

A Meta-Analysis of School-Based Bullying Prevention Programs' Effects on Bystander Intervention Behavior

Joshua R. Polanin
Loyola University Chicago

Dorothy L. Espelage
University of Illinois Urbana—Champaign

Therese D. Pigott
Loyola University Chicago

Abstract. This meta-analysis synthesized bullying prevention programs' effectiveness at increasing bystander intervention in bullying situations. Evidence from 12 school-based programs, involving 12,874 students, indicated that overall the programs were successful (Hedges's $g = .20$, 95% confidence interval [CI] = .11 to .29, $p < .001$), with larger effects for high school (HS) samples compared to kindergarten through eighth-grade (K-8) student samples (HS effect size [ES] = 0.43, K-8 ES = 0.14; $p < .05$). A secondary synthesis from eight of the studies that reported empathy for the victim revealed treatment effectiveness that was positive but not significantly different from zero ($g = .05$, 95% CI = $-.07$ to $.17$, $p = .45$). Nevertheless, this meta-analysis indicated that programs increased bystander intervention both on a practical and statistically significant level. These results suggest that researchers and school administrators should consider implementing programs that focus on bystander intervention behavior supplementary to bullying prevention programs.

Bullying perpetration often occurs when bystanders are present (Hawkins, Pepler, & Craig, 2001; Lagerspetz, Bjorkqvist, Bertz, & King, 1982). In fact, some research has indicated that more than 80% of the time an observer witnesses victimization (O'Connell, Pepler, & Craig, 1999). Despite the presence of witnesses and bystanders, nearly 1 in 3 children report victimization by a bully in the past 2 months (Frey, Hirschstein, Edstrom, & Snell, 2009; Nansel et al., 2001; Wang, Ian-

notti, & Nansel, 2009). Consequentially, bullying occurs with an audience of members who play multiple roles (Salmavalli, Lagerspetz, Bjorkqvist, Osterman, & Kaukianen, 1996) and often fail to intervene on behalf of the victim with regularity. These bullying incidents have lasting negative effects on the bully, victim, and bystanders (Olweus, 2002; Swearer, Espelage, Villancourt, & Hymel, 2010; Sweeting, Young, West, & Der, 2006; Stevens, Oost, & Bourdeaudhuij, 2004).

Correspondence regarding this article should be addressed to Joshua R. Polanin, Loyola University Chicago, 820 N. Michigan Ave., Suite 1022C, Chicago, IL 60611; e-mail: jpolanin@luc.edu

Copyright 2012 by the National Association of School Psychologists, ISSN 0279-6015

The past 20 years have seen a burgeoning of bullying prevention programs (Ferguson, Miguel, Kilburn, & Sanchez, 2007; Ryan & Smith, 2009; Ttofi & Farrington, 2009, 2011). Researchers, school administrators, and teachers have used myriad designs, theories, and techniques in an attempt to mitigate the prevalence of bullying (Astor, Meyer, Benbenishty, Marachi, & Rosemond, 2005). Ttofi and Farrington's (2011) recent large-scale meta-analysis of over 90 studies found that the majority of these programs have been successful at slowing the rate of bullying.

Although successful bullying programs remain important accomplishments, Ttofi and Farrington (2011) found that few programs specifically target the behavior of bystanders (i.e., an individual who witnesses bullying). As such, prevention programs deemphasize a population that constitutes between 60% and 70% of primary or secondary school students (Glew, Fan, Katon, Rivara, & Kerntic, 2005; Rivers, Poteat, Noret, & Ashurst, 2009). This program oversight is unfortunate because observational research has found that when bystanders intervene on behalf of the victim, they successfully abate victimization more than 50% of the time (Craig, Pepler, & Atlas, 2000; O'Connell et al., 1999).

Supported by the knowledge that bystanders can successfully intervene on behalf of the victim, a small amount of literature has focused recently on increasing this behavior. These programs explicitly emphasize the importance of bystander intervention behavior and measure this construct. Given these conditions, the purpose of this meta-analysis is to synthesize school-based bullying prevention programs' effectiveness to change bystander intervention behavior. We also aggregated the program's influence on empathy for the victim as a secondary synthesis because it has received recent investigation (Gini, Albiero, Benelli, & Altoè, 2007). The following summarizes the relevant literature, provides a comprehensive examination of the synthesis process and quantitative analysis outcomes, and elucidates moderator analysis and publication bias results. Suggestions for future research and policy are also provided.

Bullying in the Schools

Definition

Olweus (1973) first described bullying as "mobbing" where a group or individual teases or harasses another individual. As such, early research focused solely on the physical aspects of school environment (e.g., teacher-student ratio), but found little connection to perpetration or victimization (Swearer et al., 2010). Recently, Frey et al. (2009) described bullying as a social construct that disrupts social connections among students. Ross and Horner (2009) summarized the plethora of definitions:

Common definitions of *bullying* involve repeated acts of aggression, intimidations, or coercion against a victim who is weaker in terms of physical size, psychological or social power, or other factors that result in a notable power differential. (p. 748)

The bully construct has received much research and hosts of definitions remain. Taken together, the bully generally involves an individual or group who incites physical or emotional abuse on another individual or group. Although other research exists on Internet and workplace bullying (Mishna, Cook, Saini, Wu, & MacFadden, 2010), this review focuses on school bullying.

Prevalence and Negative Effects

School bullying is not a problem local to the United States; rather, it is recognized worldwide (Espelage & Swearer, 2003). Reports from Europe to North America have indicated that anywhere from 1% to 50% of students had been bullied or victimized within the last 2 months (Wang et al., 2009). Some observations have shown that as many as 30% of students were involved in bullying as either the bully or victim (Frey et al., 2009). Hymel and Swearer (2010) recently reported that 35% of students indicated being bullied at least once in the last 2 months with as many as 11% of those sampled reported being bullied more than 2 or 3 times in the last 2 months. Moreover, some research has found that bullying roles remain relatively stable across time. In a

sample of 516 middle school students, Espelage, Bosworth, and Simon (2001) found that individuals who perpetrated bullying continued to across multiple years. This finding has been replicated from other settings and populations (McDougal, Hymel, & Vaillancourt, 2009; Scholte, Engels, Overbeek, Kemp, & Haselager, 2007; Sourander, Helstela, Helenius, & Piha, 2000).

Although bullying's pervasiveness is cause for concern, the negative consequences related to bullying remain tantamount. Researchers have indicated that bullying has been linked to anger and misconduct (Bosworth, Espelage, & Simon, 1999), criminal or delinquent behavior (Olweus, 2002), and suicidal ideations (Kaltiala-Heino, Rimpela, Marttunen, Rimpela, & Rantanen, 1999). Victimization, on the other hand, has been linked to poor physical health (Rigby, 1999), low self-esteem (Rigby & Slee, 1993), depression (Sweeting et al., 2006), anxiety (Craig, 1994), and school avoidance (Kochenderfer & Ladd, 1996). Indeed, the negative effects of bullying and being bullied are persistent and problematic.

Bystander

Definition, Roles, and Negative Effects

A significant proportion of individuals within school systems are considered individuals who are bystanders of bullying (Glew et al., 2005). Twemlow, Fonagy, and Sacco (2004) defined a bystander as an individual who lacks participation in bullying scenarios as either the bully or victim. The bystander may actively intervene to stop the bully, encourage the bully to continue, or view bullying passively; bystanders can be either boys or girls (Cowie, 2000; Smith, Twemlow, & Hoover, 1998).

There are specific roles that the bystander can demonstrate. Some authors refer to the bystander as a passerby, observer, witness, or participant (Salmivalli, Kaukianinen, & Voeten, 2005; Twemlow et al., 2004); others described roles in relation to sustaining or preventing the bullying behavior such as reinforcer (e.g., laughing or seeing what is hap-

pening), assistant (e.g., follower of the bully), defender (e.g., being supportive of the victim), or outsider (e.g., remaining away from the bullying situation; Salmivalli et al., 1996). For the purposes of this review, we defined the bystander generally as any student who witnessed a bullying episode, with the operationalizing characteristic being the witnessing presence of bullying, regardless of other characteristics.

