

Effectiveness of school-based programs to reduce bullying: a systematic and meta-analytic review

Maria M. Ttofi · David P. Farrington

Published online: 16 September 2010
© Springer Science+Business Media B.V. 2010

Abstract This article presents a systematic review and meta-analysis of the effectiveness of anti-bullying programs in schools. Studies were included if they evaluated the effects of an anti-bullying program by comparing an intervention group who received the program with a control group who did not. Four types of research design were included: a) randomized experiments, b) intervention-control comparisons with before-and-after measures of bullying, c) other intervention-control comparisons, and d) age-cohort designs. Both published and unpublished reports were included. All volumes of 35 journals from 1983 up to the end of May 2009 were hand-searched, as were 18 electronic databases. Reports in languages other than English were also included. A total of 622 reports concerned with bullying prevention were found, and 89 of these reports (describing 53 different program evaluations) were included in our review. Of the 53 different program evaluations, 44 provided data that permitted the calculation of an effect size for bullying or victimization. The meta-analysis of these 44 evaluations showed that, overall, school-based anti-bullying programs are effective: on average, bullying decreased by 20–23% and victimization decreased by 17–20%. Program elements and intervention components that were associated with a decrease in bullying and victimization were identified, based on feedback from researchers about the coding of 40 out of 44 programs. More intensive programs were more effective, as were programs including parent meetings, firm disciplinary methods, and improved playground supervision. Work with peers was associated with an increase in

M. M. Ttofi

Leverhulme and Newton Trust Early Career Fellow, Institute of Criminology, Cambridge University, Sidgwick Avenue, Cambridge CB3 9DA, UK
e-mail: mt394@cam.ac.uk

D. P. Farrington

Psychological Criminology, Institute of Criminology, Cambridge University, Sidgwick Avenue, Cambridge CB3 9DA, UK

D. P. Farrington (✉)

Institute of Criminology, Sidgwick Avenue, Cambridge CB3 9DA, UK
e-mail: dpf1@cam.ac.uk

victimization. It is concluded that the time is ripe to mount a new program of research on the effectiveness of anti-bullying programs based on these findings.

Keywords School bullying · Intervention programs · Program elements · Systematic review · Meta-analysis

Background

Given the serious short-term and long-term effects of bullying on children's physical and mental health (Ttofi and Farrington 2008), it is understandable why school bullying has increasingly become a topic of both public concern and research efforts. Many school-based intervention programs have been devised and implemented in an attempt to reduce school bullying. The first large-scale anti-bullying program was implemented nationally in Norway in 1983. A more intensive version of the national program was evaluated in Bergen by Olweus (1991). This evaluation showed a dramatic decrease of about half in victimization (being bullied) after the intervention. Since then, many other anti-bullying strategies have been implemented but less frequently evaluated.

The definition of school bullying includes several key elements: physical, verbal, or psychological attack or intimidation that is intended to cause fear, distress, or harm to the victim; an imbalance of power (psychological or physical) with a more powerful child (or children) oppressing less powerful ones; and repeated incidents between the same children over a prolonged period of time (Farrington 1993; Olweus 1993b). According to this definition, it is not bullying when two persons of the same strength (physical, psychological, or verbal) victimize each other. School bullying can occur in school or on the way to or from school.

American research is generally targeted on school violence or peer victimization rather than bullying. Bullying is different from school violence or peer victimization. For example, bullying includes being called nasty names, being rejected, ostracized, or excluded from activities, and having rumors spread about you (Baldry and Farrington 1999). Also, bullying involves an imbalance of power and repeated acts. There are a number of existing reviews of school violence programs and school-based interventions for aggressive behavior (e.g., Howard et al. 1999; Mytton et al. 2006; Wilson and Lipsey 2007). We have consulted these, but we must emphasize that our research aims to review programs that are explicitly designed to reduce bullying and that explicitly measure bullying. Bullying is a type of aggressive behavior (Andershed et al. 2001; Salmivalli and Nieminen 2002). However, it should not be equated with aggression or violence; not all aggression or violence involves bullying, and not all bullying involves aggression or violence.

The most informative single source of reports of anti-bullying programs is the book edited by Smith et al. (2004a), which contains descriptions of 13 programs implemented in 11 different countries. Baldry and Farrington (2007) reviewed 16 major evaluations in 11 different countries, of which five involved an uncontrolled methodological design. There are also some reviews containing summaries of major anti-bullying programs (e.g., Rigby, 2002; Smith et al. 2003). The most relevant existing reviews are by Smith et al. (2004), who summarized effect sizes in 14

whole-school anti-bullying programs, and by Vreeman and Carroll (2007), who reviewed 26 school-based programs. However, neither carried out a full meta-analysis measuring weighted mean effect sizes and correlations between study features and effect sizes.

Smith et al. (2004) reviewed 14 evaluations up to 2002, six of which were uncontrolled. Vreeman and Carroll (2007) reviewed 26 evaluations up to 2004, restricted to studies published in the English language and with only 15 programs specifically concerned with bullying. Another meta-analytic review was published by Ferguson et al. (2007). However, this included searches in only one database (PsycINFO) for articles published between the years 1995 and 2006. It included outcome variables that measured ‘some element of bullying behavior or aggression toward peers, including direct aggressive behavior toward children in a school setting’ (p. 407). The latest meta-analytic review was completed by Merrell et al. (2008). However, this included searches in only two databases (PsycINFO and ERIC) for studies published only in English, and it included a wide range of outcome measures; there were only eight studies where the outcome was self-reported bullying and only ten studies where the outcome was self-reported victimization.

The present review includes many more evaluations (53 in total) and aims to investigate the effectiveness of program components. Special efforts were made to avoid problems arising from duplicate publications. For example, the Flemish Antibullying Program¹ was evaluated once and the results were disseminated in four publications. However, in contrast to previous reviews (e.g., Merrell et al. 2008), we carefully coded it as only one evaluation. As another example, findings on the effectiveness of the Olweus Bullying Prevention Program were disseminated in 22 publications, but the program was tested in only eight separate evaluations.

In the present report, we go beyond previous reviews by: a) doing much more extensive searches for evaluations such as hand-searching all volumes of 35 journals from 1983 up to the end of May 2009; b) searching for international evaluations in 18 electronic databases and in languages other than English; c) carrying out much more extensive meta-analyses (including correlating effect sizes with study features, research design, and program components); and d) focusing only on programs that are specifically designed to reduce bullying and not aggressive behavior (i.e., the outcome variables specifically measure bullying). This article is based on our review for the Campbell Collaboration (Farrington and Ttofi 2009); a previous review for the Swedish National Council for Crime prevention (Ttofi et al. 2008) was based on the largest evaluations (including at least 200 students). We expected that the smaller evaluations might be affected by publication bias, but found none in the Campbell review (Farrington and Ttofi 2009, p. 68).

All the programs are described in our Campbell review, but it is useful to describe one program here. The KiVa program from Finland (Karna et al. *forthcoming*) used the Internet (including password-specific online questionnaires and Web-based forums for teachers) and visual learning environments (e.g., computer games involving bullying) to change students’ attitudes about bullying. All materials were highly structured, corresponding to teaching periods with specific aims. KiVa

¹ Not included in the meta-analysis for reasons explained in the Campbell review.

included classroom discussions, group work, short films about bullying, and role-playing exercises. Teachers were trained and issued with special vests to wear in the schoolyard to enhance their visibility. The program also included peer support groups for victims of bullying and cooperative group work among experts in dealing with children involved in bullying. Finally, parents were given information and advice about bullying in guides and during special information nights.

Methods

Measuring the effects of programs

Four main research designs have been used in evaluating the effectiveness of anti-bullying programs: a) randomized experiments, b) intervention-control comparisons with before-and-after measures of bullying, c) other intervention-control comparisons, and d) age-cohort designs, where students of age X after the intervention were compared with different students of the same age X in the same school before the intervention.

In principle, a randomized experiment has the highest possible internal validity, in demonstrating that an intervention has an effect on an outcome. This is because, providing that a sufficiently large number of units is randomly assigned, the experimental and control conditions are equated (within the limits of statistical fluctuation) on all measured and unmeasured extraneous variables (Weisburd et al. 2001). In principle, therefore, the randomized experiment eliminates the threat to internal validity of selection effects. However, other threats, such as differential attrition from experimental and control conditions, can still be problematic (Farrington 2003). In contrast, quasi-experimental methods, using matching or statistical control techniques, can control for measured confounders but not for unmeasured ones. Therefore, they are generally less able to deal with the problem of selection effects.

Some scholars raise the concern that randomized experiments are likely to have low external validity because it is often difficult to ensure cooperation from institutions that are willing to randomize participants (Weisburd et al. 2001, p. 67). Therefore, these experiments may be implemented in unrepresentative conditions. Also, with randomization at the school level, in real life it is very difficult to control for diffusion of the program elements across control schools. In our review, only two evaluations (Fekkes et al. 2006; Smith et al. 2004b) provided key information about the percentage of intervention and control schools that implemented each program component.

In research on anti-bullying programs, schools or school classes, rather than children, are usually randomly assigned to receive the program². In some evaluations, a very small number of schools (between three and seven) were randomly assigned, threatening statistical conclusion validity. It is not true in all cases that randomized experiments on anti-bullying programs are methodologically superior to quasi-experimental evaluations with before-and-after measures of

² See table 10 from our Campbell review.

bullying in intervention and control conditions. It is clear that these two designs are potentially the best methodologically. The main threat to internal validity in them is differential attrition from intervention and control conditions. In addition, if the intervention classes are worse than the control classes to start with, regression to the mean could be a problem.

Non-randomized comparisons of intervention and control classes with no prior measures of bullying are clearly inferior to non-randomized comparisons with prior measures. Where there are no prior measures of bullying, it is important to include some pretest measures that establish the comparability of intervention and control children. Otherwise, this design is vulnerable to selection and regression effects in particular.

The age-cohort design, in which children of a certain age X in year 1 before the intervention are compared with (different) children of the same age X in the same school after the intervention in year 2, was pioneered by Olweus (1991). It largely eliminates problems of selection, aging, regression, and differential attrition, but it is vulnerable to history and testing effects. However, Olweus (2005a) argued convincingly that these were unlikely, especially since the effects of programs have been investigated in many different time periods. This design is likely to have high external validity.

Overall, the intervention-control comparisons and age-cohort designs might be regarded by some researchers as methodologically inferior to the randomized experiments and intervention-control/before-and-after designs, but all designs have advantages and limitations. These are the best and most controlled designs that have been used to evaluate the effects of anti-bullying programs, and we give credence to all of them in providing useful information about the effectiveness of anti-bullying programs. However, we do present results separately for the different designs so that they can be compared.

Criteria for inclusion or exclusion of studies

We used the following criteria for inclusion of studies in our systematic review:

- (a) The study described an evaluation of a program designed specifically to reduce school (kindergarten to high school) bullying. Studies of peer aggression or violence were excluded. For example, the study by Woods and colleagues (2007)³ was excluded since the intervention aimed to reduce peer victimization in general and not bullying perpetration or victimization (being bullied). Other reports were excluded from the present review because they were focused on the impact of an anti-bullying program on other outcome measures such as educational attainment (e.g., Fonagy et al. 2005), knowledge about and attitudes towards bullying (e.g., Meraviglia et al. 2003), or children's safety awareness with regard to different types of potentially unsafe situations, including being bullied (e.g., Warden et al. 1997).

