Systematic Review Protocol

Title:

Effectiveness of Programs to Prevent School Bullying

Reviewers:

David P. Farrington
Institute of Criminology, Cambridge University

Anna C. Baldry
Department of Psychology, Second University of Naples and Intervict Institute, Tilburg University

Britta Kyvsgaard
Ministry of Justice, Copenhagen, Denmark

Maria M. Ttofi
Institute of Criminology, Cambridge University

Lead Reviewer:

David P. Farrington,
Institute of Criminology, Sidgwick Avenue, Cambridge CB3 9DT, UK.
Tel. (44) 1223 872555; Fax (44) 1223 335356;
email: dpf1@cam.ac.uk

Sources of Support

Nordic Campbell Center
Swedish National Council on Crime Prevention

Revised February 2008
Contents

1. Background for the Review
2. Objectives of the Review
3. Methods
   3.1 Criteria for inclusion and exclusion of studies
   3.2 Search strategies
   3.3 Description of methods used in the primary research
   3.4 Criteria for determination of independent findings
   3.5 Details of study coding categories
   3.6 Statistical procedures and conventions
   3.7 Treatment of qualitative research
4. Time Frame
5. Plans for Updating the Review
6. Acknowledgements
7. Statement Concerning Conflict of Interest
8. References and some Key Studies
9. Table 1: Key Bullying Prevention Projects
10. Appendix: Coding Protocol
1. **Background**

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 more powerful child (or children) oppressing less powerful ones; and repeated incidents between the same children over a prolonged period (Farrington, 1993; Olweus, 1993; Roland, 1989). School bullying can occur in school or on the way to or from school. It is not bullying when two persons of the same strength (physical, psychological, or verbal) victimize each other.

Bullying is different from aggression or violence; not all aggression/violence involves bullying, and not all bullying involves aggression/violence. For example, bullying includes being called nasty names, being rejected, ostracized or excluded from activities, having rumors spread about you, having belongings taken away, teasing and threatening (Baldry & Farrington, 1999). Our aim is to review programs that are specifically intended to prevent or reduce school bullying, not programs that are intended to prevent or reduce school aggression or violence.

School bullying is perceived to be an important social problem in many different countries. The nature and extent of the problem, and research on it, in 21 different countries, is reviewed in Smith *et al.* (1999). Special methods are needed to study bullying in different countries because of problems of translating the term “bullying” into different languages. Smith *et al.* (2002) have reviewed the meaning of bullying in 14 different countries. In studying bullying, researchers often ask about specific components such as “hit him/her on the face” or “excluded him/her from games” rather than use the actual word “bullying” in their interviews and questionnaires (Pateraki & Houndoumadi, 2001). However, they still use the word “bullying” in their reports.

Many school-based intervention programs have been devised and implemented in an attempt to reduce school bullying. These have been targeted on bullies, victims, peers,
teachers, or on the school in general. Many programs seem to have been based on
commonsense ideas about what might reduce bullying rather than on empirically-
supported theories of why children bully, why children become victims, or why bullying
events occur.

The first large-scale anti-bullying program was implemented in Norway in 1983. A
more intensive version of the national program was evaluated in Bergen by Olweus
(1991). This program aimed to increase awareness and knowledge of teachers, parents,
and students about bullying and to dispel myths about it. Booklets were distributed to
schools, folders were distributed to families, and videos about bullying were shown and
discussed in schools. Also, schools received information about the nature and extent of
bullying (based on questionnaires completed by students), teachers were encouraged to
develop explicit rules about bullying, and monitoring and supervision of children in the
playground was improved.

The evaluation by Olweus (1991) showed a dramatic decrease in victimization of
about half after the program. Since then at least 15 other large-scale anti-bullying
programs, some inspired by Olweus and some based on other principles, have been
implemented and evaluated in at least 10 other countries. Baldry and Farrington (2007)
reviewed 16 major evaluations in 11 different countries and concluded that 8 produced
desirable results, 2 produced mixed results, 4 produced small or negligible effects, and 2
produced undesirable results. Most programs were quite complex, and the effectiveness
of different components of programs was not clear. These 16 evaluations are listed after
the References.

The best single source of reports of anti-bullying programs is the book edited by
Smith, Pepler, and Rigby (2004), which contains descriptions of 13 programs in 11
different countries. 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. These two prior reviews are of high quality. However, neither carried out a full meta-analysis measuring weighted mean effect sizes and correlations between study features and effect sizes. The Smith et al. (2004) review covered only 14 evaluations up to 2002, 6 of which were uncontrolled. The Vreeman and Carroll (2007) review covered 26 evaluations up to 2004, and was restricted to studies published in the English language. We hope to go beyond these previous reviews by (a) searching for evaluations up to the end of 2007, (b) searching for international evaluations, and (c) carrying out more extensive meta-analyses, as specified in this protocol.