Despite these literature discordances, few disagree about the adverse effects witnessing bullying can have on the bystander. Bystanders felt significantly more uncomfortable in bullying situations compared to bullies (Stevens et al., 2004), and reported feelings of anxiety and insecurity (Rigby & Slee, 1993). This anxiety due to witnessing bullying has been linked to aggressive retaliation (Musher-Eizenman et al., 2004), and the fear of being bullied often prevented bystanders from seeking adult help (Unnever & Cornell, 2003). A recent large-scale study conducted in the United Kingdom found that compared to perpetrators, bystanders were at elevated risk for nonclinical outcomes (i.e., interpersonal sensitivity), and compared to victims bystanders were more likely to have elevated levels of substance abuse (Rivers et al., 2009).

Evidence of the Bystanders' Effects on Bullying

Individuals who are bystanders remain present more than 80% of bullying situations (O'Connell et al., 1999), and therefore some research has focused on a social-ecological model of bullying prevention and intervention (Frey et al., 2009; Swearer & Espelage, 2004). The social presence and pervasiveness of the bystander fosters myriad opportunities to intervene. For example, bystanders supported victims by reporting bullies to adults when participating in a setting specifically designed to change bullying behavior patterns through bystanders (Sharp, Sellors, & Cowie, 1994). Ross and Horner (2009) recently implemented a school-wide bullying intervention program that resulted in a decrease in reinforcing bystander behavior and bullying perpetration

overall. Moreover, interventions that focused on dealing with conflict through peers instead of direct interventions with adults led to positive effects (Cowie & Hutson, 2006), and an individual's willingness to intervene in bullying situations was inversely related to the amount of peer-group bullying perpetration (Espelage, Green, & Polanin, 2011).

Bystander Intervention Program Characteristics

To date, best practice guidelines to promote effective bystander intervention behaviors remain undefined because research findings varied widely with regard to their implementation focuses and approaches. Several mediums for interventions have been studied to teach children about bystander behavior, including classroom-based drama (Merrell, 2004), media such as videotaped reenactments (McLaughlin, 2009; Schumacher, 2007), and individualized computer-adaptive software to track students' progress within social scenarios and provided feedback on effective bystander behavior (Evers, Prochaska, Van Marter, Johnson, & Prochaska, 2007). However, all of these programs focus on bystander behavior perhaps, because there seems to be some support for targeting peer-group behaviors to mitigate individual bullying (Salmivalli et al., 1996). Peer-group interventions often encourage bystander intervention (Andreou, Didaskalou, & Vlachou, 2008; Frey et al., 2009; Stevens et al., 2000) or enhance bystander empathy for the victim (Gini et al., 2007; Nickerson, Mele, & Princiotta, 2008). However, few studies have examined the effects of peer-group interventions on previctim empathy (Merrell, Gueldner, Ross, & Isava, 2008) and there are resulting guidelines to promote this behavior.

Previous Meta-Analyses on Bullying Prevention Programs

A number of recent quantitative meta-analyses (Ferguson et al., 2007; Ttofi & Farrington, 2009, 2011; Merrell et al., 2008; Smith, Schneider, Smith, & Ananiadou, 2004) and qualitative systematic reviews (Ryan &

Smith, 2009) have been conducted regarding bullying and victimization intervention and prevention programs. However, none of these meta-analyses focused specifically on bystander intervention constructs. Merrell et al.'s (2008) review included three studies that measured "intervene to stop bullying behavior," which resulted in a mean effect size of 0.17, but this was a secondary analysis of a small number of studies. Therefore, the goal of the present study was to conduct a meta-analysis that would directly address bystander intervention behavior and empathy attitudes.

Given this goal, two primary research questions are addressed:

1. What is the average treatment effect, across the current literature, of bullying prevention programs on bystander intervention behavior?
2. What study characteristics produced the largest treatment effect?

A secondary research question addressed bystander empathy for the victim:

3. What is the average treatment effect, across the current literature, of bullying prevention programs on bystander empathy for the victim?

Method

Search Strategy

We used a comprehensive search to retrieve articles from the international research literature within the last 30 years (1980–2010). We searched primarily five online databases: Dissertation Abstracts International, Education Resources Information Center (ERIC), PsycINFO, Medline, and Science Direct. Combinations of the following terms were used: "bystander or participant or defender or other," "bully or victim," "school, school program, or program," "prevention or intervention," "aggression," and "not higher education or not cyber-bully." To ensure that the identified studies focused on bystander behavior as the primary goal, these terms were searched in the abstract of the study. In addi-

tion, we searched the bibliographies of all articles selected for relevant studies.

The search retrieved 360 total articles, but only 83 were unique and compared to the criteria listed in the next section. Of those 83, 53 were deemed irrelevant, 13 did not address the intervene construct, 6 failed to include a control group, and 1 was a repeat of a previous study. Finally, we corresponded with a number of experts in the field to ensure inclusion of all relevant articles. This correspondence produced 1 relevant study. Hence, we reviewed and included 11 studies total.

Criteria for Considering Studies for Review

The present study focused on school-based interventions that emphasized changing the bystander's intervention behavior. To assess the effects of these programs, we collected peer-reviewed studies published or conducted from 1980 to 2010, based solely within a school system and intended purposefully to modify bystander intervention behavior. Subsequently, we excluded studies that focused on changing bullying behaviors primarily and collected a bystander measure only as a secondary procedure. The review included interventions from the United States and Europe, but we limited inclusion to English-written studies.

We reviewed studies that included participants from the kindergarten through 12th-grade population, but interventions with school-aged children based outside the school setting were excluded. In addition, we attempted to collect studies that included "at-risk" students and the general population, but none of the studies distinguished between these populations. It should be mentioned that one study attempted to deconstruct the bystander into several types of bystanders to observe treatment effects (Evers et al., 2007). Although the deconstruction was informative, it was the only study to implement such a procedure. As such, for that study we used the average intervention effects across all bystander types.

Furthermore, we included only studies that used a treatment-control research design. These designs included true experimental randomly assigned groups, nonrandom quasi-experimental designs, and nonrandomly assigned matched group. We also included all control group types; these included wait list, treatment-as-usual, and "straw-man." However, single-group pre/post-test (e.g., gain scores) and cohort designs were excluded.

Outcomes

Studies must have included a bystander intervention measure. We operationalized this outcome as a measure that assessed the contribution of the bystander to a bullying situation (Frey et al., 2005, 2009). Therefore, we included studies that measured intention to intervene, intention to stop bullying, direct intervention, or conversely, difficulty in responding assertively to a bullying situation. For example, Andreou et al. (2008) included items that assessed students' intention to intervene on a 5-point Likert scale. We included items that concerned students' intention to "seek teacher's help," "react against bullying," and "support the victims of bullying" (p. 241). Table 1 provides the measures used, the study's stated construct, and the number of items combined to create the measure.

In addition to the intervention outcome, we collected results of the program's effects on changes in attitudes of empathy toward the victim. We operationalized the empathy outcome as a measure that indicated empathy for the victim. For instance, Stevens et al. (2000) used an empathy measure that included "feeling sad about students who are bullied" and "unpleasantness when another student is being bullied" (p. 26). Measures that included items similar to this construct constituted an empathy scale. We should also mention that this outcome constituted a secondary outcome measure and thus was not a criterion for synthesis inclusion. Studies that included a bystander intervention behavior measure but failed to include a measure of empathy were included.

