize to any particular program, and the effectiveness of a new antibullying program must be established by empirical research. Second, the

---

[2] The meta-analysis included the KiVa program results for Grades 4–6.

meta-analysis results (Farrington & Ttofi, 2009) challenge the view that peers should be used in tackling bullying. When an intervention (like the KiVa program) is based on influencing the bystanders, it is important to demonstrate that the program is effective. Third, there are some conflicting views (Farrington & Ttofi, 2009; P. K. Smith, 2010) on at which grade levels the programs are most effective, and further research is needed to provide evidence about this issue. Fourth, because gender is known to be an important factor in bullying-related phenomena (e.g., Salmivalli & Voeten, 2004), it should be investigated whether the effects of antibullying programs differ by gender. Fifth, additional research is needed to find out whether the antibullying programs have beneficial effects on other bullying-related outcomes as well (e.g., by increasing empathy for the victim).

The purpose of the Studies II–IV in this thesis was to investigate the issues identified above. Specifically, we examined (a) whether the KiVa program is effective and (b) whether the KiVa program effects depend on students' age, gender, or both of them. In all these studies the main outcome variables were bullying and victimization, but in Studies II and III also several other bullying-related outcomes were included.

2. PURPOSE OF THE STUDY

The main purpose of this thesis was (a) to provide a preliminary test of the ideas behind the participant role approach and (b) to investigate the effectiveness of the new KiVa antibullying program based on this approach. More specifically, the study questions were as follows:

1. Do bystander behaviors in bullying situations influence the risk for victimization of vulnerable students? That is, are the effects of risk factors for victimization (social anxiety and peer rejection) stronger in classrooms with a high level of reinforcing the bully and a low level of defending the victims? (Study I)

2. Does the KiVa antibullying program reduce bullying and victimization and does it have positive effects on other related outcomes in Grades 4–6? (Study II)

3. Does the KiVa antibullying program reduce bullying and victimization in Grades 1–3 and 7–9? Does the Kiva program have positive effects on other related outcomes in Grades 7–9? (Study III)

4. Does the Kiva program reduce bullying and victimization when it is disseminated widely throughout Finland? (Study IV)

In Studies II–IV it was further investigated, whether the KiVa program effects depend on gender or age of the students.

3. METHOD

3.1. STUDY SAMPLES

The samples of the four studies consisted of elementary and lower secondary school students in Finland, a country with approximately 5 million inhabitants. The students were located in schools that took part in the evaluation study of the KiVa antibullying program (2007–2009), the first year of the national KiVa program dissemination (2009–2010), or both of them. The participants were in Grades 1–9, and they were, on average, 7–15 years old in the beginning of the school year.

3.1.1. Sampling and design: Evaluation study

To recruit schools, letters were sent 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 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. Altogether 275 schools volunteered to participate in the study. A few enrolled schools provided only special education, and they were excluded from the sample.

The first phase of program evaluation (2007–2008) focused on Grades 4–6. The volunteering schools were stratified by province and language and 78 of them were randomly assigned to intervention or control condition. In 15% of the sample schools, teaching was given in Swedish (cf. 9% in the population).

The second phase of the program evaluation (2008–2009) focused on both Grades 1–3 and 7–9. Altogether 31 control schools from the previous phase participated now in the intervention group. In addition, the volunteering schools were stratified by province and language and 125 of them were randomly assigned to intervention (47 schools) or control condition (78 schools). This procedure resulted in 156 participating schools: 79 schools (40 control and 39 intervention) for Grades 1–3 and in 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. In 13% of the sample schools, teaching was given in Swedish (cf. 9% in the population).

3.1.2. Sampling and design: Dissemination study

The school recruitment for the first year of national dissemination (2009–2010) took place in October 2008. All basic education schools in mainland Finland received letters that provided information about

the program, and they explained what kind of efforts the schools should be ready to invest in the implementation. Of the 3,218 schools contacted, 1,827 were willing to adopt the program. Because of limited resources, the first 1,450 schools were included into the first round of teacher training and program implementation. Of these schools, 1,189 organized the web-based survey in May 2009 and 888 (74.7%) of them responded also in May 2010. In order to compare the situation at pretest and posttest in the same schools, only those who participated in both measurements were included into the final sample. In 9% of these schools instruction was given in the Swedish language (cf. 9% in the population.

3.2. PROCEDURE

3.2.1. Data collection: Evaluation study

The school year in Finland ranges from mid-August to the end of May. Data were collected during two school years: in 2007–2008 for Grades 4–6 and in 2008–2009 for Grades 1–3 and 7–9. There were three waves of measurement: in May, in December–February, and in May (one year after the pretest). Students' parents indicated their consent for participation with forms, which were collected by the teachers. Participants filled out Internet-based questionnaires in the schools' computer labs during regular school hours. The process was administered by the teachers, who were supplied with detailed instructions about two weeks prior to data collection. The teachers were told to act in such ways that the confidentiality of the responses was secured to a maximum extent, and both younger and older students were assured that their answers are not revealed to teachers or parents. In addition, teachers were offered support through phone or e-mail prior to and during data collection.

At the beginning of the session, teachers distributed individual passwords to the students, who used them to log in to the questionnaire. The questionnaire started with demographic questions (e.g., age and gender). The term bullying was next defined for the students in the way formulated in the revised Olweus' bully/victim questionnaire (Olweus, 1996), which emphasizes the repetitive nature of bullying and the power imbalance between the bully and the victim. 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 bullying-related questions. The order of questions, items, and scales was extensively randomized to alleviate any systematic order effect.

The questionnaire was accommodated in some respects for the youngest subsample involving Grades 1–3: (a) The definition was shortened and simplified to facilitate understanding of the concept

measured; (b) The order of the questions was the same for all respondents, and the teacher read out loud the questions and the answering options in order to facilitate answering; (c) For self-reported bullying and victimization, the five answering options were provided with different colors. In this way, the teacher who was giving the instructions could also refer to the five colors when helping students pick the right alternative.

3.2.2. Data collection: Dissemination study

In May 2009, the students from all participating schools were invited to fill in a short, anonymous web-based questionnaire about bullying others, being bullied, telling about bullying, attitudes related to bullying, and classroom and school climate. The survey took place before the summer break, and the school year started again in August. One year later (May 2010) the schools were asked to organize the web-based survey again. The second assessment took place in the same time of the following school year, in order to avoid any season effects and to have same-aged students in the grade-wise comparisons. At that time, the KiVa program had been implemented in the participating schools for nine months (i.e., one entire school year). Students filled in the questionnaires in the schools' computer labs during regular school hours, while the teachers administered the process. The teachers had been provided with detailed instructions about the procedure well in advance, including guidelines on how to protect the confidentiality of the responses. No consent forms were collected from the students, because the anonymous KiVa questionnaires were now a part of normal educational practices (American Psychological Association, 2010, p. 233).

