The Effectiveness of School-Based Anti-Bullying Programs

A Meta-Analytic Review

Christopher J. Ferguson
Claudia San Miguel
John C. Kilburn, Jr.
Patricia Sanchez
Texas A&M International University, Laredo

Youth violence, including bullying and other serious violent behaviors, has received increased political and scientific attention over the past several decades. Although violent behavior among youth and in the schools has declined over the past decade, the victimization of children by other children in school settings remains a major issue of concern. In response, a number of prevention and intervention programs have been developed to reduce the incidence of bullying and violence in schools. This meta-analysis attempted to examine the effect of school-based anti-bullying programs. Results suggested a significant effect for anti-bullying programs ($r = .12$). However, this result seemed to be slightly influenced by publication bias and did not meet the adopted threshold for "practical significance." The effect for programs targeted specifically at at-risk youth was slightly better, but overall, anti-bullying programs produce little discernible effect on youth participants. Reasons that anti-bullying programs may produce little effective change are discussed.

Keywords: bullying; school-based programs; violence prevention interventions; violence; prevention; program evaluation

Violence in the schools, which can encompass a continuum of behaviors from bullying to more serious violence, is an issue that has attracted increased attention not only from the public but also from educators, social scientists, and politicians over the past decade (Phillips, 2007). Children have arguably been viewed as innocent and without malice throughout most of American culture, although research spanning several decades disputes this common assumption (Phillips, 2007), and recent literature has documented that lesser forms of violence, such as bullying, are common among children (Nansel, Overpeck, Pilla, Ruan, Simons-Morton, & Scheidt., 2001).

Although a multitude of school anti-bullying programs have been developed and implemented, it is uncertain whether these programs are successful in achieving their intended outcomes. This uncertainty may be due to the fact that many anti-bullying programs have not been subjected to systematic and empirical review. Of those programs that have been subjected to review, publication bias may be to blame for inflated significant effects. This study focuses on determining actual success (as indicated by effect size) in anti-bullying prevention programs.

The salience of this study cannot be overlooked, especially because most schools have implemented bullying prevention programs, and monies are being allocated to create a safer school environment (Sherman, 2000). With limited knowledge of programs' success, the public and politicians are "lulled into falsely believing that they are addressing this problem [bullying and school violence] when the resources committed to such efforts could be better used to develop more effective programs" (Farrell, Meyer, Kung, & Sullivan, 2001, p. 207). This study, therefore, focuses on the degree of actual success these anti-bullying and nonviolence programs espouse to have achieved in schools over the past two decades.

Literature Review

Bullying as a Problem Behavior

Bullying has been defined as repeated, intentional, harmful, and aggressive behavior inflicted by a person or group with seemingly more power on a person or group with lesser power (Nansel et al., 2001). Yet, a consensus on the precise definition of bullying has yet to be reached, which complicates research, in particular with measuring the extent of the phenomenon. Some researchers focus on studying specific forms of bullying such as acts inflicted because of race and/or gender, and others have focused on maturational differences among youth (Kyriakides, Kaloyirou, & Lindsay, 2006). Yet, others have focused on direct and physical acts of violence as opposed to indirect forms of violence such as intimidation and harassment (Marini, Dane, Bosacki, & YLC-CURA, 2006). There are some researchers who take a holistic approach and define bullying in broad terms. Olweus, Limber, and Mihalic (1999), for example, define bullying as the exposure, repeatedly and over time, to negative actions on the part of one or more other students. Such negative actions may include attempting or intentionally inflicting injury or discomfort on others. The injury or discomfort may be physical (e.g., hitting, kicking, pushing, choking) or verbal (e.g., calling names, threatening, taunting, malicious teasing, spreading nasty rumors) or may manifest itself in other ways, such as making faces or obscene gestures or the intentional exclusion from a group. The latter (usually more subtle) forms are usually termed "indirect bullying," whereas "direct bullying" includes behaviors that represent relatively open (usually verbal or physical) attacks on the victim. In general, Nansel et al. (2001) argue that bullying occurs where there is an imbalance of power between the perpetrator and the victim.

Although research has found that bullying can occur in different social settings such as the workplace (see Parkins, Fishbein, & Ritchey, 2006, for a discussion of workplace bullying), this form of violence is typically found in school settings among children and adolescents. According to Ambert (1994), peer-to-peer abuse, such as bullying, is a distressing reality in schools and more common than child abuse. The 2001 School Crime Supplement to the National Crime Victimization Survey reported that 14% of teens between the ages of 12 and 18 disclosed being bullied in the past 6 months. About half of those reporting being bullied claimed that it was indirect bullying—nonphysical social exclusion—whereas the other half of those being bullied faced direct violence or threats (DeVoe, Dean, Traube, & McKay, 2005). Other research has found that 26% of boys reported bullying others "sometimes" to "weekly," whereas 21% of boys reported being bullied "sometimes" to "weekly" (Nansel et al., 2001).