Table 1
Characteristics of the Studies used in the Meta-Analysis

Study (DoP)	Type	Location	N (%Male)	Grade Setting	Program Title	Program Characteristics	Facilitator; (Length)	Intervention Measurement	Design	IN ES (95% CI)	EM ES (95% CI)
Andreou (2008)	J	Greece	418 (60)	4 th -6 th Urban	Curriculum- Based Anti- Bullying	Awareness-building (Awareness); Self- reflection; Behavior modification (BehMod)	Teacher (1)	Bystander intention to intervene	Q	-.01 (-.20, .19)	-.19 (-.38, .01)
Evers (2007)	J	Multiple US States	710 (41)	9 th -12 th Mixed	Build Respect	BehMod; Individualized computer software; Parent component (Parent)	Computer Based (2)	Passive bystanding	Q	.46 (.27, .64)	N.M.
Fonagy (2009)	J	Kansas	578 (46)	3 rd -5 th Rural	CAPSLE	Psychiatric consultation; Awareness; Parent	Teacher (12)	Helpful bystanding	E	.05 (-.11, .22)	-.23 (-.40, -.07)
Frey (2005)	J	Washington	913 (51)	3 rd -6 th Suburban	Steps to Respect	Awareness; Social- emotional skill- building; Parent	Teacher (6)	Bystander responsibility	E	.11 (-.02, .24)	.18 (.05, .31)
Karna (2011)	J	Finland	8166 (50)	4 th -6 th Mixed	KiVa	Awareness; Role- playing; Modeling; Parent	Teacher (12)	Bystander defending	E	.14 (.10, .19)	.15 (.10, .19)
McLaughlin (2009)	D	Ohio	41 (41)	6 th Urban	Effective Bully Prevention	Awareness; Modeling with media (ModMed)	Researcher (1)	Bystander traits	Q	.21 (-.42, .83)	-.17 (-.81, .46)
Menesini (2003)	J	Italy	293 (53)	6 th -8 th Urban	Befriending intervention	Awareness; Enhanced capacity; Responsibility- training	Not Reported (12)	Defender traits	Q	.03 (-.21, .26)	Not Measured

(Table 1 continues)

Table 1 Continued

Study (DoP)	Type	Location	N (%Male)	Grade Setting	Program Title	Program Characteristics	Facilitator; (Length)	Intervention Measurement	Design	IN ES (95% CI)	EM ES (95% CI)
Merrell (2004)	D	New York	56 (30)	9 th Urban	5 W's Approach to Bullying	Awareness; ModMed	Researcher (2)	Bystander intention to intervene	E	.60 (-1.17, 2.36)	Not Measured
Schumacher (2007)	M	Pennsylvania	825 (42)	9 th -12 th Mixed	Bullying Video Program	Awareness; ModMed	Researcher (1)	Bystander intention to intervene	E	.43 (.29, .57)	.29 (.15, .43)
Stevens "A" (2000)	J	Belgium	301 (50*)	4 th -8 th Not Reported	Anti-bullying Intervention	Social-cognitive training; ModMed	Teacher (1)	Bystander intention to intervene	E	.06 (-.17, .29)	-.05 (-.28, .18)
Stevens "B" (2000)	J	Belgium	401 (50*)	9 th -11 th Not Reported	Anti-bullying Intervention	Social-cognitive training; ModMed	Teacher (1)	Bystander intention to intervene	E	.39 (.19, .59)	.14 (-.06, .33)
Whitaker (2004)	B	Texas	1763 (50)	5 th Mixed	Expect Respect	Awareness; Psycho- education; Parent	Teacher (12)	Bystander intention to intervene	E	0.25 (.15, .34)	Not Measured

Notes: DoP = Date of Publication; J = Journal Article; D = Dissertation; M = Master's Thesis; B = Book Chapter; * = imputed percent male; Parent = Parent Training; Length = Time to Posttest in months; Q = Quasi-experimental; E = Experimental; T = Treatment groups; C = Control groups; IN ES = Intervention Effect Size; EM ES = Empathy Effect Size.

As shown in Table 1, 10 of the 83 unique articles met the inclusion criteria and were included in the meta-analysis. We discontinued the literature search on May 20, 2010, but added 1 article brought to our attention 3 months later, which brought the total number to 11 studies.

Coding

Study details, appropriate program, and sample information were coded directly into an EXCEL (2010) database. This included publication year, publication type, funding provided, country of origin and publication, program location, treatment and control sample characteristics (e.g., age, gender, race, SES, disabilities), program characteristics (e.g., length of time, intervention details), and program facilitator. In addition, intervention and empathy outcome measures were transcribed. By coding directly into the EXCEL database, we eliminated errors that might have occurred during the normal transcription phase (Lipsey & Wilson, 2001).

The first author coded all 12 studies, but one independent rater coded a randomly selected portion of studies (5) for reliability purposes. The two raters agreed 92% of the time. The coders came to an agreement for all discrepancies prior to completion.

Analysis

Quantitative synthesis, or meta-analysis, is a statistical technique that combines related research studies to estimate an overall treatment effect (Cooper, Hedges, & Valentine, 2009; Glass, 1976; Hedges & Olkin, 1985). Often, and in the case of the present review, meta-analysis aggregates treatment effect sizes to assess an intervention's effectiveness. The purpose of a meta-analysis, then, is to generalize findings across multiple treatment and setting types, participants, and times (Matt & Cook, 2009). We conducted analyses using SPSS (2010) and Comprehensive Meta-Analysis (Borenstein, Hedges, Higgins, & Rothstein, 2005) software.

Independent findings. A paramount meta-analytic assumption is independence of findings. Cooper (2010) discussed several occurrences that constitute nonindependence and their effects on subsequent findings. Thus, we conducted several common procedures to ensure independent findings. To ensure that only one effect size was derived from each intervention, we used only the first treatment outcome reported for studies that reported multiple post-treatment outcomes (see Andreou et al., 2008). If studies implemented interventions with two groups but only one control group, we synthesized the treatment effects prior to calculating the study effect size (see McLaughlin, 2009). Finally, if one author implemented an intervention and published multiple articles on the same sample, then we reviewed only the first article published (see Frey et al., 2005, 2009).

Effect size metrics. The majority of effect sizes calculated used a continuous scale. As such, the appropriate effect size metric was the standardized mean difference (Equation 1):

$$d = \frac{X_{G1} - X_{G2}}{S_p} \quad (1)$$

where the numerator is the mean difference between treatment and control group posttests, and the denominator is the pooled standard deviation for the intervention and comparison groups. Further, all d metrics were bias corrected using Hedges's (1981) small sample correction (g). This correction as well as the sampling variance is represented by Equations 2 and 3:

$$g = \left[1 - \left(\frac{3}{4N - 9} \right) \right] * d \quad (2)$$

$$Var_g = \sqrt{\frac{n_{G1} + n_{G2}}{n_{G1}n_{G2}} + \frac{g^2}{2(n_{G1} + n_{G2})}} \quad (3)$$

where N is the total sample, d is the original standardized mean difference, and n_{G1}

and n_{G2} represent the treatment and control group sample sizes, respectively.

In addition, we calculated logged odds ratio effect sizes for two studies that used a categorical outcome measure (Evers et al., 2007; Merrell, 2004). Both measures were observations of treatment and control participants during free period times. The authors observed how many times treatment children intervened (or intended to intervene) in a bullying situation compared to children in the control group. Standard odds ratio calculations were first used (Sanchez-Meca et al., 2003). We then converted the logged odds ratio into a standardized mean difference as outlined by Lipsey & Wilson (2001) (Equations 4 and 5):

$$d = \frac{\sqrt{3} * \ln(OR)}{\pi} \quad (4)$$

$$SE = \frac{\sqrt{3} * SE_{\ln(OR)}^2}{\pi^3} \quad (5)$$

where $\ln(OR)$ represents the original logged odds ratio and $SE_{\ln(OR)}$ represents the original sampling variance.