At the beginning of the session, the participants logged into the questionnaire with their school-specific passwords, so each student's school identification was automatically saved into the dataset. The participants were asked to choose their gender and grade level from the options appearing on the computer screen. They were then provided with the standard definition of bullying (Olweus, 1996), followed by questions about bullying others and being bullied by others.

3.3. MEASURES

Of the measures described below, peer rejection and social anxiety were used only in Study I. All other variables were used as criterion measures in the evaluation (II–III) or in the dissemination (IV) studies of the KiVa program. Self-reports of bullying and victimization were among the criterion measures for Grades 1–9 in all studies on the KiVa program effects (Studies II–IV), but in the evaluation studies for

Grades 4–9 also other outcomes were assessed. The measurements consisting of responses to several items showed acceptable reliability (Cronbach's α coefficients .69–.95). Exploratory factor analyses were conducted for antibullying attitudes, empathy toward victims, and well-being at school, and they supported one-factor solutions for these scales. For all multi-item measures, scores across different items were averaged to create single indicators of the constructs.

3.3.1. Self-reported bullying and victimization

Self-reported bullying and victimization were measured with 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, 4 = Several times a week). In dichotomizing the measures, students who reported they had been bullied 2–3 times a month, every week, or several times a week (response alternatives 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 (cf. Solberg & Olweus, 2003). The school-level correlations between the global questions and the questions concerning the respective forms of bullying or victimization were substantial and fairly similar in all grade levels 1–9 of the KiVa data (dichotomized global items and averages of forms $r = .65–.87$, $p < .001$; nondichotomized global items and averages of forms $r = .74–.88$, $p < .001$). Furthermore, in Grades 4–9 the correlations between the global self-reported items and peer-reported bullying and victimization were large as well ($r = .52–.77$, $p < .001$ and $r = .46–.75$, $p < .001$ for nondichotomized and dichotomized items, respectively). These results indicate that the measurements with the global items have construct validity.

3.3.2. 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 that 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," "Always finds new ways of harassing the victim"), assisting the bully ("Joins in the bullying, when someone else has started it," "Assists the bully," "Helps the bully, maybe by catching the victim"), reinforcing the bully ("Comes around to watch the situation," "Laughs," "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," "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," "Nasty rumors are spread about him/her". 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.

3.3.3. Antibullying attitudes

The original 20-item Provictim scale (Rigby & Slee, 1991) was modified into a 10-item version to better fit the present context. Students responded on a 5-point scale (0 = I disagree completely, 4 = I agree completely) to items such as: "It's okay to call some kids nasty names."

3.3.4. Empathy toward victims

We used a 7-item empathy scale consisting of items such as "When a bullied child is sad I feel sad as well". Students evaluated how often the statements were true for them, responding on a 5-point scale (0 = never, 4 = always).

3.3.5. Self-efficacy for defending behavior

With the Self-efficacy for defending scale (Pöyhönen et al., 2010), students evaluated how easy or difficult it would be for them to defend and support the victim of bullying. The three items used in the scale were derived from the Participant role questionnaire items for defending behavior, for instance "Trying to make the others stop the bullying would be…" The answers were given on a four-point scale (0 = very difficult for me, 3 = very easy for me).

3.3.6. Students' well-being at school

Students' well-being at school was measured with items that were initially developed by the Finnish National Board of Education (Metsämuuronen & Svedlin, 2004), including general liking of school (e.g., "My school days are generally nice"), academic self-concept (e.g., "Learning brings me joy"), classroom climate (e.g., "There is a good climate in our class"), and school climate (e.g., "I feel safe at school"). Students responded to 14 items on a 5-point scale (0 = I disagree completely, 4 = I agree completely).

3.3.7. Social anxiety

To measure students' social anxiety, we used a 9-item version of the Social Anxiety Scale for Adolescents (La Greca & Lopez, 1998). Students responded on a 5-point Likert-type scale (0 = not at all; 4 = all the time) to items such as "I worry about what others think of me" and "I'm afraid to invite others to do things with me because they might say no."

3.3.8. Peer rejection

Participants were asked to choose from a list of classmates three peers they liked the least. The number of nominations received was divided by the number of peers doing the evaluation, resulting in proportion scores ranging from 0.00 to 1.00.

3.4. PARTICIPANTS

3.4.1. Participants in Study I

The data were collected during the pretest measurements of the evaluation study concerning Grades 4–6 (May 2007). The target sample consisted of 78 schools with 429 classrooms and a total of 8,248 students. At the time of the measurements students were finishing Grades 3–5, and they were, on average, 11 years old. To recruit the children, guardians were sent information letters including a consent form. A total of 7,564 students (92% of the target sample) received active consent to participate, and 7,312 students (89% of the target sample) from 408 classrooms in 77 schools responded to the questionnaire. One entire school dropped out because of problems related to its facilities. Of the respondents, 50.3% were girls and 49.7% were boys. Most students were native Finns (i.e., Caucasian) with only 1.9% being immigrants. Because some questions requiring peer nominations were gathered only in classrooms with at least five students, all classes below this limit were excluded

from this study. Finally, when all the cases with missing data on the explanatory variables were excluded from the analyses, the final sample included 6,980 primary school children from 378 classrooms in 77 schools.

3.4.2. Participants in Study II

Details concerning the pretest measures (Wave 1) are described above (Participants in Study I). By Waves 2 and 3, however, some changes in the student composition had taken place, with 251 students leaving the schools and 463 entering them. Between Waves 1 and 2 two control schools (51 students) dropped out, and five more (640 students) between Waves 2 and 3. There were no missing values in predictor variables, and for outcome variables percentages of missing values were not high, except for control schools at Wave 3 (27–37%). Missing data were imputed using the SAS Proc MI (SAS 9.2; SAS Institute, Cary, NC) utility employing dummy codes for classrooms and for cross-classifications of classrooms as well as all interactions of these dummy codes with study variables. We conducted 100 imputations using the Markov Chain Monte Carlo algorithm, and the means of these 100 imputations were used in the analyses. Students were excluded from the analyses if: (a) they were denied permission to participate in the study but had somehow answered the questionnaire and (b) they left school after Wave 1. The final sample size for the analyses was 8,166 (4,201 in the intervention and 3,965 in the control condition). Altogether, 50.1% of the respondents were girls and 49.9% boys. Most students were native Finns (i.e., Caucasian), but 2.4% of them were immigrants.

3.4.3. Participants in Study III

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 dataset of 74 schools (38 intervention and 36 control). In these 74 schools, there were 7,739 students, of which 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 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, and the final analysis sample consisted of 16,503 students in 1,000 classrooms in 73 schools.

3.4.4. Participants in Study IV

Altogether 1,450 schools started implementing KiVa in August 2009. Of these schools, 1,189 organized a web-based survey in May 2009 and 888 (74.7%) of them responded also in May 2010. To be able to compare the situation at pretest and posttest in the same schools, only those who participated in both measurements were included in the final sample. The sample included practically equal numbers of boys (51%) and girls (49%). Data on socioeconomic status or ethnic background of the students were not collected, but in the Finnish population about 3% of the students are immigrants, whereas the vast majority of Finns are Caucasian. A group of students was excluded from the analyses because of obviously contradictory responding (in total 403 respondents at Waves 1 and 2). This left us with control and intervention samples of 156,634 and 141,103 for victimization and 156,629 and 141,099 for bullying. There were approximately 200,000 students in these 888 schools included in the present study, which resulted in response rates of 78% and 70% at Waves 1 and 2, respectively.