³ See Woods et al. (2007: 379) for the outcome measures of the evaluation which did not include any measure of school bullying.

- (b) A clearly stated definition of bullying was included in the evaluation report and bullying was specified as the outcome measure of interest.
- (c) Bullying (specifically) was measured using self-report questionnaires, peer ratings, teacher ratings, or observational data.
- (d) The effectiveness of the program was measured by comparing students who received it (the intervention condition) with a comparison group of students who did not receive it (the control condition). We require that there must have been some control of extraneous variables in the evaluation (establishing the equivalence of conditions) by (i) randomization, or (ii) pre-test measures of bullying, or (iii) choosing some kind of comparable control condition. Because of low internal validity, we exclude uncontrolled studies that only had before-and-after measures of bullying in intervention schools or classes. However, we include studies that controlled for age and school. As mentioned, in the Olweus (1991) evaluation, all students received the anti-bullying program, but Olweus compared students of age X after the program (the intervention condition) with different students of the same age X in the same schools before the program (the control condition). We include this kind of age-cohort design because arguably the intervention and control students are comparable (at least in age and in attending the same schools).
- (e) Published and unpublished reports of research conducted in developed countries between 1983 and May 2009 are included. We believe that there was no worthwhile evaluation research on anti-bullying programs conducted before the pioneering research of Olweus, which was carried out in 1983.
- (f) It was possible to measure the effect size. The main measures of effect size are the odds ratio, based on numbers of bullies/non-bullies (or victims/non-victims), and the standardized mean difference, based on mean scores on bullying and victimization (being bullied). These measures are mathematically related (Lipsey and Wilson 2001: 202). Where the required information is not presented in reports, we have tried to obtain it by contacting the authors directly. Some evaluations of programs involving controlled methodological designs were included in the systematic review but not in the meta-analysis because they did not provide enough data to allow us to calculate an effect size (see Appendix). Some other controlled studies are included (e.g., the KiVa project: Karna et al. forthcoming) even though the final evaluation is not completed. In this case, we use the available evaluation data with the caveat that the final results may be different.

Searching strategies

- (a) We started by searching for the names of established researchers in the area of bullying prevention (e.g., Australia, Ken Rigby; England, Peter K. Smith; Finland, Christina Salmivalli; Greece, Eleni Andreou; Spain, Rosario Ortega; Norway, Dan Olweus). This searching strategy was used in different databases in order to initially obtain as many evaluations of known research programs as possible.

- (b) We then searched by using several keywords in different databases. In total, we carried out the same searching strategies in 18 electronic databases (see Campbell review, Table 1). In all databases, the same key words were used with different combinations. More specifically:
Bully/Bullies/Anti-Bullying/Bully-Victims/Bullying
AND: School
AND: Intervention/Program/Outcome/Evaluation/Effect/Prevention/Tackling/Anti-bullying
We did not include ‘violence’ or ‘aggression’ as keywords along with Bully/Bullies/Anti-Bullying/Bully-Victims because we knew that this would identify many studies that were not relevant to the present review, which focuses specifically on studies designed to reduce school bullying.
- (c) We also hand-searched 35 journals either online or in print, from 1983 until end of May 2009 (see Campbell review, Table 2).
- (d) We sought information from key researchers on bullying and from international colleagues in the Campbell Collaboration. In March 2008, we had a meeting with key educational users of the information in Copenhagen, organized by the Nordic Campbell Centre (SFI Campbell). Where we identified a report in a language other than English⁴, we asked colleagues to provide us with a brief translation of key features that were needed for our coding schedule. We believe that, with the cooperation of colleagues in the Campbell Collaboration, we are able potentially to include research in many different developed countries.
- (e) A stipulation was made that the title or abstract of each paper would have to include one of the essential keywords that were searched. However, some book chapters, mainly from edited books on bullying prevention, were included even though their titles and/or abstracts (if provided) did not include any of our keywords.

Results of searches

Studies found

A total of 622 reports that were concerned with interventions to prevent school bullying, as indicated in either the title or the abstract, are included in our systematic review. All reports were categorized based on a relevance scale that we constructed (Table 1). Table 2 shows the percentage of studies within each category. The vast majority (40.7%) were somewhat relevant (category 2), making general suggestions about reducing bullying or, more rarely, reviewing anti-bullying programs. With regard to the reports that we did not obtain (16, or 2.6%), most of them were Masters or PhD theses. Moving on to the obtained reports, 89 (14.3%) were eligible for inclusion in our Campbell review (categories 5 and 6). It is regrettable that a fair number of evaluations of anti-bullying programs were excluded from our review (category 4: 11.4%) because of their (uncontrolled) methodological design.

⁴ For example: Ciucci and Smorti 1998; Gini et al. 2003; Martin et al. 2005; Sprober et al. 2006.

Table 1 Categorization of reports based on their relevance to the present review

1: minor relevance; recommendations for integration of survey results into anti-bullying policies; and/or talk generally about the necessity for bullying interventions.
2: weak relevance; talking more specifically about anti-bullying programs [description of more than one anti-bullying program]; and/or reviews of anti-bullying programs; and/or placing emphasis on suggestions/recommendations for reducing bullying.
3: medium relevance; description of a specific anti-bullying program.
4: strong relevance; evaluation of an anti-bullying program, but not included because it has no experimental versus control comparison, or no outcome data on bullying.
5: included in the Campbell review; evaluation of an anti-bullying program that has an experimental and control condition [n may be <200 ; teacher and peer nominations may also be included as outcome measures].
6: also included in the Swedish review; evaluation of an anti-bullying program that has an experimental and control condition [$n >200$, self-reported bullying as outcome measure].

Of the 89 reports (of 53 evaluations) that were eligible for inclusion in our comprehensive Campbell review, 62 reports involved 32 evaluations of programs with a sample size more than 200, and 15 reports involved 12 evaluations of programs with a sample size less than 200. Twelve reports of nine evaluations did not provide enough data to allow the calculation of an effect size and were, therefore, not included in the meta-analysis (see Table 2).

Included evaluations

The 89 reports of 53 evaluations were divided into four categories of research design: randomized experiments, before-and-after quasi-experimental designs, other quasi-experimental designs, and age-cohort designs. The [Appendix](#) lists the 89 reports included in the present systematic review. For each evaluation, all relevant reports are presented so that readers can follow up according to their own interests. Within each of the four categories of research design, reports were grouped based on

Table 2 Percentage of reports and evaluations of programs within each category

Category	No of reports	No of evaluations	Percentage
Not Obtained	16	—	2.6%
Category 1	100	—	16.1%
Category 2	253	—	40.7%
Category 3	93	—	15.0%
Category 4	71	—	11.4%
Category 5	18	15 [3 excluded] ^a	2.9%
Category 6	71	38 [6 excluded] ^b	11.4%

a. 3 evaluations presented in 3 reports were excluded from the meta-analysis (see Table 3 for relevant references)

b. 6 evaluations presented in 9 reports were excluded from the meta-analysis (see Table 3 for relevant references)

the program evaluated. It was quite possible for different reports from a particular project to be placed in different categories, depending on the content of the report.

Analysis of included evaluations

Analysis of effect sizes

Table 3 summarizes key results of the 44 program evaluations that presented effect size data. Our aim was to identify the best available effect size measures in each evaluation⁵. The measure of effect size is the odds ratio (OR) with its associated 95% confidence interval (CI). Where the CI includes the chance value of 1.0, the OR is not statistically significant. Smaller studies ($n < 200$ students) are indicated with an asterisk. In all cases, the effect sizes for smaller studies were non-significant. Random-effects models were used to calculate the weighted mean effect sizes. Figure 1 shows the accompanying forest graph for bullying effect sizes. In this figure, the measure of effect size is the logarithm of OR (LOR). Figure 1 shows that the majority of evaluations found that the program was followed by a reduction in bullying.

Only one of the nine randomized experiments (Fonagy et al. 2009) found a significant effect of the program on bullying, although one other evaluation (Hunt 2007) reported a near-significant effect. Overall, the nine randomized experiments yielded a weighted mean OR of 1.10, indicating a non-significant effect of these programs on bullying. In contrast, five of the 14 evaluations with before-and-after/intervention-control designs found a significant effect, and one other (Olweus/Bergen 2) reported a near-significant result. Overall, these 14 studies yielded a large weighted mean OR of 1.60 ($p < .0001$).

One of the four other intervention-control comparisons found significant effects on bullying (Ortega et al. 2004), and the weighted mean OR for all four studies was 1.20 ($p = .010$). Seven of nine age-cohort designs yielded significant effects, with an overall weighted mean OR of 1.51 ($p < .0001$). Over all 41 studies, the weighted mean OR was 1.36 ($p < .0001$), indicating a substantial effect of these programs on bullying. To give a concrete example, if there were 20 bullies and 80 non-bullies in the intervention condition and 26 bullies and 74 non-bullies in the control condition, the OR would be 1.41. If there were 25 bullies and 75 non-bullies in the control condition, $OR = 1.33$. Hence, $OR = 1.36$ can correspond to 25–30% more bullies in the control condition (or conversely 20–23% fewer bullies in the intervention condition).

Table 3 also shows the analysis of effect sizes for victimization (being bullied). Only three of the randomized experiments found significant effects of the program on victimization and the weighted mean OR of 1.17 was just significant ($p = .050$). Five of the 17 studies with before-and-after/intervention-control designs yielded significant results, and the weighted mean OR of 1.22 was significant ($p = .007$). Three of the four studies with other intervention-control designs found significant results, yielding a significant weighted mean OR of 1.43 ($p < .0001$). Seven of the nine age-

⁵ See Campbell review, table 7, for a detailed presentation of the key results of each evaluation (i.e., outcome measures before and after the implementation of each program), and for the explanation of how all effect sizes were calculated in the Technical Appendix.