Missing data. Only one of the studies failed to provide appropriate descriptive statistics. Schumacher (2007) provided only a t statistic, as well as sample sizes, that compared the treatment and control groups. Lipsey and Wilson (2001) provided an appropriate conversion (Equation 6):

$$ES_{sm} = t \sqrt{\frac{n_1 + n_2}{n_1 n_2}} \quad (6)$$

where t represents the t statistic the study provided, n_1 represents the sample size of the treatment group, and n_2 represents the sample size of the control group.

Random effects model. We assumed that the treatments were derived from a random sample of the literature but lacked a common effect size. Given this assumption, Borenstein, Hedges, Higgins, and Rothstein (2010) posited that the random effects model

was most appropriate. We further assumed that an underlying distribution of effect sizes was plausible; thus our goal, given the random effects framework, was to estimate the distribution's mean and confidence interval.

To estimate a random effects mean and confidence interval, we calculated the weighted treatment effect of each study. The weighted effect estimation synthesized both within-study error variance and common between-study variance. We represented this weight calculation as Equation 7:

$$W_i = \frac{1}{V_i + T^2} \quad (7)$$

where W represents the i th study weight, V_i indicates the within-study error variance, and T^2 represents the between-study variance. We used these weights then to estimate the combined treatment effect. This can be represented by Equation 8:

$$M = \frac{\sum W_i * g_i}{\sum W_i} \quad (8)$$

where M represents the combined effect, W_i represents the i th study weight, and g_i indicates Hedges's effect size g for the i th study. Further, we calculated confidence intervals and p values by taking the square root of the inverse of the sum of the weights.

Moderator analysis. A critical next step to the investigation of effect size distribution is moderator analysis. We started by calculating the homogeneity statistic Q . This statistic provided information about the distribution of effect sizes, and a large test statistic (i.e., rejecting the null hypothesis of study homogeneity) indicated that moderator analyses were appropriate (Raudenbush, 2009). Given this statistical confirmation, we used procedures analogous to analysis of variance (ANOVA), where one attempts to model effect-size heterogeneity associated with categorical study-level variables. Further, because of the small number of studies that constituted the review, we calculated the variance compo-

ment across all groups rather than within groups as is generally conducted (Hedges & Vevea, 1998).

Lipsey (2009) discussed three types of independent variables common to meta-analytic practice: extrinsic variables, method variables, and substantive variables. Extrinsic and method variables relate to the study's dissemination (i.e., published or unpublished) or methodological constraints (i.e., randomized or nonrandomized). Substantive variables, on the other hand, should be regarded as variables of interest and generally include characteristics of the population or treatment. For this review, we coded substantive independent variables to reflect the participant's age, length of treatment, and treatment type (e.g., individual or group).

Finally, we conducted a moderator analysis procedure analogous to regression, weighted meta-regression. This statistical procedure allows for the simultaneous estimation of study-level effects, but shares the problems of typical regression (Cooper, 2010). For this review, we modeled two independent variables simultaneously, treatment population (categorical variable) and the percent of males in the treatment group.

Sensitivity analysis. Two of the reviewed studies contributed a measure that used an odds ratio. To ensure that study findings were not biased by including these measures, we conducted a sensitivity analysis. The analysis consisted of removing the 2 studies (Evers et al., 2007; Merrell, 2004) and recalculating the weighted effect size. We hypothesized that no difference would be found between the two types of measurements.

Publication bias. Publication bias remained an important consideration during the literature search and analyses. Rosenthal (1979) introduced the "file-drawer problem," which stated that studies with small or nonsignificant effect sizes tended to remain unpublished. To combat this problem, we included unpublished works from three dissertations and one master's thesis (Rothstein, Sutton, & Borenstein, 2005).

To avoid and interpret the overestimation of the random-effects estimate, we used the nonparametric trim and fill procedure (Duval & Tweedie, 2000) to assess the sensitivity of results to publication bias. This procedure estimates the number of publications theoretically missing because of funnel plot asymmetry, and then recalculates the random-effects mean and confidence interval to include the imputed missing studies. We also reported Rosenthal's fail-safe *N* (1979) and Egger, Davey Smith, Schnieder, & Minder's (1997) regression coefficient.

Results

Meta-Analysis Literature

Table 1 provides characteristics for each study included in the review. The studies reviewed include 7 published journal articles, 1 book chapter, and 3 unpublished papers (two dissertations and one master's thesis). Seven of the 11 studies were conducted within the United States and the other 4 were conducted in Belgium, Finland, Greece, and Italy, respectively. All studies were completed between 2000 and 2010.

One article contributed two effect sizes because it included two mutually exclusive interventions from two separate populations (Stevens et al., 2000). Therefore, we synthesized a total of 12 interventions. Each of the 12 interventions included a treatment and control group; 4 of the 12 programs used quasi-experimental design and the other 8 used a randomized experimental design. A total of 12,874 students participated in the 12 interventions.

Outcome Effect Sizes

Bystander intervention outcome. As delineated in Lipsey and Wilson (2001), we estimated the random effects weighted mean by using Equation 8. The results revealed a statistically significant positive weighted average ($g = .20, p < .001, 95\% \text{ CI} = .11 \text{ to } .29$). In other words, the treatment increased bystander intervention behavior 20% of one standard deviation more than individuals in the

Table 2
Summary Statistics for Bystander Intervention Effect Sizes

Study Name	Hedges' <i>g</i>	Standard Error	Lower Limit	Upper Limit	Z-Value	<i>p</i> -Value
Andreou	-.01	.10	-.20	.19	-.04	.97
Evers	.46	.09	.27	.64	4.88	.01
Fonagy	.05	.08	-.11	.22	.62	.54
Frey	.11	.07	-.02	.24	1.68	.09
Karna	.14	.02	.10	.19	6.51	.01
McLaughlin	.21	.32	-.42	.83	.65	.52
Mensini	.03	.12	-.21	.26	.22	.82
Merrell	.60	.90	-1.17	2.24	.66	.51
Schumacher	.43	.07	.29	.57	6.08	.01
Stevens "A"	.06	.12	-.17	.29	.53	.60
Stevens "B"	.39	.10	.19	.59	3.89	.01
Whitaker	.25	.05	.15	.34	5.16	.01
Overall	.20	.04	.11	.28	4.54	.01

control group. Table 2 provided a forest plot of the random-effects model's relevant statistics.

Empathy outcome. Of the 12 interventions used to calculate the bystander intervention outcome weighted average, 8 included a measure on victim empathy. As previously conducted, we used a random effects model to estimate the weighted treatment mean. The results revealed a very small, nonstatistically significant result ($g = .05$, $p = .38$) with a confidence interval that included zero (95% CI = $-.07$ to $.17$). The small number of

studies that included a measure of empathy may not provide enough power to detect a small effect. Therefore, the results of this analysis should be considered inconclusive (see Table 3).