3.5. INTERVENTION

As described above, the participant role approach (Salmivalli et al., 1996) implies that to decrease bullying, it is essential to change the bystander behaviors in bullying situations. The KiVa program consists of both universal and indicated interventions to accomplish these goals (for the program manuals, see Salmivalli, Poskiparta, Tikka, & Pöyhönen, 2009; Salmivalli, Pöyhönen, & Kaukiainen, 2009; and Sainio et al., 2009). The core of the universal components consists of student lessons and virtual learning environments. With these components the KiVa program aims (a) to raise awareness of the role that the group plays in maintaining bullying; (b) to increase empathy towards victims; (c) to promote children's strategies of supporting the victim and their self-efficacy to do so; and (d) to increase children's coping skills when victimized. There are two elementary school versions of the KiVa program (for Grades 1 and 4); both include 10 double lessons (2×45 minutes), which classroom teachers are instructed to implement according to the teacher's manual during the school year (one per month from August till May; Table 1). In the lower secondary school version (for Grade 7), four

themes are described in the teachers' manual that can be introduced to students as a series of lessons, whole theme days, or otherwise. The themes are (a) "Group interaction", (b) "Me and the others", (c) "Forms of bullying", and (d) "The consequences and counterforces of bullying". The recommended time to be dedicated for the Kick-Off session, the four themes, and the concluding session, is 13–23 hours in total.

Table 1

The lesson titles in the two elementary school versions of the KiVa program

Grade 1	Grade 4
10 double lessons	10 double lessons
1. Let's get to know each other	1. Respect is for everyone
2. Emotions	2. In a group
3. Our class — everyone is included!	3. Recognize bullying!
4. Difference is a richness	4. Hidden forms of bullying
5. There is no bullying in KiVa school	5. Consequences of bullying
6. We won't join in bullying!	6. Group involvement in bullying
7. The victim needs your support	7. Confronting bullying as a group
8. I will not be bullied!	8. What to do if I get bullied?
9. Literature lesson	9. KiVa school — let's do it together!
10. KiVa contract	10. How are we doing?

Note. During the evaluation study, the grade-1 and grade-4 versions were implemented also in Grades 2–3 and 5–6, respectively (see below).

In addition to student lessons, the KiVa program utilizes virtual learning environments: two different versions of an antibullying computer game (Grades 1 and 4) and an Internet forum "KiVa Street" (Grade 7). The virtual learning environments are closely connected to the topics of the lessons and themes, enhancing the learning process and motivating students to apply the learned skills in everyday interactions with peers. Furthermore, the universal actions include a parents' guide as well as symbols (posters, highly visible vests for teachers supervising recess time) reminding both students and school personnel about KiVa.

The indicated actions involve discussions with bullies and victims, and with selected prosocial classmates as well, who are invited to support the victimized classmate. The discussions with the bullies and victims are conducted by a team of three teachers or other school personnel, called a KiVa team, whereas the classroom teacher organizes separate meetings with potential supporters of the victim (for a more detailed description of the program contents, see Salmivalli et al., 2010a, 2010b).

There were three notable differences between the evaluation and dissemination studies in how the intervention was implemented. First, during the evaluation study, the student lessons and the virtual learning environments were used in all grades in the intervention schools, with the different age-appropriate versions in Grades 1–3, 4–6, and 7–9, respectively. In the dissemination study, however, the lessons and the virtual learning environments were intended only for grades one, four, and seven, which is in line with the present recommendations for implementing the KiVa program.[3] In both evaluation and dissemination studies the discussions with bullies and victims were prescribed for each bullying case detected. Second, during the evaluation study, school networks were created to enhance program implementation. Three teachers or other personnel formed a KiVa team for each school, and the KiVa teams of three schools in a same geographical area formed a school network. The school network met, together with a key person representing the program providers, three times during the intervention year. The meetings were designed to facilitate sharing ideas and experiences, provide support in implementation, and to enhance the motivation to implement the program with fidelity. During the dissemination period, there was no regular face-to-face contact between the schools and the program developers, except for the two-day preimplementation training. Third, during the evaluation study the schools were randomly assigned to two conditions utilizing different approaches to discussions with the bullies: the No Blame Approach and the Confronting Approach. During the dissemination the school teams could freely choose between the two.

---

[3] The actual implementation during the first dissemination year may have differed from the recommendations. For instance, some schools told that they started implementing KiVa in all grade levels.

3.6. STATISTICAL ANALYSES

Statistical analyses and the software used (IBM SPSS Statistics, MLwiN, and Mplus) varied according to the purpose of the particular study. Because of the nested structure of the data, multilevel modeling (e.g., Snijders & Bosker, 1999) was used in Studies I–III, and the standard errors were corrected for clustering in Study IV. Multilevel regression models are 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 and because variables at different levels can be adequately dealt with (Snijders & Bosker, 1999). Depending on the study design and practical considerations, we used varying strategies to deal with missing data: listwise deletion (Study I), missing value imputation (Study II), full information maximum likelihood estimation (FIML; Study III), and sensitivity analysis (Study IV).

4. OVERVIEW OF EMPIRICAL STUDIES

STUDY I

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.**

The main purpose of the study was to examine whether the bystander behaviors in bullying situations influence risk for victimization of vulnerable students. Specifically, victimization was predicted with its two risk factors, social anxiety and peer rejection, and we investigated whether their effects are stronger in classrooms with a high level of reinforcing the bully and a low level of defending the victims. The sample consisted of 6,980 Finnish primary school children in the end of Grades 3–5 (aged 10–12 years), who were nested within 378 classrooms in 77 schools. In spring 2007, these students filled out Internet-based questionnaires in their schools' computer labs. The term bullying was defined to the students in the way formulated in the Olweus' (1996) bully/victim questionnaire with an emphasis on the repetitive nature of bullying and on the power imbalance between the bully and the victim. Victimization, reinforcing, defending, and peer rejection were measured with peer nominations, and classroom averages were calculated from the reinforcing and defending scores to represent the levels of these behaviors in a classroom. Social anxiety was measured with self-reports using a 5-point Likert-type scale. We used two-level regression analysis with random slopes to model the between-classroom variance of the slopes of victimization on its risk factors. The results from multilevel models indicated that, in accordance with the hypotheses, the associations between victimization and social anxiety and peer rejection were strongest in classrooms that were high in reinforcing bullying and low in defending the victims. This suggests that bystander behaviors in bullying situations can strengthen or weaken the effects of individual and interpersonal risk factors for victimization. It might therefore be possible to protect vulnerable children from victimization by decreasing reinforcing behaviors and by increasing defending behaviors.