American research is generally targeted on school violence or peer victimization rather than bullying. 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 et al., 2003; Wilson & Lipsey, 2007). We will consult these, but we must emphasize that our research aims to review anti-bullying programs specifically.

2. Objectives of the Review

The main objective is to assess the effectiveness of school-based anti-bullying programs in reducing school bullying. Our aim is to locate and summarize all the major evaluations of programs in developed countries. Bullying has been studied in Australia, Austria, Belgium, Canada, Cyprus, Denmark, England and Wales, Finland, France, Germany, Greece, Iceland, Ireland, Israel, Italy, Luxembourg, Japan, Malta, New Zealand, Northern Ireland, Norway, Portugal, Scotland, Spain, Sweden, Switzerland, The Netherlands, and the United States (Smith et al., 1999). We aim (potentially) to include research in all these countries. We aim to measure effect sizes in each evaluation and to investigate which features (e.g. of programs, students, and schools) are related to effect sizes. We hope to make recommendations about which components of programs are most effective in which circumstances, and hence about how future anti-bullying programs
might be improved. We also hope to make recommendations about how the design and analysis of evaluations of anti-bullying programs might be improved in future.

3. **Methods**

3.1 **Criteria for inclusion and exclusion of studies**

We propose 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 aggression or violence will be excluded.

(b) Bullying was defined as including: physical, verbal, or psychological attack or intimidation that is intended to cause fear, distress, or harm to the victim; and an imbalance of power, with the more powerful child (or children) oppressing less powerful ones. Many definitions also require repeated incidents between the same children over a prolonged period, but we do not require that, because many studies of bullying do not specifically measure or report this element of the definition.

(c) Bullying (specifically) was measured using self-report questionnaires, peer ratings, teacher ratings, observational data, or school records.

(d) The effectiveness of the program was measured by comparing students who received it (the experimental condition) with students who did not receive it (the control condition). We require that there must have been some pretest control of extraneous variables in the evaluation (establishing the prior equivalence of conditions) by (i) randomization, or (ii) pretest measures of bullying, or (iii) matching on pretest bullying or risk factors/scores for bullying, or (iv) pretest measurement of risk factors or risk scores for bullying. Because of low internal validity, we will exclude uncontrolled studies that only had before and after measures of bullying in experimental schools or classes. However, we propose to include studies that controlled for age. For example, in the Olweus (1991)
evaluation, all students received the anti-bullying program, but Olweus compared students of age X after the program with different students of the same age X in the same schools before the program. We will include this kind of evaluation.

(e) Published and unpublished reports of research conducted in developed countries between 1983 and the present will be included. We believe that there was no worthwhile evaluation research on antibullying programs conducted before the pioneering research of Olweus carried out in 1983.

There is an additional criterion for inclusion in our meta-analysis (but not in our systematic review):

(f) It was possible to measure the effect size. The main measures of effect size are the odds ratio, based on numbers of bullies/nonbullies (or victims/nonvictims), and the standardized mean difference, based on mean scores on bullying and victimization. Where the required information is not presented in reports, we will seek to obtain it by contacting the authors directly.

(g) We have agreed not to set a minimum initial sample size for including studies. Initially, we wished to set a minimum initial sample size of 200, for the following reasons: First, larger studies are usually better-funded and of higher methodological quality (e.g. Jolliffe & Farrington, 2007). Second, we are very concerned about the frequently-found negative correlations between sample size and effect size (e.g. Farrington & Welsh, 2003). We think that this correlation probably reflects publication bias. Smaller studies that yield statistically significant results are published, whereas those that do not are left in the file drawer. In contrast, larger studies (often funded by some official agency) tend to be published irrespective of their results. Excluding smaller studies minimizes problems of publication bias and therefore yields a more accurate estimate of the true effect size. Third, we think that larger studies are likely to have higher external validity or
generalizability. Fourth, attrition (e.g. between pretest and posttest) is less problematic in larger studies. A study with 100 children that suffers 30% attrition will end up with only 35 boys and 35 girls: these are very small samples (with associated large confidence intervals) for estimating the prevalence of bullying and victimization. In contrast, a study with 300 children that suffers 30% attrition will end up with 105 boys and 105 girls: these are much more adequate samples.

However, the reviewers of our protocol argued that useful information was contained in smaller studies, that the minimum sample size of 200 was arbitrary, that methodological quality and publication bias should be addressed in other ways, that attrition is equally important in smaller and larger studies, and that larger studies do not necessarily have higher external validity. Based on Vreeman and Carroll (2007), we estimate that about half of all evaluations have an initial sample size of 200 or more, and conversely half have an initial sample size less than this. Therefore, not setting a minimum sample size doubles the number of studies included in our systematic review. We plan to analyze larger and smaller studies separately to investigate whether the results obtained with the two types of studies are concordant or discordant.

