Check for
updates

Addressing Specific Forms of Bullying: A Large-Scale Evaluation of the Olweus Bullying Prevention Program

Dan Olweus[1] · Susan P. Limber[2] · Kyrre Breivik[3]

Abstract
The purpose of this study was to evaluate the effectiveness of the Olweus Bullying Prevention Program (OBPP) in reducing specific forms of bullying—verbal bullying, physical bullying, and indirect/relational bullying, as well as cyberbullying and bullying using words or gestures with a sexual meaning. This large-scale longitudinal study, which involved more than 30,000 students in grades 3–11 from 95 schools in central and western Pennsylvania over the course of 3 years, employed a quasi-experimental extended age-cohort design to examine self-reports of being bullied, as well as bullying others. Findings revealed that the OBPP was successful in reducing all forms of being bullied and bullying others. Analyses by grade groupings (grades 3–5, 6–8, and 9–11) revealed that, with only a few exceptions, there were significant program effects for all forms of bullying for all grade groupings. For most analyses, program effects were stronger the longer the program was in place. Most analyses indicated similar and substantial effects for both boys and girls, but a number of program by gender interactions were observed. Program effects for Black and White students were similar for most forms of being bullied and bullying others. Although Hispanic students showed results that paralleled the development for Black and White students for particular grade groups and variables, they were overall somewhat weaker. The study provided strong support for the effectiveness of the OBPP among students in elementary, middle, and early high school grades. Program effects were broad, substantial, and largely consistent, covering all forms of bullying—verbal, physical, indirect, bullying through sexual words and gestures, and, with somewhat weaker effects, cyberbullying—both with regard to being bullied and bullying others. Strengths and limitations of the study, as well as future research directions, are discussed.

Keywords Bullying victimization · Bullying perpetration · Forms of bullying · USA Evaluation · Olweus Bullying Prevention Program · Anti-bullying programs

Bullying among children and youth is an age-old phenomenon, but it is only relatively recently that bullying has come to be

Dan Olweus and Susan P. Limber contributed equally to this work.

✉ Susan P. Limber
slimber@clemson.edu

Dan Olweus
olweus@uib.no

Kyrre Breivik
kybr@norceresearch.no

[1] Department of Health Promotion and Development, Faculty of Psychology, University of Bergen, Bergen, Norway

[2] Institute on Family & Neighborhood Life, Clemson University, Clemson, SC, USA

[3] Regional Centre for Child and Youth Mental Health and Child Welfare, NORCE Norwegian Research Centre, Bergen, Norway

viewed as an international public health concern (Masiello and Schroeder 2014; National Academies of Science, Engineering, and Medicine 2016). Research on bullying began in the 1970s in Scandinavia (Olweus 1973, 1978). Since this time, extensive research has documented the nature and extent of bullying, as well as its consequences. Bullying is a form of aggressive behavior that involves a power imbalance between a target and his or her perpetrator(s), and typically is repeated over time (Gladden et al. 2014; Olweus 1993). Several different forms of bullying have been identified, including physical bullying, verbal bullying, indirect or relational bullying, and cyberbullying.

In the USA, a recent nationally representative survey of students aged 12–18 indicated that 21% had been bullied at school during the 2015 school year (Musu-Gillette et al. 2018). Thirteen percent had been verbally bullied, 12% had been the subject of rumors, 5% had been physically bullied, 5% had been purposefully excluded from activities, 4% had been threatened with harm, 3% had been forced to do things

against their will, and 2% had property destroyed (Musu-Gillette et al. 2018). In the most recent version of this survey to address cyberbullying, 7% of 12–18-year-olds indicated that another student had done one or more of the following to them during the 2013 school year: posted hurtful information about them on the Internet; purposely shared private information about them on the Internet; threatened or insulted them through instant messaging; threatened or insulted them through text messaging; threatened or insulted them through e-mail; threatened or insulted them while gaming; or excluded them online (Musu-Gillette et al. 2018).

A robust body of research has identified many negative mental health, psychosocial, physiological, and behavioral effects of bullying on children and youth who are targeted (for overviews, see Cook et al. 2010; National Academies 2016; Olweus 2013; Ttofi et al. 2011). Bullying others has also been associated with short- and long-term negative characteristics (for overviews, see Cook et al. 2010; National Academies 2016; Olweus 2013; Ttofi et al. 2012). However, whereas bullied youth commonly experience internalizing problems (including depression, poor self-esteem, and anxiety), the behavior of children and youth who bully others is more typically characterized by externalizing behaviors, including rule-breaking, violence, and delinquency.

In view of the high personal and societal costs of bullying, and recognizing that bullying violates an individual's fundamental human rights to be safe in school (Olweus 1993), numerous efforts have been launched in recent years to prevent and reduce bullying (National Academies 2016). Several meta-analyses and systematic reviews of bullying prevention programs have been published, which have reached somewhat different conclusions about the effectiveness of these efforts. The most comprehensive meta-analysis, conducted by Ttofi and Farrington (2009, 2011) included 44 evaluations of school-based prevention programs. The authors found that bullying prevention programs were effective in reducing bullying victimization and/or perpetration by an average of 17–23%, although the effects were typically small and there was great variation in results. They further observed that programs implemented and evaluated in Europe were more effective than those in the USA, and that "programs inspired by the work of Dan Olweus worked best" (Ttofi and Farrington 2011, pp. 41–42).

Description of the Olweus Bullying Prevention Program

The Olweus Bullying Prevention Program (OBPP, Olweus 1991, 1993, Olweus and Limber 2010a, b), the oldest and most researched school-based bullying prevention program (National Academies 2016), was first developed and evaluated by Dan Olweus in Norway in the mid-1980s. Initially designed for students in elementary, middle, and junior high school grades, the goals of the OBPP are to reduce bullying among children and youth, prevent new bullying problems, and more generally, achieve better relations among peers at school (Olweus 1993; Olweus and Limber 2010b). To achieve these goals, school personnel focus on restructuring the school environment to reduce opportunities and rewards for bullying and on building a sense of community. The OBPP is built upon four key principles, which were derived from theory and research on aggression in youth (Olweus 1993). Adults within a school environment should (a) show warmth and positive interest in students, (b) set limits to unacceptable behavior, (c) use consistent positive consequences to reinforce positive behavior, and consistent, non-hostile consequences when rules are broken, and (d) act as positive role models for appropriate behavior (Olweus 1993; Olweus et al. 2007). These principles are translated into school-level interventions, classroom-level interventions, individual interventions, and community-level interventions. There are eight school-level components, which are implemented schoolwide, including the development of a building-level coordinating team that is responsible for ensuring that all components of the OBPP are implemented with fidelity and sustained over the long-term, yearly administration of the Olweus Bullying Questionnaire (OBQ, Olweus 2007a), training and ongoing consultation for members of the coordinating team and all school staff, the adoption of clear rules and policies related to bullying, and the review and refinement of the school's system of student supervision. Classroom-level program components include holding class meetings that are designed to build understanding of bullying and related issues through discussion and role play and build class cohesion, supporting school-wide rules against bullying, and holding periodic class-level meetings with parents. Teachers also are encouraged to integrate bullying prevention themes throughout the curriculum. Individual-level interventions include careful supervision of students, particularly in known hotspots for bullying, training for all staff to intervene on-the-spot when bullying happens or is suspected, and follow-up interventions with children and youth involved in bullying (and their parents, when appropriate). Community-level interventions include involvement of one or more community member on the school's coordinating team, other activities to help ensure community support of the school's bullying prevention work, and efforts to spread bullying prevention messages and strategies into other community settings where children and youth gather (Olweus and Limber 2010b). Training and continued consultation are provided by certified OBPP Trainer-Consultants, who help to address challenges and maintain fidelity to the program. Print, online, and video resources are provided for administrators, members of the coordinating team, teachers, and parents. For a more detailed description of program elements and resources to support program implementation, see previous articles by the authors (e.g., Limber and Olweus 2017; Limber et al. 2018; Olweus et al. 2007; Olweus and Limber 2010b).

Previous Evaluations of the OBPP

The OBPP has been evaluated in a number of studies in Norway, as well as the USA. In the first evaluation, which took place in Bergen, Norway, between 1983 and 1985, Olweus (Olweus 1991, 1993, 1997) developed what he called an "extended age cohort design" (Olweus 2005; Olweus and Limber 2010b) to compare same-aged students across time. Nearly 2500 students in grades 5–8 in 42 schools were followed over two and a half years. Results revealed significant reductions in students' self-reports of being bullied and bullying others, reductions in teachers' and students' ratings of bullying among peers in the classroom, and improvements in students' and teachers' assessments of the school climate (Olweus 1991, 1993, 1997). A measure of program fidelity was positively associated with program outcomes (Olweus and Kallestad 2010).