Moderator Analysis

We categorized the effect sizes into several relevant groups and conducted the random-effects ANOVA-like analysis (Table 4). A statistically significant Q value indicated appropriate heterogeneity between studies that

Table 3
Summary Statistics for Empathy Effect Sizes

Study Name	Hedges' <i>g</i>	Standard Error	Lower Limit	Upper Limit	Z-Value	<i>p</i> -Value
Andreou	-.19	.10	-.38	.01	-1.90	.06
Fonagy	-.23	.09	-.40	-.07	-2.75	.01
Frey	.18	.07	.05	.31	2.72	.01
Karna	.15	.02	.10	.19	6.62	.01
McLaughlin	-.17	.32	-.80	.45	-.05	.59
Schumacher	.29	.07	.15	.43	4.13	.01
Stevens "A"	-.05	.12	-.28	.18	-.40	.69
Stevens "B"	.14	.10	-.06	.33	1.35	.18
Overall	.05	.06	-.07	.17	.74	.46

Table 4
Moderator & Meta-regression Analysis of Bystander Intervention
Effect Sizes

Moderator	K	Hedges' <i>g</i>	95% C.I.	<i>p</i> -value
Population				
3 rd –8 th Grade	8	.14	.11, .18	.04*
9 th –12 th Grade	4	.43	.33, .52	
Location				
United States	7	.26	.14, .38	.17
Europe	5	.13	–.01, .26	
Treatment length				
1–2 months	5	.31	.16, .45	.09
6–12 months	7	.16	.06, .25	
Parent Component				
Yes	5	.19	.08, .31	.92
No	7	.20	.07, .34	
Facilitator^a				
Teacher	7	.15	.09, .22	.01*
Other	4	.43	.30, .56	
Assignment				
Non-Random	4	.17	–.01, .35	.74
Random	8	.21	.10, .31	
Publication type				
Peer-review	8	.16	.06, .25	.07
Non-peer review	4	.32	.17, .48	
Meta-Regression				
Percent Male	β	S.E.	Z-score	p-value
HS (1 = Yes)	–.01	.009	1.04	.15
	.25	.13	1.98	.02*

* $p < .05$; a-Menesini et al. (2003) did not indicate the facilitator; HS = High School sample.

measured the bystander intervention construct ($Q = 39.81$, $df = 11$, $p < .001$; $I^2 = 72.36$). However, because of the relatively small number of studies that included an empathy measure, we chose not to conduct moderator analyses to protect against findings of chance.

The results of these analyses were decomposed into substantive and methodological characteristics. Samples that consisted of high school students only generated a significantly greater treatment effect (ES = 0.43, CI = .33 to .52) compared to samples of primary schools only (ES = 0.14, CI = .11 to .18). We further assessed sample differences by conducting an analysis to evaluate location differences. The results revealed that pro-

grams located in the United States (US) did not differ significantly from those located in Europe (EU; US ES = 0.26, EU ES = 0.13, $p = .17$).

Another substantive moderator, treatment length, failed to produce significantly greater treatment effects (1–2 months ES = 0.31, 6–12 months ES = 0.16; $p = .09$). Similarly, treatments that included a parental component (e.g., parent guides, parent training sessions) failed to influence the treatment significantly compared to programs without the component (parent included ES = 0.19, parent excluded ES = 0.20, $p = .92$).

We also grouped the studies by who facilitated the treatment programs. The teach-

ers implemented a significant portion of the programs for 7 of the 12 programs. Four other programs were facilitated by the researcher, a counselor, or in one case (Evers et al., 2007), with computer software. One study failed to indicate who implemented the program (Menesini, Codecasa, Benelli, & Cowie, 2003). The results of this moderator analysis revealed significantly greater treatment effects for programs that implemented the program with facilitators other than the teacher (teacher ES = 0.15, other ES = 0.43, $p < .01$). However, serious caution should be given to this finding because two of the four programs that used researchers as facilitators had the smallest sample sizes, and therefore this could be a reflection of imprecision or biased effects because of small samples (Levine, Asada, & Carpenter, 2009).

Finally, we conducted moderator analyses with two methodological groupings. We first observed mean group differences between randomly assigned (RA) and nonrandomly assigned treatments (NRA) groups. The results indicated that there were no statistically significant differences (RA ES = 0.21, NRA ES = 0.17, $p = .74$). We also estimated group differences between peer-reviewed (PR) and nonpeer-reviewed studies (NPR). The results of this calculation revealed that nonpeer-reviewed studies did not produce a greater treatment effect (PR ES = 0.16, NPR ES = 0.32, $p = .07$).

In addition to the ANOVA-like modeling, we conducted a weighted regression analyses. This analysis allowed us to estimate the effects of several predictors simultaneously. We hypothesized that the percentage of males in the treatment interventions and programs conducted in a high school would be significantly related to the treatment effect (i.e., Hedges's g). The results of this analysis again revealed that, after controlling for the percentage of males in the treatment group, high school samples produced a greater treatment effect compared to middle or elementary school interventions ($\beta = 0.25$, $Z = 1.98$, $p < .05$). These results bolstered the previous findings.

To ensure that the overall effect size was not upwardly biased by including different measurements, we conducted a sensitivity analysis removing studies that used a dichotomous outcome. The results of the analysis found that the overall weighted effect size decreased slightly (original ES = 0.20, modified ES = 0.18). The overall effect remained statistically and practically significant.

Publication Bias

We applied Duval and Tweedie's (2000) trim and fill procedure to address publication bias. This procedure revealed that one negative result was missing from the bystander intervention outcomes. However, the imputed missing values would only slightly change the overall fixed effect size (ES = 0.20); more important, it remained practically and statistically significant (see Figure 1). We also used Rosenthal's fail-safe N procedure; the results of this calculations indicated that 236 null studies would be required to result in a non-significant finding. Egger's regression intercept coefficient calculation also produced non-significant results ($\beta_0 = 0.57$, $p = .26$). Taken together, we concluded that the review's results were not affected significantly by publication bias.

Discussion

The purpose of this quantitative synthesis was to examine the treatment effects of bullying prevention programs on bystander intervention behavior. Empathy for the victim was also synthesized as a secondary outcome, but was not of primary purpose for the current review. In total, we reviewed 11 studies (12 effect sizes) from the United States and Europe that included 12,874 children.

Using meta-analytic techniques, the results revealed that the intervention behavior of bystanders increased (i.e., bystanders indicated greater intervention behavior in bullying situations) compared to control groups ($g = .20$). The results of a secondary analysis revealed that intervention programs did not have a similar effect on empathy for the victim ($g = .05$), but this finding should be viewed as