3.2 Search strategies

(a) We propose to search as many as possible of the following electronic databases using the following keywords:

Bully*/Bullies/Anti-Bullying

AND

School;

AND

Intervention*/Program*/Outcome*/Evaluation*/Effect*

We will not include Violence or Aggression as key words along with Bully/Bullies/Anti-Bullying because we know that this will identify many studies that are not
relevant to the present review. We have considered whether to experiment with key words such as Intimidation/Harrassment/Teasing/Victim* but, as mentioned above, our aim is to review studies of interventions designed specifically to reduce school bullying.

(b) **List of Databases**

Australian Criminology Database (CINCH)
Cochrane Controlled Trials Register
C2-SPECTR
Criminal Justice Abstracts
Database of Abstracts of Reviews of Effectiveness (DARE)
Dissertation Abstracts
Educational Resources Information Clearinghouse (ERIC)
EMBASE
Google Scholar
MEDLINE
National Criminal Justice Reference Service (NCJRS)
PsychInfo/Psychlit
Sociological Abstracts
Social Sciences Citation Index (SSCI)

Our experience is that electronic searches are very time-consuming and identify few previously unknown studies.

We propose to handsearch the following key journals since 1983:

(c) **Aggression and Violent Behavior**

**Aggressive Behavior**

**Archives of Pediatrics and Adolescent Medicine**

**British Journal of Educational psychology**
We propose to seek information from key researchers on bullying and from international colleagues in the Campbell Collaboration, including Catherine Blaya, Vincente Garrido, Peter van der Laan, Friedrich Lösel, Dan Olweus, Debra Pepler, Ken Rigby and Peter Smith. Our experience is that this method identifies the most relevant studies. Where international colleagues identify a report in a language that we cannot understand, we will ask them to provide us with a brief translation of key features that are needed for our coding schedule. We believe that, with the cooperation of colleagues in the Campbell Collaboration, we can include research in many different developed countries.

We will search reference lists of review articles and primary studies.
3.3 Description of methods used in the primary research

Very few studies randomly assign students to experimental or control groups. A few studies randomly assign schools or school classes (e.g. Cross et al., 2004). Some studies have before and after measures of bullying in experimental and control schools or school classes (e.g. Alsaker & Valkenover, 2001). Some studies compare students of a particular age before the intervention with students of the same age in the same schools after the intervention (e.g. Olweus, 2005). Many studies have before and after measures of bullying in experimental schools but no control schools or school classes (e.g. Pitts & Smith, 1995). These latter studies will be excluded.

3.4 Criteria for determination of independent findings

Most studies measure bullying and victimization using student self-report questionnaires. Information is usually collected separately about whether students bully and whether students are bullied. Some studies also use teacher ratings, peer ratings, or (more rarely) playground observations or school records. We propose to calculate effect size measures of bullying and victimization, based on self-report questionnaires, in as many studies as possible. Our main meta-analysis will be based on this measure. Where self-report questionnaires are not available, we will calculate effect size measures based on other sources. Where short-term and long-term follow-up measures are available, we will calculate effect sizes separately for the shortest and longest follow-up periods.

3.5 Details of study coding categories

We propose to cover the following topics at least:

Author(s)
Dates of research
Date of publication(s)
Country and place of research

Age of students (Experimental, Control)

Gender composition (E, C)

Ethnic composition (E, C)

Initial sample size (E, C)

Final sample size (E, C)

Number of schools (E, C)

Number of classes (E, C)

Research design and methodological quality (taking account of randomization, generalizability and attrition)

Type and components of intervention (e.g. “whole-school”, curriculum; work with bullies, victims, peers, teachers, improving playground supervision, etc.)

Target of intervention (e.g. students, teachers, parents)

Duration of intervention

Follow-up time period

Outcome measures (self-report questionnaires, teacher ratings, peer ratings, systematic observation, school records)

The Appendix contains a draft coding schedule. A random sample of 20 studies will be coded by two persons in order to measure reliability.

3.6 Statistical procedures and conventions

The main measure of effect size will be the odds ratio (OR), calculated by comparing experimental and control conditions on bullies/nonbullies (and victims/nonvictims). Where scores are reported, the standardized mean difference \( d \) will be calculated and will be converted into LOR using the equation:

\[
d = 0.5513 \times \text{LOR} \quad (\text{see Lipsey & Wilson, 2001, p. 202}).
\]

Where before and after information is provided, the OR and \( d \) will be adjusted for
the before data.