Six additional large-scale studies of the OBPP in Norway (involving more than 30,000 students from more than 300 schools) have produced consistently positive effects for students in grades 4–7 (with reductions in bullying in the range of 35–50% after 8 months) and positive, although somewhat less consistent and immediate results for students in grades 8–10 (Olweus and Limber 2010b). In an evaluation of long-term effectiveness of the OBPP, Olweus followed students from 14 schools (with 3000 students at each assessment) over 5 years and observed reductions in self-reported victimization of 40% and self-reported bullying of 51%. A very recent study with many more schools has largely extended these results, showing positive long-term school-level effects of the program over a period of up to 8 years after the original implementation (Olweus et al. 2018).

The OBPP has also been evaluated in several large-scale studies in diverse areas of the USA. The first, which took place in the mid-1990s, involved elementary and middle schools in six rural school districts in South Carolina (Limber et al. 2004; Olweus and Limber 2010b). After 7 months of implementation of the OBPP, there were significant differences between intervention and comparison schools with respect to students' self-reports of bullying of peers and self-reports of delinquency, vandalism, and school misbehavior, but there were no significant program effects for students' reports of being bullied. In a non-randomized controlled evaluation of the OBPP with seven intervention and three control schools in Washington State, Bauer et al. (2007) reported significant program effects for both relational and physical bullying among White students but no program effects among students of other races/ethnicities. Although these US findings are encouraging, it is also clear that the results from these studies in the USA have not been uniformly positive.

However, these results are by no means unique. There actually are very few, if any, studies with US students that have documented clear-cut, convincing results as a consequence of an anti-bullying program, at least if the evaluation is based on student reports. Serious concerns about the lack of positive program effects in the USA were also echoed in a recent review of evaluations of bullying prevention programs (Evans et al. 2014) conducted after Ttofi and Farrington's comprehensive meta-analysis (Ttofi and Farrington 2009, 2011). In this review, six of the eight studies examining bullying victimization with nonsignificant results were conducted in the USA, as were six of the ten nonsignificant studies examining bullying perpetration. These concerns have also taken the form of issuing a call for the need to completely rethink society's anti-bullying efforts and strategies (Cohen et al. 2015; Espelage et al. 2018; Hong and Espelage 2012), often combined with a skeptical view of the value and usefulness of anti-bullying programs such as the OBPP which was developed outside of the USA. In view of such concerns, there was obviously a need for a new large-scale study of the OBPP in the USA, in which the program had been systematically implemented over a period of at least 2 years and data from a sample of appropriate size had been adequately analyzed with multilevel techniques taking account of cluster effects.

A first response to these concerns was recently published. In this evaluation of the OBPP, Limber et al. (2018) used an extended age cohort design to examine the effects of the program on students in grades 3–11 in the southern and central Pennsylvania. The researchers conducted two related studies: one followed 210 schools over 2 years and a second followed a subsample of 95 schools over 3 years. For almost all grades, there were significant reductions in being bullied and bullying others, whether measured with single items or scale scores, which comprised nine specific forms of bullying. The longitudinal analysis indicated that program effects were generally larger, the longer the program was in place, and it also documented increases in students' expressions of empathy for bullied peers, decreases in students' willingness to join in bullying, and perceptions that their teachers were actively addressing bullying in their classrooms. These changes were observed for both boys and girls and for students in elementary, middle, and high school, although program effects on being bullied were somewhat stronger in elementary and middle school grades. Where possible (where sample sizes were sufficiently large), program effects were examined for students of different races/ethnicities, and significant reductions in involvement in bullying were found for 8 of 14 possible analyses using global variables of being bullied and bullying others. Program effects were typically larger for White students but were significant for Black and Hispanic middle school students with respect to bullying others. No significant program effects were observed for Black or Hispanic elementary or middle school students being bullied or for Black or Hispanic elementary students bullying others. This study provided strong support for the

effectiveness of the OBPP among students in elementary, middle, and early high school grades in the USA, but it left a number of questions unanswered, including the extent to which the OBPP may be effective in reducing specific forms of bullying, and for what subgroups of students.

A second response to the concerns noted above is the current follow-up study, which represents much more detailed analyses of the program effects on various forms of being bullied and bullying others based on the 3-year longitudinal sample from the published study (Limber et al. 2018). Examination of the effectiveness of bullying prevention efforts on specific forms of bullying is needed to advance the field. With rare exceptions (Salmivalli et al. 2011), evaluations of bullying prevention programs have not examined whether the interventions equally affect different forms of bullying. Some (Woods and Wolke 2003) have suggested that bullying prevention interventions may more readily reduce direct, visible forms of bullying, such as verbal bullying and physical bullying, and have less effect on forms of bullying that are more likely to be "hidden" from view (such as cyberbullying or indirect forms of bullying).

The Present Study

The purpose of this study was to evaluate the effectiveness of the OBPP in addressing several specific forms of bullying: direct verbal bullying, direct physical bullying, indirect bullying, electronic bullying, and sexual bullying. We hypothesized reductions in students' reports of being bullied and bullying others for all forms of bullying over the course of 3 years as a consequence of the OBPP. We further anticipated that these effects would be documented for both boys and girls and for all age groups (grades 3–5, 6–8, and 9–11). Based on earlier evaluations of the OBPP (Limber et al. 2018; Olweus and Kallestad 2010; Olweus and Limber 2010b), we expected somewhat stronger effects of being bullied among elementary and middle school grades vs. high school grades. With regard to bullying others, previous research had provided a somewhat inconsistent picture of program effects at different grade levels (Limber et al. 2018). As a result, although grade-level differences in bullying others were examined in this study, we did not have specific hypotheses about the strength of program effects for younger versus older students. Consistent with previous research on the OBPP (Limber et al. 2018; Olweus and Limber 2010b), we also hypothesized that all program effects would be stronger with the longer implementation of the OBPP. Research is scant on the effectiveness of bullying prevention on students of different races and ethnicities. As noted above, a previous evaluation (Limber et al. 2018) had suggested that program effects for being bullied and bullying others, as measured through global questions,

were somewhat weaker and less consistent for Black and Hispanic school youth, but it was unknown whether students of different races and ethnicities might experience reductions in specific forms of bullying, as a consequence of the OBPP.

Method

Participants

The sample included students in grades 3–11 from schools that were involved in a wide-scale effort to implement the OBPP in elementary, middle, and high schools in 49 counties in southern and central Pennsylvania (Limber et al. 2018). A total of 95 schools had taken the Olweus Bullying Questionnaire (Olweus 2007a) four consecutive years and the survey results were used in the present study to examine changes in being bullied and bullying others over the course of 3 years. A total of 31,620 students completed the measures at baseline (T0), and 29,814 (94.1%) completed the measures 3 years later (T3). Demographic information for participants is presented in Table 1. Additional information about the participating schools is available in our previous study (Limber et al. 2018).

Table 1 Characteristics of the participants at baseline

	N (%)
Sex	
Female	15,560 (49.2)
Male	15,937 (50.4)
Missing (sex)	123 (0.4)
Grade	
Grade 3	4447 (14.1)
Grade 4	4402 (13.9)
Grade 5	4446 (14.1)
Grade 6	4502 (14.2)
Grade 7	4156 (13.1)
Grade 8	4185 (13.2)
Grade 9	1900 (6.3)
Grade 10	1555 (4.9)
Grade 11	1997 (6.1)
Race/ethnicity	
White	19,291 (61.0)
Black or African-American	1432 (4.5)
Hispanic or Latino	1006 (3.2)
Other	3559 (11.2)
Missing data/do not know	6532 (20.1)

Measures

Participants completed the Olweus Bullying Questionnaire (OBQ), a 40-item questionnaire designed to assess students' self-reports of being bullied, bullying others, their own actions and reactions when they witness bullying, their attitudes about bullying, and their perceptions of their teachers' efforts to address bullying (Olweus 2007a). The measure is designed to be used by students in grades 3–12. Most questions ask students about their experiences over the last couple of months (Olweus 2007a; Olweus 2013). As the focus of this study was on students' experiences of being bullied and bullying others, these questions and relevant scales are described in detail below and in Table 2. Additional questions on the OBQ relate to circumstances surrounding the bullying (e.g., where it has occurred, how long it has lasted), students' responses to the bullying (e.g., whom, if anyone, they have told about being bullied), and students' perceptions of peer and adult responses to bullying at school, but such questions were not a focus of the current study. The present version of the OBQ (with the addition of items about cyberbullying in 2006) has been used for more than 20 years (Olweus 1996) and has been extensively validated (e.g., Breivik and Olweus 2015; Olweus 2013; Solberg and Olweus 2003).