STUDY II

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.**

In this study we examined whether the KiVa antibullying program (a) is effective in reducing bullying and victimization and (b) has positive effects on several other related outcomes. The sample consisted of 8,166 youth from Grades 4–6 (aged 10–12 years) who were nested in 77 schools that had been randomly assigned to intervention (39 schools, 4,201 students) and control conditions (38 schools, 3,965 students). Students filled out Internet-based questionnaires in the schools' computer labs during regular school hours. They responded to questions concerning self- and peer-reported bullying and victimization, and they also gave peer nominations on assisting the bully, reinforcing the bully and defending the victims. Furthermore, the students reported their antibullying attitudes, empathy toward victims, self-efficacy for defending, and well-being at school. The intervention effects were examined with four-level regression models. The results indicated that after 9 months of implementation, the intervention had beneficial effects on 7 of the 11 dependent variables: self-reported victimization (Cohen's $d = 0.17$), peer-reported victimization ($d = 0.33$), self-reported bullying ($d = 0.10$), assisting the bully ($d = 0.14$), reinforcing the bully ($d = 0.17$), self-efficacy for defending ($d = 0.08$), and well-being at school ($d = 0.10$). Overall, the results indicate that the KiVa program is effective in Grades 4–6 in reducing bullying and victimization and that it has positive effects on several other related outcomes as well.

STUDY III

Kärnä, A., Voeten, M., Little, T. D., Alanen, E., Poskiparta, E., & Salmivalli, C. (2012). Effectiveness of the KiVa antibullying program: Grades 1–3 and 7–9. Manuscript submitted for publication.

We conducted a study to investigate the effectiveness of the KiVa antibullying program for Grades 1–3 (aged 7–9 years) and Grades 7–9 (13–15 years). The program effects on self-reported bullying and victimization were examined in both age groups, whereas the outcomes for Grades 7–9 included also peer-reported behaviors. The two samples included a large number of students from Grades 1–3 ($N = 6,927$; 74 schools) and Grades 7–9 ($N = 16,503$; 73 schools). The schools had been randomly assigned to intervention and control conditions. Participants filled out Internet-based questionnaires in the

schools' computer labs during regular school hours. Students in Grades 1–3 responded to questions concerning self-reported bullying and victimization. In addition to the same self-reports, students in Grades 7–9 gave peer nominations on bullying, victimization, assisting the bully, reinforcing the bully and defending the victim. For Grades 1 and 7, we had only posttest measurements, and the effects were examined with three-level regression models comparing intervention and control schools on the posttests only. For Grades 2–3 and 8–9 we had pretest measurements as well, and to this data four-level models were fitted as in Study II. After nine months of implementation, the intervention had desired effects in Grades 1–3 on self-reported victimization and bullying (*OR*s about 1.5), but some effects differed by gender. In Grades 7–9, statistically significant positive results were obtained on five out of seven variables, with results often depending on gender and sometimes on age of the students. The largest effects were obtained for boys' peer reports: bullying, assisting the bully and reinforcing the bully (Cohen's *d*s 0.11–0.19). The findings from Study III indicate that the KiVa program is effective in reducing bullying and victimization in Grades 1–3, but the results are more modest and mixed in Grades 7–9.

STUDY IV

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.

The evaluation studies of school-based antibullying programs have usually involved a small number of schools, ranging from 1 to 78. It may be the case that the results obtained in these studies are not generalizable to a large-scale dissemination of an antibullying program. In this study, we investigated the effectiveness of the KiVa antibullying program during the first year of its nationwide implementation in Finland. At pre- and posttest, the participants included 888 schools with approximately 150,000 students in 11,200 classrooms in Grades 1–9 (aged 8–16 years; 51% boys and 49% girls). The sample thus included 28% of the school population and 30% of the student population in the country. Victims and bullies were identified with self-report items: Students who reported having been victimized (or having bullied others) repeatedly were categorized as victims (bullies). The program effects were examined by calculating odds ratios based on a cohort-longitudinal design. In this design, posttest data from students in each grade level are compared with data from students in the same grade levels in the same schools (i.e., the previous cohort), who have not yet been exposed to the

intervention. The odds ratios indicate the odds of being victimized (bullying others) in the control group as compared with the corresponding odds in the intervention group. The standard errors were corrected for clustering by multiplying them with the design effect. The results indicated that, during the first nine months of implementation, the KiVa program reduced both victimization and bullying, with a control/intervention group odds ratio of 1.22, 95% CI [1.19, 1.24] for victimization and 1.18, 95% CI [1.15, 1.21] for bullying. If these results are generalizable to the Finnish student population of about 500,000 students, this would mean to a reduction of approximately 7,500 bullies and 12,500 victims in one school year.

5. GENERAL DISCUSSION

The participant role approach views bullying as a group phenomenon, in which the bystanders contribute to the continuation of bullying by reinforcing and assisting the bully, whereas their defending behaviors can put an end to it (Salmivalli et al., 1996). Although there have been studies investigating what makes students behave in different ways as bystanders in bullying situations (e.g., Salmivalli & Voeten, 2004), there is little evidence on whether bystander behaviors actually matter. Furthermore, it has not been investigated whether the KiVa program (based on the participant role approach) is effective. In this thesis, studies were conducted to examine (a) the influence of bystander behaviors on the effects of risk factors for victimization and (b) the effectiveness of the KiVa antibullying program.

Results show that bystander behaviors in bullying situations have an effect on risk for victimization of vulnerable students (Study I). Specifically, social anxiety and peer rejection had stronger effects in classrooms where bystanders reinforced bullying to a high degree and rarely defended the victims. If classmates reinforced bullying, this increased significantly the risk for the vulnerable students, whereas the protective effect of defending was weaker. The results add to the existing literature (Craig, 1998; Hodges & Perry, 1996; Olweus, 1993; Rigby & Slee, 1993; Salmon, James, Cassidy, & Javaloyes, 2000; Slee, 1995) by showing that the way classmates generally behave as bystanders may either enhance risk or foster resiliency in vulnerable students. These results also suggest that bystander behaviors in bullying situations make a difference, implying that it might be beneficial to target the bystanders in antibullying interventions.

The evaluation of the KiVa program (Studies II and III) demonstrates that the KiVa program is effective in reducing bullying and victimization, as indicated by the control/intervention group odds ratio for victimization ($OR = 1.28$, 95% CI [1.17, 1.40]) and for bullying ($OR = 1.30$, 95% CI [1.15, 1.48]). These results mean that 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. In terms of percentages, this points to a reduction of about 20% in the prevalence of bullying and victimization. The odds ratios for the KiVa program are in line with the meta-analysis by Farrington and Ttofi (2009), which found that the average odds ratios for antibullying intervention programs (excluding the KiVa program) were 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, 8th of March, 2010). Furthermore, it can be noted that the KiVa

program effects are larger than the average effects from studies with a randomized design for both victimization ($OR = 1.17$, 95% CI [1.00, 1.37]) and bullying ($OR = 1.10$, 95% CI [0.97, 1.26]; David P. Farrington, personal communication, 8th of March, 2010).

