School Psychology Quarterly
2008, Vol. 23, No. 1, 26-42

1045-3830/08/$12.00 DOI: 10.1037/1045-3830.23.1.26

How Effective Are School Bullying Intervention Programs? A Meta-Analysis of Intervention Research

Kenneth W. Merrell,
Barbara A. Gueldner, and Scott W. Ross
University of Oregon

Duane M. Isava
Anne Arundel County Public Schools

Research on effectiveness of school bullying interventions has lagged behind descriptive studies on this topic. The literature on bullying intervention research has only recently expanded to a point that allows for synthesis of findings across studies. The authors conducted a meta-analytic study of school bullying intervention research across the 25-year period from 1980 through 2004, identifying 16 studies that met our inclusion criteria. These studies included 15,386 K through 12 student participants from European nations and the United States. Applying standard meta-analysis techniques to obtain averaged effect size estimates across similar outcomes, the authors found that the intervention studies produced meaningful and clinically important positive effects for about one-third of the variables. The majority of outcomes evidenced no meaningful change, positive or negative. The authors conclude that school bullying interventions may produce modest positive outcomes, that they are more likely to influence knowledge, attitudes, and self-perceptions rather than actual bullying behaviors; and that the majority of outcome variables in intervention studies are not meaningfully impacted.

Keywords: schools, bullying, intervention, violence prevention, meta-analysis

As a subset of the larger construct of antisocial-aggressive behavior, bullying behavior represents a unique and distinctly defined phenomenon. Bullying is usually defined as repeated acts of aggression, intimidation, or coercion against a victim who is weaker than the perpetrator in terms of physical size, psychological/social power, or other factors that result in a notable power differential (Carney & Merrell, 2001; Smith & Ananiadou, 2003). The key features of bullying include the *intent to harm*, the *repeated aspect of the harmful acts*, and the *power imbalance between bully and victims*. Bullying behavior may be manifest in a variety of ways. In addition to acts of physical aggression, bullying may also be exhibited through acts of relational aggression (i.e., social exclusion or injuring the reputation of another person), as well as verbal harassment or intimidation (e.g., threats, psychological intimidation).

A growing body of research evidence has demonstrated convincingly that bullying is associated with negative outcomes and troubling trajectories for both bullies and their victims. Among many examples in the literature, a range of reviews related to characteristics of bullies and victims can be found in articles by Carney (2000); Kumpulainen, Raesaenen, and Henttonen (1999); Pelligrini (2001), and Rodkin and Hodges (2003); and, in books on this topic edited by Espalage and Swearer (2004); Juvonen and Graham (2001), and Sharp and Smith (1994).

Although the research outcomes in this area are complex, some of the more frequently stated descriptive findings are worth considering. Bullies tend to have poorer academic skills and grades than the majority of their classmates, often are lacking in the characteristic of empathy, and may have cognitive distortions and social perception biases related to perceived threats in their environment and with respect to how aggression is viewed as an effective way to solve problems. Bullies also tend to be at heightened risk for substance use and later crim-

Kenneth W. Merrell, Barbara A. Gueldner, and Scott W. Ross, Department of Special Education and Clinical Sciences, University of Oregon; Duane M. Isava, Anne Arundel County Public Schools, MD.

Correspondence concerning this article should be addressed to Kenneth W. Merrell, School Psychology Program, 5208 University of Oregon, Eugene, OR 97403-5208. E-mail: kmerrell@uoregon.edu

inal behavior, are likely to become increasingly unpopular with peers as they get older, and tend to come from homes where there is poor parental role modeling in the form of coercive and aggressive means of problem solving and a lack of consistent and effective discipline. Some studies have shown that bullies are often physically larger than their peers, especially in the early grades. Although these characteristics border on creating a caricature of the schoolyard bully that is deeply embedded in the American culture and media (e.g., the "Nelson Muntz" character from the popular TV program *The Simpsons*, an archetypical bully, budding sociopath, and misunderstood soul), they have also held up relatively well in descriptive studies.

Likewise, there has been consistency in the findings of descriptive research on victims of bullying, who tend to be physically smaller or weaker in some other way than the perpetrators, and are often anxious, fearful, insecure, depressed, and have poor self-esteem. A high percentage of victims tend to engage in school avoidance behaviors, and many repeated victims of bullying at school end up dropping out of the school system. Victims are also more likely than perpetrators to bring weapons to school, for the purpose of revenge. A disturbing element of some of the high profile school shootings in the United States during the past few years has been that some of these youthful shooters were repeat victims of bullying and peer harassment, were unpopular, and they ultimately went on a shooting spree as a way of exacting revenge.

Although bullying may occur in almost any context or setting where people gather and interact, schools have been the most frequently studied environment in which bullying occurs. Because school environments provide a microcosm of sorts of the broader society and culture, and because schools are the only setting in which almost all children and adolescents participate, they provide an ideal naturalistic laboratory in which to study bullies, victims, and bullying behavior; to develop bullying prevention and intervention programs, and to investigate the effectiveness of these programs. Schools are also of special interest as research and innovation sites because of social justice concerns, and because of the developmental trajectory of bullying, which is found at all ages, but tends to peak during the middle school years (Hazler, 1996; Rios-Ellis, Bellamy, & Shoji, 2000). Bullying and other forms of antisocial-aggressive behavior in American schools are such a significant public concern that federal initiatives such as *No Child Left Behind* have specifically identified school safety and acts of aggressive behavior as data collection and reporting targets.

Although there has been a notable increase in concern and interest in school bullying within the United States in recent years, it is a mistake to think that this increasing focus reflects a new problem, or a problem that is uniquely American. To the contrary, school bullying has been a concern for generations—perhaps for as long as there have been school systems—and is an international phenomenon (Carney & Merrell, 2001). With respect to research and innovations in prevention and intervention efforts for bullying, American educators and mental health professionals have been relatively recent players at the international table. With a few exceptions, the major impetus on research and prevention-intervention innovation related to bullying has been led by researchers and practitioners from Europe, Canada, and Australia, where the problem has been studied and addressed directly for several decades (Swearer & Espelage, 2004).