Table 3 Effect sizes for bullying and victimization

Project	Bullying			Victimization		
	OR	CI	<i>p</i> value	OR	CI	<i>p</i> value
Randomized experiments						
Baldry and Farrington (2004)	1.14	0.51–2.58	ns	1.69	0.76–3.78	ns
Beran and Shapiro (2005) ^a	1.14	0.53–2.46	ns	—	—	—
Boulton and Flemington (1996) ^a	0.93	0.38–2.27	ns	—	—	—
Cross et al. (2004)	0.77	0.51–1.15	ns	1.07	0.79–1.43	ns
DeRosier (2004)	0.87	0.63–1.21	ns	1.04	0.75–1.45	ns
Fekkes et al. (2006)	1.12	0.74–1.69	ns	1.25	0.95–1.65	ns
Fonagy et al. (2009)	1.66	1.10–2.50	.016	1.39	1.02–1.91	.038
Frey et al. (2005)	1.04	0.81–1.34	ns	1.09	0.76–1.56	ns
Hunt (2007)	1.46	0.93–2.28	.097	1.26	0.67–2.36	ns
Jenson and Dieterich (2007)	1.17	0.57–2.41	ns	1.63	0.78–3.41	ns
Karna et al. (forthcoming)	1.38	0.92–2.06	ns	1.55	1.28–1.88	.0001
Meyer and Lesch (2000) ^a	0.68	0.16–2.90	ns	—	—	—
Rosenbluth et al. (2004)	0.99	0.63–1.58	ns	0.70	0.50–0.97	.032
Sprober et al. (2006) ^a	0.95	0.63–1.45	ns	1.15	0.64–2.09	ns
Weighted mean	1.10	0.97–1.26	ns	1.17	1.00–1.37	.050
Before-and-after, Intervention-Control						
Alsaker and Valkanover (2001)	1.15	0.55–2.40	ns	3.14	1.52–6.49	.002
Andreou et al. (2007)	1.75	1.20–2.57	.004	1.48	1.01–2.16	.047
Bauer et al. (2007)	—	—	—	1.01	0.85–1.18	ns
Beran et al. (2004) ^a	—	—	—	1.04	0.28–3.88	ns
Ciucci and Smorti (1998)	1.20	0.58–2.47	ns	1.21	0.70–2.12	ns
Evers et al. (2007)	1.65	1.15–2.36	.007	1.79	1.23–2.60	.002
Fox and Boulton (2003) ^a	—	—	—	0.71	0.14–3.71	ns
Gini et al. (2003) ^a	0.76	0.15–3.84	ns	0.40	0.12–1.40	ns
Gollwitzer et al. (2006) ^a	1.23	0.63–2.42	ns	1.38	0.70–2.72	ns
Martin et al. (2005) ^a	2.56	0.33–19.63	ns	1.97	0.23–16.78	ns
Melton et al. (1998)	1.52	1.24–1.85	.0001	1.06	0.91–1.23	ns
Menard et al. (2008)	1.74	1.45–2.09	.0001	1.26	1.05–1.51	.013
Menesini et al. (2003)	1.60	0.81–3.16	ns	1.42	0.84–2.39	ns
Olweus/Bergen 2	1.79	0.98–3.26	.057	1.43	1.04–1.95	.026
Pepler et al. (2004)	1.69	1.22–2.35	.002	0.94	0.71–1.24	ns
Rahay and Craig (2002)	1.19	0.70–1.99	ns	0.79	0.47–1.33	ns
Rican et al. (1996)	2.52	0.60–10.52	ns	2.46	0.62–9.73	ns
Weighted mean	1.60	1.45–1.77	.0001	1.22	1.06–1.40	.007
Other experimental-control						
Galloway and Roland (2004)	1.20	0.91–1.59	ns	1.59	1.20–2.11	.001
Kaiser-Ulrey (2003) ^a	0.76	0.33–1.76	ns	0.65	0.28–1.50	ns
Ortega et al. (2004)	1.63	0.84–3.14	ns	2.12	1.15–3.91	.016
Raskauskas (2007)	1.20	1.01–1.42	.035	1.35	1.14–1.60	.0004
Weighted mean	1.20	1.04–1.38	.010	1.43	1.11–1.85	.006

Table 3 (continued)

Project	Bullying			Victimization		
	OR	CI	<i>p</i> value	OR	CI	<i>p</i> value
Age-Cohort Designs						
Ertesvag and Vaaland (2007)	1.34	1.13–1.58	.0008	1.18	0.99–1.39	.060
Olweus/Bergen 1	1.69	1.25–2.28	.0006	2.89	2.14–3.90	.0001
Olweus/Oslo1	2.14	1.18–3.87	.012	1.81	1.23–2.66	.002
Olweus/New National	1.78	1.54–2.06	.0001	1.59	1.45–1.73	.0001
Olweus/Oslo2	1.75	1.35–2.26	.0001	1.48	1.25–1.77	.0001
O'Moore and Minton (2004)	2.12	0.81–5.55	ns	1.99	0.98–4.07	.059
Pagliocca et al. (2007)	1.30	0.93–1.83	ns	0.92	0.71–1.21	ns
Salmivalli et al. (2005)	1.31	1.07–1.61	.010	1.30	1.06–1.60	.014
Whitney et al. (1994)	1.33	1.12–1.60	.002	1.14	1.00–1.29	.044
Weighted mean	1.51	1.35–1.70	.0001	1.44	1.21–1.72	.0001
Total weighted mean	1.36	1.26–1.47	.0001	1.29	1.18–1.42	.0001

OR odds ratio; CI confidence interval; ^a Initial $n < 200$

cohort designs yielded significant results, and another one (O'Moore and Minton 2004) was nearly significant. The weighted mean OR of 1.44 was significant ($p < .0001$).

Over all 41 studies, the weighted mean OR was 1.29 ($p < .0001$), indicating significant effects of these programs on victimization. To give a concrete example, if there were 20 victims and 80 non-victims in the intervention condition, and 25 victims and 75 non-victims in the control condition, then OR = 1.33. If there were 24 victims and 76 non-victims in the control condition, then OR = 1.26. Hence, this value of the OR can correspond to 20–25% more victims in the control condition (or conversely, 17–20% fewer victims in the intervention condition). Figure 2 shows the accompanying forest graph for victimization effect sizes. In this figure, the measure of effect size is the logarithm of OR (LOR). It is obvious from Fig. 2 that the majority of evaluations found that the program was followed by a reduction in victimization.

An overall reduction in bullying of 17–23% seems substantial and practically important in our opinion. However, it should perhaps be pointed out that the overall OR values of 1.36 and 1.29 correspond to standardized mean difference (*d*) values of .17 and .14, respectively. These *d* values are conventionally reviewed as 'small' (Lipsey and Wilson 2001: 147). Nevertheless, they correspond to a substantial amount of bullying prevented.

Effect size versus research design

Table 3 shows that the weighted mean odds ratio effect size measure varied across the four types of research design. In order to test whether this variation is statistically significant, it is necessary to calculate the heterogeneity between groups or QB (Lipsey and Wilson 2001: 135–138). For bullying, QB = 31.88 (3 df, $p < .0001$). For

victimization, $QB = 19.85$ (3 df, $p = .0002$). Therefore, we can conclude that effect sizes varied significantly across research designs.

Coding of study features

Key features of the evaluation

We have already discussed one feature of the evaluation, namely the research design. In order to investigate the relationship between evaluation features and

Fig. 1 Forest graph for bullying (*LOR* logarithm of the odds ratio)

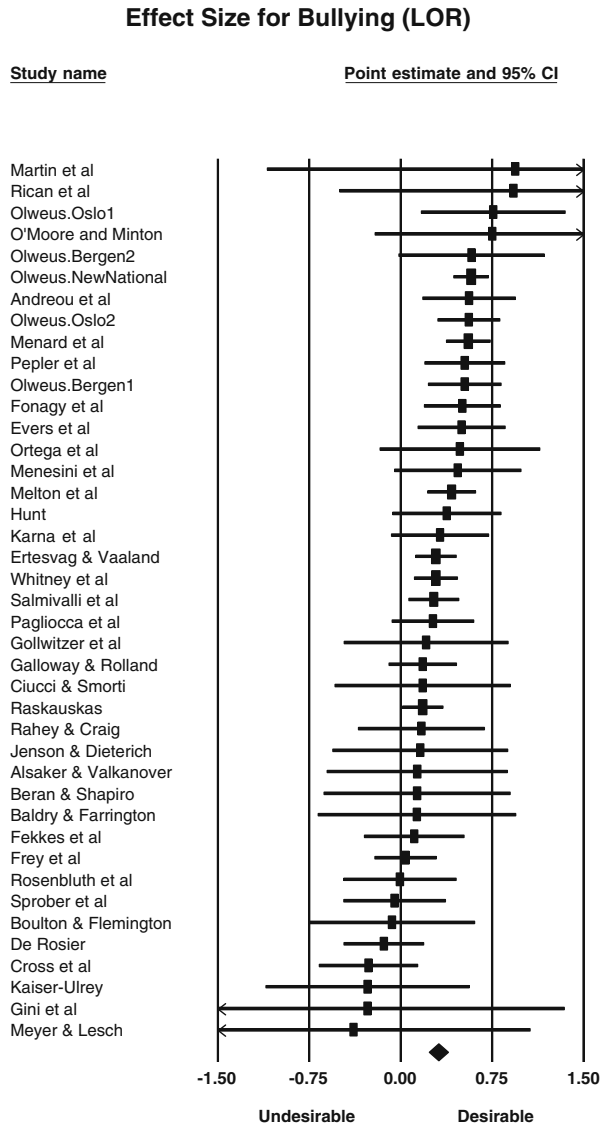
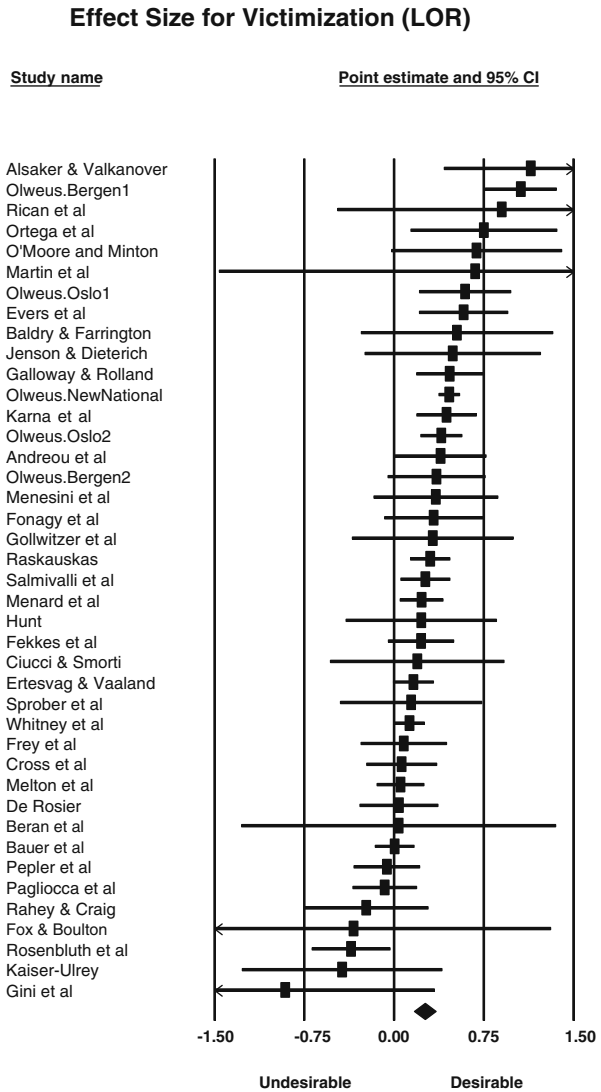


Fig. 2 Forest graph for victimization (*LOR* logarithm of the odds ratio)



effect size in a comparable way, all features were dichotomized (in order to produce roughly equal groups, as much as possible). For example, research design was dichotomized into (1) randomized experiments plus before-after/intervention-control designs (31 studies) versus (2) other intervention-control designs plus age-cohort designs (13 studies). Other features of the evaluation that were investigated were as follows:

- (a) Sample size (intervention plus control conditions), dichotomized into 900 children or more (22) versus 899 children or less (22). Several researchers (e.g., Farrington and Welsh 2003; Weisburd 1993) have found a negative relationship between effect size and sample size.