Effect size measures will be calculated for bullying and victimization separately. Weighted mean effect size measures will be calculated using the procedures described in Lipsey and Wilson (2001). Where appropriate, fixed effects or random effects models will be used (depending on the heterogeneity Q). Correlations between features of studies and effect sizes will be investigated, in an attempt to assess which components of programs might be the most effective in which circumstances. Meta-analytic regressions will also be carried out to investigate independent influences of program components, methodological quality, features of participants, and design features. As mentioned, we hope to make recommendations about how to make anti-bullying programs more effective and about how to improve the design and evaluation of such programs, including recommendations about analytic techniques that the researchers should use. Ideally, we could produce a “CONSORT” statement for reports of anti-bullying programs that specifies what information should always be reported.

Statistical problems arise when the unit of allocation (e.g. schools or school classes) is different from the unit of analysis (e.g. students). This problem has rarely been addressed in the bullying literature. We will address it using formulae to correct test statistics for clustering developed by Cathleen McHugh based on Hedges (2007). We will carry out a sensitivity analysis assuming intraclass correlations between .10 and .20, as used by the What Works Clearinghouse of the U.S. Department of Education (Lipsey, 2008). We will also address the possibility of publication bias by using techniques described by Rothstein, Sutton, and Borenstein (2005), such as the funnel plot.

3.7 Treatment of qualitative research

The review will be focussed on quantitative studies but will include qualitative
information (e.g. on the degree of implementation of interventions) where this is helpful in discussing explanations for findings or conflicting results.

4. **Time Frame**

Our tentative time schedule is as follows. We have already completed most of the searches.

February, 2008: Develop/test coding scheme

March-April: Code included studies, analyze results

May-June 2008: Complete first draft of review and submit it to Campbell Collaboration

Later in 2008: Receive feedback from Campbell Collaboration. Submit final draft of review for electronic publication on the Campbell website.

5. **Plans for Updating the Review**

The review will be updated every 3 years. The lead reviewer will take the lead in arranging this.

6. **Acknowledgements**

We are very grateful to David Wilson, Jeff Valentine and two anonymous reviewers for helpful comments.

7. **Statement Concerning Conflict of Interest**

The only possible (minor) conflict of interest is that one of the evaluations of anti-bullying programs was conducted by Baldry and Farrington (2004). Otherwise, none of us has any interest in the conclusions of the review and none of us stands to benefit in any way from any source that has any interest in the conclusions of the review. We are not aware of any personal, political, academic, or financial factors that might bias our judgment.
8. References and Some Key Studies


Farrington, D. P. & Welsh, B. C. (2003). Family-based prevention of offending: A meta-


In P. K. Smith, D. Pepler, & K. Rigby (Eds.), *Bullying in schools: How successful can interventions be?* (pp. 125-140). Cambridge: Cambridge University Press.


### Table 1

Key Bullying Prevention Projects (from Baldry & Farrington, 2007)