Being Bullied Students were provided a detailed definition of bullying (Olweus 2013; Solberg and Olweus 2003) and were asked how often they had been bullied at school during the past couple of months (Olweus 2007a). Following this global question, students were asked about the frequency with which they had experienced nine specific variants of bullying. There were five response options for each of these nine questions: "It has not happened to me in the past couple of months" (coded 1); "Only once or twice" (coded 2), "2 or 3 times a month" (coded 3), "About once a week" (coded 4), or "Several times a week" (coded 5). In order to examine students' experiences of being verbally bullied, physically bullied, and indirectly bullied, three scale scores were created by taking the average of items belonging to the scale (see Table 2 for a listing of all items in the scales, as well as the aggregate reliability estimates for each scale [Kallestad et al. 1998, Table 5; Snijders and Bosker 2012, p. 26]). In addition to the three being-bullied scales, two individual items that were considered to have particular public interest were analyzed: being bullied with names, comments, or gestures with a sexual meaning (referred to here as being sexually bullied); and being bullied electronically (also referred to as cyberbullying). See Table 2 for a description of these two items.

Bullying Others Students were also asked about the frequency with which they had bullied other students at school in the past couple of months (i.e., the global question), followed by questions about the nine specific variants of bullying others, using the same response alternatives as items for being bullied (above). Three bullying others scales were created to capture verbally bullying others, physically bullying others, and indirectly bullying others, which were parallel scales to those for being bullied. In addition to the bullying-others scales, two

Table 2 Variables for being bullied and bullying others

Variables	Number of items	Description of items	Cronbach's alpha
Being verbally bullied scale (direct verbal)	3	(1) Was called mean names, was made fun of, or teased; (2) Was bullied with mean names or comments about their race or color; (3) Was bullied with mean names, comments, or gestures with a sexual meaning	$\alpha = 0.89$
Being physically bullied scale (direct physical)	3	(1) Was hit, kicked, pushed, shoved; (2) Had money or belongings taken or damaged; (3) Was threatened or forced to do things against will	$\alpha = 0.94$
Being indirectly bullied scale	2	(1) Was purposefully excluded or ignored; (2) Had lies or rumors spread about them	$\alpha = 0.89$
Being sexually bullied	1	(1) Was bullied with mean names, comments, or gestures with a sexual meaning	N/A
Being electronically bullied	1	(1) Was bullied with mean or hurtful messages, calls, or pictures or in other ways on cellphone or over the Internet	N/A
Verbally bullying others scale	3	(1) Called others mean names, made fun of, or teased others; (2) Bullied others with mean names or comments about race or color; (3) Bullied others with mean names, comments, or gestures with a sexual meaning	$\alpha = 0.94$
Physically bullying others scale	3	(1) Hit, kicked, pushed, shoved others; (2) Took money or damaged belongings; (3) Threatened others or forced others do things	$\alpha = 0.94$
Indirectly bullying others scale	2	(1) Purposefully excluded or ignored others; (2) Spread lies or rumors against others	$\alpha = 0.90$
Sexually bullying others	1	(1) Bullied others with mean names, comments, or gestures with a sexual meaning	N/A
Electronically bullying others	1	(1) Bullied others with mean or hurtful messages, calls or pictures, or in other ways on cell phone or over the Internet	N/A

individual items were analyzed: bullying others with names, comments, or gestures with a sexual meaning (referred to here as sexually bullied others); and electronically bullying others (also referred to as cyberbullied others). See Table 2 for a description of all bullying others variables.

Demographic Questions Students answered questions about their sex, grade in school, and their race or ethnicity. With regard to race/ethnicity, students were asked, "How do you describe yourself?" and could designate as many of the following categories that applied to them: American-Indian, Black or African-American, Arab or Arab-American, Hispanic or Latino, Asian-American, White, other, or I do not know.

Procedure

The OBPP was implemented in participating schools following standard practices (Limber et al. 2018; Olweus and Limber 2010b; Masiello and Schroeder 2014). Local certified OBPP trainer-consultants provided a 2-day training and monthly in-person or telephone consultation to members of each school's Bullying Prevention Coordinating Committee (BPCC) throughout the implementation of the program. Subsequently, members of the BPCCs, with support from certified OBPP trainer-consultants, provided one full-day training to all staff prior to the start of the program. Schools received all OBPP support materials (e.g., handbooks, DVD and CD-ROM resources, and training manuals) prior to the start of the program. For more details about the program resources and implementation, see Olweus and Limber 2010b.

As further detailed by Limber et al. (2018), evaluation of the OBPP involved distribution of a pencil/paper scannable version of the Olweus Bullying Questionnaire (Olweus 2007a) to students approximately 3–4 months prior to the launch of the program. Classroom teachers administered the OBQ in anonymous format. Schools began the program in the fall (shortly after the beginning of the school year) or winter (shortly after winter holidays). Although the month in which the OBQ was administered varied among schools, the dates of baseline survey administration were carefully recorded so that subsequent measurements with the OBQ were made at the same time of the year, 1, 2, and 3 years after baseline administration. In keeping with the standard implementation of the OBPP, all schools received a detailed school-level report of findings from the OBQ (Olweus 2007b) to help with their planning and internal evaluation of progress/lack of progress. When schools submit the scannable OBQ forms to be processed, they are asked to indicate whether or not they agree to share their data with Dan Olweus and his fellow researchers. In this study, all agreed to do so. No schools required active parental consent for students to complete the OBQ; rather, parents had the option to have their child complete an alternate

activity. No parents and no youth participants declined to participate. The study was reviewed and conducted in compliance with the human subjects review board of Clemson University.

Study Design

This study used a quasi-experimental "extended age cohort design" as developed by Olweus (Limber et al. 2018; Olweus 2005; Olweus and Limber 2010b; Shadish et al. 2002). In such a design, same-age students within the same schools are compared over time. For example, students in 5th grade at baseline (T0) are compared with students in 5th grade from the same school at time 1 (T1), 1 year later. These students were in 4th grade at T0, and at T1, they had been exposed to the program for approximately 9 months. In making such a comparison, possible age-related maturational differences between the comparison groups are controlled. For detailed information about this design, its strengths, and uses, see Olweus 2005; Olweus and Limber 2010b).

When the groups to be compared belong to the same schools, there are good grounds for assuming that a grade cohort differs in only minor ways from its contiguous cohort. Usually, the majority of the members in the various grade cohorts have been recruited from the same relatively stable populations and are also likely to have been students in the same schools for several years. The schools thus serve as their own controls and in this way, the problem with initial differences between the groups to be compared can be largely reduced or avoided. In other words, the "pretest" (T0) values for the individual schools can be considered good answers to the critical counterfactual question in all evaluation research: How do we obtain reasonable estimates of what the result would have been if the intervention subjects had not been exposed to the intervention? As has been repeatedly documented, attempts to statistically correct for preexisting initial differences in common quasi-experimental designs (with nonequivalent control and intervention groups) are fraught with great difficulties (see e.g., Judd and Kenny 1981; Shadish et al. 2002; Weisberg 1979).

A possible threat to the internal validity of conclusions about program effects in this design is what is usually called "history effects." Such effects can occur due to general time trends or some irrelevant (subject or environmental) factor that may have affected the intervention group(s) and not the baseline/comparison group(s). The current study, which used two consecutive cohorts of similar schools, permitted us to examine possible history effects by comparing initial assessments on key measures of interest from the adjacent cohorts. If there are differences in initial assessments among these cohorts, this *might* suggest that history effects are a partial explanation for the findings. If there are no such differences, history effects are clearly less likely.

Analytical Plan

The Mplus 8.0 program (Muthén and Muthén 2012) was used to analyze the multilevel (two-level) data, which consisted of individual students nested within schools. Program effects were based on school-aggregated outcome variables (Spybrook et al. 2011) and the reliabilities for the school-aggregated versions of these variables (based on an average "school" size of 310 students) were very high, ranging between .89 and .94 for the six scaled variables, and between .79 and .85 for the sexual bullying and cyberbullying variables (Kallestad et al. 1998, Table 5; Snijders and Bosker 2012, p. 26).

Our general model can be described as a multi-site block design (Spybrook et al. 2011), where schools represent the blocks or sites. As evident from the description of the study design noted above, same-aged students (in the same grade) from the same schools are compared across periods in time. This blocking is likely to considerably increase the power of the analyses.