- (b) Publication date, dichotomized into 2004 or later (27) versus 2003 or earlier (17).
- (c) Average age of the children, dichotomized into 10 or less (19) versus 11 or more (25). The calculation of average age is problematic. For example, students in grade 4 (age 10–11) could range from 10.000 to 11.999, and we therefore estimated their average age as 11. Researchers who calculated average ages based on integer values of age (rather than exact values to several decimal places) might have reported an average age of 10.5 in this case.
- (d) Location in the USA or Canada (15) versus other places (29).
- (e) Location in other places (37) versus Norway (7).
- (f) Location in other places (19) versus Europe (25).
- (g) Outcome measure, classified into a dichotomous measure of two or more times per month (10) versus other measures (34). This dichotomous measure was associated with larger effect sizes than mean scores or simple prevalences.

In the Campbell review (Fig. 2), key features of the evaluation are specified for each study.

Key elements of the program

Each anti-bullying program included a variety of intervention elements. We coded 20 elements of the intervention in different programs. We consulted the evaluators of the various programs, and sent them our coding of the elements of the intervention for their program. We received feedback on 40 out of 44 evaluations and relevant changes were made to the coding where appropriate. A detailed description of all program elements can be found in the Campbell review.

We also coded other features of the intervention programs:

- (a) The number of elements included out of 20, dichotomized into ten or less (25 programs) versus 11 or more (19 programs).
- (b) The extent to which the program was not (27) or was (17) inspired by the work of Dan Olweus.
- (c) The duration of the program for children, dichotomized into 240 days or less (23) versus 270 days or more (20).
- (d) The intensity of the program for children, dichotomized into 19 h or less (21) versus 20 h or more (14).
- (e) The duration of the program for teachers, dichotomized into 3 days or less (21) versus 4 days or more (20). Where programs did not include teacher training, teacher duration was coded as zero.
- (f) The intensity of the program for teachers, dichotomized into 9 h or less (18) versus 10 h or more (21). Where programs did not include teacher training, teacher intensity was coded as zero.

In the Campbell review (Fig. 3), intervention components for each study are specified. In cases (c) to (f) above, they were not known in all cases.

Effect size versus study features

There have been few other attempts to relate effect size to program elements (see e.g., Kaminski et al. 2008). Table 4 shows the program elements and design features that were significantly related to effect sizes for bullying. Because of small numbers in one category, four of the 20 program elements could not be investigated (information for teachers, restorative justice approaches, school tribunals/bully courts, and virtual reality computer games). As explained before, the significance test is based on the heterogeneity between groups QB. The weighted mean odds ratio effect sizes are given for the different categories.

The most important program elements that were associated with a decrease in bullying were parent training/meetings, improved playground supervision, disciplinary methods, classroom management, teacher training, classroom rules, a whole-school anti-bullying policy, school conferences, information for parents, and cooperative group work. In addition, the total number of elements and the duration and intensity of the program for teachers and children were significantly associated with a decrease in bullying. Also, programs inspired by the work of Dan Olweus

Table 4 Significant relationships with bullying

	Cat (n) OR	Cat (n) OR	<i>p</i> value
Program elements			
Parent training/meetings	No (24) 1.25	Yes (17) 1.57	.0001
Playground supervision	No (30) 1.29	Yes (11) 1.53	.0001
Intensity for children	19- (19) 1.25	20+ (13) 1.62	.0001
Intensity for teachers	9- (16) 1.19	10+ (20) 1.52	.0001
Duration for children	240- (20) 1.17	270+ (20) 1.49	.0001
Disciplinary methods	No (28) 1.31	Yes (13) 1.59	.0003
Duration for teachers	3- (19) 1.22	4+ (19) 1.50	.0004
Classroom management	No (13) 1.15	Yes (28) 1.44	.005
Teacher training	No (13) 1.24	Yes (28) 1.46	.006
Classroom rules	No (11) 1.15	Yes (30) 1.44	.006
Whole-school policy	No (17) 1.19	Yes (24) 1.44	.008
School conferences	No (21) 1.30	Yes (20) 1.49	.008
Total elements	10- (23) 1.30	11+ (18) 1.48	.009
Based on Olweus	No (25) 1.31	Yes (16) 1.50	.011
Information for parents	No (13) 1.21	Yes (28) 1.44	.013
Cooperative group work	No (19) 1.31	Yes (22) 1.48	.019
Design Features			
Age of children	10- (18) 1.22	11+ (23) 1.50	.0001
Outcome measure	Other (31) 1.32	2+ M (10) 1.64	.0002
Publication date	04+ (25) 1.31	03- (16) 1.56	.0009
In Norway	Rest (34) 1.33	Nor (7) 1.58	.001

Cat category of variable; *OR* weighted mean odds ratio; duration in days; intensity in hours; outcome measure 2 + M: two times per month or more (versus other measures)

worked best. Regarding the design features, the programs worked better with older children and in Norway specifically. Older programs, and those in which the outcome measure of bullying was two times per month or more, also yielded better results.

Table 5 shows the program elements and design features that were significantly related to effect sizes for victimization (being bullied). Weighted effect sizes for bullying and victimization were significantly correlated ($r=.51, p<.0001$). The most important program elements that were associated with a decrease in victimization were disciplinary methods, parent training/meetings, videos and cooperative group work. In addition, the duration and intensity of the program for children and teachers were significantly associated with a decrease in victimization. Work with peers was associated with a significant increase in victimization. Work with peers was also associated with an increase in bullying, but not significantly so (OR = 1.42 for no work with peers, OR = 1.35 for work with peers). Regarding the design features, the programs worked better with older children, in Norway specifically and in Europe more generally, and they were less effective in the USA and Canada. Older programs, those in which the outcome measure of victimization was two times per month or more, and those with other intervention-control and age-cohort designs, also worked better.

Our finding that anti-bullying programs work better with older children (age 11 or older) conflicts with the arguments of Peter Smith (2010). Therefore, we examined

Table 5 Significant relationships with victimization

	Cat (n) OR	Cat (n) OR	<i>p</i> value
Program elements			
Work with peers	No (25) 1.39	Yes (16) 1.13	.0001
Disciplinary methods	No (28) 1.21	Yes (13) 1.44	.0001
Parent training/meetings	No (24) 1.20	Yes (17) 1.41	.0001
Duration for teachers	3- (18) 1.18	4+ (20) 1.41	.0003
Videos	No (22) 1.17	Yes (19) 1.38	.0004
Cooperative group work	No (18) 1.20	Yes (23) 1.38	.001
Duration for children	240- (20) 1.15	270+ (20) 1.35	.001
Intensity for children	19- (18) 1.21	20+ (14) 1.42	.002
Intensity for teachers	9- (15) 1.22	10+ (21) 1.37	.028
Design features			
Outcome measure	Other (31) 1.18	2+ M (10) 1.57	.0001
In Norway	Rest (34) 1.18	Nor (7) 1.55	.0001
Not in US or Canada	US/Can (14) 1.06	Rest (27) 1.42	.0001
In Europe	Rest (17) 1.11	EU (24) 1.44	.0001
Design	12 (28) 1.16	34 (13) 1.41	.0001
Publication date	04+ (26) 1.21	03- (15) 1.42	.0001
Age of children	10- (18) 1.22	11+ (23) 1.34	.047

Cat category of variable; *OR* weighted mean odds ratio; duration in days; intensity in hours; outcome measure 2 + M: two times per month or more (versus other measures)

this finding in more detail, by dividing the average age into four categories: 6–9 (12 programs), 10 (seven programs), 11–12 (14 programs), and 13–14 (11 programs). The weighted mean OR for bullying steadily increased with age: 1.21 (6–9), 1.23 (10), 1.44 (11–12) and 1.53 (13–14); $QB = 15.65$, 3 df, $p = .001$. Similarly, the weighted mean OR for victimization steadily increased with age: 1.17 (6–9), 1.25 (10), 1.26 (11–12) and 1.37 (13–14); $QB = 7.24$, 3 df, $p = .065$. These results confirm our conclusion that the effectiveness of programs increases with the age of the children.

Variables that might help to explain differential treatment effects in meta-analysis (e.g., elements of the intervention) cannot be assumed to be statistically independent. Researchers should try to disentangle the relationships among them and identify those that truly have significant independent relationships with effect sizes (Lipsey 2003: 78). Multivariate techniques can be used to solve this problem in meta-analysis (Hedges 1982). Weighted regression analyses (Lipsey and Wilson 2001: 138–140) were carried out to investigate which elements of the programs were independently related to bullying and victimization effect sizes (LORs).

These analyses were severely limited by the small number of studies. Nevertheless, they showed that the most important elements of a program that were related to a decrease in bullying were parent training/meetings and disciplinary methods (Table 6). When all the intensity and duration factors from Table 4 were added, the most important program elements were intensity for children and parent training/meetings.

The most important elements of the program that were associated with a decrease in victimization were videos and disciplinary methods. Work with peers was associated with an increase in victimization (Table 6). When all the intensity and duration factors from Table 5 were added, the most important elements were work with peers (negatively related), the duration of the program for children, and videos.

Discussion

Summary of main findings

The present systematic review shows that school-based anti-bullying programs are often effective, and that particular program elements were associated with a decrease in bullying and victimization. More intensive programs were more effective, as were programs including parent meetings, firm disciplinary methods, and improved playground supervision. One program element (work with peers) was significantly associated with an increase in victimization. Work with peers referred to the formal engagement of peers in tackling bullying. This could include peer mediation, peer mentoring, and encouraging bystander intervention to prevent bullying.

We conclude that, on average, bullying decreased by 20–23% and victimization by 17–20%. The effects were generally highest in the age-cohort designs and lowest in the randomized experiments. Weisburd et al. (2001) also found that the strongest designs, using randomized experiments, generally yielded the weakest treatment

Table 6 Results of weighted regression analyses

	B	SE (B)	Z	p value
Bullying effect size				
(a) 20 Elements only				
Parent training/meetings	.1808	.0557	3.25	.001
Disciplinary methods	.1178	.0582	2.02	.043
(b) All elements				
Intensity for children	.1726	.0675	2.56	.010
Parent training/meetings	.1594	.0635	2.51	.012
Victimization effect size				
(a) 20 Elements only				
Work with peers	-.2017	.0478	4.22	.0001
Videos	.1285	.0505	2.55	.011
Disciplinary methods	.1102	.0469	2.35	.019
(b) All Elements				
Work with peers	-.2362	.0480	4.93	.0001
Duration for children	.1498	.0536	2.79	.005
Videos	.1338	.0491	2.73	.006

effects, and suggested that nonrandomized designs in criminology might have a positive bias in favor of finding treatment effects. It was not clear, however, that the randomized experiments in this review were methodologically superior to other designs, because of very small numbers of schools randomized in some cases, and because of other methodological problems such as differential attrition. For example, there was differential attrition in the evaluation of KiVa, with many more students lost in the control condition (27%) than in the experimental condition (13%). This differential attrition created higher effect sizes than when (as in the present article) the analysis was based only on students known both before and after (OR for bullying = 1.47 in the Swedish Report, 1.38 here; OR for victimization = 1.66 in the Swedish Report, 1.55 here).