Compared with the evaluation results (Studies II–III), the program effects during the broad dissemination (Study IV) were somewhat smaller, the control/intervention group odds ratio being 1.22, 95% CI [1.19, 1.24] for victimization and 1.18, 95% CI [1.15, 1.21] for bullying. The results do indicate, however, that the KiVa antibullying program is effective even when disseminated at the national level. The smaller effect sizes at the dissemination phase may partly result from the fact that some of the most motivated schools joined in already during the evaluation study, whereas the schools participating in the dissemination study were probably a more heterogeneous group. In addition, schools also got less support during the national dissemination. These factors may have resulted in a lower quality of implementation in the dissemination phase, with fewer actions taken and may possibly have led to local adaptations of the program. These conclusions are to some extent supported by the results concerning program implementation and its associations with outcomes. Specifically, the program was implemented less compared with the evaluation study, and there was considerable variation in the implementation of its components. Also the intervention dosage was positively associated with reductions in bullying and victimization.

On the basis of the evaluation and dissemination studies (II–IV), the intervention effects on self-reported victimization and bullying are generally larger in Grades 1–6 than in Grades 7–9. The dissemination study (Study IV) provides the most accurate results, as evidenced by the narrow confidence intervals. There the effect sizes grow until grade four and diminish thereafter. Comparing the effects for elementary (Grades 1–6) and lower secondary schools (Grades 7–9), it can be seen that the KiVa program is more effective in the elementary school. Farrington and Ttofi (2009) concluded that the effects of antibullying programs increase steadily as a function of age, and they recommended that the programs should be targeted at children aged 11 years or older. This recommendation is in stark contrast with our results concerning the KiVa program, in which the effects start to decrease after the age of 11. Perhaps the discrepancy can be explained by the fact that studies included in the meta-analysis by Farrington and Ttofi (2009) varied not only in the age of the students but also in other aspects as well, and the association between age and intervention effects may be confounded with these other study characteristics (e.g., program contents). Actually, several studies comparing the effects of

one and the same program across age groups have found that the programs actually work better for younger rather than older students (Menesini et al., 2003; Salmivalli, Kaukiainen, & Voeten, 2005), and better in primary rather 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). These studies were reviewed by P. K. Smith (2010, p. 138–139), who concluded that antibullying programs often have smaller effects in secondary than in primary schools. This may be due to (a) developmental changes due to puberty and adolescence (e.g., in attitudes to victims) and (b) organizational changes resulting from larger and more complex organization of secondary schools compared with primary schools. These organizational factors may make it more difficult to implement the intervention well (P. K. Smith, 2010).

In addition to self-reported victimization and bullying, the evaluation study included investigation of the intervention effects on other outcomes as well: peer-reported behaviors and self-reported attitudinal and motivational constructs. The program effects on these variables were evaluated in Grades 4–9 only. In general, the intervention effects were larger and more consistent in Grades 4–6 than in Grades 7–9. By comparing the effects across outcomes, it can be noted that 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, and the intervention effects on antibullying attitudes, empathy towards victims, and self-efficacy for defending were among the smallest in both samples.[4]

The KiVa program reduced in Grades 4–9 peer-reported bullying, victimization, assisting and reinforcing. The magnitude of the intervention effects, however, depended on gender and age. In Grades 4–6, the intervention had statistically significant effects on peer-reported bullying only among the 5th and 6th graders. There were no statistically significant effects in Grade 7 on the peer-reported outcomes, and most of the effect sizes were close to zero. Nevertheless, in Grade 8 positive and significant effects were found for peer-reported victimization, whereas in Grades 8 and 9 the intervention had reduced peer-reported bullying, assisting, and reinforcing as well. The effects for these latter behaviors were stronger for boys and in classrooms with a high proportion of boys. The intervention effects on peer-reported defending were in Grades 4–7 statistically nonsignificant, and in

---

[4] In Grades 7-9, the effect sizes for these outcomes and for wellbeing in school were practically zero, and to conserve space, the models were not reported in Study III.

Grades 8 and 9, the effects were even negative: The intervention seemed to decrease defending the victims. This is a surprising finding that should be examined more closely in future studies. If the result is replicated, it is an important task to investigate in detail the processes involved. As such, the present results suggest that it is a difficult task to turn students into defenders of victims, and it may be particularly challenging in Grades 8 and 9. These results are in contradiction with a meta-analysis by Polanin and his colleagues (2012) who found that the average intervention program effect size for bystander intervention was larger for older (Grades 9–12; $g = .43$) than for younger students (Grades 3–8; $g = .14$). Because only eleven studies were included in the meta-analysis, its results can only be considered tentative, and the observed difference may be due to some other factors besides age. In any case, the results from Study III contribute to the relatively scarce literature on this topic.

In Grades 4–6, statistically significant intervention effects were found for self-efficacy for defending and well-being at school. For these outcomes in Grades 7–9 and for other self-reported outcomes in Grades 4–9, the effect sizes were close to zero. Perhaps these results provide a partial explanation for the modest results concerning defending: Had the intervention enhanced attitudes, empathy and self-efficacy for defending, it might have increased the actual defending behaviors as well (Pöyhönen et al., 2010; Salmivalli & Voeten, 2004). Increasing students' empathy for the victim may nevertheless be a difficult task: A recent meta-analysis of intervention effects also found that the average effect size was small ($g = 0.05$; Polanin et al., 2012). Four studies of the eight included reported negative effects, of which one was statistically significant, whereas three positive effects reached statistical significance (Polanin et al., 2012).

In Grades 1–6, the results did not show any consistent gender-specific patterns for the intervention effects. In Grades 8 and 9, however, the intervention effects on peer-reported bullying, assisting and reinforcing were larger for boys and in classrooms with a high proportion of boys. This may be a consequence of boys' high scores that make them suitable targets for the intervention. In classrooms with a high proportion of boys there may be the largest potential for improvement: A large concentration of boys may lead to an increase of problematic behaviors, and this increase may be counteracted by the intervention.

It is possible that the intervention effects differ for different types of students. If that is true, it may be a reason for the somewhat modest average effect sizes. This is apparent in the interaction effects with student and classroom characteristics that were found in the data from Grades 8–9. The

lack of large effects may also be related to the high proportion of consistently noninvolved students. Students who never bullied nor were victimized during the school year cannot show an intervention effect; it can only be detected when the proportion of noninvolved students is lower in intervention schools than in control schools. In other words, intervention effects in the form of reducing negative behaviors can only be found for those students who bullied others or were victimized. This may be associated with the finding that the intervention effects for some variables were stronger for boys than girls and stronger in classrooms with higher rather than lower proportions of boys.

The results for self-reported and peer-reported bullying and victimization were consistent in Grades 4–6, but there were some differences in Grades 8 and 9. This may be due to the fact that, in Grades 8 and 9, there was a decrease in the control group for the self-reports but not for the peer reports. If the problematic behaviors decrease to some extent by time or with schools' ordinary antibullying activities, this makes it difficult to detect the effects of the KiVa program.