Published research on school bullying intervention efforts has certainly lagged behind the more voluminous literature in which the problem is described and analyzed. The first widely disseminated published research on school bullying interventions stemmed from the pioneering work of Norwegian researcher Dan Olweus in the 1970s (e.g., Olweus, 1978), whose anti-bullying prevention/intervention program served as the prototype for most efforts that were developed during the 1980s and 1990s, and still exerts great influence on contemporary intervention models and programs. The research base on bullying intervention programs and practices has slowly accumulated during the past two decades. Although this body of intervention research is not yet plentiful, it has recently reached a size and level of sophistication to begin to allow for evaluative synthesis of the outcomes that have been identified.

When making public policy and educational programming decisions, it is desirable to draw conclusions from a collective body of evidence rather than from individual studies conducted in isolation (Cooper & Hedges, 1994). Given that

the body of research evidence on school bullying interventions is finally beginning to reach a critical enough mass to allow for systematic review of the collective body of work, we conducted this meta-analysis. In our background research, we located only one published meta-analytic review of bullying interventions, a recent study by Smith, Schneider, Smith, and Ananiadou (2004), which focused on whole school antibullying programs. The study by Smith and colleagues, who concluded that the majority of programs evaluated yielded nonsignificant outcomes on self-report measures of bullying and victimization, is an important step in drawing conclusions from the collective body of research. It is important to note that their effort focused on the important topic of *whole school* antibullying programs. However, many interventions designed to prevent bullying are implemented with small groups of targeted students, in individual classrooms, or in clusters of selected classrooms, rather than in whole schools. Thus, an essential aspect of synthesizing the body of research on antibullying interventions in schools is to evaluate studies targeting smaller sectors of the school environment, as well as studies that focus on whole school interventions. Our aim for the present study was to focus our meta-analytic review of school bullying interventions broadly and on a range of intervention environments and conditions, using an international sample of studies from a 25-year-period.

Method

Selection of Studies

Psychological and educational publication databases were searched using the computer-assisted location tools PsycINFO and Educational Resources Information Center (ERIC). Descriptions of published research studies, dissertation abstracts, and related research documents (e.g., chapters in edited books) for the 25-year period from 1980 through 2004 were retrieved using the following key words or combinations of these words: *bullying*, *intervention(s)*, *peer victimization*, *schools*, *programs.* The reference sections of publications obtained through this search also were inspected to locate additional studies not otherwise identified through the computer-assisted search. The cut-off point of 1980 was selected because most of the empirical research on bullying interventions has been published after this time, and because we wished to limit the meta-analysis to a relatively manageable period of 25 years after which this school bullying phenomenon had been adequately defined and described.

During the course of our search for potential articles, abstracts were initially screened based on three criteria. First, the abstract had to indicate that a school-based intervention was conducted. Second, this intervention had to include bullying behavior (or peer-related antisocial behavior that clearly could be interpreted as bullying) as a focus of the intervention. Third, the study had to be available in English. The initial screening procedure yielded 40 studies. The manuscripts of these studies were retrieved and examined in greater detail.

Studies were included in our analysis based on five primary criteria. First, each study had to evaluate the effectiveness of a school-based intervention using an experimental or quasi-experimental group design. Single case, descriptive, and qualitative studies were excluded from the analysis. Second, the intervention had to address bullying behavior. Because bullying behavior and its effects include students in a variety of roles related to bullying (i.e., bully, bystander, defender, enforcer, victim), the intervention had to address at least one of these roles. Third, the research report had to indicate that intervention for bullying behavior was either the primary focus of an intervention program or a main component. Although more general violence prevention and intervention programs are commonly used in schools, we were interested in interventions that specifically identified bullying behavior and/or victimization as primary targets for the intervention. Fourth, each study had to report data in a statistical format necessary to calculate an effect size (e.g., means, standard deviation, group sizes, and percentages). Fifth, although the most stringent inclusion criteria might allow only the use of studies that appeared in peer-reviewed scholarly journals, we chose to broaden the scope of inclusion to include studies from retrievable doctoral dissertations, and chapters in edited books. Our decision to expand the inclusion criteria beyond the strict use of peer-reviewed journal articles was made partially to increase the potential number of usable studies

because the database is still relatively small, and partially because these other two sources (dissertations and chapters in edited books), although not necessarily peer-reviewed, are subject to quality control from supervising committees and editors, and thus had an element of external scrutiny beyond the work of the authors.

Using these criteria to review the 40 studies obtained through the initial screening procedure, 16 studies were included for use in the final quantitative meta-analysis. Studies not included in the analysis were rejected because of the following reasons: lack of an experimental or quasi-experimental design (5 studies), bullying was not the focus of the intervention (3 studies), statistical information was not reported in a format from which an effect size could be calculated (12 studies), statistical information was previously reported in a study already included in studies that met our criteria (3 studies), and no prior peer or committee review of any kind was involved (1 study).

Study Characteristics

A coding form was developed that included a brief description of the study, demographic information (number of participants; gender, age, and grade range of participant; and the geographical location of the research), the design of the study, the name and/or type of intervention that was implemented, outcome variables and measures used, statistical information used to calculate effect sizes (means, standard deviations, percentages), and a brief description of the outcomes and conclusions of the study.

A total of 15,386 student participants were included in the 16 studies that we used in our meta-analysis. Although gender representation was not consistently reported, all studies reported either an age range or grade-level range for participants. Each study included child or adolescent participants or a combination of the two, with the exception of one study that focused exclusively on an adult informant sample (i.e., teachers). The child and adolescent participants ranged from kindergarten through secondary school. With the exception of three studies that measured the effects of interventions on samples that combined primary and secondary level participants, data from primary/elementary (Grades K–6) and secondary grade (Grades 7–12) samples were organized separately for our analysis. Table 1 presents a summary regarding the number of participants for the primary, secondary, combined, and adult sample groups. Note that two of the studies (by Stevens and colleagues) included data from separate breakdowns of elementary and secondary grade students, which allowed us to code each of these studies into two separate sets.