The general (combined) model is:

$$Y_{ij} = y_{00} + y_{10}X_{ij} + u_{0j} + u_{1j} + e_{ij}$$

where Y_{ij} is the outcome variable, y_{00} is the average school mean at baseline (T0), X_{ij} is the treatment indicator (coded with sets of dummy variables reflecting program year), y_{10} is the main effect of the treatment (the average difference between the treatment conditions), u_{0j} the random error associated with the level-2 means, u_{1j} the random error associated with the treatment effects, and e_{ij} the random error associated with the students. Program year (T0 = baseline, T1 = 1 year after program start, T2 = 2 years after program start, T3 = 3 years after program start) is treated as a within-school (grade or grade grouping) treatment indicator (X_{ij}).

Six outcome variables (being verbally bullied, being physically bullied, being indirectly bullied, verbally bullying others, physically bullying others, and indirectly bullying others) were treated as continuous as they were average scores of two to three items. The robust maximum likelihood (MLR) estimator was used in all multilevel analyses except for analyses of possible interactions between program year and gender or race/ethnicity. In these cases, the Bayes estimator was used because it is less computationally demanding than MLR (Muthen and Asparouhov 2012). When a significant interaction effect was found, analyses were rerun separately for the groups involved. In order to facilitate identification of possible main developmental trends, we combined the students into sets of three grade levels (grades 3–5, 6–8, and 9–11), which roughly correspond to elementary, middle, and high school grades.

Four outcome variables, based on highly skewed individual items (being sexually bullied, being cyberbullied, sexually bullying others, and cyberbullying others) were treated as ordinal variables. These variables were analyzed with multilevel logistic regression using a probit link (Heck and Thomas 2015). A key indicator of program effects for the global variables is the unstandardized regression coefficients which, in our case with scales expressed in the same metric, is a measure of absolute change (the difference between the means of the groups compared). This measure has the advantage of being largely independent of the levels of baseline values of the groups compared and is easy to interpret.

With the continuous variables used in this study, a common individual-level effect size measure such as Cohen's d gives results that are misleadingly low, since a majority of the participants have scores of zero (= not been bullied, etc.) and cannot improve. Also, because program effects are based on school-aggregated variables, we calculated school-level Cohen's d's with the between-school standard deviation in the denominator (Hedges 2007, 2011) rather the individual-level variant. The school-level measure can give complementary information but is, like the individual-level variant, affected by variations in the standard deviations of the groups compared which can result in somewhat inflated or deflated values.

Results

Overall Program Effects for Different Forms of Bullying

Results for the five being bullied and the five bullying others variables are presented in Tables 3, 4, and 5. Since the total sample and most subsamples are very large, and because of the sizable number of analyses that have been undertaken, it is important not to over-interpret significance tests. As a result, we have chosen to focus our presentation of results on overall patterns of findings and trends, including our presentation of program by gender interactions and program by race/ethnicity interactions. Our interpretation of the findings is primarily based on the unstandardized regression coefficients (which are a measure of absolute change).

Being Bullied Beginning with the overall grade 3–11 analyses in Table 3 (for scale scores) and Table 4 (for ordinal variables), there were significant program effects for all five outcome variables. For four of them (being verbally, physically, indirectly, and sexually bullied), the effects for all three time comparisons (T0 vs. T1, T0 vs. T2, and T0 vs. T3, respectively) were highly significant ($p < .001$), and effects were generally stronger at a later time point than at the preceding point (although not always significantly so). Cohen's school-level d-values (column six in each table) were large/relatively large, ranging from .67 to .96, for the three variables with scale scores. Program effects for being cyberbullied were clearly

Table 3 Changes in various forms of being bullied (scale scores) across four time periods by grade groupings

Grades	Total N	B T0 vs. T1	B T0 vs. T2	B T0 vs. T3	d (S) T0 vs. T3	Additional contrasts
Being verbally bullied						
3–11	121,898	− 0.081***	− 0.114***	− 0.128***	$d = 0.96$	T1 vs. T2*** T1 vs. T3***
3–5	51,603	− 0.093***	− 0.142***	− 0.149***	$d = 1.09$	T1 vs. T2*** T1 vs. T3***
6–8	Girls, 24,583	− 0.099***	− 0.122***	− 0.127***	$d = 1.05$	T1 vs. T3*
	Boys, 25,062	− 0.057*	**− 0.070****	− 0.092***	$d = 0.73$	
	W and B, 34,168	− 0.095***	− 0.113***	− 0.123***	$d = 0.96$	
	H, 2444	0.026	**0.040**	− 0.055	$d = 0.27$	
9–11	20,384 [a]	− 0.039	− 0.061*	− 0.106***	$d = 1.39$	T1 vs. T3** T2 vs. T3***
Being physically bullied						
3–11	121,729	− 0.054***	− 0.076***	− 0.094***	$d = 0.67$	T1 vs. T2** T1 vs. T3*** T2 vs. T3**
3–5	Girls, 25,276	− 0.067***	− 0.084***	− 0.101***	$d = 0.93$	T1 vs. T3*
	Boys, 26,086	− 0.072***	**− 0.119****	**− 0.140****	$d = 1.03$	T1 vs. T2*** T1 vs. T3***
6–8	49,844	− 0.041**	− 0.056***	− 0.069***	$d = 0.67$	T1 vs. T3**
	W and B, 34,12[a]	− 0.048***	− 0.062***	− 0.064***	$d = 0.96$	T1 vs. T3*
	H, 2442	0.007	**0.071**	0.015	$d = − 0.15$	
9–11	20,366[a]	− 0.022	− 0.018	− 0.052***	$d = 0.98$	T1 vs. T3* T2 vs. T3**
Being indirectly bullied						
3–11	121,796	− 0.070***	− 0.101***	− 0.108***	$d = 0.74$	T1 vs. T2** T1 vs. T3***
3–5	51,545	− 0.085***	− 0.130***	− 0.132***	$d = 0.92$	T1 vs. T2** T1 vs. T3**
6–8	Girls, 24,573	− 0.094***	− 0.098***	− 0.086**	$d = 0.63$	
	Boys, 25,038	**− 0.039***	− 0.070***	− 0.086***	$d = 0.73$	T1 vs. T3*
	W, 32,140	− 0.088***	− 0.109***	− 0.096***	$d = 0.70$	
	B, 2008	**0.039**	− 0.063	− 0.147	$d = 0.67$	
	H, 2444	**0.091**	**0.085**	**0.144**	$d = − 1.08$	
9–11	Girls, 10,191[a]	− 0.036	− 0.053*	− 0.068**	$d = 0.76$	
	Boys, 10,026[a]	− 0.047*	− 0.063*	**− 0.126****	$d = 1.34$	T1 vs. T3*** T2 vs. T3**

T0 = baseline, T1 = Time1 (1 year later), T2 = Time 2 (2 years later), T3 = Time 3 (3 years later); B = unstandardized regression coefficient; d (S) = school-level effects

W = White, B = Black, H = Hispanic; **Bold** = significant interaction effect (compared to girls or White students)

*$p < .05$

**$p < .01$

***$p < .001$

[a] Between slope variances set to zero due to estimation problems because of low slope variance

weaker and more variable: Reductions emerged at T1 ($p < .05$) were lost at T2 but re-emerged at T3 ($p < .05$).