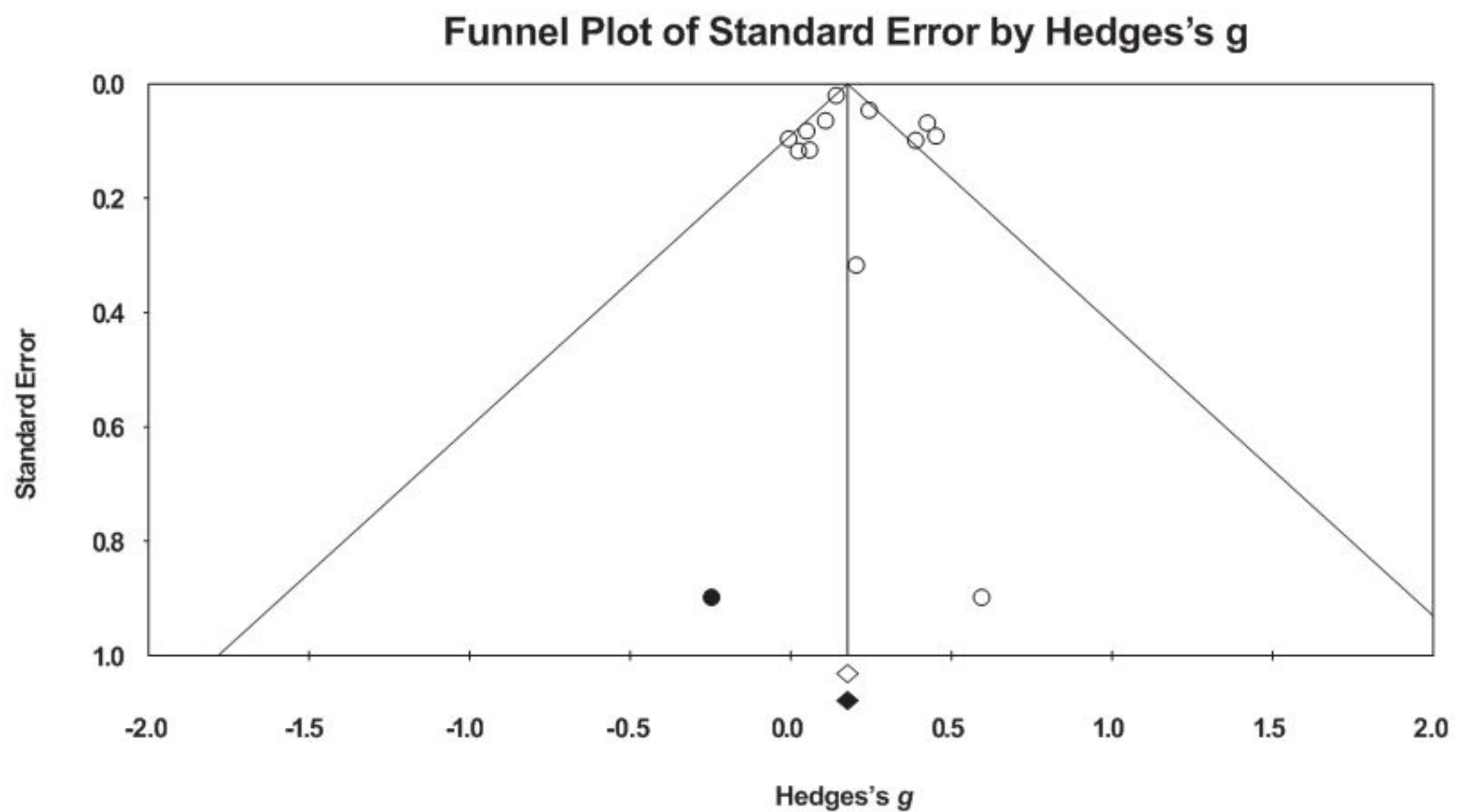


Figure 1. Trim and fill funnel plot of intervention effect sizes. This figure illustrates each effect size relative to its standard error; the shaded dot indicates an imputed effect size.

inconclusive because of the small number of studies that reported this outcome and its secondary nature.

These overall results mirrored the findings of a previously conducted small synthesis. Merrell et al. (2008) synthesized bullying prevention programs to investigate the effects of the programs on bullying perpetration, but also included several secondary measures (e.g., bystander intervention behavior, empathy). The authors reported small but significant treatment effects for bystander intervention behavior ($k = 3, g = .17$) and nonsignificant negative effects with regard to empathy ($k = 3, g = -.10$). Taken together, these replications results provided evidence against mono-operation bias and thus greater validity (Shadish, Cook, & Campbell, 2002).

Moderator analyses also revealed several findings of interest. Results of both ANOVA-like and weighted-regression analyses revealed that the treatment effects were greater for high school only samples. This is somewhat surprising because some scholars have postulated that bullying prevention programs are more effective for middle school-

aged children (Williford et al., 2011). These results may indicate that bystander intervention behavior is a developmental process and programs may not influence younger students as intended.

The purpose of meta-analytic research is to generalize findings across populations, treatments, outcomes, and designs (Matt & Cook, 2009). Although this synthesis aggregated a smaller number of studies, its findings rendered generalizability of bullying prevention program's effects on the bystander intervention construct. A few factors bolster this belief.

The populations assessed varied across ages, locations, and treatments. The largest effects were found for high school only samples; however, an overall significant treatment effect was also found for the total sample. With regard to location, no significant differences were found between U.S. and European samples. In addition, the studies employed a wide variety of treatment programs that proved efficacious. Moderator analyses also suggested that study design, publication type, and parental components produced similar

findings across studies. These results constituted a test of effects holding “across presumed irrelevancies” (Shadish et al., 2002, p. 455) and increases the findings’ external validity. Further investigation and evaluation is certainly required, but these results should cautiously suggest program generalizability.

Limitations

Several limitations should be noted. First, this meta-analysis included only 11 studies and 12 effect sizes. Although we took precautions to ensure unbiased effect sizes and findings, a great deal of caution should be used when interpreting the findings. The findings from a small collection of studies, no matter the statistical technique or number of students surveyed, should not enact immediate policy and practical changes. This becomes especially clear when one considers the effect of studies across time. Recent research in the field of meta-analysis publication bias has indicated that as programs increase in size and fidelity, effect sizes tend to decrease (Trikalinos & Ioannidis, 2005). Therefore, we plan to update these results periodically to observe the effect of time.

Second, because of the nature of meta-analyses, causal inferences should be stated cautiously. Quantitative meta-analysis, although statistically sophisticated and important, remains essentially an observational study (Lipsey & Wilson, 2001; Cooper & Hedges, 2009). On the other hand, the sample of studies we synthesized contained only those that used a treatment and control group, and these research designs constitute the most efficient measure of treatment effect (Shadish et al., 2002). Therefore, the synthesis of these results should partially reflect the nature of the primary studies.

Third, because of the low number of studies per factor, the ANOVA-like moderator analyses generated relatively low statistical power and the results should be interpreted with caution (Hedges & Pigott, 2004). Moreover, the low number of studies per group could easily capitalize on chance, or the groupings may reflect some other unforeseen

variable. This becomes especially clear with regard to the studies grouped by treatment length. As alluded to previously, a majority of the studies grouped in the 1–2 month category were the smallest of studies. Smaller studies tend to produce larger and more unstable effect sizes (Levine et al., 2009), and therefore the grouping could reflect this phenomenon.

Fourth, although we made efforts to collect all primary studies that focused on bystander intervention behavior, it is quite possible that studies failed to be included. New material published postsearch, misspecified search terms, or simple human error could all cause inadvertent omission of extant literature. As such, we must temper our inferences with regard to the extrapolation of this information.

Implications for Future Policy and Practice

This meta-analysis should help cautiously to shift the emphasis of policy and practice. The results of this meta-analysis revealed two implications for policy. First, state and national bullying legislation should implement and evaluate programs that address bullying behaviors as a group process. Prevention frameworks and programs that attempt to abate bullying within schools are increasingly emphasizing changes in school climate that desist reinforcing bystander behavior or bullying perpetration (Cohen, 2006). The results of this study support these efforts to raise awareness about the participant roles, to encourage active and prosocial behavior, and to provide opportunities to role-play and practice bystander intervention *in vivo*.

Second, the results of this meta-analysis revealed that bullying prevention programs might be effective at encouraging prosocial bystander intervention when the framework, program, and/or curriculum explicitly target bystander attitudes and behaviors. It is simply not sufficient to only define prosocial bystander behaviors, such as walk away, get help, or stand up to those engaged in bullying. Policy must encourage the adoption of programs and interventions that shift attitudes

supportive of intervention (willingness to intervene) and behaviors through a consistent message about intervention and ample support from adults and administrators.

Future Research

This research suggests important future projects. First, primary research should focus on designing programs, implementing change, and measuring the bystander construct. As mentioned previously, researchers should focus on changing the behavior of the bystander. Second, further work is required to evaluate the effects of bystander behavior on bullying. This meta-analysis merely demonstrated that explicitly stated bystander programs have the ability to increase bystander intervention behavior. However, future research must continue to assess how bystanders implement these processes and the direct effects on active bullying. Third, the results from the empathy review revealed inconclusive findings. Future research is required to elucidate the effects of prevention programs on this outcome. Finally, continued quantitative syntheses that focus on bystander behavior are required. Certainly only 11 studies from the last 10 years cannot accurately describe the scope of this issue. Future meta-analyses should incorporate, hopefully, new relevant literature.

Footnote

*Article used in meta-analysis.