Table 1
Summary of Participants in 16 Studies, by Grade Level or Status

Grade Level or Status	*N*
Primary/elementary students (K–6)	4,786
Secondary students (7–12)	1,889
Combined students (K–12)	8,681
Adult informants (teachers)	30
Total	15,386

Participant samples from the 16 studies were from Europe and North America, with the following six nations represented: Belgium, Canada, United Kingdom, Italy, Norway, and the United States. Table 2 includes a breakdown of the participants in the 16 included studies by nation, as well as the number of studies from each nation. By far, the largest number of participants were from the United Kingdom (6,675, or 44% of the entire meta-analysis sample), followed by Norway and the United States, with approximately 2,500 participants (about 16%) each. The largest number of studies were from the United States (6), followed by the United Kingdom (4). Although we consider this selection of studies to represent a broad and inclusive international sample, we recognize that our necessary use of only studies that were available in English may have reduced the potential pool of studies, and limited the cross-cultural generalizability of our conclusions.

Description of School-based Bullying Interventions

Table 3 presents an overview of the descriptive information for each of the 16 studies included in our review, including the authors names and date of publication, the number of participants, gender distribution, age and grade range, location of study, brief description of the intervention program that was used, measure-

Table 2
Summary of Participants in 16 Studies, Divided by Nation and Number of Studies

Nation	Number of participants	Number of studies
Belgium	1,868	2
Canada	1,473	2
United Kingdom	6,675	4
Italy	289	1
Norway	2,500	1
United States	2,581	6
Total	15,386	16

ment type, dependent variables examined, research design, and outcome summary. Of the 16 studies, 12 used a two-group, prepost, quasi-experimental design. Of these, four studies included experimental and control groups. Also included were three studies implementing a mixed design with two-groups, prepost, and wait-list or control groups. Finally, three of the studies implemented true experimental designs.

Classification of Outcome Measures

Outcome measures (dependent measures) were classified by the primary measurement method that was used to report study outcomes. Five types of measurement methods were used across the 16 studies: *student self-report*, *teacher self-report*, *teacher report of child behavior*, *peer report* (using peer nominations or ratings), and *school discipline records* (using data from the Osiris program). Each study used one or more of these five methods to measure outcomes. For each type of method, classification categories were developed to link or consolidate common dependent variables across studies. This procedure was done to capture the effects that a variety of interventions had on commonly measured dependent variables using similar methods of measurement. For example, several studies examined the effect of a school-based bullying intervention using student self-reports (measurement method) to measure bullying others (dependent variable). A classification variable was created labeled "Bullying Others" and used to combine all the effect sizes calculated across the eight studies that used student self-report to measure the outcome of student self-report of bullying others. In Table 4, where the results of the meta-analysis are presented, the data are organized by the five measurement methods and the classification categories identified for each method. Table 4 also indicates the number of studies measuring dependent variables within a particular measurement method, in addition to the number of effect sizes that were calculated for each classification variable. Some of the particular dependent variables were relatively common (e.g., 10 studies used student self-report to identify "being bullied"), whereas other dependent variables within a particular measurement method were found with less frequency (e.g., only one study used student self-report to identify "global self-esteem").

Data Analysis Procedures

Meta-analytic procedures were used for synthesizing the findings across the many variables and outcomes over the 16 studies. The meta-analysis technique is designed to determine the main potential sources of variability in the program effects, assessing the effects of various moderator variables. There are many specific techniques for defining and deriving meta-analysis results, but most are derived from the techniques advocated in the seminal texts on the subject by Glass, McGaw, and Smith (1981) and Hedges and Olkin (1985). The basic unit of analysis in these procedures is effect size (ES). For prepost intervention studies, ES is defined as the difference between the mean score of an outcome measure at the end of the intervention and the mean score on that measure before the start of the intervention, divided by the pooled or harmonic standard deviation of the of the two scores. For experimental and quasi-experimental studies which include treatment and control groups, ES is defined as the difference between the mean scores of the treatment and control group mean scores following intervention, divided by the pooled harmonic standard deviation of the outcome scores of the two groups.

Many variations on the ES calculation method have been developed since the technique was first widely advocated in the early 1980s. For this study, we used one of the basic mean difference procedures, the *standardized mean difference procedure* described by Lipsey and Wilson (2001). Some of the studies in our review did not include means and standard deviations, but reported results in proportions or

percentages, such as the percentage of students in a classroom who received discipline referrals, before and after treatment, or across treatment and control conditions. To calculate ES for these types of studies in our review, we used a more recent procedure, described by Lipsey and Wilson (2001) as the *proportion difference*. This method involves a computation of the differenced proportions (or percentages) from two observations (prepost or treatment-control), which is then further computed by using the weighted sample sizes upon which each proportion was based. For the few studies in our analysis that required the proportion difference method for ES, we used a The Effect Size Determination Program, computer analysis procedure developed by Wilson, and widely used for this purpose in social science research.

After ES statistics were computed and coded, the averaged effect sizes for each classification variable and method were examined, and when appropriate, converted to either a positive or negative value based on the outcome that would be desirable because of the bullying intervention program. That is, if bullying behavior decreased after the implementation of bullying intervention programs, the averaged ES would be a negative value. However, this is an effect in a desirable direction. To clearly represent this effect, the averaged ES was changed to a positive value to reflect this positive and desired outcome. Likewise, if bullying behavior increased after implementation of intervention programs, the averaged ES was changed from a positive to a negative number to reflect a negative and undesirable outcome. Table 4 presents the effect sizes in their final, converted form, which was standardized to assist readers in interpretation of our results. ES statistics are interpreted as the proportion of a standard deviation unit. For example, an ES of 1.0 would indicate a difference in scores between two observations of approximately one standard deviation.

Results

Table 4 presents the effects of the bullying interventions according to our classification of 28 major outcome types within the five measurement typologies. This table also lists the number of studies in which each outcome type was included, and the number of effect sizes and average effect size per outcome type. In addition, the number of significant positive effects and significant negative effects are listed for each type of outcome.