These overall results were largely confirmed in the analyses by grade groupings (Tables 3 and 4). In addition, there were 12 programs by gender interactions (out of 90 possible; in these two tables, a regression coefficient in bold signifies a significant interaction and the nature of the interaction is shown in the associated analyses on the next lines in the table). In five of the program by gender interactions, effects were stronger for girls and for the other seven, results were in the opposite direction. Program effects were somewhat stronger for girls for being verbally bullied and were somewhat stronger for boys for being physically bullied and cyberbullied. For being indirectly bullied, program effects were stronger for

Table 4 Changes in sexual bullying and cyberbullying across four time periods by grade groupings

Grades	Total N	B T0 vs. T1	B T0 vs. T2	B T0 vs. T3	d (S) T0 vs. T3	Additional contrasts
Victimization						
Being sexually bullied						
3–11	120,511	− 0.114***	− 0.164***	− 0.210***	d = 0.99	T1 vs. T2*
						T1 vs. T3***
						T2 vs. T3*
3–5	Girls, 24,828	− 0.134***	− 0.178***	− 0.194***	d = 0.82	
	Boys, 25,688	− 0.139***	**− 0.225***	**− 0.265***	d = 1.04	T1 vs. T2*
						T1 vs. T3**
	B/W, 30,475	− 0.139***	− 0.215***	− 0.244***	d = 0.98	T1 vs. T2*
						T1 vs. T3*
	H, 1147	− 0.187	− 0.195	**− 0.080**	d = 0.24	
6–8	Girls, 24,442[a]	− 0.156***	− 0.187***	− 0.252***	d = 1.01	T1 vs. T3*
	Boys, 24,863	− 0.073	**− 0.114**	− 0.171***	d = 0.66	T1 vs. T3*
	B/W, 33,959	− 0.130**	− 0.159***	− 0.230***	d = 0.92	T1 vs. T3*
	H, 2421	− 0.025	**0.009**	− 0.110	d = 0.28	
9–11	Girls, 10,156[a]	− 0.135*	− 0.173*	− 0.282**	d = 0.98	
	Boys, 9972	**− 0.076**	**− 0.058**	− 0.185*	d = 0.60	
Being cyberbullied						
3–11	119,938	− 0.046*	− 0.033	− 0.046*	d = 0.20	
3–5	Girls, 24,720	− 0.033	0.024	0.028	d = − 0.10	
	Boys, 25,495	− 0.064	**− 0.066**	**− 0.080**	d = 0.34	
6–8	49,324	− 0.087*	− 0.079*	− 0.080*	d = 0.33	
	B/W, 33,817	− 0.095*	− 0.087*	− 0.071	d = 0.30	
	H, 2405	0.228	**0.306**	0.151	d = − 0.39	
9–11	20,248	− 0.014	− 0.044	− 0.092	d = 0.30	
Perpetration						
Sexual bullying others						
3–11	118,916	− 0.154***	− 0.234***	− 0.356***	d = 1.38	T1 vs. T2**
						T1 vs. T3***
						T2 vs. T3***
3–5	49,806	− 0.206***	− 0.284***	− 0.397***	d = 1.38	T1 vs. T2*
						T1 vs. T3***
						T2 vs. T3**
6–8	Girls, 24,245	− 0.265***	− 0.360***	− 0.469***	d = 1.66	T1 vs. T3*
	Boys, 24,546	**− 0.105***	**− 0.209***	− 0.346***	d = 1.34	T1 vs. T2*
						T1 vs. T3***
						T2 vs. T3*
	B/W, 33,725	− 0.211***	− 0.279***	− 0.465***	d = 1.78	T1 vs. T3***
						T2 vs. T3**
	H, 2388	− 0.035	− 0.209	**− 0.154**	d = 0.36	
9–11	20,062	− 0.026	− 0.102	− 0.225**	d = 0.75	T1 vs. T3*
Cyberbullying others						
3–11	117,176	− 0.102***	− 0.096***	− 0.192***	d = 0.58	T1 vs. T3**
						T2 vs. T3**
3–5	Girls, 24,028	− 0.172**	− 0.046	− 0.114*	d = 0.41	T1 vs. T2*
	Boys, 24,704	− 0.082	− 0.066	**− 0.261***	d = 0.79	T1 vs. T3**
						T2 vs. T3**
6–8	48,398	− 0.116**	− 0.177***	− 0.247***	d = 0.99	T1 vs. T3**
	B/W, 33,283	− 0.150***	− 0.221***	− 0.256***	d = 1.03	T1 vs. T3*
	H, 2359	0.103	**− 0.040**	− 0.288	d = 0.69	T1 vs. T3*
9–11	19,901	− 0.082	− 0.084	− 0.179*	d = 0.61	

T0 = baseline, T1 = Time 1 (1 year later), T2 = Time 2 (2 years later), T3 = Time 3 (3 years later); B = unstandardized probit coefficient; d (S) = school-level effects

W = White, B = Black, H = Hispanic; **Bold** = significant interaction effect (compared to girls or White students)

*p < .05

**p < .01

***p < .001

[a] b convergence set to Mplus default convergence criterion to make it converge. Otherwise it was set to the more conservative 0.1; one tailed significance tests are reported for the ordinal analyses (as Bayesian estimator is used)

middle school girls vs. boys, but effects were stronger for high school boys vs. girls. For being sexually bullied, results were somewhat contradictory with three interactions favoring girls (grades 6–8 and 9–11) and two favoring boys (grades 3–5).

Table 5 Changes in various forms of bullying others (scale scores) across four time periods by grade groupings

Grades	Total N	B T0 vs. T1	B T0 vs. T2	B T0 vs. T3	d (S) T0 vs. T3	Additional contrasts
Verbally bullying others						
3–11	120,535	− 0.042***	− 0.071***	− 0.098***	d = 0.77	T1 vs. T2***
						T1 vs. T3***
						T2 vs. T3***
3–5	Girls, 24,932	− 0.033**	− 0.057***	− 0.066***	d = 0.67	T1 vs. T2**
						T1 vs. T3**
	Boys, 25,761	− 0.042***	− 0.074***	**− 0.090***	d = 0.79	T1 vs. T2***
						T1 vs. T3***
						T2 vs. T3*
6–8	49,461	− 0.055***	− 0.089***	− 0.122***	d = 1.06	T1 vs. T2**
						T1 vs. T3***
						T2 vs. T3**
9–11	Girls, 10,121 [a]	− 0.035	− 0.051	− 0.089***	d = 1.34	T1 vs. T3**
						T2 vs. T3**
	Boys, 9920 [a]	− 0.039	− 0.107**	**− 0.179***	d = 1.44	T1 vs. T2**
						T1 vs. T3***
						T2 vs. T3**
Physically bullying others						
3–11	120,195	− 0.012*	− 0.029***	− 0.045***	d = 0.49	T1 vs. T2***
						T1 vs. T3***
						T2 vs. T3***
3–5	Girls, 24,862	− 0.012	− 0.023***	− 0.019*	d = 0.31	
	Boys, 25,671	− 0.020**	**− 0.041***	**− 0.055***	d = 0.57	T1 vs. T2**
						T1 vs. T3***
6–8	49,332	− 0.007	− 0.034**	− 0.053***	d = 0.60	T1 vs. T2***
						T1 vs. T3***
						T2 vs. T3*
	W, 31,888	− 0.014	− 0.032***	− 0.046***	d = 0.89	T1 vs. T2**
						T1 vs. T3***
	B, 1984[a]	− 0.029	**− 0.091*	− 0.098*	d = 0.42	
	H, 2411[a]	− 0.028	**0.030	− 0.057	d = 0.46	T1 vs. T2*
						T2 vs. T3**
9–11	Girls, 10,094[a]	− 0.008	0.008	− 0.018	d = 0.50	T2 vs. T3*
	Boys, 9900 [a]	− 0.028	− 0.046	**− 0.093**	d = 0.96	T1 vs. T3**
Indirectly bullying others						
3–11	121,695	− 0.061***	− 0.102***	− 0.118***	d = 0.79	T1 vs. T2***
						T1 vs. T3***
3–5	Girls, 25,294	− 0.041	− 0.096***	− 0.106***	d = 1.32	T1 vs. T3*
						T1 vs. T2**
	Boys, 26,088	− 0.080***	**− 0.150***	**− 0.175***	d = 1.72	T1 vs. T2***
						T1 vs. T3***
	W and H, 30,147	− 0.079***	− 0.123***	− 0.140***	d = 1.52	T1 vs. T2**
	B, 2016 [a]	**0.086	− 0.133*	− 0.100	d = 1.40	T1 vs. T2**
						T1 vs. T3**
6–8	49,805	− 0.084***	− 0.097***	− 0.103***	d = 0.95	
9–11[c]	20,347	− 0.018	− 0.056*	− 0.079**	d = 0.84	T1 vs. T2*
						T1 vs. T3**
						T2 vs. T3*

T0 = baseline, T1 = Time 1 (1 year later), T2 = Time 2 (2 years later), T3 = Time 3 (3 years later); B = unstandardized regression coefficient; d (S) = school-level effects; W = White, B = Black, H = Hispanic; **Bold** = significant interaction effect (compared to girls or White students)

*p < .05

**p < .01

***p < .001

[a] Between slope variances set to zero due to estimation problems due low slope variance

Tables 3 and 4 also contain results of possible interactions of program effects with race/ethnicity. The three major groupings by race/ethnicity consisted of White students (62% of the population), Black students (4.5%), and Hispanic students (3.2%). The most striking overall result was that there was only one significant interaction that revealed a difference between White and Black students (students in grades 6–8 being indirectly bullied at T1). In all other analyses, Black and White students had similar results, which could be described by the same regression coefficients. Students with Hispanic background showed a somewhat different developmental pattern than White and

Black students, which resulted in some program year by ethnicity interactions, but none of the time comparisons for this group reached significance.