If the intervention is implemented widely, the intervention effects for the KiVa program (Studies II–IV) can be considered important and practically significant. Already during the first year of the nationwide implementation of KiVa, the program may have resulted in reduction of about 3,900 victims and 2,300 bullies. This result alone suggests that the program has positively influenced the lives of a large number of children and adolescents. In addition, we can estimate how much KiVa would have reduced bullying and victimization, had all Finnish comprehensive schools implemented KiVa with the dissemination study results (ORs ≈ 1.2). In the Finnish student population of around 500,000 students, the reductions would have amounted to about 7,500 bullies and 12,500 victims during one school year.

5.1. CONCLUSIONS

This thesis demonstrates that bystander behaviors in bullying situations may influence the vulnerable students' risk for victimization. This has the practical implication that influencing the bystanders can be an effective way to decrease victimization. In addition, the KiVa antibullying program, based on influencing the bystanders, is effective in reducing victimization and bullying. The program effects are larger in elementary schools than in lower secondary schools, and in Grades 8 and 9 for some peer-reported outcomes the program effects are larger for boys than for girls. The magnitude of the overall

effects can be considered practically significant, when obtained in a large-scale dissemination of the program.

5.2. STRENGTHS

Several strengths of this thesis should be mentioned. First, the evaluation and dissemination samples used in this thesis included large numbers of students, classrooms, and schools, and all grade levels 1–9 of the basic education system were represented. On the basis of the sample characteristics, the dissemination sample in particular seemed quite representative of Finnish schools in general. Second, schools in the evaluation study were randomly assigned to intervention and control conditions, and this facilitates inferences about the KiVa program effects. Third, the nested data structures were taken into account in all of the analyses: in Studies I–III by multilevel modeling and in Study IV by correcting the standard errors for clustering. Fourth, in the evaluation of the KiVa program (Studies II–III) we used several outcome measurements and multiple informants in the assessments. Results concerning psychometric properties of all measures were provided. Fifth, in Studies III and IV basic results concerning the program implementation were reported. Sixth, both the extent and patterns of missing data were described in all studies, and their potential effects were taken into account in Studies II–IV.

5.3. LIMITATIONS

The most important limitations of the present study need to be addressed as well. First, an obvious limitation of Study I is the cross-sectional design, because it does not as such provide evidence about the direction of the effects. Prior research has, however, provided evidence that peer rejection is more likely a cause than a consequence of victimization (Ladd & Troop-Gordon, 2003; Salmivalli & Isaacs, 2005), and this is in accordance with our conceptualization of these processes. Second, the scope of the individual risk factors and their moderators was limited, and the model should be expanded in future studies (see below).

Missing data created some uncertainty about the effects of the bystander behaviors (Study I) and of the KiVa program (Studies II–IV). Unintentional missing data were present because of (a) absence during a measurement period, (b) lacking parental consent or (c) not completing the whole questionnaire. In the evaluation and dissemination studies, the longitudinal design with two consecutive school years caused some missingness as well. Missing data occurred more in self-reports than in peer reports, and more in the dependent variables than the predictors, which in the models for the

intervention studies had hardly any or only few missing values. To deal with missing data, we used in the KiVa program evaluation (Studies II and III) the best-practice methods (Schafer & Graham, 2002), and in the dissemination study (IV), missingness was shown to be an implausible alternative explanation for the effects by conducting a sensitivity analysis. In addition to unintentional missing data, there also were some intentionally missing data: Due to school transitions and due to the fact that the pretest was to be administered at the end of the school year preceding the school year when the intervention took place, there were only posttest data in Study III for students in Grades 1 and 7. This makes it impossible to control for potential preexisting differences between the intervention and the control group, and it therefore weakens the evidence for these grade levels.

In Study III for students in Grades 1–3 and in Study IV for students in Grades 1–9, the outcome variables included only self-reports of bullying and victimization. This limits the evidence to some extent. The self-report measures nevertheless correlated well with other corresponding self-reports (in Grades 1–9) and peer reports (measured in Grades 4–9), as described above in the Methods section. The correlations indicate that the measurements with self-reports do have construct validity.

A further limitation of Studies I–IV is that the outcomes were assessed by questionnaire data only. It is possible that the more modest results in the higher grades may partially be a consequence of the measurement method: In the data for Grades 7–9, there were signs (i.e., implausible or impossible responses) suggesting that the students were not always sincerely answering the questions. Finally, student surveys were administered by teachers. When students provide reports about undesirable behaviors in the presence of their teacher, they may be influenced to answer the way they think the teacher would like. In order to prevent this, teachers were given detailed instructions on how to act when administering the survey, and the students were told that their answers remain confidential.

5.4. DIRECTIONS FOR FUTURE RESEARCH

As can be noted from Studies II–IV in the present dissertation, field trials of intervention programs involve many methodological challenges (see also Farrington, 2003; Shadish, Cook, & Campbell, 2002), and some important themes can be highlighted that should be taken into account in further research as well. These issues concern measurements, missing data, and effect size measures.

The estimates for the intervention effects may, to some extent, depend on the measurement method (Studies II and III). It is therefore advisable to use not only self-reports of bullying and

victimization but peer reports and possibly teacher reports as well. Each of these methods has its strengths and weaknesses, and their functionality can depend on the developmental level of students (Ladd & Kochenderfer-Ladd, 2002). This provides some challenges for research, because (a) the measurement methods need to be age-appropriate and (b) for comparing results across grades, the questionnaire items should function in an invariant way (for measurement invariance, see Meredith, 1993). In any case, the use of multiple measures and sources is a desirable standard for evaluation studies, because it provides evidence about the validity and reliability of the measurements and the robustness of findings (Flay et al., 2005).

Missing values should be minimized as far as possible during data collection. At minimum, values for basic demographic variables like age and gender should be available for each student. It may be then possible to use FIML to deal with missing data provided that (a) there are missing values only in the dependent variables and (b) the distributional assumptions of maximum likelihood are satisfied (Enders, 2010). Alternatively, multiple imputation can be used even with missing values in the predictor variables (Enders, 2010; Schafer & Graham, 2002). Both methods require that the data are missing at random (MAR). An important advantage of multiple imputation is that it enables including a large number of variables into the imputation model, which makes the required MAR assumption more plausible (Schafer, 1997). FIML usually makes use of a smaller subset of variables, consisting of the variables needed in the analysis model. With complex nested data structures and interactions imputing can, however, present difficult practical problems, and therefore FIML may be preferable.