<table>
<thead>
<tr>
<th>Project</th>
<th>Components of the program</th>
<th>Participants</th>
<th>Assessment methods</th>
<th>Research Design</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bergen, Norway (Olweus, 1991, 1993)</td>
<td>Individual, classroom and school components. Each school had to set up anti-bullying rules</td>
<td>2,500 students aged 10-15 in 42 primary and secondary schools.</td>
<td>Student self-report questionnaires, teacher ratings</td>
<td>Age cohort design with two post-tests after 8 and 20 months</td>
</tr>
<tr>
<td>Rogaland, Norway (Roland, 1989, 1993)</td>
<td>Similar to the Bergen evaluation but less detailed</td>
<td>7,000 students aged 8-16 in 37 primary and secondary schools</td>
<td>Student self-report questionnaires, teacher ratings</td>
<td>Pre-test, post-test design. Follow-up after 3 years. No control schools</td>
</tr>
<tr>
<td>Sheffield, England (Smith &amp; Sharp, 1994; Eslea &amp; Smith, 1998)</td>
<td>Whole school approach with several components.</td>
<td>6,500 students aged 8-16 from 16 primary and 7 secondary schools (intervention). 4 control schools</td>
<td>Student self-report questionnaires, interviews with teachers</td>
<td>Pre-test and post-test 18 months later. 3 year follow-up in 4 intervention schools</td>
</tr>
<tr>
<td>Liverpool and London, England (Pitts &amp; Smith, 1995)</td>
<td>Anti-bullying policy, focusing on peer support and training in assertiveness skills</td>
<td>1,284 students aged 8-16 in 4 primary and secondary schools.</td>
<td>Student self-report questionnaires</td>
<td>Pre-test and post-test (follow-up after 2 years). No control schools.</td>
</tr>
<tr>
<td>Toronto, Canada (Pepler et al., 1994)</td>
<td>Olweus components based on the individual, the classroom, parents, and the school</td>
<td>First study: 898 students aged 8-14 from 4 elementary schools. Second study: 922 students from 3 schools.</td>
<td>Student self-report questionnaires, playground observations</td>
<td>Pre-test, post test (after 18 and 30 months), with no control schools.</td>
</tr>
<tr>
<td>New South Wales, Australia (Peterson &amp; Rigby, 1999)</td>
<td>Peer support and method of shared concern by staff.</td>
<td>758 students aged 12-17 in 1995, and 657 in 1997</td>
<td>Student self-report questionnaires</td>
<td>Pre-test, post-test design with no control group</td>
</tr>
<tr>
<td>Donegal, Ireland (O’Moore &amp; Minton, 2004)</td>
<td>Training teachers, Information pack, work with students</td>
<td>527 students from 42 schools. For only 22 it was possible to match pre and post data</td>
<td>Student self-report questionnaires</td>
<td>Pre-test, post-test design with no control group</td>
</tr>
<tr>
<td>Turku and Helsinki, Finland (Salmivalli, 2001)</td>
<td>Individual, classroom, school and teacher interventions</td>
<td>1,220 students aged 9-12 in 16 schools</td>
<td>Student self-report questionnaires, peer nominations</td>
<td>Baseline, low, high implementation, age cohort design; follow-up after 6 months.</td>
</tr>
<tr>
<td>Country/Region</td>
<td>Intervention Description</td>
<td>Participants</td>
<td>Methodology</td>
<td>Outcome Measures</td>
</tr>
<tr>
<td>------------------------------</td>
<td>------------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------</td>
<td>------------------------------------------------------------------------------</td>
<td>---------------------------------------</td>
</tr>
<tr>
<td>Flanders, Belgium (Stevens et al., 2000)</td>
<td>Individual, classroom and school interventions</td>
<td>1,104 students aged 10-16 from 18 schools.</td>
<td>Student self-report questionnaires</td>
<td>Full, part implementation or control. Pre-test, 6 months later, and one year later.</td>
</tr>
<tr>
<td>Expect Respect, Texas, US (Whitaker et al., 2004)</td>
<td>Classroom education, staff training, policy/procedure development, parent education, support services</td>
<td>929 experimental and 834 control students aged 11-12 in 12 schools</td>
<td>Student self-report questionnaires</td>
<td>Pre-test, post-test, intervention and control groups.</td>
</tr>
<tr>
<td>Berne, Switzerland (Alsaker &amp; Valkanover, 2001)</td>
<td>Whole school approach to bullying focussed on teachers</td>
<td>152 experimental children aged 5-7 in 8 classes. 167 control children in 8 classes.</td>
<td>Teacher ratings, peer nominations</td>
<td>Pre-test, post-test, control group design</td>
</tr>
<tr>
<td>“Bulli &amp; Pupe”, Rome, Italy (Baldry &amp; Farrington, 2004)</td>
<td>Work with the class, with booklets and videos on bullying and family violence</td>
<td>239 students aged 10-16 in 13 schools</td>
<td>Student self-report questionnaires</td>
<td>Intervention and control groups, random assignment, pre-test and post-test measures</td>
</tr>
<tr>
<td>SAVE Seville, Spain (Ortega &amp; Lera, 2000)</td>
<td>Community approach focussing on democratic values, cooperative group work, empathy.</td>
<td>910 students aged 8-18 in primary and secondary schools.</td>
<td>Student self-report questionnaires</td>
<td>5 intervention schools had pre-test and post-test measures, compared to 3 control schools with only post-test measures. Follow-up after 4 years</td>
</tr>
<tr>
<td>Friendly Schools, Perth (Cross et al., 2004)</td>
<td>Friendly schools approach addressing individuals, classes and schools, and teacher training.</td>
<td>2,068 students aged 9-10 from 29 schools</td>
<td>Student self-report questionnaires</td>
<td>Pre-test and post-test data from intervention and control schools, randomly assigned</td>
</tr>
<tr>
<td>Steps to Respect, Pacific Northwest, US (Frey et al., 2005)</td>
<td>Staff training, promotion of prosocial beliefs</td>
<td>1,126 students aged 8-12 in 6 schools</td>
<td>Teacher and student ratings, playground observation</td>
<td>Pre-test, post-test, experimental and control groups, randomly assigned.</td>
</tr>
<tr>
<td>Norway (Olweus, 2005)</td>
<td>As for the Bergen evaluation</td>
<td>20,709 students aged 9-12</td>
<td>Student self-report questionnaires</td>
<td>Pre-test and post-test measures</td>
</tr>
</tbody>
</table>
Appendix

Coding Protocol

Use one coding sheet for each distinct research project that is included in the systematic review.