Looking at the size of the unstandardized regression coefficients in Table 3 for being bullied, program effects for the youngest grade group were somewhat larger than for the 6–8 group and the 9–11 group, in particular. However, for the other two being bullied variables, being sexually and cyberbullied (Table 4), no such clear grade-level trends emerged.

Bullying Other Students For all grade 3–11 analyses, there were marked program effects on the five outcome variables for bullying others (see Tables 4 and 5). All of them showed gradually stronger effects with increasing program exposure and the majority of the time comparisons were highly significant ($p < .001$). Cohen's school-level d-values varied between .49 for physically bullying others and .79 for indirectly bullying others. In contrast to what was found for being cyberbullied, program effects for reducing cyberbullying others were clearer with highly significant changes for all three time comparisons ($p > .001$), even though marked effects were most salient for the T0 vs T3 comparison.

These results were largely confirmed in the analyses based on grade groupings (Tables 4 and 5). In addition, there were a total of ten program by gender interactions. In two of the ten analyses (sexual bullying among 6th–8th graders at T1 and T2), program effects were greater for girls than boys. In the remaining eight analyses, program effects were greater for boys with regard to verbal bullying (observed at two grade levels, 3–5 and 9–11), physical bullying among students in grades 3–5 and 9–11), indirect bullying (among students in grades 3–5), and cyberbullying (among students in grades 3–5).

As revealed in Tables 4 and 5, there were few program year by race/ethnicity interactions for the variables measuring bullying others, and again, results for White students and Black students were generally quite similar. There were three instances in which program effects were significantly weaker for Hispanic students, all in grades 6–8.

With regard to possible grade/age level differences, no clear and consistent patterns were visible. For example, whereas, indirect bullying of others showed somewhat weaker effects in higher grades, an opposite trend was found for verbally bullying others.

Analyses of Possible History Effects To examine possible historical effects, we did separate multilevel analyses on the baseline assessments of all the outcome variables which were regressed on a dummy-coded cohort variable (with program start in 2008 = 0 or 2009 = 1), while controlling for age. The analyses revealed no significant differences across cohorts. Thus, there were no indications that the observed reductions might be a consequence of history effects.

Discussion

The purpose of this study was to evaluate the effectiveness of the Olweus Bullying Prevention Program in reducing all major forms of bullying—verbal bullying, physical bullying, and indirect/relational bullying, as well as cyberbullying and bullying using words or gestures with a sexual meaning. We examined students' self-reports of being bullied, as well as bullying others in a large-scale, longitudinal study involving more than 30,000 students in grades 3–11 from 95 schools over the course of 3 years. Previous research on the OBPP in Norway has produced consistently positive program effects (Olweus 1991, 2005; Olweus and Limber 2010a, 2010b) as has our recent large-scale evaluation of the program in Pennsylvania among youth in grades 3–11. The latter study, which used the same dataset as the present study, documented clear and largely consistent reductions in being bullied and bullying other students, measured with global questions and scale scores, for both male and female students across a wide range of grades/ages. The study also found improvements in several aspects of the school climate related to bullying (Limber et al. 2018).

Program Effects for Different Forms of Bullying

The results of the present study extend these previous findings of the effectiveness of the OBPP in US schools. The consistency of the results is striking, showing that the OBPP was successful in reducing all forms of being bullied and bullying others—verbal, physical, indirect, sexual, and electronic/cyberbullying. Analyses by grade groupings revealed that, with only a few exceptions (self-reports of cyber victimization among students in grades 3–5 and 9–11, and self-reports of physically bullying others by girls in grades 9–11), there were significant program effects for all forms of bullying, based on reports by youth who were bullied and youth who bullied others for all grade groupings. For most analyses, program effects were stronger the longer the program was in place. Most school-level program effects were large. As noted earlier, a possible threat to internal validity of conclusions about program effects in an extended age cohorts design is "history effects," which may result from general time trends or a factor unrelated to the program that may have affected the intervention group but not the baseline/control groups. Analyses revealed no indications of history effects in this study.

With regard to the three being bullied variables with scale scores (verbal, physical, and indirect forms), all of them showed substantial reductions, with somewhat larger effects for being verbally bullied. The counterpart of this variable, verbally bullying others, also showed substantial reductions. It is not surprising that the largest effects were observed with regard to verbal bullying. Verbal bullying has consistently

been found to be the most prevalent form of bullying (Musu-Gillette et al. 2018). Moreover, bullying with degrading, hostile comments is a component in almost all instances of bullying. The OBPP (through its training and print and video resources for educators) emphasizes the harms that verbal bullying can cause and the importance of addressing it. It is also worth emphasizing that generally positive results were also observed for being both exposed to, and actively participating in, more subtle, indirect/relational forms of bullying which are often difficult for school personnel to observe (Salmivalli et al. 2011). Results for being sexually bullied and sexually bullying others, although not directly comparable with the scaled variables, also showed clear and systematic reductions for all grade groupings.

Overall program effects (grades 3–11) were smallest for being cyberbullied ($d = 0.20$). Although most of the regression coefficients for the various grade groupings indicated reductions in being cyberbullied, not all of them were significant, and much of the positive change was carried by the students in grades 6–8. Generally, it was not unexpected that the program had weaker and somewhat less consistent effects on students' reports of being cyberbullied than on being bullied in other ways. Although various OBPP materials and OBPP training and consultation include attention to cyberbullying, prevention of cyberbullying was not an area of particular focus in the schools at that time. In spite of this, there were clear and significant reductions in cyberbullying others in all grade groupings. These results are in line with previous research that has demonstrated positive effects on cyberbullying of a general anti-bullying program without particular focus on reduction of cyberbullying (KiVa; Salmivalli et al. 2011). Later variants of the OBPP have included various program materials, including class resources (Limber et al. 2008; Limber et al. 2009), with a special focus on cyberbullying. We would expect even stronger program effects would emerge with such additional supports and this should be a focus of future research.

Program Effects for Different Grade Groupings

With respect to the relative strength of program effects at different grade groupings, there was, as expected, a trend for students in the youngest grade group to have somewhat stronger program effects than students in the higher age groups and in the 9th–11th grade group, in particular. The results are largely consistent with the overall findings on bullying that were obtained in our previous analyses on the same sample (Limber et al. 2018) and with the results of large-scale evaluations of the OBPP in Norway (e.g., Olweus 2013; Olweus and Limber 2010a, b). These findings may reflect differences in the environments of middle or high schools vs. elementary schools (such as the size of the student body and staff, multiple teachers for each student, less flexible schedules, greater

difficulties in finding time to hold class meetings, and somewhat different definitions of the roles of teachers (Olweus and Kallestad 2010)). At the same time, it is worth noting that for two of these being bullied variables (verbal and indirect forms), there was basically no difference in program strength between the two oldest grade groupings. The conclusion about a trend in favor of lower grades is also tempered by the fact that the other two being bullied variables (sexual bullying and cyberbullying) were not characterized by such a pattern. In addition and in line with findings from our previous analyses of the bullying others variable on the same sample (Limber et al. 2018), no clear and consistent grade/age trends were identified for the various forms of bullying other students.

Program Effects for Boys and Girls

Consistent with the findings from our previous studies (Limber et al. 2018; Olweus 2010) and those of other bullying prevention programs (Salmivalli et al. 2011; Williford et al. 2013) and in line with our expectations, we generally observed significant reductions in the different forms of being bullied for both boys and girls. This finding is very positive and suggests that the messages and components of the program have resonated for boys and girls alike. Although most analyses indicated similar and usually substantial effects for both boys and girls, there were 12 program by gender interactions as well. For about half of them, results were in favor of girls and for the other half, program effects were stronger for boys. Girls seemed to benefit more from the program with regard to verbal forms of being bullied, whereas the particularly positive program effects applied to physical forms of bullying for boys. As physical bullying is more common (Musu-Gillette et al. 2018) and likely more readily identified among boys versus girls, it is not surprising that some of the program effects on physical forms of bullying were more pronounced among boys. But even though these interactions are based on large numbers of students, they need to be tempered by the observation that several of these interactions were found for only one time point comparison, such as T0 vs T3, and none of them applied to all time point comparisons for a variable.

Although there were positive program effects for both boys and girls for most forms of bullying others, a very clear, dominant trend emerged in the program year by gender interactions. Of ten such interactions, eight applied to boys and only two to girls. The program effects for boys were in several comparisons/grade groups twice the size of the effects for girls (even when the latter effect was significant), such as for verbally bullying others (grade group 9–11) and physically bullying others (grade group 3–5). There was also a greater program effect for boys with regard to cyberbullying other students. The only variable where girls showed stronger effects than boys concerned sexually bullying others in grades 6–8.