Effect sizes from the trials of intervention programs need to be interpreted in the context of previous research (e.g., Cooper, 2008). For between-study comparisons, the meta-analysis by Farrington & Ttofi (2009) provides an important benchmark, which can be used for assessing the magnitude of the program effects. The odds ratio used in this meta-analysis is a natural choice for an effect size measure in future research as well. In addition, the effect sizes need to be interpreted not only with regard to previous research but also to their practical implications (i.e., whether the intervention is worth implementing). For this purpose, percentages may be preferable to odds ratios, because they are easier to interpret for practitioners, policy-makers and the general public. It is important, however, to be clear about whether the reported figures are percentages or percentage points (i.e., the unit for the arithmetic difference of two percentages). Perhaps the percentage points could also be compared to normative expectations for change (Hill, Bloom, Black, & Lipsey, 2008): For

example, during Grade 2 the reduction in victimization associated with age (based on cross-sectional data) and intervention were 2.4% and 3.5%, respectively (Study IV). This would mean that the intervention effect is 1.5 times larger than the effect of maturation of one year at Grade 2. If the natural reduction of victimization is larger in early grades, a specific number of percentage points would then indicate a smaller substantive change in early than in later grades (cf. Hill et al., 2008).

Substantively, many important questions remain for future research. Investigators should examine further the influence of the bystander behaviors on the risk for victimization by adding other risk and protective factors to the models. At student level, these factors include, for instance, externalizing behavior and physical weakness (Hodges, Malone, & Perry, 1997; Olweus, 1978), and being a special-education student (e.g., Nabuzoka & Smith, 1993; Kärnä, Voeten, Poskiparta, Little, & Salmivalli, 2009). The effects of these risk factors may depend on how bystanders typically behave in bullying situations as well as on other moderating characteristics at student or classroom levels (e.g., the proportion of special education students). The various risk and protective factors should preferably be integrated into a single theoretical framework that would specify the proper time lag between the causes and their effects. This would enable testing competing theories about the processes that lead to victimization.

Future research could also examine 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. For example, it would be interesting to test whether the program effects on bullying and victimization are mediated by changes in bystander behaviors at the classroom level. If this kind of results were found, it would provide additional support for the participant role approach developed by Salmivalli et al. (1996). Furthermore, this mediation mechanism might be moderated by age, and in that case it would provide a partial explanation why the effects differ in primary and lower secondary schools. As discussed above, it is possible that the KiVa program has larger effects for certain types of students, classrooms and schools than for other types. Identifying their characteristics (in addition to age) would be valuable for both theory and practice. It would help (a) to understand risk factors for bullying and victimization, (b) to develop even more effective interventions, and (c) to target the intervention components in an efficient way. Moreover, previous studies have found that the degree of program implementation may 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 could give some idea, how much the intervention results can be improved by providing support for schools in program implementation. To some extent, it might help to identify the effective intervention components as well. Another important task is to investigate the predictors of implementation (Kallestad & Olweus, 2003), because this may inform us about what kind of schools need additional resources for high-quality implementation of the KiVa program. This knowledge would be useful for ensuring an effective use of the program. Finally, as the KiVa program is at present implemented in about 90% of the schools in Finland, it will be possible to investigate its long-term effects in a nationwide dissemination. Actually, Ryan and Smith (2009) recommended the use of a three-year follow-up period to properly investigate the effects of an intervention program. Enough time is needed so that the initial implementation problems can be solved and the program effects can fully unfold. It may even happen that the program effects become stronger over time (Olweus & Alsaker, 1991).

6. REFERENCES

American Psychological Association. (2010). *Publication Manual of the American Psychological Association,* (6th ed.). Washington, DC: Author.

Andreou, E., & Metallidou, P. (2004). The relationship of academic and social cognition to behaviour in bullying situations among Greek primary school children. *Educational Psychology, 24*, 27–41.

Baldry, A. C., & Farrington, D. P. (2007). Effectiveness of programs to prevent school bullying. *Victims & Offenders, 2*, 183–204.

Bandura, A. (1977). *Social learning theory*. Oxford, England: Prentice Hall.

Boulton, M., Bucci, E., & Hawker, D. (1999). Swedish and English secondary school pupils' attitudes towards, and conceptions of, bullying: Concurrent links with bully/victim involvement. *Scandinavian Journal of Psychology, 40*, 277–284.

Camodeca, M., & Goossens, F. A. (2005). Children's opinions on effective strategies to cope with bullying: The importance of bullying role and perspective. *Educational Research, 47*, 93–105.

Caravita, S., Di Blasio, P., & Salmivalli, C. (2009). Unique and interactive effects of empathy and social status on involvement in bullying. *Social Development, 18*, 140–163

Cook, C. R., Williams, K. R., Guerra, N. G., & Kim, T. E. (2010). Variability in the prevalence of bullying and victimization: A cross-national and methodological analysis. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *Handbook of bullying in schools: An international perspective* (pp. 347–362). New York: Routledge.

Cooper, H. (2008). The search for meaningful ways to express the effects of interventions. *Child Development Perspectives, 2*, 181–186.

Craig, W. M. (1998). The relationship among bullying, victimization, depression, social anxiety, and aggression in elementary school children. *Personality and Individual Differences, 24*, 123–130.

Craig, W., & Harel, Y. (2004). Bullying, physical fighting, and victimization. In C. Currie (Ed.), *Young people's health in context: International report from the HBSC 2001/02 survey. WHO Policy*

Series: Health policy for children and adolescents, Issue 4, Copenhagen: WHO Regional Office for Europe.

Farrington, D. P. (2003). Methodological quality standards for evaluation research. *Annals of the American Academy of Political and Social Science, 587*, 49–68.

Farrington, D. P., & Ttofi, M. M. (2009). School-based programs to reduce bullying and victimization. *Campbell Systematic Reviews,* 2009:6. Retrieved from http://www.d300.org/files/School-based%20Anti-Bullying%20Programs%20v2_R.pdf

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.

Flay, B. R., Biglan, A, Boruch, R. F., Castro, F. G., Gottfredson, D., Kellam, S., … Ji. P. (2005). Standards of evidence: Criteria for efficacy, effectiveness and dissemination. *Prevention Science, 6*, 151–175.

Goossens, F. A., Olthof, T., & Dekker, P. H. (2006). New participant role scales: Comparison between various criteria for assigning roles and indications for their validity. *Aggressive Behavior, 32*, 343–357.

Hanewinkel R. (2004). Prevention of bullying in German schools: An evaluation of an antibullying approach. In P. K. Smith, D. Pepler, & K. Rigby (Eds.)*, Bullying in schools: How successful can interventions be?* (pp. 81–97). Cambridge, UK: Cambridge University Press.

Hawker, D. S. J., & Boulton, M. J. (2000). Twenty years' research on peer victimization and psychosocial maladjustment: A meta-analytic review of cross-sectional studies. *Journal of Child Psychology and Psychiatry, 41*, 441–455.

Hawkins, D. L., Pepler, D. J., & Craig, W. M. (2001). Naturalistic observations of peer interventions in bullying. *Social Development*, 10, 512–527.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives, 2*, 172–177.

Hodges, E. V. E., Malone, M. J., & Perry, D. G. (1997). Individual risk and social risk as interacting determinants of victimization in the peer group. *Developmental Psychology, 33*, 1032–1039.

Hodges, E. V. E., & Perry, D. G. (1996). Victims of peer abuse: An overview. *Journal of Emotional and Behavioral Problems, 5*, 23–28.