Policy implications

In developing new policies and practices to reduce bullying, policy-makers and practitioners should draw upon high-quality evidence-based programs that have been proved to be effective. New anti-bullying initiatives should be inspired by existing successful programs but should be modified in light of the key program elements that we have found to be most effective (or ineffective). For example, it seems from our results that work with peers should not be used, in agreement with other research showing that programs targeting delinquent peers tend to cause an increase in offending (e.g., Dishion et al. 1999; Dodge et al. 2006). It should be borne in mind, however, that we have discovered the program elements that are most highly correlated with effectiveness. This does not prove that they cause effectiveness, but this is the best evidence we have at present.

We recommend that a system of accrediting effective anti-bullying programs should be developed. In England and Wales in 1996, a system of accrediting effective programs in prison and probation was established (McGuire 2001). For a program to be accredited, it had to meet explicit criteria based on knowledge about what worked to reduce offending. Only accredited programs can be used in England and Wales, and similar systems have been developed in other countries such as Scotland and Canada. A similar system should be developed for accrediting anti-bullying programs in schools to ensure that programs contain elements that have been proved to be effective in high-quality evaluations. This accreditation system could perhaps be organized by an international body such as the International Observatory on Violence in Schools.

New anti-bullying programs should be disseminated using high quality standards of implementation in a way that ensures that the program is more likely to have an impact. The quality of a program is undoubtedly important, but so is the way in which it is implemented. Implementation procedures should be transparent in order to enable researchers to know whether effects are related to key features of the intervention or key feature of the evaluation. It is sad, for instance, that only two of the 44 evaluations included in our meta-analytic review (Fekkes et al. 2006; Smith et al. 2004) provided key information about the percentage of intervention and control schools that implemented each program component.

Our results show that the intensity and duration of a program is directly linked to its effectiveness, and other researchers (Olweus 2005a; Smith 1997: 198) also found a 'dose-response' relationship between the number of components of a program that were implemented in a school and the effect on bullying. Our findings show that programs need to be intensive and long-lasting to have an impact on this troubling problem. It could be that a considerable time period is needed in order to build up an appropriate school ethos that efficiently tackles bullying.

New anti-bullying initiatives should also pay attention to enhancing playground supervision. For bullying, playground supervision was one of the elements that were most strongly related to program effectiveness. It is plausible that this is effective since a lot of bullying occurs during recess time. Improving the school playground environment (e.g., through reorganization and/or identification of 'hot spots') may also be a promising and low-cost intervention component.

Disciplinary methods (i.e., firm methods for tackling bullying) was an intervention component that was significantly related to reductions in both bullying and victimization. To some extent, this finding was attributable to the big effects of the Olweus program, which included a range of firm sanctions, including serious talks with bullies, sending them to the principal, making them stay close to the teacher during recess time, and depriving them of privileges.

Contrary to the arguments of Peter Smith (2010) the results of our review show that programs have a bigger impact on bullying for older children. This is an age range when bullying is decreasing anyway. Peter Smith argued that programs were less effective in secondary schools because negative peer influence was more important and because secondary schools were larger and students did not spend most of their time with one teacher who could be very influential. We speculate that programs may be more effective in reducing bullying by older children because of their superior cognitive abilities, decreasing impulsiveness, and increasing likelihood

of making rational decisions. Many programs are based on social learning ideas of encouraging and rewarding prosocial behavior and discouraging and punishing bullying. These programs are likely to work better, for example, in building empathy and perspective-taking skills with older students.

Establishing a whole-school anti-bullying policy was significantly related to effect sizes for bullying but not for victimization (being bullied). There was no evidence that individual work with bullies or victims was effective. We recommend that more efforts should be made to implement effective programs with individual bullies and victims, perhaps based on child skills training (Losel and Beelman 2003). Most current programs, with some exceptions⁶, are not based on this.

New anti-bullying initiatives should go beyond the scope of the school and target wider systemic factors such as the family. Studies indicate that bullied children often do not communicate their problem to anyone while parents and teachers often do not talk to bullies about their conduct (e.g., Fekkes et al. 2005). In our systematic review, parent training/meetings was significantly related to a decrease in both bullying and victimization. These findings suggest that efforts should be made to sensitize parents about the issue of school bullying through educational presentations and teacher-parent meetings. Future anti-bullying initiatives should also bring together experts from various disciplines and make the most of their expertise. In our review, cooperative group work among experts was significantly related to the reduction of both bullying and victimization.

Future evaluations of anti-bullying programs should be designed in light of our results. Attention should be paid not only to the quality of the program but also to the way it is implemented. The present review has shown that different features of the evaluation were significantly related to a decrease in bullying and victimization. In particular, the way bullying was measured and the age of the children were important. Programs should be targeted on children aged 11 or older rather than on younger children. The outcome measure of bullying or victimization should be two times per month or more. It would be regrettable if some evaluations of anti-bullying programs did not establish the effectiveness of the program only because of the way the outcome variable was measured. Programs implemented in Norway seem to work best and this could be related to the long tradition of bullying interventions and research in Scandinavian countries. Other factors are that Scandinavian schools are of high quality, with small classes and well-trained teachers, and there is a Scandinavian tradition of state intervention in matters of social welfare (J.D. Smith et al. 2004: 557).

Importantly, cost-benefit analyses of anti-bullying programs should be carried out, to investigate how much money is saved for the money expended (Welsh et al. 2001). Saving money is a powerful argument to convince policy-makers and practitioners to implement intervention programs (Farrington 2009: 59). There never has been a cost-benefit analysis of an anti-bullying program. For example, the benefits of reducing bullying might include less delinquency, less anxiety and depression, less truancy, less medical or psychological treatment, and more successful lives generally. All of these benefits could be monetized and compared with the financial costs of anti-bullying programs.

⁶ For example, see DeRosier 2004; Fox and Boulton 2003; Gollwitzer et al. 2006 from the Campbell review.

Conclusions

Some previous reviews (Ferguson et al. 2007; Merrell et al. 2008) concluded that anti-bullying programs had little effect on school bullying. We attribute their results to the relatively limited searches done and also to their inclusion criteria (e.g., not clearly focusing on bullying; including uncontrolled evaluations); for more details, see the 'Background' section above. The present, more extensive, systematic review indicates that school-based anti-bullying programs are effective. There are many implications of our review for future research.

Future evaluations should have before-and-after measures of bullying and victimization in experimental and control schools. Bullying and victimization should be carefully defined and measured. Depending on the nature of the anti-bullying intervention, schools, classes, or students should be randomly assigned to conditions. For example, if the intervention involves interpersonal skills training, students or classes could be assigned. Since it is difficult to randomly assign a large number of schools, it may be best to place schools in matched pairs and randomly assign one member of each pair to the experimental condition and one member to the control condition. It seems unsatisfactory to randomly assign school classes because of the danger of contamination of control children by experimental children. Only students who are tested both before and after the intervention should be analyzed, in order to minimize problems of differential attrition. In order to investigate the effects of different program elements, students could be randomly assigned to receive or not receive them. Research is needed on the best methods of measuring bullying, on what time periods to enquire about, and on seasonal variations.

It is important to develop methodological quality standards for evaluation research that can be used by systematic reviewers, scholars, policy makers, the mass media, and the general public in assessing the validity of conclusions about the effectiveness of interventions in reducing crime (Farrington 2003: 66). Such quality standards could include guidelines to program evaluators with regard to what elements of the intervention should be included in published reports, perhaps under the aegis of the Campbell Collaboration Crime and Justice Group (Farrington and Petrosino 2001; Farrington and Weisburd 2007). If these guidelines had been in existence, they would have been very helpful in the ambitious task we have undertaken to fully code the elements of the intervention in all studies.

With a positive response from researchers regarding our coding for 40 out of 44 evaluations of anti-bullying programs, we have been quite successful. Future reports should provide key information about features of evaluations, according to a checklist that should be developed, inspired perhaps by the CONSORT Statement for medical research (see Altman et al. 2001; Moher et al. 2001; Perry and Johnson 2008). Information about key elements of programs, and about the implementation of programs, should be provided. Where bullying and victimization are measured on five-point scales, the full 5×2 table should be presented, so that the Area Under the ROC Curve (AUC) could be used as a measure of effectiveness (Farrington et al. 2008). This would avoid the problem of results varying according to the particular cut-off points that are chosen. Providing more complete tables would also make it possible to investigate whether the cut-off point of two or more times per month was really associated with larger effect sizes, and if so why.

In conclusion, results obtained so far in evaluations of anti-bullying programs are encouraging. The time is ripe to mount a new program of research on the effectiveness of anti-bullying programs, based on our findings.

Appendix

89 reports of 53 different evaluations*

Randomized experiments

- (1) ViSC Training Program [Atria and Spiel 2007]; category 5 => excluded due to many missing values
- (2) Bulli and Pupe [Baldry 2001; Baldry and Farrington 2004]; category 6
- (3) Project Ploughshares Puppets for Peace [Beran and Shapiro 2005]; category 5
- (4) Short Video Intervention [Boulton and Flemington 1996]; category 5
- (5) Friendly Schools [Cross et al. 2004; Pintabona 2006]; category 6
- (6) S.S.GRIN [DeRosier 2004; DeRosier and Marcus 2005]; category 6
- (7) Dutch Anti-bullying Program [Fekkes et al. 2006]; category 6
- (8) SPC and CAPSLE Program [Fonagy et al. 2009]; category 6
- (9) Steps to Respect [Frey, Edstrom and Hirschstein 2005; Frey et al. 2005; Hirschstein et al. 2007]; category 6
- (10) Anti-bullying Intervention in Australian Secondary Schools [Hunt 2007]; category 6
- (11) Youth Matters [Jenson and Dieterich 2007; Jenson et al. 2005a, 2005b, 2006a, 2006b]; category 6
- (12) Kiva [Karna et al. forthcoming; Salmivalli et al. 2009]; category 6
- (13) Korean Anti-bullying Program [Kim 2006]; category 5 => excluded; data produced implausible effect size
- (14) Behavioral Program for Bullying Boys [Meyer and Lesch 2000]; category 5
- (15) Expect Respect [Rosenbluth et al. 2004; Whitaker et al. 2004]; category 6
- (16) Pro-ACT + E [Sprober 2006; Sprober et al. 2006]; category 5
- (17) The Peaceful Schools Experiment [Twemlow et al. 2005]; category 6 => excluded; part of a larger evaluation by Fonagy et al. 2009

Before-and-after, intervention-control comparisons

- (1) Be-Prox [Alsaker and Valkanover 2001; Alsaker 2004]; category 5
- (2) Greek Anti-bullying Program [Andreou et al. 2007]; category 6
- (3) Seattle Trial of the Olweus Program [Bauer et al. 2007]; category 6
- (4) Dare to Care: Bully Proofing your School Program [Beran et al. 2004]; category 5
- (5) Progetto Pontassieve [Ciucci and Smorti 1998]; category 6
- (6) Cooperative Group Work Intervention [Cowie et al. 1994]; category 5 => excluded due to lack of data