As noted previously, ES estimates in this meta-analysis were converted into a uniform format (positive or negative valence), and are thus interpreted by their positive or negative value. Again, positive ES values reflect a change in the dependent variable in a desired direction if the bullying intervention was successful in producing the type of change that was intended, whereas negative ES values reflect a change in the dependent variable in an undesired or unexpected direction, and indicate that the bullying intervention was associated with undesirable changes in behaviors, attitudes, and affect. For example, student self-reports of being bullied averaged across 10 studies revealed that overall, students reported a small, meaningful change in the direction desired (ES = .27). That is, students reported a meaningful decrease in being bullied following participation in a bullying intervention program. In contrast, teacher reports of child behavioral and emotional problems from the one study that used this outcome type indicated that teachers reported very large, meaningful change in an undesired direction (ES = −3.81), or a worsening in child behavior and emotional problems following the implementation of a bullying intervention program. Although it is certainly possible that the worsening in teacher-reported behavioral and emotional problems of students could be attributed to something other than the bullying intervention, the computation of effect size indicates an association between the two.

The average effect size for each intervention classification outcome type was evaluated using Cohen's (1988) widely used criteria of .20 being the smallest effect size that has any significant practical or clinical meaning, with three specific categories or levels of power: small (.20 to .49), medium (.50 to .79), and large (.80 and higher). For the average or mean effect sizes across the 28 intervention outcome types, 10 were associated with positive meaningful changes: 4 in the small range, 2 in the medium range, and 4 in the large range. These outcome variables where meaningful or clinically significant average effects were found include student self reports of *being bullied*, *witnessing bullying*, and *global self-esteem*; teacher self-reports

Table 3
Description of Studies Used in Meta-Analysis

Study	*N* (gender)	Age Range (grade levels)	Location	Intervention program	Measurement type	Dependent variable(s)	Research design	Outcome summary
Bagley & Prichard (1998)	1,200 (n.r.)	Ages 5–11 (primary and secondary)	U.K.	Assigned social workers to building	Student self-report	Number of bullying incidents	Mixed design, two group prepost, experimental and control	Significant decreases in bullying among primary students receiving intervention
Cowie & Olafsson (2000)	207 (all boys)	n.r. (high school)	England	Antibullying policy using peer supporters	Survey; staff and student interviews	Self-report of having been bullying, bullied others, reporting bullies	Quasi-experimental, two group prepost	Significant increases in bullying behavior among students receiving intervention
Fox & Boulton (2003)	28 (x boys, x girls)	Ages 9–11 (n.r.)	U.K.	Eight week social skills program	Sociometrics and student self-report	Victimization, social skills problems, number of best friends, peer acceptance, emotional problems, self-esteem, social acceptance	Mixed design, two group prepost wait list control	Students receiving intervention showed significant increases in self esteem, number of friends, anxiety; significant decreases in depression
Kaiser-Ulrey (2003)	125 (60 n.r. boys, 65 girls)	(Grade 7)	USA (Florida)	BEST program: education, empathy training, problem-solving	Student self-report	Bullying, victimization, empathy, prosocial behavior, global self-esteem, parental involvement	Quasi-experimental, two group prepost, with comparison group	Students receiving intervention reported significant increases in victimization
Leadbeater, Hoglund, & Woods (2003)	432 (220 boys, 212 girls)	Age 6 (Grades 1–2)	Western Canada	WITS program: teacher delivered peer victimization prevention program	Student self-report, teacher report	Teacher reported social competence, emotional & behavioral problems; student reported relational and physical aggression	Quasi-experimental, two group prepost	Students receiving intervention showed significant decreases in relational aggression and social competence, significant increases in behavioral and emotional problem symptoms

Table 3 (*continued*)

Study	*N* (gender)	Age range (grade levels)	Location	Intervention program	Measurement type	Dependent variable(s)	Research design	Outcome summary
Menesini, Codecasa, Belenni, & Cowie (2003)	289 (153 boys, 136 girls)	Ages 11–14 (Grades 6–8)	Italy	Peer-support model: befriending intervention	Sociometrics	Roles: bully, reinforcer, assistant, outsider, defender, victim	Quasi-experimental two groups (tx and control)	Significantly fewer peer nominations in bullying roles, significantly more peer nominations in "defender" role
Mueller & Parisi (2002)	28 (n.r.)	n.r. (Grades 3–4)	USA (Midwest)	Anti-bullying incentive-combination of open discussion, role-play, reflective journals	Student self-report	Physical bullying, victimization, and social aggression	Quasi-experimental, two groups prepost	Students receiving intervention displayed significant decreases in bullying behavior
	30 (n.r.)	n.r. (Grades 6–8)	USA (Georgia)	Teacher in service training plus consultation (*Bully Busters program*)	Teacher self-report, Osiris	Teacher knowledge, use of intervention skills, teaching efficacy, discipline referrals	Quasi-experimental, two group prepost	Teachers receiving intervention showed significant increases in knowledge, use of intervention skills, teaching efficacy, and significant decreases in discipline referrals
Newman (1999)								
Olweus (1997)	2,500 (n.r.)	Ages 11–14 (Grades 4–7)	Norway (Bergen)	Olweus program (increase in authoritative adult involvement, warmth, across settings)	Student self-report	Bullying behavior	Quasi-experimental, two group prepost	Students receiving intervention showed significant decreases in victimization and participation in bullying behavior

(*table continues*)

Table 3 (*continued*)