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: Guilford.

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.

Kärnä, A., Voeten, M., Poskiparta, E., Little. T. D., & Salmivalli, C. (2009, August). The effects of the KiVa anti-bullying program on special education students. In I. Vermaes (Chair), *Bullying of youths with special needs*. Symposium conducted at the 14th ESDP European Conference on Developmental Psychology, Vilnius, Lithuania.

Ladd G. W., & Kochenderfer-Ladd, B. J. (2002). Identifying victims of peer aggression from early to middle childhood: Analysis of cross-informant data for concordance, estimation of relational adjustment, prevalence of victimization, and characteristics of identified victims. *Psychological Assessment*, *14*, 74–96.

Ladd, G. W., & Troop-Gordon, W. (2003). The role of chronic peer difficulties in the development of children's psychological adjustment problems. *Child Development, 74*, 1344–1367.

La Greca, A. M., & Lopez, N. (1998). Social anxiety among adolescents: Linkages with peer relations and friendships. *Journal of Abnormal Child Psychology, 26*, 83–94.

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*, 10–14.

Menesini, E., Eslea, M., Smith, P. K., Genta, M. L., Giannetti, E., Fonzi, A., & Costabile, A. (1997). Cross-national comparison of children's attitudes towards bully/victim problems in school. *Aggressive Behavior, 23*, 245–257.

Meredith, W. (1993). Measurement invariance, factor analysis, and factorial invariance. *Psychometrika, 58*, 525–543.

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.

Metsämuuronen, J., & Svedlin, R. (2004). Kouluviihtyvyyden muuttuminen peruskoulussa ja lukiossa iän funktiona [Change in perceived well-being in primary and secondary schools as a function of age]. Manuscript submitted for publication.

Nabuzoka, D., & Smith, P. K. (1993). Sociometric status and social behavior of children with and without learning difficulties. *Journal of Child Psychology and Psychiatry, 34*, 1435–1448.

O'Connell, P., Pepler, D., & Craig, W. (1999). Peer involvement in bullying: Insights and challenges for intervention. *Journal of Adolescence, 22*, 437–452.

Olweus, D. (1978). *Aggression in the schools: Bullies and whipping boys*. Washington, DC: Hemisphere.

Olweus, D. (1993). *Bullying at school: What we know and what we can do*. Malden, MA: Blackwell.

Olweus, D. (1994). Bullying at school: Long-term outcomes for the victims and an effective school-based intervention program. In L. R. Huesmann (Ed.), *Aggressive Behavior: Current Perspectives* (pp. 97–130). New York: Plenum.

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

Olweus, D. (2005). New positive results with the Olweus Bullying Prevention Program in 37 Oslo schools. Unpublished report. Bergen, Norway: HEMIL-center.

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: Cambridge University Press.

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

Polanin, J., 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 defend the victimized peer? The interplay between personal and social factors. *Merrill-Palmer Quarterly, 56*, 143–163.

Rigby, K., & Slee, P. T. (1991). Bullying among Australian school children: Reported behavior and attitudes towards victims. *Journal of Social Psychology, 131*, 615–627.

Rigby, K., & Slee, P. T. (1993). Dimensions of interpersonal relation among Australian children and implications for psychological well-being. *Journal of Social Psychology, 133*, 33–42.

Ryan W., & Smith J. D. (2009). Antibullying programs in schools: How effective are evaluation practices? *Prevention Science, 10*, 248–259.

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. (1999). Participant role approach to school bullying: Implications for intervention. *Journal of Adolescence, 22*, 453–459.

Salmivalli, C., Huttunen, A., & Lagerspetz, K. M. J. (1997). Peer networks and bullying in schools. *Scandinavian Journal of Psychology, 38*, 305–312.

Salmivalli, C., & Isaacs, J. (2005). Prospective relations among victimization, rejection, friendlessness, and children's self- and peer-perceptions. *Child Development,76*, 1161–1171.

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

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: 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: Routledge.

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.

Salmivalli, C., Lappalainen, M., & Lagerspetz, K. M. J. (1998), Stability and change of behavior in connection with bullying in schools: A two-year follow-up. *Aggressive Behavior, 24*, 205–218.

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: Guilford.

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, Psychology Department.

Salmivalli, C., Pöyhönen, V., & Kaukiainen, A. (2009). *KiVa: Teacher's guide, unit 2* (Research Into Practice Publication Series, No. 3). Turku, Finland: University of Turku, Psychology Department.

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.

Salmon, G., James, A., Cassidy, E. L., & Javaloyes, M. A. (2000). Bullying a review: Presentations to an adolescent psychiatric service and within a school for emotionally and behaviourally disturbed children. *Clinical Child Psychology and Psychiatry, 5*, 563–579.

Schafer, J. L. (1997). *Analysis of Incomplete Multivariate Data*. London: Chapman and Hall.

Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods, 7*, 147–177.

Schäfer, M., & Korn, S. (2004). Bullying als gruppenphänomen: Eine adaptation des 'participant role' - ansatzes. *Entwicklungspsychologie und Pädagogische Psychologie, 36*, 19–29.

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. Boston: Houghton-Mifflin.

Slee, P. T. (1995). Peer victimization and its relationship to depression among a Australian primary school students. *Personality and Individual Differences, 18*, 57–62.

Smith, J. D., Schneider, B. H., Smith, P. K., & Ananiadou, K. (2004). The effectiveness of whole-school antibullying 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: Routledge.

Smith, P. K., Ananiadou, K., & Cowie, H. (2003). Interventions to reduce school bullying. *Canadian Journal of Psychiatry, 48*, 591–599.

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

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

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

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

Sourander A., Jensen P., Rönning J. A., Niemelä, S., Helenius, H., Sillanmäki, L., … Almqvist, F. (2007). What is the early adulthood outcome of boys who bully or are bullied in childhood? The Finnish "From a Boy to a Man" study. *Pediatrics, 120*, 397–404.

Sutton, J., & Smith, P. K. (1999). Bullying as a group process: An adaptation of the participant role approach. *Aggressive Behavior, 25*, 97–111.

Ttofi, M. M., Farrington, D. P., Lösel, F., & Loeber, R. (2011a). Do the victims of school bullies tend to become depressed later in life? A systematic review and meta-analysis of longitudinal studies. *Conflict and Peace Research, 3*, 63–73.

Ttofi, M. M., Farrington, D. P., Lösel, F., & Loeber, R. (2011b). The predictive efficiency of school bullying versus later offending: A systematic/meta-analytic review of longitudinal studies. *Criminal Behaviour and Mental Health, 21*, 80–89.

O'Connell, P., Pepler, D., & Craig, W. (1999). Peer involvement in bullying: Insights and challenges for intervention. *Journal of Adolescence, 22*, 437–452.

Vreeman, R. C., & Carroll, A. E. (2007). A systematic review of school-based interventions to prevent bullying. *Archives of Pediatrics & Adolescent Medicine, 161*, 78–88.

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: Routledge.

APPENDIX: THE ORIGINAL PUBLICATIONS