Throughout, zero = not known, 1 = Yes, 2 = No

Name of coder: ____________________________
Date of coding: ____________________________

Study

Study no. identifier: ____________________________
Author(s) name: ____________________________
Author(s) affiliation: ____________________________
Date(s) research was conducted: ____________________________
Date of primary publication: ____________________________
Place of research: ____________________________
Country of research: ____________________________

Source of funding: ____________________________

Details of primary report:

_____________________________________________________________________________

Publication type:

1 book
2 book chapter
3 Journal article
4 technical report
5 unpublished conference paper
6 dissertation

Pub type: ____________________________

Details of other reports:

_____________________________________________________________________________

_____________________________________________________________________________

_____________________________________________________________________________
2. Eligibility criteria

Are the following inclusion criteria present? (if not clear, attempt to find out from the author)

The report deals with bullying defined as physical, verbal, or psychological attack
or intimidation that is intended to cause fear, distress, or harm to the victim;
an imbalance of power, with the more powerful child (or children) oppressing
less powerful ones.  

Types of bullying measured _______________________________________________________
The report analyzes repeated incidents between the same children over
a prolonged period.  
The report describes an evaluation of a program designed to reduce school
bullying.  
The report indicates which methods were adopted to gather information on
bullying (self-report questionnaires, peer ratings, teacher ratings, observational
data, or school records).  

Before data on bullying and victimization as defined above (specifically) are
reported.  

After data on bullying and victimization as defined above (specifically) are
reported

The effectiveness of the program was measured by comparing students
who received it (the experimental group) with students who did not receive
it (the control group).

The Study had some control of extraneous variables (establishing the
prior equivalence of groups) by (i) randomization, or (ii) pretest measures
of bullying, or (iii) matching, or (iv) pretest measurement of risk factors or
risk scores for bullying.
Numbers of bullies/nonbullies and victims/nonvictims are reported.

Scores on bullying and victimization are reported.

The research was conducted in one of the following countries since 1983:

<table>
<thead>
<tr>
<th>Country</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Australia</td>
<td>01</td>
</tr>
<tr>
<td>Austria</td>
<td>02</td>
</tr>
<tr>
<td>Belgium</td>
<td>03</td>
</tr>
<tr>
<td>Canada</td>
<td>04</td>
</tr>
<tr>
<td>Cyprus</td>
<td>05</td>
</tr>
<tr>
<td>Denmark</td>
<td>06</td>
</tr>
<tr>
<td>England and Wales</td>
<td>07</td>
</tr>
<tr>
<td>Finland</td>
<td>08</td>
</tr>
<tr>
<td>France</td>
<td>09</td>
</tr>
<tr>
<td>Germany</td>
<td>10</td>
</tr>
<tr>
<td>Greece</td>
<td>11</td>
</tr>
<tr>
<td>Iceland</td>
<td>12</td>
</tr>
<tr>
<td>Ireland</td>
<td>13</td>
</tr>
<tr>
<td>Israel</td>
<td>14</td>
</tr>
<tr>
<td>Italy</td>
<td>15</td>
</tr>
<tr>
<td>Japan</td>
<td>16</td>
</tr>
<tr>
<td>Luxembourg</td>
<td>17</td>
</tr>
<tr>
<td>Malta</td>
<td>18</td>
</tr>
<tr>
<td>New Zealand</td>
<td>19</td>
</tr>
<tr>
<td>Northern Ireland</td>
<td>20</td>
</tr>
<tr>
<td>Norway</td>
<td>21</td>
</tr>
<tr>
<td>Portugal</td>
<td>22</td>
</tr>
<tr>
<td>Scotland</td>
<td>23</td>
</tr>
<tr>
<td>Spain</td>
<td>24</td>
</tr>
<tr>
<td>Sweden</td>
<td>25</td>
</tr>
<tr>
<td>Switzerland</td>
<td>26</td>
</tr>
<tr>
<td>The Netherlands</td>
<td>27</td>
</tr>
<tr>
<td>United States</td>
<td>28</td>
</tr>
<tr>
<td>Other</td>
<td>29</td>
</tr>
</tbody>
</table>

3. Sample

Mean age of students:

<table>
<thead>
<tr>
<th></th>
<th>Experimental</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cage</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Gender composition:

<table>
<thead>
<tr>
<th></th>
<th>Experimental</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Epcf</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpcf</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Ethnic composition:

<table>
<thead>
<tr>
<th></th>
<th>Experimental</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Initial sample size:

<table>
<thead>
<tr>
<th></th>
<th>Experimental</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>En1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cn1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sample size in short follow-up:

<table>
<thead>
<tr>
<th></th>
<th>Experimental</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>En2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cn2</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sample size in long follow-up:

<table>
<thead>
<tr>
<th></th>
<th>Experimental</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>En3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cn3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Number of schools:

<table>
<thead>
<tr>
<th></th>
<th>Experimental</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eschools</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cschools</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Number of school classes: Experimental Eclasses________________
Control Cclasses________________

4. **Research design**

1. randomized
2. before-after with control condition
3. before-after with age cohort comparison
4. only after with control condition

Who were randomized? 1. Schools, 2. Classes, 3. Students

How were units allocated to experimental and control conditions?