References

- *Andreou, E., Didaskalou, E., & Vlachou, A. (2008). Outcomes of a curriculum-based anti-bullying intervention program on students' attitudes and behavior. *Emotional & Behavioural Difficulties, 13*, 235–248.
- Astor, R. A., Meyer, H. A., Benbenishty, R., Marachi, R., & Rosemond, M. (2005). School safety interventions: Best practices and programs. *Children and Schools, 27*(1), 17.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2005). *Comprehensive meta-analysis, Version 2*. Englewood, NJ: Biostat.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods, 1*, 97–111.
- Bosworth, K., Espelage, D., & Simon, T. (1999). Factors associated with bullying behavior in middle school students. *The Journal of Early Adolescence, 19*, 341–352.
- Cohen, J. (2006). Social, emotional, ethical, and academic education: Creating a climate for learning, participation in democracy, and well-being. *Harvard Educational Review, 76*, 201–237.
- Cooper, H. (2010). *Research synthesis and meta-analysis* (4th ed.). Thousand Oaks, CA: Sage Publications.
- Cooper, H., & Hedges, L.V. (2009). Potentials and limitations. In H. Cooper, L.V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 562–571). New York: Sage Publication.
- Cooper, H., Hedges, L. V., & Valentine, J. C. (2009). *The handbook of research synthesis* (2nd ed.). New York: Russell Sage Foundation.
- Cowie, H. (2000). Bystanding or standing by: Gender issues in coping with bullying in English schools. *Aggressive Behavior, 26*, 85–97.
- Cowie, H., & Hutson, N. (2006). Peer support: A strategy to help bystanders challenge school bullying. *Pastoral Care in Education, 23*(2), 40–44.
- Craig, W. M. (1994). The relationship among bullying, victimization, depression, anxiety, and aggression in elementary school children. *Personal Individual Differences, 24*(1), 123.
- Craig, W. M., Pepler, D., & Atlas, R. (2000). Observations of bullying in the playground and in the classroom. *School Psychology International, 21*(2), 22–36.
- Duval, S., & Tweedie, R. (2000). A nonparametric “trim and fill” method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95*(449), 89–98.
- Egger, M., Davey Smith, G., Schnieder, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315*, 629–634.
- Espelage, D. L., Bosworth, K., & Simon, T. R. (2001). Short-term stability and prospective correlates of bullying in middle-school students: An examination of potential demographic, psychosocial, and environmental factors. *Violence and Victims, 16*, 411–426.
- Espelage, D. L., & Swearer, S. M. (2003). Research on school bullying and victimization: What have we learned and where do we go from here? *School Psychology Review, 32*, 365–385.
- Espelage, D. L., Green, H. D., & Polanin, J. R. (in press). Willingness to intervene in bullying episodes among middle school students: Individual and peer-group influences. *Journal of Early Adolescence*.
- *Evers, K., Prochaska, J., Van Marter, D., Johnson, J., & Prochaska, J. (2007). Transtheoretical-based bullying prevention effectiveness trials in middle schools and high schools. [references]. *Educational Research, 49*, 397–414.
- Ferguson, C. J., Miguel, C. S., Kilburn, J. C., & Sanchez, P. (2007). The effectiveness of school-based anti-bullying programs: A meta-analytic review. *Criminal Justice Review, 32*, 401–414.
- *Fonagy, P., Twemlow, S. W., Vernberg, E. M., Nelson, J. M., Dill, E. J., Little, T. D., et al. (2009). A cluster randomized controlled trial of child-focused psychiatric consultation and a school systems-focused intervention to reduce aggression. *Journal of Child Psychology and Psychiatry, 50*, 607–616.
- Frey, K. S., Hirschstein, M. K., Edstrom, L. V., & Snell, J. L. (2009). Observed reductions in school bullying, nonbullying aggression, and destructive bystander behavior: A longitudinal evaluation. *Journal of Educational Psychology, 101*, 466–481.

- *Frey, K. S., Hirschstein, M. K., Snell, J. L., Edstrom, L., Van Schoiack, L., MacKenzie, E. P., & Broderick, C. J. (2005). Reducing playground bullying and supporting beliefs: An experimental trial of the steps to respect program. *Developmental Psychology, 41*, 479–491.
- Glass, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher, 5*(10), 3–8.
- Glew, G. M., Fan, M., Katon, W., Rivara, F. P., & Kermic, M. A. (2005). Bullying, psychosocial adjustment, and academic performance in elementary school. *Pediatric Adolescent Medicine, 159*, 1026–1031.
- Gini, G., Albiero, P., Benelli, B., & Altoè, G. (2007). Does empathy predict adolescents' bullying and defending behavior? *Aggressive Behavior, 33*, 467–476.
- Hawkins, D. L., Pepler, D. J., Craig, W. M. (2001). Naturalistic observations of peer interventions in bullying. *Social Development, 10*, 543–555.
- Hedges, L. V. (1981). Distribution theory for glass's estimator of effect size and related estimators. *Journal of Educational Statistics, 6*, 107–128.
- Hedges, L.V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.
- Hedges, L. V., & Pigott, T. D. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods, 9*, 426–444.
- Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3*, 486–504.
- Howell, D. C. (2010). *Statistical methods for psychology* (7th ed.). Pacific Grove, CA: Duxbury.
- Hymel, S., & Swearer, S. (2010). Bullying: An age-old problem that needs new solutions. *Education.com*. Retrieved from <http://www.education.com/reference/article/bullying-about-power-and-abuse-of-power/>
- Kaltiala-Heino, R., Rimpela, M., Marttunen, M., Rimpela, A., & Rantanen, P. (1999). Bullying, depression, and suicidal ideation in Finnish adolescents: School survey. *British Medical Journal, 319*, 348–351.
- *Karna, A., Voeten, M., Little, T. D., Poskiparta, E., Kaljonen, A., & Salmivalli, C. (2011). A large-scale evaluation of the KiVa antibullying program: Grades 4–6. *Child Development, 82*, 311–330.
- Kochenderfer, B. J., & Ladd, G. W. (1996). Peer victimization: Manifestations and relations to school adjustment in kindergarten. *Journal of School Psychology, 34*, 267–283.
- Lagerspetz, K. J., Bjorkqvist, K., Bertz, M., & King, E. (1982). Group aggression among school children in three schools. *Scandinavian Journal of Psychology, 23*, 45–53.
- Levine, T., Asada, K. J., & Carpenter, C. (2009). Sample sizes and effect sizes are negatively correlated in meta-analyses: Evidence and implications of a publication bias against nonsignificant findings. *Communication Monographs, 76*, 286–302.
- Lipsey, M. W. (2009). Identifying interesting variables and analysis opportunities. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 147–158). New York: Sage Publication.
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis* (2nd ed.). Thousand Oaks, CA: Sage Publications.
- Matt, G. E., & Cook, T. D. (2009). Threats to the validity of generalized inferences. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 538–557). New York: Sage Publication.
- McDougal, P., Hymel, S., & Vaillancourt, T. (2009). What happens over time to those who bully and those who are victimized? *Education.com*. Retrieved from http://www.education.com/reference/article/Ref_What_Happens_Over/
- *McLaughlin, L. P. (2009). *The effect of cognitive behavioral therapy and cognitive behavioral therapy plus media on the reduction of bullying and victimization and the increase of empathy and bystander response in a bully prevention program for urban sixth-grade students*. Unpublished doctoral dissertation, University of Toledo, Digital Dissertations.
- *Menesini, E., Codecasa, E., Benelli, B., & Cowie, H. (2003). Enhancing children's responsibility to take action against bullying. *Aggressive Behavior, 29*, 1–14.
- *Merrell, R. (2004). *The impact of drama intervention program on response of the bystander to bullying situations*. Unpublished doctoral dissertation, University of Rochester, Digital Dissertations.
- Merrell, K. W., Gueldner, B. A., Ross, S. W., & Isava, D. M. (2008). How effective are school bullying intervention programs? A meta-analysis of intervention research. *School Psychology Quarterly, 23*, 26–42.
- Mishna, F., Cook, C., Saini, M., Wu, M. J., & MacFadden, R. (2010). Interventions to prevent and reduce cyber abuse of youth. *Research on Social Work Practice, 1*(1), 1–10.
- Musher-Eizeman, D. R., Boxer, P., Danner, S., Dubow, E. F., Goldstein, S. E., & Herretick, D. M. I. (2004). Social-cognitive mediators of the relation of environmental and emotion regulation factors to children's aggression. *Aggressive Behavior, 30*, 389–408.
- Nansel, T. R., Overpeck, M., Pilla, R. S., Ruan, W. J., Simons-Morton, B., & Scheidt, P. (2001). Bullying behaviors among US youth: Prevalence and association with psychosocial adjustment. *Journal of American Medical Association, 285*, 2094–2100.
- Nickerson, A. B., Mele, D., & Princiotta, D. (2008). Attachment and empathy as predictors of roles as defenders or outsiders in bullying interactions. *Journal of School Psychology, 46*, 687–703.
- O'Connell, P., Pepler, D., & Craig, W. (1999). Peer involvement in bullying: Insights and challenges for intervention. *Journal of Adolescence, 22*, 437–452.
- Olweus, D. (1973). Personality and aggression. In J. K. Cole & D. D. Jensen (Eds.), *Nebraska Symposium on Motivation* (pp. 260–321). Lincoln: University of Nebraska Press.
- Olweus, D. (2002). *Bullying at school*. Malden, MA: Blackwell Publishers.
- Raudenbush, S. W. (2009). Analyzing effect sizes: Random-effects models. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 295–314). New York: Sage Publication.
- Rigby, K. (1999). Peer victimization at school and the health of secondary school students. *The Journal of School Psychology, 69*, 95–104.
- Rigby, K., & Slee, P. T. (1993). Dimensions of interpersonal relation among Australian children and implica-