Study	*N* (gender)	Age Range (grade levels)	Location	Intervention program	Measurement type	Dependent variable(s)	Research design	Outcome summary
Orpinas, Horn, & Staniszeski (2003)	520 (254 boys, 266 girls)	n.r. (Grades K-5)	USA (Southeast)	Collaborative school-wide program; positive environment, education, staff training	Student self-report	Aggression, victimization	Quasi-experimental, two group prepost	Students in Grades K-2 who received intervention showed significant decreases in aggression and victimization; students in Grades 3–5 showed significant decreases in victimization only
Pepler, Craig, Ziegler, & Charach (1994)	1,041 (502 boys, 539 girls	Ages 5–14 (n.r.)	Canada (Toronto)	Toronto Antibullying Intervention (multi-level across schools, parents, class/peers, individuals)	Student self-report, teacher report of student	Teachers talking about and/or stopped bullying (school level), parents talked to bullies and victims about bullying (parent level), peers stopped bullying (peer level), own bullying behavior (individual level: bullies), recipients of bullying (individual level: victims)	Quasi-Experimental, two group prepost	Students who received intervention were significantly more likely to engage bullying behavior
Stevens, Van Oost, & de Bourdeaudhuij (2000a)	323 (n.r.)	Ages 10–12 (primary)	Belgium	Social cognitive curriculum	Student self-report	Probully score, provictim score, self-efficacy, intention, behavior	Quasi-Experimental, two groups (experimental and control)	Students receiving intervention were significantly less likely to have probully attitudes

Table 3 (*continued*)

Study	*N* (gender)	Age Range (grade levels)	Location	Intervention program	Measurement type	Dependent variable(s)	Research design	Outcome summary
Stevens, Van Oost, & de Bourdeaudhuij (2000b)	441 (n.r.)	Ages 13–16 (secondary)	Belgium	Social cognitive curriculum	Student self-report	Probully score, provictim score, self-efficacy, intention, behavior	Quasi-Experimental, two groups (experimental and control)	Students receiving intervention were significantly more competent in dealing with bullying, had greater intention to react to bullying, and were more likely to react to bullying
Stevens, de Bourdeaudhuij, Van Oost (2000a)	392 (n.r.)	n.r. (primary)	Belgium	Flemish antibullying intervention (authoritative focus on social system, curriculum, and peer group)	Student self-report	Bullying, being bullied, positive peer interactions	Experimental with three groups: with support, without support, and control	Students receiving intervention were significantly less likely to bully others, be victimized, and have positive interactions when treatment included support, but significantly more likely to bully others, be victimized, and have positive interactions when treatment did not include support
Stevens, de Bourdeaudhuij, & Van Oost (2000b)	712 (n.r.)	n.r. (secondary)	Belgium	Flemish antibullying intervention	Student self-report	Bullying, being bullied, positive peer interactions	Experimental with 3 groups: with support, without support, and control	No significant effects

(*table continues*)

Table 3 (*continued*)

Study	*N* (gender)	Age Range (grade levels)	Location	Intervention program	Measurement type	Dependent variable(s)	Research design	Outcome summary
Turpeau (1998)	115 (66 boys, 49 girls)	n.r. (Grades 6–8)	USA (Georgia)	Antibullying classroom component from *Bullyproofing Your School*	Sociometrics, student self-report, teacher report, Osiris	Student victimization and aggression	Mixed design, two group prepost with waitlist control	No significant effects
Whitaker, Rosenbluth, Valle & Sanchez (2003)	1,763 (50.3% boys, 48.3% girls)	n.r. (Grade 5)	USA (Texas)	*Expect-Respect Project,* theoretically	Student self report, teacher report based on Olweus program, and adapted 12 sessions from Bully-Proofing program	Bullying, student response to bullying, student intended actions, staff typical actions, student perceptions of staff actions, staff perception of other staff reactions, staff perception of student reactions, student and staff awareness student and staff attitudes	Experimental	Students receiving intervention were significantly more likely to proactively respond to physical bullying
Whitney, Rivers, Smith, & Sharp (1994)	5,140 (n.r.)	n.r. (elementary to middle school)	England (Sheffield)	One day staff meeting with follow up, including optional teacher-selected interventions	Student self-report	Bullying, Victimization, Refusing to join in bullying, reporting to teachers or others about bullying, witnessing teacher stopping bullying	Quasi-experimental, two group prepost	No meaningful effects

Table 4
Results of Meta-Analysis: Summary of Effect Sizes by Assessment Method and Classification Variable

Method/classification variables	No. of studies	No. of effect sizes	Average effect size	No. of significant positive effects	No. of significant negative effects
Student self-report					
Bullying others	8	11	.04	1	2
Positive attitude toward bullying	4	6	.15	3	0
Being bullied	10	14	.27	6	2
Witnessed bullying	3	7	.35	5	0
Intervene to stop bullying	3	10	.17	6	0
Bullies were talked to by adult	2	2	−.04	0	0
Teacher action/response	3	4	.06	0	0
Ignore/refuse to join bullying	4	5	.06	1	0
Reported bullying, or likely to	4	10	.07	1	0
Feeling safe at school	1	1	−.13	0	1
Feelings of anxiety/depression	1	2	−.06	1	0
Global self-esteem	1	1	1.08	1	0
Social skills	1	1	.06	0	0
Sympathy for victims	2	2	−.10	0	0
Positive interactions with peers	4	4	−.10	0	1
Teacher self-report					
Witnessed students being bullied	1	1	−.16	0	0
Knowledge of bullying prevention	1	1	1.52	1	0
Staff appropriate responses to bullying	2	5	.30	1	0
Efficacy of intervention skills	1	2	.99	2	0
Attitude about school safety	1	1	−.16	0	0
Teacher report of child behavior					
Student behavior/emotional problems	1	2	−3.81	0	2
Student social competence	1	1	3.31	1	0
Peer report					
Participation in bullying roles	1	4	.32	4	0
Identified victims	2	2	−.03	0	0
Identified aggressors	3	3	.04	1	0
Peer acceptance	1	3	.61	3	0
Social skills problems	1	1	.16	0	0
School records					
Osiris teacher discipline referrals	1	1	.79	1	0

of *knowledge of bullying prevention*, *appropriate staff responses to bullying*, and *efficacy of intervention skills*; teacher reports of *student's social competence*; peer reports of *participation in bullying roles* and *peer acceptance*; and, school records of *teacher discipline referrals*. Again, the 16 bullying intervention studies we reviewed for this meta-analysis showed evidence of meaningful changes in several educational, social, and behavioral outcomes associated with the interventions.