- (7) Transtheoretical-based Tailored Anti-bullying Program [Evers et al. 2007]; category 6
- (8) Social Skills Training (SST) Program [Fox and Boulton 2003]; category 5
- (9) Stare bene a scuola: Progetto di prevenzione del bullismo [Gini et al. 2003]; category 5
- (10) Viennese Social Competence (ViSC) Training [Gollwitzer et al. 2006]; category 5
- (11) Conflict Resolution Program [Heydenberk et al. 2006]; category 6 => excluded due to lack of data
- (12) Granada Anti-bullying Program [Martin et al. 2005]; category 5
- (13) South Carolina Program; implementation of OBPP [Melton et al. 1998; Limber et al. 2004]; category 6
- (14) 'Bullyproofing your School' Program [Menard et al. 2008]; category 6
- (15) Befriending Intervention Program [Menesini and Benelli 1999; Menesini et al. 2003]; category 5
- (16) New Bergen Project against Bullying; 'Bergen 2' [1997–1998]; category 6
- (17) Toronto Anti-bullying Program [Pepler et al. 2004]; category 6
- (18) Ecological Anti-bullying Program [Rahey and Craig 2002]; category 6
- (19) Short Intensive Intervention in the Czech Republic (Rican et al. 1996); category 6
- (20) Flemish Anti-bullying Program [Stevens, De Bourdeaudhuij and Van Oost 2000; Stevens, Van Oost and De Bourdeaudhuij 2000; Stevens et al. 2001, 2004]; category 6 => excluded due to nature of data
- (21) Anti-bullying Intervention in the Netherlands [Wiefferink et al. 2006]; category 6 => excluded due to lack of data

Other intervention-control comparisons

- (1) Norwegian Anti-bullying Program [Galloway and Roland 2004]; category 6
- (2) BEST [Kaiser-Ulrey 2003]; category 5
- (3) SAVE [Ortega and Del Rey 1999; Ortega et al. 2004]; category 6
- (4) Kia Kaha [Raskauskas 2007]; category 6

Age-cohort designs

- (1) Respect [Ertesvag and Vaaland 2007]; category 6
- (2) Anti-bullying Intervention in Schleswig-Holstein, Germany [Hanewinkel 2004]; category 6 => excluded due to lack of data
- (3) Anti-bullying Intervention in Kempele Schools [Koivisto 2004]; category 6 => excluded due to lack of data
Olweus Bullying Prevention Program [OBPP]; category 6:
- (4) First Bergen Project against Bullying; 'Bergen 1' [1983–1985]; category 6
- (5) First Oslo Project against Bullying; 'Oslo 1' [November 1999–November 2000]; category 6
- (6) New National Initiative Against Bullying in Norway; 'New National' [2001–2007]; category 6

- (7) Five-year Follow-up in Oslo; ‘Oslo 2’ [2001–2006]; category 6 [Olweus 1991, 1992, 1993a, 1994a, 1994b, 1994c, 1995, 1996a, 1996b, 1996c, 1997a, 1997b, 1997c, Olweus 2004a, 2004b, 2005a, 2005b, Olweus and Alsaker 1991]
- (8) Donegal Anti-Bullying Program [O’Moore and Minton 2004; O’Moore 2005]; category 6
- (9) Chula Vista OBPP [Pagliocca et al. 2007]; category 6
- (10) Finnish Anti-bullying Program [Salmivalli et al. 2004; 2005]; category 6
- (11) Sheffield Anti-bullying Program [Whitney et al. 1994; Smith 1997; Smith et al. 2004b]; category 6

* Nine evaluations [presented in 12 reports] were excluded from the meta-analysis

References included in the systematic review⁷

- * Alsaker F.D. (2004) Bernese program against victimization in kindergarten and elementary schools. In P.K. Smith, D. Pepler, K. Rigby (Eds.), *Bullying in schools: How successful can interventions be?* (pp. 289–306). Cambridge: Cambridge University Press.
- * Alsaker F.D., & Valkanover S. (2001) *Early diagnosis and prevention of victimization in kindergarten. In J. Juvonen, S. Graham (Eds.) Peer harassment in school* (pp.175–195). New York: Guilford.
- * Andreou E, Didaskalou E, Vlachou A (2007) Evaluating the effectiveness of a curriculum-based anti-bullying intervention program in Greek primary schools. *Educational Psychology*, 27, 693–711.
- Atria, M., & Spiel, C. (2007). Viennese Social Competence (ViSC) training for students: program and evaluation. In J. E. Zins, M. J. Elias, & C. A. Maher (Eds.), *Bullying, victimization and peer harassment: a handbook of prevention and intervention* (pp. 179–197). New York: Haworth.
- * Baldry, A.C. (2001). *Bullying in schools: Correlates and intervention strategies*. PhD Thesis, Cambridge University. Index to Theses Database, 51–8145.
- * Baldry, A.C., & Farrington, D.P. (2004). Evaluation of an intervention program for the reduction of bullying and victimization in schools. *Aggressive Behavior*, 30, 1–15.
- * 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.
- * Beran, T., & Shapiro, B. (2005). Evaluation of an anti-bullying program: Student reports of knowledge and confidence to manage bullying. *Canadian Journal of Education*, 28 (4), 700–717.
- * Beran, T., Tutty, L., Steinrath, G. (2004). An evaluation of a bullying prevention program for elementary schools. *Canadian Journal of School Psychology*, 19 (1/2), 99–116.
- * Boulton, M.J., & Flemington, I. (1996). The effects of a short video intervention on secondary school pupils’ involvement in definitions of and attitudes towards bullying. *School Psychology International*, 17, 331–345.
- * Ciucci, E., & Smorti, A. (1998). Il fenomeno delle pretonenze nella scuola: problemi e prospettive di intervento [The phenomenon of bullying in school: problems and prospects for intervention]. *Psichiatria dell’infanzia e dell’adolescenza*, 65, 147–157.
- Cowie, H., Smith, P. K., Boulton, M., & Laver, R. (1994). *Cooperation in the multi-ethnic classroom: The impact of cooperative group work on social relationships in middle schools*. London: David Fulton.
- * Cross, D., Hall, M., Hamilton, G., Pintabona, Y., Erceg, E. (2004). Australia: The Friendly Schools project. In P.K. Smith, D. Pepler, & K. Rigby (Eds.), *Bullying in schools: How successful can interventions be?* (pp. 187–210). Cambridge: Cambridge University Press.
- * DeRosier, M.E. (2004). Building relationships and combating bullying: Effectiveness of a school-based social skills group intervention. *Journal of Clinical Child and Adolescent Psychology*, 33, 196–201.
- * DeRosier, M.E. & Marcus, S.R. (2005). Building friendships and combating bullying: Effectiveness of S.S.GRIN at one-year follow-up. *Journal of Clinical Child and Adolescent Psychology*, 34, 140–150.

⁷ References with an asterisk indicate reports on evaluations that were included in the meta-analysis.

- * Ertesvag, S. K. & Vaaland, G. S. (2007). Prevention and reduction of behavioural problems in school: An evaluation of the Respect program. *Educational Psychology*, 27, 713–736.
- * Evers, K.E., Prochaska, J.O., Van Marter, D.F., Johnson, J.L., & Prochaska, J.M. (2007). Transtheoretical-based bullying prevention effectiveness trials in middle schools and high schools. *Educational Research*, 49, 397–414.
- * Fekkes, M., Pijpers, F.I.M., & Verloove-Vanhorick, S.P. (2006). Effects of antibullying school program on bullying and health complaints. *Archives of Pediatrics and Adolescent Medicine*, 160, 638–644.
- * Fonagy, P., Twemlow, S.W., Vernberg, E.M., Nelson, J.M., Dill, E.J., Little, T.D., & Sargent, J.A. (2009). A cluster randomized controlled trial of child-focused psychiatric consultation and a school systems-focused intervention to reduce aggression. *Child Psychology and Psychiatry*, 50 (5), 607–616.
- * Fox, C. & Boulton, M. (2003). Evaluating the effectiveness of a social skills training (SST) programme for victims of bullying. *Educational Research*, 45, 231–247.
- * Frey, K.S., Edstrom, L.V.S., & Hirschstein, M. K. (2005). The Steps to Respect program uses a multi-level approach to reduce playground bullying and destructive playground behaviours. In D.L. White, M.K. Faber, & B.C. Glenn (Eds.), *Proceedings of Persistently Safe Schools 2005*, (pp. 47–55). Washington, DC: Hamilton Fish Institute, George Washington University.
- * Frey, K., Hirschstein, M.K., Snell, J. L., van Schoiack Edstrom, L., MacKenzie, E.P., & Broderick, C.J. (2005). Reducing playground bullying and supporting beliefs: An experimental trial of the Steps to Respect program. *Developmental Psychology*, 41, 479–491.
- * Galloway, D. & Roland, E. (2004). Is the direct approach to reducing bullying always the best? In P.K. Smith, D. Pepler, & K. Rigby (Eds.), *Bullying in schools: How successful can interventions be?* (pp. 37–53). Cambridge: Cambridge University Press.
- * Gini, G., Belli, B., & Casagrande, M. (2003). Le prepotenze a scuola: una esperienza di ricerca-intervento antibullismo [Bullying at school: an experience of research-intervention against bullying], *Eta Evolutiva*, 76, 33–45.
- * Gollwitzer, M., Eisenbach, K., Atria, M., Strohmeier, D., & Banse, R. (2006). Evaluation of aggression-reducing effects of the ‘Viennese Social Competence Training’. *Swiss Journal of Psychology*, 65, 125–135.
- Hanewinkel, R. (2004). Prevention of bullying in German schools: an evaluation of an anti-bullying approach. In P. K. Smith, D. Pepler, & K. Rigby (Eds.), *Bullying in schools: How successful can interventions be?* (pp. 81–97). Cambridge: Cambridge University Press.
- Heydenberk, R. A., Heydenberk, W. R., & Tzenova, V. (2006). Conflict resolution and bully prevention: skills for school success. *Conflict Resolution Quarterly*, 24, 55–69.
- * Hirschstein, M. K., Van Schoiack Edstrom, L. Frey, K. S. Snell, J. L. Mackenzie, E. P. (2007). Walking the talk in bullying prevention: Teacher implementation variables related to initial impact of Steps to Respect Program. *School Psychology Review*, 36, 3–21.
- * Hunt, C. (2007). The effect of an education program on attitudes and beliefs about bullying and bullying behaviour in junior secondary school students. *Child and Adolescent Mental Health*, 12 (1), 21–26.
- * Jenson, J.M. & Dieterich, W.A. (2007). Effects of a skills-based prevention program on bullying and bully victimization among elementary school children. *Prevention Science*, 8, 285–296.
- * Jenson, J.M., Dieterich, W.A., Rinner, J.R. (2005a). *Effects of a skills-based prevention program on bullying and bully victimization among elementary school children*. Paper given at the Annual Meeting of the American Society of Criminology. Toronto, Canada (November).
- * Jenson, J.M., Dieterich, W.A., & Rinner, J.R. (2005b). *The prevention of bullying and other aggressive behaviors in elementary school students: Effects of the Youth Matters curriculum*. Paper given at the Annual Meeting of the Society for Prevention Research, Washington, DC (May).
- * Jenson, J.M., Dieterich, W.A., Powell, A., & Stoker, S. (2006a). *Effects of a skills-based intervention on aggression and bully victimization among elementary school children*. Paper given at the Annual Meeting of the Society for Prevention Research, San Antonio, Texas (June).
- * Jenson, J.M., Dieterich, W.A., Powell, A., & Stoker, S. (2006b). *Effects of the Youth Matters prevention curriculum on bullying and other aggressive behaviors in elementary school students*. Paper given at the Annual Meeting of the Society for Social Work and Research, San Antonio, Texas (January).
- * Kaiser-Ulrey, C. (2003). *Bullying in middle schools: A study of B.E.S.T. —Bullying Eliminated from Schools Together— an anti-bullying programme for seventh grade students*. PhD Thesis. Tallahassee: Florida State University, College of Education.
- * Karna, A., Voeten, M., Little, T.D., Poskiparta, E., Kaljonen, A., & Salmivalli, C. (forthcoming). A large-scale evaluation of the KiVa anti-bullying program. *Child Development*, in press.