Was Control condition comparable? 1 = Yes, 2 = No

Variables measured to establish matching or comparability

To what population can the results be generalized?

Is there a potential generalizability threat from overall attrition? 1 = Yes, 2 = No

Is there a potential threat to internal validity from differential attrition? 1 = Yes, 2 = No

5. **Pretest Measures**

What measures were used?

- self-report questionnaires
  - Yes = 1
  - No = 2

- teacher ratings
  - Yes = 1
  - No = 2

- peer ratings
  - Yes = 1
  - No = 2

- systematic observation
  - Yes = 1
  - No = 2

- school records
  - Yes = 1
  - No = 2

Where available, effect size measures will be based on self-report questionnaires. If these are not available, code in the order: teacher ratings, peer ratings, systematic observation, school records.
Scales:

- a) What scale was used to measure bullying? Bulscale
- b) What scale was used to measure victimization? Vicscale

What reference period was used for bullying? (in days) Bulref
What reference period was used for victimization? (in days) Vicref
Was there any information about reliability or validity of measures? Rel
Val

6. Intervention (experimental condition)

Type and components of intervention: (Yes = 1, No = 2)

<table>
<thead>
<tr>
<th>Component (Yes = 1, No = 2)</th>
<th>Int</th>
</tr>
</thead>
<tbody>
<tr>
<td>Whole-school policy and procedures</td>
<td>Intwhole</td>
</tr>
<tr>
<td>Classroom rules</td>
<td>Intrules</td>
</tr>
<tr>
<td>School conferences/providing information about bullying</td>
<td>Intconf</td>
</tr>
<tr>
<td>Curriculum materials</td>
<td>Intcurr</td>
</tr>
<tr>
<td>Classroom management</td>
<td>Intclass</td>
</tr>
<tr>
<td>Cooperative group work</td>
<td>Intcoop</td>
</tr>
<tr>
<td>Work with bullies (e.g. empathy, skills training)</td>
<td>Intbull</td>
</tr>
<tr>
<td>Work with victims (e.g. assertiveness training)</td>
<td>Intvic</td>
</tr>
<tr>
<td>Work with peers (e.g. peer support, mentoring)</td>
<td>Intpeers</td>
</tr>
<tr>
<td>Work with teachers</td>
<td>Intteach</td>
</tr>
<tr>
<td>Work with parents</td>
<td>Intparent</td>
</tr>
<tr>
<td>Improving playground supervision</td>
<td>Intsup</td>
</tr>
<tr>
<td>Disciplinary methods</td>
<td>Intdisc</td>
</tr>
<tr>
<td>Non-punitive methods (e.g. Pikas)</td>
<td>Intnonpun</td>
</tr>
<tr>
<td>Restorative justice approaches</td>
<td>Intrj</td>
</tr>
<tr>
<td>Bully courts, school tribunals</td>
<td>Intcourt</td>
</tr>
<tr>
<td>Other</td>
<td>Int other</td>
</tr>
</tbody>
</table>

Was the intervention highly structured, that is did it follow a protocol or manual? (Yes = 1, No = 2) Struct

Mode of program delivery: (Yes = 1, No = 2)

- Video
- Role play
- Booklets
- Other

Duration of intervention (No. of days): Durint
What time of the year (month) did the program start? Start ______________

What time of the year (month) did the program end? End ______________

Who delivered the intervention? (Yes = 1, No = 2)

- external researchers Delres__________
- teachers Delteach__________
- psychologists or professionals within the school system Delpsych__________
- other (specify ________________________________) Deloth__________

Were there any problems of implementation?
(specify) ________________________________

Was there a measure of treatment integrity? Integ__________

What happened to the control group?

1. No intervention
2. Wait-list control
3. Placebo control
4. Given some information on bullying
5. Management as usual
6. Other (__________________________) Contgp__________

Were any materials delivered to control participants (video, booklets)?
(Yes = 1, No = 2) Contdel__________

7. Post-test measures

Follow-up time period (No. of days):
Short FU__________
Long FU__________

What measures were used? (Yes = 1, No = 2)

<table>
<thead>
<tr>
<th></th>
<th>Short FU</th>
<th>Long FU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-report questionnaires</td>
<td>SSRQ</td>
<td>LSRQ</td>
</tr>
<tr>
<td>Teacher ratings</td>
<td>STR</td>
<td>LTR</td>
</tr>
<tr>
<td>Peer ratings</td>
<td>SPR</td>
<td>LPR</td>
</tr>
<tr>
<td>Systematic observation</td>
<td>SSO</td>
<td>LSO</td>
</tr>
<tr>
<td>School records</td>
<td>SSR</td>
<td>LSR</td>
</tr>
<tr>
<td>Other</td>
<td>SOTH</td>
<td>LOTH</td>
</tr>
</tbody>
</table>
Scales:

a) What scale was used to measure bullying in the short follow-up? Ssbul________ 

b) What scale was used to measure bullying in the long follow-up? Lsbul________ 

c) What scale was used to measure victimization in the short follow-up? Ssvic________ 

d) What scale was used to measure victimization in the long follow-up? Lsvic________ 