Conversely, 1 of the 28 average effect sizes (teacher report of student behavioral and emotional problems, ES = −3.81) was associated with negative meaningful change, or a worsening of the outcome. However, it is important to note that this particular outcome type was used in only one study, and was based on an average of only two effects. In addition, the majority of average effect sizes by intervention type—17 of the 28 we identified for this study—were not strong enough to be considered meaningful or practically significant, regardless of their positive or negative valence.

Although the meta-analysis technique is generally based on averages or arithmetic means of outcomes across aggregated studies, it is important to recognize that means can sometimes be misleading, as they may be obscured by outliers, they are highly influenced by the number of

values that are summed and divided to obtain the mean, and they may be minimized or maximized by other factors such as the regression phenomenon. For this reason it is worthwhile to consider the actual number (rather than the mean) of significant positive and negative effects across each of the 28 intervention outcome types in our analysis, particularly in terms of their relation to the total number of effect sizes identified for each outcome type.

By totaling the columns in Table 4 to indicate number of effect sizes, number of significant positive effects, and number of significant negative effects, a broader view of the findings across these bullying intervention studies can be obtained. For the 15 intervention outcome types that were classified through student self-report, 80 effect sizes were computed. Of these, 25 were considered to be significant positive effects, and 6 were considered to be significant negative effects. The majority of these effect sizes that we derived in this domain—49 out of 80—were not strong enough to be considered meaningful or significant. For the five intervention outcome types that were classified through teacher-self report, 10 effect sizes were computed. Of these, four were significant positive effects, and none of the negative effects was considered significant. Again, the majority of effect sizes we computed (6 out of 10) were not strong enough to be considered meaningful. For the two-intervention outcome types that were classified through teacher report of child behavior, three effect sizes were computed. Of these, one was significant and positive, whereas the other two were both significant and negative. In the peer report domain, 13 effect sizes were computed. Of these, eight were positive and meaningful, none was negative and meaningful, and six were not strong enough to be considered meaningful. For the school records outcome domain, only one effect was computed in the one study that included this domain, and it was both positive and significant. Overall, across all 28-intervention outcome domains, 107 effect sizes were computed. Of these, 39 were positive and meaningful, 8 were negative and meaningful, and the remainder (60) were not strong enough to be considered meaningful. In this analysis of individual effect sizes rather than means, positive meaningful effects outnumber negative meaningful effects by a wide margin (about a 5:1 ratio), but the majority of the effects were not strong enough to meet Cohen's (1988) minimum criterion of .20 to be considered meaningful and clinically important or significant.

Discussion

To summarize briefly the results of this meta-analysis, meaningful positive average effects for school bullying interventions across the 16 studies were found for slightly more than one third of the outcome classification variables (10 out of 28, or 36%), and for slightly more than one-third of the individual effect sizes within the studies (39 out of 107, or 36%). There was no apparent pattern to these significant effects, which seemed to be dispersed somewhat uniformly across types of measurement methods and classification variables, and across various types of school bullying interventions. These results lead us to conclude—somewhat tentatively—that there is some evidence supporting the effectiveness of school bullying interventions in enhancing students social competence, self-esteem, and peer acceptance; in enhancing teachers knowledge of effective practices, feelings of efficacy regarding intervention skills, and actual behavior in responding to incidences of bullying at school; and, to a lesser extent, in reducing participation by students in bully and victim roles.

The tentative nature of our conclusions is because of the fact that the majority of average effects for the school bullying interventions across studies (17 of 28, approximately 60%) were too weak to be considered meaningful, as were the majority of individual effect sizes within studies (60 of 107, approximately 56%). This aspect of our findings indicates that although a substantial percentage of intervention outcomes linked with the school bullying interventions were associated with meaningful positive impact, such instances did not constitute a plurality of intervention effects. Rather, the majority of intervention effects did not evidence sufficient power to be considered meaningful or clinically important. If our findings were generalized to the overall state of school bullying intervention efficacy, then we could conclude that such interventions are likely to produce some important and meaningful positive effects, but the majority of the effects that will be mea-

sured will lack the power to be considered meaningful.

Our finding that a small minority of intervention outcomes were associated with significant negative effects (1 out of 28 mean effects across studies, or slightly less than 4%; 8 out of 107 individual effects within studies, or about 7%) is difficult to interpret, and somewhat perplexing. It is indeed possible that some well-intended interventions may actually produce adverse or negative effects with students, particularly interventions that group together deviant peers for treatment (Dishion, McCord, & Poulin, 1999). Given that the evidence we found for such negative effects constituted a very small proportion of the overall intervention effects, our view is that interventionists and educational policymakers should not conclude that school-bullying interventions would harm students. Especially given that the only average effect that was negative and meaningful was derived from a single study (Leadbetter, Hoglund, & Woods, 2003), and that this same study also found positive and meaningful effects in some areas, it is possible that the negative findings were sample-specific or could be attributed to something other than the intervention. However, these somewhat troubling findings do not preclude the possibility of harm. Researchers and practitioners should use high standards of care and always consider the possibility that their actions may produce unanticipated negative effects in addition to positive consequences that are anticipated.

Related interesting limitations to consider involve research designs and types of measurement methods. In our meta-analysis, antibullying studies produced meaningful effects for 39% of the outcome variables measured, with the remaining majority producing no meaningful change. Of the 16 studies evaluated, 13 implemented quasi-experimental or mixed two-group prepost designs, and many of these did not include a control group. Thus, only 3 of the 16 studies implemented true experimental designs. The use of nonexperimental designs, while admittedly more feasible to implement in school-based research, raises *history* as a specific threat to internal validity, and points to the possibility of something other than the intervention leading to the program results. This possibility is especially plausible in bullying prevention programs that begin at the beginning of the school year or at designated calendar changes, because of the multitude of other changes that schools may be made at these times of the year (e.g., curriculum, scheduling, classroom assignments, etc.).