- Kim, J. U. (2006). The effect of a bullying prevention program on responsibility and victimization of bullied children in Korea. *International Journal of Reality Therapy*, 26(1), 4–8.
- Koivisto, M. (2004). A follow-up survey of anti-bullying interventions in the comprehensive schools of Kempele in 1990–1998. In P. K. Smith, D. Pepler, & K. Rigby (Eds.), *Bullying in schools: how successful can interventions be?* (pp. 235–251). Cambridge: Cambridge University Press.
- * Limber, S.P., Naton, M., Tracy, A. J., Melton, G.B., & Flerx, V. (2004). Implementation of the Olweus Bullying Prevention Program in the Southeastern United States. In P. K. Smith, D. Pepler, & K. Rigby (Eds.), *Bullying in schools. How successful can interventions be?* (pp. 55–80). Cambridge: Cambridge University Press.
- * Martin, F.D.F., del. Martinez, M.C.P., & Tirado, J.L.A. (2005). Design, implementation and evaluation of a bullying prevention pilot program [Spanish: Diseno, aplicacion y evaluacion de un Programa Piloto para la Prevencion del Maltrato entre companeros]. *Revista Mexicana de Psicologia*, 22, 375–384.
- * Melton, G.B., Limber, S.P., Flerx, V., Nation, M., Osgood, W., Chambers, J., Henggeler, S., Cunningham, P., & Olweus, D. (1998). *Violence among rural youth*. Final report to the Office of Juvenile Justice and Delinquency Prevention, Washington DC.
- * Menard, S., Grotper, J., Gianola, D., & O’Neal, M. (2008). Evaluation of Bullyproofing your School: final report. Downloaded from NCJRS: <http://www.ncjrs.gov/pdffiles1/nij/grants/221078.pdf>.
- * Menesini, E., & Benelli, B. (1999). *Enhancing children's responsibility against bullying: Evaluation of a befriending intervention in Middle School Children*, Paper presented at the IXth European Conference on Developmental Psychology, Spetses, Greece, September, 1999.
- * Menesini, E., Codecasa, E., Benelli, B., & Cowie, H. (2003). Enhancing children’s responsibility to take action against bullying: Evaluation of a befriending intervention in Italian Middle schools. *Aggressive Behavior*, 29, 1–14.
- * Meyer, N. & Lesch, E. (2000). An analysis of the limitations of a behavioral programme for bullying boys from a sub-economic environment. *Southern African Journal of Child and Adolescent Mental Health*, 12 (1), 59–69.
- * Olweus, D. (1991). Bully/victim problems among school children: 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, NJ: Erlbaum.
- * Olweus, D. (1992). Bullying among school children: Intervention and prevention. In R.D. Peters, R.J. McMahon, & V.L. Quinsey (Eds.), *Aggression and violence throughout the lifespan* (pp. 100–125). London: Sage.
- * Olweus, D. (1993a). Bully/victim problems among schoolchildren: Long-term consequences and an effective intervention program. In S. Hodgins (Ed.), *Mental disorder and crime* (pp. 317–349). Thousand Oaks, CA: Sage.
- * Olweus, D. (1994a). Bullying at school: Basic facts and effects of a school-based intervention program. *Journal of Child Psychology and Psychiatry*, 35, 1171–1190.
- * Olweus, D. (1994b). Bullying at school: Basic facts and an effective intervention programme. *Promotion and Education*, 1, 27–31.
- * Olweus, D. (1994c). Bullying at school: long-term outcome 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. (1995). Peer abuse or bullying at school: Basic facts and a school-based intervention programme. *Prospects*, 25 (1), 133–139.
- * Olweus, D. (1996a). Bullying or peer abuse in school: Intervention and prevention. In G. Davies, S. Lloyd-Bostock, M. McMurran, & C. Wilson (Eds.), *Psychology, law and criminal justice: International developments in research and practice* (pp. 248–267). Berlin: Walter de Gruyter.
- * Olweus, D. (1996b). Bullying at school: Knowledge base and effective intervention. *Annals of the New York Academy of Sciences*, 784, 265–276.
- * Olweus, D. (1996c). Bully/victim problems at school: Facts and effective intervention. *Reclaiming Children and Youth: Journal of Emotional and Behavioral Problems*, 5 (1), 15–22.
- * Olweus, D. (1997a). Bully/victim problems in school: Knowledge base and an effective intervention project. *Irish Journal of Psychology*, 18, 170–190.
- * Olweus, D. (1997b). Bully/victim problems in school: Facts and intervention. *European Journal of Psychology of Education*, 12, 495–510.
- * Olweus, D. (1997c). Tackling peer victimization with a school-based intervention program. In D.P. Fry & K. Bjorkqvist (Eds.), *Cultural variation in conflict resolution: Alternatives to violence* (pp. 215–232). Mahwah, NJ: Erlbaum.

- * Olweus, D. (2004a). The Olweus Bullying Prevention Programme: Design and implementation issues and a new national initiative in Norway. In P.K. Smith, D. Pepler, & K. Rigby (Eds.), *Bullying in schools: How successful can interventions be?* (pp. 13–36). Cambridge: Cambridge University Press.
- * Olweus, D. (2004b). Bullying at school: Prevalence estimation, a useful evaluation design, and a new national initiative in Norway. *Association for Child Psychology and Psychiatry Occasional Papers*, 23, 5–17.
- * Olweus, D. (2005a). A useful evaluation design, and effects of the Olweus bullying prevention program. *Psychology, Crime and Law*, 11, 389–402.
- * Olweus, D. (2005b). *Bullying in schools: Facts and intervention*. Paper presented at the IX International Meeting on Biology and Sociology of Violence, Under the Honorary Presidency of H.M. Queen Sofia, Valencia, Spain (October).
- * Olweus, D. & Alsaker, F.D. (1991). Assessing change in a cohort-longitudinal study with hierarchical data. In D. Magnusson, L.R. Bergman, G. Rudinger, & B. Torestad (Eds.), *Problems and methods in longitudinal research* (pp. 107–132). Cambridge: Cambridge University Press.
- * O'Moore, M. (2005). *Prevention programme for teachers*. Paper presented at the IX International Meeting on Biology and Sociology of Violence, Under the Honorary Presidency of H.M. Queen Sofia, October, 6–7, 2005, Valencia.
- * O'Moore, A.M., & Minton, S. J. (2004). Ireland: The Donegal primary school antibullying project. In P. K. Smith, D. Pepler, & K. Rigby (Eds.), *Bullying in schools: How successful can interventions be?* (pp. 275–288). Cambridge: Cambridge University Press.
- * Ortega, R. & Del Rey, R. (1999). *The use of peer support in the S.A.V.E. project*. Paper presented at the IXth European Conference on Developmental Psychology, Spetses, Greece (September).
- * Ortega, R., Del-Rey, R., & Mora-Mercan, J. A. (2004). SAVE Model: An anti-bullying intervention in Spain. In P.K. Smith, D. Pepler, & K. Rigby (Eds.), *Bullying in schools: How successful can interventions be?* (pp. 167–186). Cambridge: Cambridge University Press.
- * Pagliocca, P.M., Limber, S.P., & Hashima, P. (2007). *Evaluation report for the Chula Vista Olweus Bullying Prevention Program*. Final report prepared for the Chula Vista Police Department.
- * Pepler, D.J., Craig, W.M., O'Connell, P., Atlas, R., & Charach, A. (2004). Making a difference in bullying: Evaluation of a systemic school-based program in Canada. In P.K. Smith, D. Pepler, & K. Rigby (Eds.), *Bullying in schools: How successful can interventions be?* (pp. 125–140). Cambridge: Cambridge University Press.
- * Pintabona, Y.C. (2006). *Frequently bullied students: Outcomes of a universal school-based bullying preventive intervention on peer victimization and psychological health*. PhD Thesis, Curtin University of Technology, Bentley, Western Australia.
- * Rahey, L. & Craig, W.M. (2002). Evaluation of an ecological program to reduce bullying in schools. *Canadian Journal of Counselling*, 36, 281–295.
- * Raskauskas, J. (2007). *Evaluation of the Kia Kaha anti-bullying programme for students in years 5-8*. Wellington: New Zealand Police.
- * Rican, P., Ondrova, K., & Svatos, J. (1996). The effect of a short, intensive intervention upon bullying in four classes in a Czech Town. *Annals of the New York Academy of Sciences*, 794, 399–400.
- * Rosenbluth, B., Whitaker, D.J., Sanchez, E., & Valle, L.A. (2004). The Expect Respect Project: Preventing bullying and sexual harassment in US elementary schools. In P.K. Smith, D. Pepler & K. Rigby (Eds.), *Bullying in schools: How successful can interventions be?* (pp. 211–233). Cambridge: Cambridge University Press.
- * Salmivalli, C., Kama, A., & Poskiparta, E. (2009). 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: Guilford.
- * Salmivalli, C., Kaukiainen, A., Voeten, M., & Sinisammal, M. (2004). Targeting the group as a whole: The Finnish anti-bullying intervention. In P.K. Smith, D. Pepler & K. Rigby (Eds.), *Bullying in schools: How successful can interventions be?* (pp. 251–275). Cambridge: Cambridge University Press.
- * Salmivalli, C., Kaukiainen, A., Voeten, M. (2005). Anti-bullying intervention: Implementation and outcome. *British Journal of Educational Psychology*, 75, 465–487.
- * Smith, P.K. (1997). Bullying in schools: The UK experience and the Sheffield Anti-bullying Project. *Irish Journal of School Psychology*, 18, 191–201.
- * Smith, P.K., Sharp, S., Eslea, M., & Thompson, D. (2004b). England: the Sheffield project. In P.K. Smith, D. Pepler & K. Rigby (Eds.), *Bullying in schools: How successful can interventions be?* (pp. 99–123). Cambridge: Cambridge University Press.