- tions for psychological well-being. *British Journal of Educational Psychology*, 133, 33–42.
- Rivers, I., Potteat, V. P., Noret, N., & Ashurst, N. (2009). Observing bullying at school: The mental health implications of witness status. *School Psychology Quarterly*, 24, 211–223.
- Rosenthal, R. (1979). The “file drawer problem” and tolerance for null results. *Psychological Bulletin*, 86, 638–641.
- Ross, S. W., & Horner, R. H. (2009). Bullying prevention in positive behavior support. *Journal of Applied Behavior Analysis*, 42, 747–759.
- Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005). *Publication bias in meta-analysis*. West Sussex, England: John Wiley & Sons Ltd.
- Ryan, W., & Smith, D. J. (2009). Antibullying programs in schools: How effective are evaluation practices? *Prevention Science*, 10, 248–262.
- Salmivalli, C., Lagerspetz, K., Bjorkqvist, K., Osterman, K., & Kaukianen, A. (1996). Bullying as a group process: Participant roles and their relations to social status within the group. *Aggressive Behavior*, 22, 1–12.
- Salmivalli, C., Kaukianen, A., & Voeten, M. (2005). Anti-bullying intervention: Implementation and outcome. *British Journal of Educational Psychology*, 75, 465.
- Sanchez-Meca, J., Marin-Martinez, F., & Chacon-Moscoso, S. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods*, 8, 448–462.
- Scholte, R. H. J., Engels, R. C. M. E., Overbeek, G., Kemp, R. A. T., & Haselager, G. J. T. (2007). Stability in bullying and victimization and its association with social adjustment in childhood and adolescence. *Journal of Abnormal Child Psychology*, 35, 217–228.
- *Schumacher, P. J. (2007). *To what extent will a sixty-second video on bullying produced by high school students increase students’ awareness of bullying and change their attitudes to reduce acceptance of bullying in a high school environment?* Unpublished master’s thesis, University of Pennsylvania, Digital Dissertations.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs*. New York: Houghton Mifflin.
- Sharp, S., Sellors, A., & Cowie, H. (1994). Time to listen: Setting up a peer counseling service to help tackle the problem of bullying in schools. *Pastoral Care Education*, 12, 3–6.
- Smith, J. D., Twemlow, S. W., & Hoover, D. W. (1999). Bullies, victims, and bystanders: A method of in-school intervention and possible parental contributions. *Child Psychiatry and Human Development*, 30, 29–37.
- 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.
- Sourander, A., Helstela, L., Helenius, H., & Piha, J. (2000). Persistence of bullying from childhood to adolescence—A longitudinal 8-year follow-up study. *Child Abuse & Neglect*, 24, 873–881.
- SPSS. (2010). *Statistical package for social sciences*, Version 19.0.0. Chicago, IL: IBM.
- Sutton, A. J. (2005). Evidence concerning the consequences of publication and related biases. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis* (pp. 175–192). West Sussex, England: Wiley & Sons.
- *Stevens, V., Oost, P. V., & 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., Vaillancourt, T., & Hymel, S. (2010). What can be done about school bullying? Linking research to educational practice. *Educational Researcher*, 39, 38–47.
- Sweeting, H., Young, R., West, P., & Der, G. (2006). Peer victimization and depression in early-mid adolescence: A longitudinal study. *British Journal of Educational Psychology*, 76, 577–594.
- Trikalinos, T. A., & Ioannidis, J. P. A. (2005). Assessing the evolution of effect sizes over time. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis*. West Sussex, England: John Wiley & Sons Ltd.
- Ttofi, M. M., & Farrington, D. P. (2009). What works in preventing bullying: Effective elements of anti-bullying programmes. *Journal of Aggression*, 1, 13–24.
- Ttofi, M. M., & Farrington, D. P. (2011). Effectiveness of school-based programs to reduce bullying: A systematic and meta-analytic review. *Journal of Experimental Criminology*, 7, 27–56.
- Twemlow, S. W., Fonagy, P., & Sacco, F. C. (2004). The role of bystander in social architecture of bullying and violence in schools and communities. *Annals of the New York Academy of Sciences*, 1036, 215–232.
- Unnever, J. D., & Cornell, D. G. (2003). The culture of bullying in middle school. *Journal of School Violence*, 2, 5–27.
- Wang, J., Iannotti, R. J., & Nansel, T. R. (2009). School bullying among adolescents in the United States: Physical, verbal, relational, cyber. *Journal of Adolescent Health*, 45, 368–375.
- *Whitaker, D. J., Rosenbluth, B., Valle, L. A., & Sanchez, E. (2004). Expect respect: A school-based intervention to promote awareness and effective responses to bullying and sexual harassment. In D. L. Espelage & S. M. Swearer (Eds.), *Bullying in American schools* (1st ed., pp. 327–350). Mahwah, NJ: Lawrence Erlbaum Associates.
- Williford, A., Boulton, A., Noland, B., Little, T. D., Karna, A., & Salmivalli, C. (2011). Effects of the KiVa anti-bullying program on adolescents’ depression, anxiety, and perception of peers. *Journal of Abnormal Child Psychology*. Retrieved from <http://www.springerlink.com/content/2532380182216828/>

Date Received: February 1, 2011

Date Accepted: September 7, 2011

Action Editor: Joseph Betts ■

Joshua R. Polanin is a research methodology doctoral student at Loyola University Chicago. His interests include methodological improvements to meta-analysis and hierarchical linear modeling. He currently serves as the methodologist for a large, school-based bullying prevention program and as the managing editor of the Campbell Collaboration's Methods Group.

Dorothy L. Espelage is a professor in the Department of Educational Psychology at the University of Illinois, Urbana-Champaign. Her research programs include investigations of bullying, sexual harassment, and dating violence among adolescents for almost two decades. She is engaged in a large randomized clinical trial of a school-based bullying prevention program.

Therese D. Pigott is a professor at the School of Education, Loyola University Chicago. Her research interests are statistical methods for meta-analysis and methods for handling missing data in statistical analysis. She is a co-chair and editor of the Methods Group of the Campbell Collaboration, an international organization that produces systematic reviews of social interventions.