Figure 1 portrays one of many program years by gender interactions, where program effects on verbally bullying others were greater for boys than girls (grades 9–11). The curves also visualize the well-documented fact that boys, overall, are much more involved in bullying others than girls with regard to direct forms of bullying. This conclusion applies both to targets of their own and of the opposite gender (Olweus 1993, 2010). In the present total sample, the gender difference for verbally, physically, and sexually bullying other students (at T0) were all quite large, with t-values in the 15.00 to 17.00 range ($p < .001$). Against this background, these highly consistent results with stronger program effects for boys must be seen as highly desirable since effective interventions on those who do most of the bullying are likely to have the greatest positive effects overall. However, this insight does not in any way negate the need to also focus vigorously on more subtle, and often less visible forms of bullying among girls.

Program Effects for Youth of Different Racial/Ethnic Groups

Considering possible racial/ethnic differences, the overriding message was that program effects for Black and White students were quite similar for most forms of being bullied and bullying others. These more detailed results were clearly stronger than those obtained in our previous analyses (Limber et al. 2018). This may in part reflect use of a statistically more powerful analytic strategy (including the various ethnic groups in the same analyses rather than conducting separate analyses for each ethnic group). Although these results are quite encouraging, it must be kept in mind that the total number of Black (and Hispanic) students in the studied sample were relatively small and underrepresented from a national perspective. Accordingly, our results cannot be directly generalized to

schools with a majority or a substantially larger proportion of Black students. Generalizations of this kind will have to wait until our findings may have been replicated in such schools. Although students with Hispanic background showed significant program effects for some grade groups and variables that paralleled the development for Black and White students, in several other cases (particularly in grades 6–8), they did not make much progress, which resulted in several program year by ethnicity interactions. There are a number possible explanations for these findings. First, since examination of the regression coefficients reveals that in some cases, trends were in positive directions, it is possible that a larger sample size of Hispanic students would have resulted in significant effects. Second, there may be differences in how Hispanic students understand, experience, and engage in bullying (Wang 2013), the degree to which they are receptive to prevention and intervention strategies, and the extent to which adults interact effectively with them. Future research is needed to evaluate these and other possible explanations.

Summary of Program Effects and Implications

In summary, this large-scale longitudinal study provides additional, strong support for the effectiveness of the OBPP among US students in elementary, middle, and high school grades. Program effects were broad, substantial, and highly consistent, covering all forms of bullying—verbal, physical, indirect, bullying through sexual words and gestures, and cyberbullying—both with regard to being bullied and bullying others. The fact that a program such as the OBPP, which was developed and tested out in Norway, has documented substantial effects with a large group of schools/students in the USA suggests that children and youth in the two countries have a good deal of similarities, at least with regard to bullying problems.
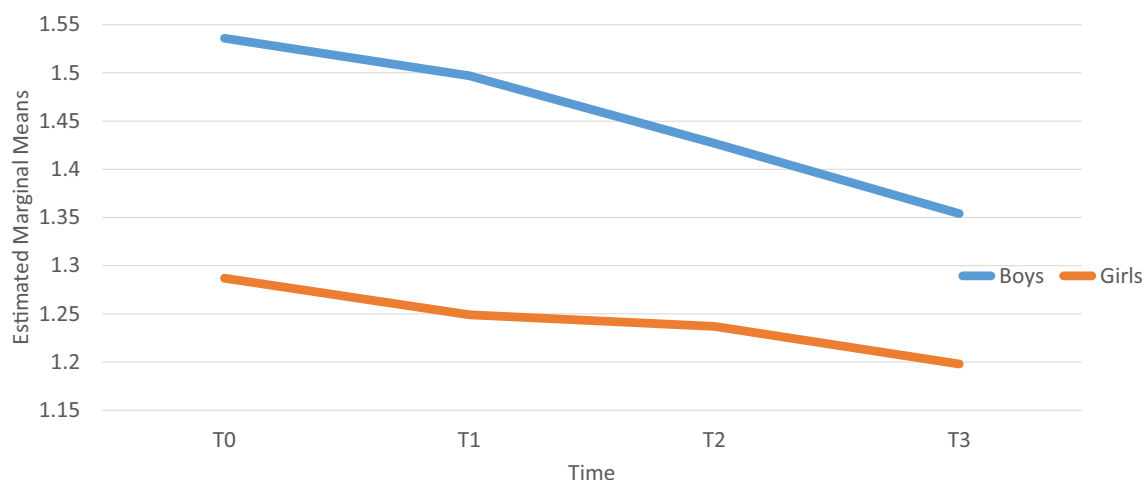

Fig. 1 Program by gender interaction effects for verbally bullying others

The positive and similar program effects in both countries also suggest that the program with its implementation model is built on a reasonably realistic view of the characteristics and mechanisms of the basic problem as manifested among children and youth in schools. In addition, the results indicate that the program has actually captured some important principles and mechanisms for changing and preventing the targeted problems, although it may be difficult at present to specify exactly what these mechanisms are since the program is a whole-school "package" with several levels and components. In this context, we want to emphasize that the OBPP is not a "program" in a narrow sense but should rather be seen as a coordinated collection of research-based components that form a unified whole-school approach to bullying. In our view, most of these components should be in place in all schools that want to create a safe and productive learning environment for their students. Lastly, reflecting on the call for a complete reorientation of society's anti-bullying efforts mentioned in the introduction (Cohen et al. 2015; Espelage et al. 2018; Hong and Espelage 2012), we cannot escape the impression that this call was a bit premature.

Strengths and Limitations

This study had a number of strengths, including a very large sample size, multiple forms of bullying across a wide range of grade levels, and a strong quasi-experimental design (the extended age cohort design). It also had a longitudinal nature, which allowed us to follow students over 3 years and examine year-by-year changes in self-reports of being bullied and bullying others. In addition, program effects were evaluated at the aggregate level (level 2) and the aggregate reliabilities of the scales and also individual items were quite high, in the .80's and .90's.

Outcome variables were limited to students' self-reports. However, the Olweus Bullying Questionnaire (Olweus 1996) has been extensively validated (e.g., Breivik and Olweus 2015; Olweus 2013; Solberg and Olweus 2003). Although questions have been raised about the effectiveness of the OBPP with minority youth (Bauer et al. 2007; Limber et al. 2018), findings from this study revealed that program effects were significant and basically similar for Black and White students. Additional research would be valuable to further examine the effectiveness of the OBPP with larger and more representative samples of Black and Hispanic youth, as well as children of other races/ethnicities. Building upon this study, additional research is also needed to examine the association between program fidelity (e.g., the fidelity with which specific components of the program were implemented) to program outcomes.

References

Bauer, N. S., Lozano, P., & Rivara, F. P. (2007). The effectiveness of the Olweus Bullying Prevention Program in public middle schools: a controlled trial. *Journal of Adolescent Health, 40*, 266–274. https://doi.org/10.1016/j.jadohealth.2006.10.005.

Breivik, K., & Olweus, D. (2015). An item response theory analysis of the Olweus Bullying scale. *Aggressive Behavior, 41*, 1–13. https://doi.org/10.1002/ab.21571.

Cohen, J., Espelage, D. L., Twemlow, S. W., Berkowitz, M. W., & Comer, J. P. (2015). Rethinking effective bully and violence prevention efforts: promoting healthy school climates, positive youth development, and preventing bully-victim-bystander behavior. *International Journal of Violence and schools, 15*(1), 2–40.

Cook, R. C., Williams, K. R., Guerra, N. G., Kim, T. E., & Sadek, S. (2010). Predictors of bullying and victimization in childhood and adolescence: a meta-analytic investigation. *School Psychology Quarterly, 25*, 65–83. https://doi.org/10.1037/a0020149.

Espelage, D. L., King, M. T., & Colbert, C. L. (2018). Emotional intelligence and school-based bullying prevention and intervention. In *Emotional Intelligence in Education* (pp. 217–242). Springer.

Evans, C. B., Fraser, M. W., & Cotter, K. L. (2014). The effectiveness of school-based bullying prevention programs: A systematic review. *Aggression and Violent Behavior, 19*(5), 532–544. https://doi.org/10.1111/jora.12060.

Gladden, R. M., Vivolo-Kantor, A. M., Hamburger, M. E., & Lumpkin, C. D. (2014). Bullying surveillance among youths: uniform definitions for public health and recommended data elements, version 1.0. Atlanta, GA: National Center for Injury Prevention and Control, Centers for Disease Control and Prevention and U.S. Department of Education.