- * Spröber, N. (2006). Entwicklung und Evaluation eines Mehrebenen-Programms zur Prävention von „Bullying“ an weiterführenden Schulen und zur Förderung der positiven Entwicklung von Schülern. *Dissertation der Fakultät für Informations- und Kognitionswissenschaften der Eberhard-Karls-Universität Tübingen zur Erlangung des akademischen Grades eines Doktors der Naturwissenschaften.*
- * Spröber, N., Schlotke, P.F., & Hautzinger, M. (2006). ProACT + E: Ein Programm zur Prävention von "bullying" an Schulen und zur Förderung der positiven Entwicklung von Schülern: Evaluation eines schulbasierten, universalen, primär-präventiven Programms für weiterführende Schulen unter Einbeziehung von Lehrern, Schülern und Eltern. [German: ProACT + E: A programme to prevent bullying in schools and to increase the positive development of students. Evaluation of a school-based, universal, primary preventive programme for secondary schools that includes teachers, students and parents]. *Zeitschrift für Klinische Psychologie und Psychotherapie: Forschung und Praxis*, 35, 140–150.
- Stevens, V., De Bourdeaudhuij, I., & Van Oost, P. (2000). Bullying in Flemish Schools: an evaluation of anti-bullying intervention in primary and secondary schools. *The British Journal of Educational Psychology*, 70, 195–210.
- Stevens, V., Van Oost, P., & De Bourdeaudhuij, I. (2000). The effects of an anti-bullying intervention programme on peers' attitudes and behaviour. *Journal of Adolescence*, 23, 21–34.
- Stevens, V., Van Oost, P., & De Bourdeaudhuij, I. (2001). Implementation process of the Flemish anti-bullying intervention and relation with program effectiveness. *Journal of School Psychology*, 39(4), 303–317.
- Stevens, V., Van Oost, P., & De Bourdeaudhuij, I. (2004). Interventions against bullying in Flemish schools. In P. K. Smith, D. Pepler, & K. Rigby (Eds.), *Bullying in schools: How successful can interventions be?* (pp. 141–165). Cambridge: Cambridge University Press.
- Twemlow, S. W., Fonagy, P., & Sacco, F. C. (2005). A developmental approach to mentalizing communities: II. The Peaceful Schools experiment. *Bulletin of the Menninger Clinic*, 69(4), 282–304.
- * Whitaker, D.J., Rosenbluth, B., Valle, L.A., & Sacher, E. (2004). Expect Respect: A school-based intervention to promote awareness and effective responses to bullying and sexual harassment. In D.L. Espelage & S.M. Swearer (Eds.), *Bullying in American Schools: A social-ecological perspective on prevention and intervention* (pp. 327–350). Mahwah, NJ: Erlbaum.
- * 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.
- Wiefferink, C., Hoekstra, H., Beek, J. T., & Van Dorst, A. (2006). Effects of an anti-bullying programme in elementary schools in the Netherlands. *European Journal of Public Health*, 16, 76–76.

Other references

- Altman, D. G., Schulz, K. F., Moher, D., Egger, M., Davidoff, F., Elbourne, D., et al. (2001). The revised CONSORT Statement for reporting randomized trials: explanation and elaboration. *Annals of Internal Medicine*, 8, 663–694.
- Andershed, H., Kerr, M., & Stattin, H. (2001). Bullying in schools and violence on the streets: are the same people involved? *Journal of Scandinavian Studies in Criminology and Crime Prevention*, 2, 31–49.
- Baldry, A. C., & Farrington, D. P. (1999). Types of bullying among Italian school children. *Journal of Adolescence*, 22, 423–426.
- Baldry, A. C., & Farrington, D. P. (2007). Effectiveness of programs to prevent school bullying. *Victims and Offenders*, 2, 183–204.
- Dishion, T. J., McCord, J., & Poulin, F. (1999). When interventions harm: peer groups and problem behavior. *The American Psychologist*, 54, 755–764.
- Dodge, K. A., Dishion, T. J., & Lansford, J. E. (Eds.). (2006). *Deviant peer influences in programs for youth*. New York: Guilford.
- Farrington, D. P. (1993). Understanding and preventing bullying. In M. Tonry (Ed.), *Crime and justice* (pp. Vol. 17, pp. 381–458). Chicago: University of Chicago Press.
- Farrington, D. P. (2003). Methodological quality standards for evaluation research. *Annals of the American Academy of Political and Social Sciences*, 587, 49–68.
- Farrington, D. P. (2009). Conduct disorder, aggression and delinquency. In R. M. Lerner & L. Steinberg (Eds.), *Handbook of adolescent psychology* (3rd ed., Vol. 1, pp. 683–722). Hoboken: Wiley.

- Farrington, D. P., Jolliffe, D., & Johnstone, L. (2008). *Assessing violence risk: A framework for practice*. Edinburgh: Risk Management Authority Scotland.
- Farrington, D. P., & Petrosino, A. (2001). The Campbell Collaboration Crime and Justice Group. *The Annals of the American Academy of Political and Social Science*, 578, 35–49.
- Farrington, D.P. & Ttofi, M.M. (2009). School-based programs to reduce bullying and victimization. *Campbell Systematic Reviews*, 2009:6.
- Farrington, D. P., & Weisburd, D. (2007). The Campbell Collaboration Crime and Justice Group. *The Criminologist*, 32(1), 1–5.
- Farrington, D. P., & Welsh, B. C. (2003). Family-based prevention of offending: a meta-analysis. *Australian and New Zealand Journal of Criminology*, 36, 127–151.
- Fekkes, M., Pijpers, F. I. M., & Verloove-Vanhorick, P. S. (2005). Bullying: Who does what, when and where? Involvement of children, teachers and parents in bullying behavior. *Health Education Research*, 20, 81–91.
- Ferguson, C. J., Miguel, C. S., Kilburn, J. C., & Sanchez, P. (2007). The effectiveness of school-based anti-bullying programs: A meta-analytic review. *Criminal Justice Review*, 32, 401–414.
- Fonagy, P., Twemlow, S. W., Vernberg, E., Sacco, F. C., & Little, T. D. (2005). Creating a peaceful school learning environment: the impact of an anti-bullying program on educational attainment in elementary schools. *Medical Science Monitor*, 11(7), 317–325.
- Hedges, L. V. (1982). Fitting continuous models to effect size data. *Journal of Educational Statistics*, 7, 245–270.
- Howard, K. A., Flora, J., & Griffin, M. (1999). Violence-prevention programs in schools: State of science and implications for future research. *Applied and Preventive Psychology*, 8, 197–215.
- Kaminski, J. W., Valle, L. A., Filene, J. H., & Boyle, C. Y. (2008). A meta-analytic review of components associated with parent training program effectiveness. *Journal of Abnormal Child Psychology*, 36, 567–589.
- Lipsey, M. W. (2003). Those confounded moderators in meta-analysis: Good, bad, and ugly. *The Annals of the American Academy of Political and Social Science*, 587, 69–81.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks: Sage.
- Losel, F., & Beelman, A. (2003). Effects of child skills training in preventing antisocial behavior: a systematic review of randomized evaluations. *The Annals of the American Academy of Political and Social Science*, 587, 84–109.
- McGuire, J. (2001). What works in correctional intervention? Evidence and practical implications. In G. A. Bernfeld, D. P. Farrington, & A. W. Leschied (Eds.), *Offender rehabilitation in practice: Implementing and evaluating effective programs* (pp. 3–19). Chichester, UK: Wiley.
- Meraviglia, M. G., Becker, H., Rosenbluth, B., Sanchez, E., & Robertson, T. (2003). The Expect Respect Project: creating a positive elementary school climate. *Journal of Interpersonal Violence*, 18, 1347–1360.
- Merrell, K. W., Gueldner, B. A., Ross, S. W., & Isava, D. M. (2008). How effective are school bullying intervention programs? a meta-analysis of intervention research. *School Psychology Quarterly*, 23(1), 26–42.
- Moher, D., Schulz, K. F., & Altman, D. (2001). The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomized trials. *Journal of the American Medical Association*, 285, 1987–1991.
- Mytton, J., DiGiuseppi, C., Gough, D., Taylor, R., & Logan, S. (2006). School-based secondary prevention programs for preventing violence. *Cochrane Database of Systematic Reviews*, Issue 3. Art No. CD 004606.
- Olweus, D. (1993b). *Bullying at school: What we know and what we can do*. Oxford: Blackwell.
- Perry, A. E., & Johnson, M. (2008). Applying the Consolidated Standards of Reporting Trials (CONSORT) to studies of mental health provision for juvenile offenders. *Journal of Experimental Criminology*, 4, 165–185.
- Rigby, K. (2002). *A meta-evaluation of methods and approaches to reducing bullying in preschools and early primary school in Australia*. Canberra: Attorney General's Department, Crime Prevention Branch.
- Salmivalli, C., & Nieminen, E. (2002). Proactive and reactive aggression among school bullies, victims and bully-victims. *Aggressive Behavior*, 28, 30–44.
- Smith, J. D., Schneider, B., Smith, P. K., & Ananiadou, K. (2004). The effectiveness of whole-school anti-bullying programs: A synthesis of evaluation research. *School Psychology Review*, 33, 548–561.
- 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., Pepler, D., & Rigby, K. (Eds.). (2004a). *Bullying in schools: How successful can interventions be?* Cambridge: Cambridge University Press.
- Ttofi, M. M., & Farrington, D. P. (2008). Bullying: Short-term and long-term effects, and the importance of Defiance Theory in explanation and prevention. *Victims and Offenders*, *3*, 289–312.
- Ttofi, M. M., Farrington, D. P., & Baldry, C. A. (2008). *Effectiveness of programs to reduce school bullying: a systematic review*. Stockholm: Swedish National Council for Crime Prevention.
- 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.
- Warden, D., Moran, E., Gillies, J., Mayes, G., & Macleod, L. (1997). An evaluation of a children's safety training programme. *Educational Psychology*, *17*, 433–448.
- Weisburd, D. (1993). Design sensitivity in criminal justice experiments. In M. Tonry (Ed.), *Crime and justice* (Vol. 17, pp. 337–380). Chicago: University of Chicago Press.
- Weisburd, D., Lum, C. M., & Petrosino, A. (2001). Does research design affect study outcomes in criminal justice? *The Annals of the American Academy of Political and Social Science*, *578*, 50–70.
- Welsh, B. C., Farrington, D. P., & Sherman, L. W. (Eds.). (2001). *Costs and benefits of preventing crime*. Boulder: Westview.
- Wilson, S. J., & Lipsey, M. W. (2007). School-based interventions for aggressive and disruptive behavior: update of a meta-analysis. *American Journal of Preventive Medicine*, *33*, S130–S143.
- Woods, T., Coyle, K., Hoglund, W., & Leadbeater, B. (2007). Changing the contexts of peer victimization: The effects of a primary prevention program on school and classroom levels of victimization. In J. E. Zins, M. J. Elias, & C. A. Maher (Eds.), *Bullying, victimization, and peer harassment: A handbook of prevention and intervention* (pp. 369–388). New York: Haworth.

Maria M. Ttofi completed her MPhil and PhD Degrees at the Institute of Criminology, University of Cambridge, under the supervision of Professor David P. Farrington. Her PhD research focused on the effectiveness of bullying prevention programs and on the empirical testing of theories of teacher victimization, school bullying, and aggressive behavior in general. Maria's interests are focused on systematic/meta-analytic reviews, experimental research, school bullying/aggression and juvenile delinquency. She is currently conducting a new systematic/meta-analytic review on adverse health and criminal outcomes of children involved in school bullying, based on longitudinal studies, in collaboration with professors Farrington, Losel and Loeber.

David P. Farrington O.B.E., is Professor of Psychological Criminology at the Institute of Criminology, Cambridge University. He is co-chair of the U.S. National Institute of Justice Study Group on Transitions from Juvenile Delinquency to Adult Crime and co-chair of the U.S. Centers for Disease Control Expert Panel on Promotive and Protective Factors for Youth Violence. His major research interest is in developmental criminology, and he is Director of the Cambridge Study in Delinquent Development, which is a prospective longitudinal survey of over 400 London males from age 8 to age 48. In addition to over 500 published journal articles and book chapters on criminological and psychological topics, he has published over 75 books, monographs and government publications.