A related concern is that studies on the effects of bullying interventions focus primarily on *indirect measures* of the behavior. Student self reports, teacher self reports, and sociometrics were primary measures used in all but one of the studies, which often accounts for what the participants know about bullying rather than how often they actually engage in the behavior. For example, the strongest effect sizes we found were for student social competence (ES = 3.31), knowledge of bullying prevention (ES = 1.52), and global self esteem (ES = 1.08). Each of these variables measured how well participants *knew* the intervention program and how they *thought* they would respond to bullying, instead of how they actually *did* respond.

On more direct measures or reports of bullying and victimization, bullying interventions had little positive effects, and as we have noted, in some cases, actually were associated with negative effects. Student self reports of instances of bullying produced an average effect size of .04, whereas self reports of instances of being bullied produced an average effect size of .27, and teacher reports of bullying incidents produced an average effect size of −.16. Viewed in this manner, the average teacher actually reported more bullying after intervention than before! These findings may be attributable to students and teachers learning how to better recognize bullying through the bullying programs, and then acknowledging the behavior more often because of their increased knowledge.

Our results are best understood as one research synthesis within the larger body of research evidence on school bullying interventions. The practice and science in this area has only recently grown and matured to the point were such synthesis studies are even possible. In commenting on the state of research on school bullying, Swearer and Espelage (2004) noted that there are a number of societal and institutional barriers to conducting intensive bullying intervention research in school, particularly within American schools. The difficulty of designing and conducting good intervention research on school bullying, coupled with these social and institutional barriers to implementing

such research, have combined to make good research on this topic surprisingly hard to come by. The fact that we located only 16 intervention studies over a quarter of a century of published work that was strong enough to meet our inclusion criteria is troubling.

As we indicated previously, the only other meta-analysis on school bullying interventions that we have found is a recent review of whole school antibullying programs by Smith and colleagues (2004), who concluded that although some intervention studies yielded positive outcomes, "the majority of programs evaluated to date have yielded nonsignificant outcomes on measures of self-reported victimization and bullying" (p. 547). Their study, which was based on 14 research publications, included a few of the same sources that we included in our analysis, but their focus was more narrowly defined than ours, as their emphasis was on whole-school intervention programs. Our review, which included interventions aimed at individual classrooms and across schools as well as within whole-schools studies, also found that the majority of effects within and across studies were not significant. However, we found a somewhat stronger pattern of meaningful positive effects across studies than did Smith and colleagues, which may be attributed to our literature base being broader, and not limited to whole-school programs, which are quite difficult to implement. Our complementary findings lead us to agree with the assertion of these researchers that "despite the limited empirical support for the effectiveness of antibullying programs, there is not sufficient evidence to conclude that such programs should be abandoned. . .the overarching message is that intervention can succeed, but not enough is known to indicate exactly how and when" (p. 558).

There are some limitations to consider in interpreting our results and conclusions, and in generalizing these findings to the broad field of school-based bullying and violence prevention. Foremost among these limitations is that the number of studies we identified for inclusion in our analysis (16) was relatively modest, and perhaps more importantly, these studies varied widely in terms of research design, intervention models, and intensity of intervention. Although we found some notable trends and were able to draw some meaningful conclusions regarding effectiveness of school bullying interventions, our findings and conclusions must be tempered with the reality that the literature in this area has only recently reached a level of volume and sophistication that even allows for a useful meta-analysis to be conducted. Thus, our findings should be considered, along with those of Smith and colleagues (2004), as a starting point in what we hope will be continued efforts to synthesize the research in this area.

On a similar note, the widely varying bullying intervention models that were used in the studies included in our meta-analysis make drawing specific conclusions regarding "which intervention is best" a question that will need to be answered in the future, when science in this area reaches a state of complexity that will allow such questions to be explored with integrity. Considering that the interventions utilized in the studies we analyzed ranged from assigning a social worker to a school building to multilevel interventions across school buildings within a system, the reality is that the school bullying interventions that we evaluated reflect a substantial range of differing intervention theories and techniques, and that it is incorrect to view school bullying interventions as a common entity.

An additional limitation to consider is that we chose to not weight the 16 studies in the meta-analysis for sample size, degree of experimental rigor, or threats to validity when we computed effect sizes within the individual research studies. Such weighting procedures are sometimes used as a way to reduce the impact of redundancy, overweight of estimates, and occasional inverse relation of treatment effects to sample size (Hedges & Olkin, 1985; Lipsey & Wilson, 2001; Weisz, Weiss, Han, Granger, & Morton, 1995). Because of the widely varying types of intervention programs that were used across the 16 studies, and because one particular study (Whitney, Rivers, Smith, & Sharp, 1994) accounted for more than one third of all student participants across the studies, we decided against weighting, considering that the disadvantages to doing so would outweigh any potential advantages that weighting might offer. Again, we consider this meta-analysis to be an initial exploratory study, and our hope is that the scientific literature related to school bullying prevention and intervention will continue to grow to the point where future efforts might consider differences in intervention type, valid-

ity threats, and sample sizes in a meaningful way.

In summary, we conclude that the research we synthesized within our meta-analysis indicates that school-bullying interventions may produce some clinically meaningful effects on students, teachers, and school staff. However, we are not able to frame the minority of positive findings with an emphatic or dramatic spin, because most of the studies we analyzed were limited by some important issues and questions related to experimental design, and types of measurement methods that were utilized. Although antibullying interventions appear to be useful in increasing awareness, knowledge, and self-perceived competency in dealing with bullying, it should not be expected that these interventions will dramatically influence the incidence of actual bullying and victimization behaviors, or that they will positively influence even a majority of the targeted outcomes. In fact, our evidence indicates that the majority of targeted outcomes in school bullying interventions may not be significantly impacted, either positively or negatively. The continued search for effective bullying interventions can be informed by the results of this meta-analysis, but should not be limited strictly to interventions that are labeled "antibullying programs." Rather, this search should follow the data, and consider behavioral interventions that are universal in nature as well as those that target the specific problems associated with bullying in schools (Crone, Horner, & Hawken, 2004; Sprague & Walker, 2005). Future intervention studies which use solid experimental designs, and which measure impact on actual bullying behaviors as well as perceptions and knowledge, are particularly needed.