Heck, R. H., & Thomas, S. L. (2015). *An introduction to multilevel modelling techniques: MLM and SEM approaches using Mplus*. New York: Routledge.

Hedges, L. (2007). Effect sizes in cluster-randomized designs. *Journal of Educational and Behavioral Statistics, 32*, 341–370. https://doi.org/10.3102/1076998606298.

Hedges, L. (2011). Effect sizes in three-level cluster-randomized experiments. *Journal of Educational and Behavioral Statistics, 36*, 346–380. https://doi.org/10.3102/10769986103766.

Hong, J. S., & Espelage, D. L. (2012). A review of research on bullying and peer victimization in school: an ecological system analysis. *Aggression and Violent Behavior, 17*(4), 311–322.

Judd, C. M., & Kenny, D. A. (1981). *Estimating the effects of social interventions*. New York: Cambridge University Press.

Kallestad, J. H., Olweus, D., & Alsaker, F. (1998). School climate reports from Norwegian teachers: a methodological and substantive study. *School Effectiveness and School Improvement, 9*, 70–94. https://doi.org/10.1080/0924345980090104.

Limber, S. P., & Olweus, D. (2017). Lessons learned from scaling-up the Olweus Bullying Prevention Program. In C. Bradshaw (Ed.), *Handbook on bullying prevention: a life course perspective* (pp. 189–199). Washington, DC: National Association of Social Workers Press.

Limber, S. P., Nation, M., Tracy, A. J., Melton, G. B., & Flerx, V. (2004). Implementation of the Olweus Bullying Prevention programme in the southeastern United States. In P. K. Smith, D. Pepler, & K. Rigby (Eds.), *Bullying in schools: how successful can interventions be?* (pp. 55–79). Cambridge: Cambridge Press.

Limber, S. P., Agatston, P. W., & Kowalski, R. M. (2008). *Cyber bullying: a prevention curriculum for grades 6–12*. Center City: Hazelden.

Limber, S. P., Agatston, P. W., & Kowalski, R. (2009). *Cyber bullying: a prevention curriculum for grades 3–5*. Center City: Hazelden.

Limber, S. P., Olweus, D., Wang, W., Masiello, M., & Breivik, K. (2018). Evaluation of the Olweus Bullying Prevention Program: a large scale study of U.S. students in grades 3-11. *Journal of School Psychology, 69*(4), 56–72.

Masiello, M. G., & Schroeder, D. (2014). *A public health approach to bullying prevention*. Washington, DC: American Public Health Association.

Musu-Gillette, L., Zhang, A., Wang, K., Zhang, J., Kemp, J., Diliberti, M., and Oudekerk, B.A. (2018). Indicators of school crime and safety: 2017 (NCES 2018-036/NCJ 251413). National Center for education statistics, U.S. Department of Education, and Bureau of Justice Statistics, Office of Justice Programs, U.S. Department of Justice. Washington, DC.

Muthen, B. O., & Asparouhov, T. (2012). Bayesian structural equation modeling: a more flexible representation of substantive theory. *Psychological Methods, 17*, 313–335. https://doi.org/10.1037/a0026802

Muthén, L. K., & Muthén, B. O. (2012). *Mplus User's Guide, 7th Ed.* Los Angeles: Muthén and Muthén.

National Academies of Sciences, Engineering, and Medicine. (2016). *Preventing bullying through science, policy, and practice*. Washington, DC: The National Academies Press.

Olweus, D. (1973). *Hackkycklingar och översittare. Forskning om skolmobbning*. Stockholm: Almqvist & Wicksell.

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

Olweus, D. (1991). Bully/victim problems among schoolchildren: basic facts and effects of a school based intervention program. In D. J. Pepler & K. H. Rubin (Eds.), *The development and treatment of childhood aggression* (pp. 411–448). Hillsdale: Erlbaum.

Olweus, D. (1993). *Bullying at school: what we know and what we can do*. New York: Blackwell.

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

Olweus, D. (1997). Bully/victim problems in school: facts and intervention. *European Journal of Psychology of Education, 12*, 495–510. https://doi.org/10.1007/BF03172807

Olweus, D. (2005). A useful evaluation design, and effects of the Olweus Bullying Prevention Program. *Psychology, Crime & Law, 11*, 389–402. https://doi.org/10.1080/10683160500255471

Olweus, D. (2007a). *Olweus bullying questionnaire*. Center City: Hazelden.

Olweus, D. (2007b). *Olweus bullying questionnaire: standard school report*. Center City: Hazelden.

Olweus, D. (2010). Understanding and researching bullying: some critical issues. In S. S. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *Handbook of bullying in schools: an international perspective* (pp. 9–33). New York: Routledge.

Olweus, D. (2013). School bullying: development and some important challenges. *Annual Review of Clinical Psychology, 9*, 751–780. https://doi.org/10.1146/annurev-clinpsy-050212-185516

Olweus, D., & Kallestad, J. H. (2010). The Olweus Bullying Prevention Program: effects of classroom components at different grade levels. In K. Osterman (Ed.), *Indirect and direct aggression* (pp. 113–131). New York: Peter Lang.

Olweus, D., & Limber, S. P. (2010a). Bullying in school: evaluation and dissemination of the Olweus Bullying Prevention Program. *American Journal of Orthopsychiatry, 80*, 124–134. https://doi.org/10.1111/j.1939-0025.2010.01015.x.

Olweus, D., & Limber, S. P. (2010b). The Olweus Bullying Prevention Program: implementation and evaluation over two decades. In S. R. Jimerson, S. M. Swearer, & D. L. Espelage (Eds.), *Handbook of bullying in schools: An international perspective* (pp. 377–401). New York: Routledge.

Olweus, D., Limber, S. P., Flerx, V., Mullin, N., Riese, J., & Snyder, M. (2007). *Olweus Bullying Prevention Program: schoolwide guide*. Center City: Hazelden.

Olweus, D., Solberg, M.,& Breivik, K. (2018). Long-term school-level effects of the Olweus Bullying Prevention Program (OBPP). *Scandinavian Journal of Psychology* (online).

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

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

Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis (2nd edition)*. London: Sage.

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

Spybrook, J., Bloom, H., Congdon, H. Hill, C., Martinez, A., Raudenbush, S. (2011). Optimal design plus empirical evidence: documentation for the "Optimal Design" software. Retrieved from: http://hlmsoft.net/od/od-manual-20111016-v300.pdf Google Scholar.

Ttofi, M., & Farrington, D. (2009). What works in preventing bullying: effective elements of programmes. *Journal of Aggression, Conflict and Peace Research, 1*, 13–24. https://doi.org/10.1108/17596599200900003.

Ttofi, M. M., & Farrington, D. P. (2011). Effectiveness of school-based programs to reduce bullying: a systematic and meta-analytic review. *Journal of Experimental Criminology, 7*, 27–56. https://doi.org/10.1007/s11292-010-9109-1.

Ttofi, M. M., Farrington, D. P., Lösel, F., & Loeber, R. (2011). Do the victims of school bullies tend to become depressed later in life? A systematic review and meta-analysis of longitudinal studies. *Journal of Aggression, Conflict and Peace Research, 3*, 63–73. https://doi.org/10.1108/17596591111132873.

Ttofi, M. M., Farrington, D. P., & Lösel, F. (2012). School bullying as a predictor of violence later in life: a systematic review and meta-analysis of prospective longitudinal studies. *Aggression and Violent Behavior, 17*, 405–418. https://doi.org/10.1016/j.avb.2012.05.002.

Wang, W. (2013). Bullying among U.S. school children: An examination of race/ethnicity and school-level variables on bullying (Order No. 3592547). Available from Dissertations & Theses @ Clemson University. (1437006141). Retrieved from http://libproxy.clemson.edu/login?url=https://search-proquest-com.libproxy.clemson.edu/docview/1437006141?accountid=6167.

Weisberg, H. I. (1979). Statistical adjustments and uncontrolled studies. *Psychological Bulletin, 86*, 1149–1164. https://doi.org/10.1037/0033-2909.86.5.1149.

Williford, A., Elledge, L. C., Boulton, A. J., DePaolis, K. J., Little, T. D., & Salmivalli, C. (2013). Effects of the KiVa antibullying program on cyberbullying and cybervictimization frequency among Finnish youth. *Journal of Clinical Child and Adolescent Psychology, 42*, 820–833.

Woods, S., & Wolke, D. (2003). Does the content of anti-bullying policies inform us about the prevalence of direct and relational bullying behaviour in primary schools? *Educational Psychology, 23*, 381–401.