What reference period was used for bullying? Short FU: Srbul________ 

Long FU: Lrbul________ 

What reference period was used for victimization? Short FU: Srvic________ 

Long FU: Lrvic________ 

8. Effect Size Measures

Prevalence of Bullying

<table>
<thead>
<tr>
<th></th>
<th>Experimental</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of bullies before</td>
<td>EBBef</td>
<td>CBBef</td>
</tr>
<tr>
<td>No. of nonbullies before</td>
<td>ENBBef</td>
<td>CNBBef</td>
</tr>
<tr>
<td>Short: No. of bullies after</td>
<td>EBS</td>
<td>CBS</td>
</tr>
<tr>
<td>Short: No. of nonbullies after</td>
<td>ENBS</td>
<td>CNBS</td>
</tr>
<tr>
<td>Long: No. of bullies after</td>
<td>EBL</td>
<td>CBL</td>
</tr>
<tr>
<td>Long: No. of nonbullies after</td>
<td>ENBL</td>
<td>CNBL</td>
</tr>
</tbody>
</table>

Short: Odds Ratio ORBS________________________ Confidence Interval CIBS________________________

Long: Odds Ratio ORBL________________________ Confidence Interval CIBL________________________

Prevalence of Victimization

<table>
<thead>
<tr>
<th></th>
<th>Experimental</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of victims before</td>
<td>EVBef</td>
<td>CVBef</td>
</tr>
<tr>
<td>No. of nonvictims before</td>
<td>ENVBef</td>
<td>CNVBef</td>
</tr>
<tr>
<td>Short: No. of victims after</td>
<td>EVS</td>
<td>CVS</td>
</tr>
<tr>
<td>Short: No. of nonvictims after</td>
<td>ENVS</td>
<td>CNVS</td>
</tr>
<tr>
<td>Long: No. of victims after</td>
<td>EVL</td>
<td>CVS</td>
</tr>
<tr>
<td>Long: No. of nonvictims after</td>
<td>ENVL</td>
<td>CNVL</td>
</tr>
</tbody>
</table>
### Bullying scores

<table>
<thead>
<tr>
<th></th>
<th>Experimental</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Before</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>EMBB bef</td>
<td>CMB bef</td>
</tr>
<tr>
<td>SD</td>
<td>ESDB bef</td>
<td>CSDB bef</td>
</tr>
<tr>
<td>N</td>
<td>ENOB bef</td>
<td>CNOB bef</td>
</tr>
<tr>
<td><strong>After</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short: Mean</td>
<td>EMBS</td>
<td>CMBS</td>
</tr>
<tr>
<td>SD</td>
<td>ESDBS</td>
<td>CSDBS</td>
</tr>
<tr>
<td>n</td>
<td>ENOBS</td>
<td>CNOBS</td>
</tr>
<tr>
<td>Long: Mean</td>
<td>EMBL</td>
<td>CMBL</td>
</tr>
<tr>
<td>SD</td>
<td>ESDBL</td>
<td>CSDBL</td>
</tr>
<tr>
<td>n</td>
<td>ENOBL</td>
<td>CNOBL</td>
</tr>
</tbody>
</table>

Short: \(d_{BS}\) \_________________________ \ SEBS \______________

Long: \(d_{BL}\) \__________________________ \ SEBL \__________________

### Victimization Scores

<table>
<thead>
<tr>
<th></th>
<th>Experimental</th>
<th>Control</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Before</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>EMVB bef</td>
<td>CMVB bef</td>
</tr>
<tr>
<td>SD</td>
<td>ESDV bef</td>
<td>CSDV bef</td>
</tr>
<tr>
<td>N</td>
<td>ENOV bef</td>
<td>CNOV bef</td>
</tr>
<tr>
<td><strong>After</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short: Mean</td>
<td>EMVS</td>
<td>CMVS</td>
</tr>
<tr>
<td>SD</td>
<td>ESDVS</td>
<td>CSDVS</td>
</tr>
<tr>
<td>n</td>
<td>ENOVS</td>
<td>CNOVS</td>
</tr>
<tr>
<td>Long: Mean</td>
<td>EMVL</td>
<td>CMVL</td>
</tr>
<tr>
<td>SD</td>
<td>ESDVL</td>
<td>CSDVL</td>
</tr>
<tr>
<td>n</td>
<td>ENOVL</td>
<td>CNOVL</td>
</tr>
</tbody>
</table>

Short: \(d_{VS}\) \_________________________ \ SEVS \______________

Long: \(d_{VL}\) \__________________________ \ SEVL \__________________