References

*References marked with an asterisk indicate studies included in the meta-analysis.

*Bagley, C., & Prichard, C. (1998). The reduction of problem behaviours and school exclusion in at-risk youth: An experimental study of school social work with cost—Benefit analyses. *Child and Family Social Work, 3,* 219–226.

Carney, A. G., & Merrell, K. W. (2001). Bullying in schools: Perspectives on understanding and preventing an international problem. *School Psychology International, 22,* 364–382.

Carney, J. V. (2000). Bullied to death: Perceptions of peer abuse and suicidal behavior during adolescence. *School Psychology International, 21,* 213–223.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Mahwah, NJ: Erlbaum.

Cooper, H., & Hedges, L. V. (1994). Research synthesis as a scientific enterprise. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 3–14). New York: Russell Sage Foundation.

*Cowie, H., & Olafsson, R. (2000). The role of peer support in helping the victims of bullying in a school with high levels of aggression. *School Psychology International, 21,* 79–94.

Crone, D. A., Horner, R. H., & Hawken, L. S. (2004). *Responding to problem behavior in schools: The behavior education program.* New York: Guilford Press.

Dishion, T. J., McCord, J., & Poulin, F. (1999). When interventions harm: Peer groups and problem behavior. *American Psychologist, 54,* 755–764.

Espelage, D. L., & Swearer, S. M. (Eds.). (2004). *Bullying in American schools: A social-Ecological perspective on prevention and intervention.* Mahwah, NJ: Erlbaum.

*Fox, C. L., & Boulton, M. J. (2003). Evaluating the effect of a social skills training (SST) programme for victims of bullying. *Educational Research, 45,* 231–247.

Glass, G. V., McGaw, B., & Smith, M. L. (1981). Meta-analysis in social research. *Educational Researcher, 5,* 3–8.

Hazler, R. J. (1996). *Breaking the cycle of violence: Interventions for bullying and victimization.* Washington, DC: Accelerated Development.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* San Diego: Academic Press.

Juvonen, J., & Graham, S. (Eds.) (2001). *Peer harassment in schools: The plight of the vulnerable and victimized.* New York: Guilford Press.

*Kaiser-Ulrey (2003). *Bullying in middle school: A Study of B. E. S. T.—Bullying Eliminated from Schools Together—An anti-bullying program for seventh grade students.* Unpublished doctoral dissertation, Florida State University, Tallahassee.

Kumpulainen, K., Raesaenen, E., & Henttonen, I. (1999). Children involved in bullying: Psychological disturbance and the persistence of involvement. *Child Abuse and Neglect, 23,* 1253–1262.

*Leadbetter, B., Hoglund, W., & Woods, T. (2003). Changing contexts? The effects of a primary prevention program on classroom levels of peer-relational and physical victimization. *Journal of Community Psychology, 31,* 397–418.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis* (applied social research methods series, v. 49). Thousand Oaks, CA: Sage.

*Mensini, 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.

*Mueller, E. E., & Parisi, M. J. (2002). *Ways to minimize bullying.* Unpublished master's thesis, Saint Xavier University, Chicago.

*Newman, D. A. (1999). *The effectiveness of a psychoeducational intervention for classroom teachers aimed at reducing bullying behavior in middle school students.* Unpublished doctoral dissertation, University of Georgia, Athens.

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

*Olweus, D. (1997). Bully/victim problems at school: Facts and intervention. *European Journal of Psychology of Education, 12,* 495–510.

*Orpinas, P., Horne, A. M., & Staniszewski, D. (2003). School bullying: Changing the problem by changing the school. *School Psychology Review, 32,* 431–444.

Pelligrini, A. (2001). The roles of dominance and bullying in the development of early heterosexual relationships. *Journal of Emotional Abuse, 2,* 63–73.

*Pepler, D. J., Craig, W. W., Ziegler, S., & Charach, A. (1994). An evaluation of an anti-bullying intervention in Toronto schools. *Canadian Journal of Community Mental Health, 13,* 95–110.

Rios-Ellis, B., Bellamy, L., & Shoji, J. (2000). An examination of specific types of *ijime* within Japanese schools. *School Psychology International, 21,* 227–241.

Rodkin, P. C., & Hodges, V. E. (2003). Bullies and victims in the peer ecology: Four questions for psychologists and school professionals. *School Psychology Review, 32,* 384–400.

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

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

Smith, P. K., & Ananiadou, K. (2003). The nature of school bullying and the effectiveness of school-based interventions. *Journal of Applied Psychoanalytic Studies, 5,* 189–209.

Sprague, J. R., & Walker, H. M. (2005). *Safe and healthy schools: Practical prevention strategies.* New York: Guilford Press.

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

*Stevens, V., Van Oost, P., de Bourdeaudhuij, I. D. (2000). The effects of an anti-bullying intervention programme on peers' attitudes, and behaviour. *Journal of Adolescence, 23,* 21–34.

Swearer, S. M., & Espelage, D. L. (2004). Introduction: A social-ecological framework of bullying among youth. In D. L. Espelage and S. M. Swearer (Eds.), *Bullying in American schools: A social-ecological perspective on prevention and intervention* (pp. 1–12). Mahwah, NJ: Erlbaum.

*Turpeau, A. M. (1998). *Effectiveness of an antibullying classroom curriculum intervention on an American middle school.* Unpublished doctoral dissertation, University of Georgia, Athens.

Weisz, J. R., Weiss, B., Han, S. S., Granger, D. A., & Morton, T. (1995). Effects of psychotherapy with children and adolescents revisited: A meta-analysis of treatment outcomes. *Psychological Bulletin, 117,* 450–468.

*Whitaker, D. J., Rosenbluth, B., Valle, L. A., & Sanchez, E. (2003). 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., & Sharp, S. (1994). The Sheffield project: Methodology and findings. In P. Smith & S. Sharp (Eds.), *School bullying: Insights and perspectives* (pp. 20–56). London: Routledge.