Regardless of definitional constraints, victims of bullying are more likely to have negative perceptions of their school (Nansel et al., 2001), behavior problems (Haynie, Nansel, & Eitel, 2001), and difficulty concentrating on schoolwork (Sharp & Smith, 1994). Other research adds that victims receive lower grades and avoid certain areas and activities (DeVoe et al., 2005). In addition, some victims of this form of violence tend to experience stress-related symptoms such as headaches and nightmares, whereas others suffer from depression and school phobia (Sharp & Smith, 1994). Research has also shown that bystanders or witnesses to acts of bullying may also experience similar problems (Kyriakides et al., 2006). With respect to bullies, there is evidence that has shown that their likelihood to engage in criminal activity as young adults is much higher than their less aggressive counterparts (Bollmer, Harris, & Milich, 2006; Heydenberk, Heydenberk, & Tzenova, 2006), although bullies may evidence fewer mental health and social problems than was once thought (Olweus, 1995).

Programs Designed to Prevent Bullying

Anti-bullying programs in schools are as diverse as the definitions constructed to understand this form of violence. More traditional anti-bullying programs follow the Olweus model. The Olweus Bully Prevention Program (Olweus et al., 1999) was implemented in Norway after a severe incident of bullying that sparked national attention. The program is designed to help identify bullies, in elementary, middle, and high schools, and to help them as well as their victims cope with the effects of this type of school violence. The Second Step Violence Prevention Program is also a classroom-based program that has shown some success in improving social competence and reducing anti-social behaviors (Taub, 2001) as well as decreasing aggression (Van Schoiack-Edstrom, Frey, & Beland, 2002). Moreover, Responding in Peaceful and Positive Ways (RIPP) teaches students social skills and how to respond to conflict (Farrell, Meyer, Sullivan, & Kung, 2003; Farrell, Meyer, & White, 2001).

More recent programs have approached this problem from a restorative justice model, which tries to restore the relationship between the victim and the offender by using the reintegrative shaming techniques proposed by Braithwaite (1989), as well as forgiveness and reconciliation, to reduce incidents of bullying (Ahmed & Braithwaite, 2006). In addition, researchers are exploring the role of biological processes and, more specific, the role that hormones play in the ability of youth to mediate conflict (Hazler, Carney, & Granger, 2006). There is also research on understanding individual coping mechanisms, how children and adolescents process information and interpret situational cues, and how they use their past experiences to cope with aggression (Gini, 2006). Literature suggests that aggressive children tend to interpret situational cues differently from nonaggressive peers (Gini, 2006). Bullies are also considered "deficient in their social intelligence and in their ability to interpret and manage information deriving from social interactions with peers" (p. 535) and thus tend to respond with aggressive behavior. However, literature also suggests that bullies are "skilled social manipulators . . . that lack the empathic reactivity towards victims' suffering" (p. 535). Thus, programs that focus on the cognitive and emotional dimensions of bullying as well as moral sensitivity and appropriate empathic reactivity have also been implemented in schools (Jolliffe & Farrington, 2006).

Essentially, depending on the approach taken, anti-bullying programs can encompass a number of elements ranging from the implementation of conflict-resolution pedagogy

within the school system to individual accountability for the behavior, which often can involve the juvenile justice system (Phillips, 2007). Phillips (2007) suggests programs that hold society accountable for bullying. Because "masculinity and male violence are primarily socially constructed . . . we should hold society accountable both for the production of unhealthy norms and for the violent practices that affirm them" (p. 176). Although there are many other programs, it is salient to be able to discern which are most effective. Thus, the purpose of this study is not to provide a review of each anti-bullying program implemented in the United States (for a complete review of anti-bully programs, see Nansel et al., 2001; Olweus, 1995; Olweus et al., 1999) but to provide tools to analyze the effectiveness of these programs in general.

Meta-Analyses of School Violence Prevention Programs

With much concern over preventing bullying in schools and billions of dollars given to school districts in the United States to implement anti-bullying programs (see Sherman, 2000), it is imperative to determine whether these programs are effective. Two main attempts have been made to address the effectiveness of prevention programs targeting school bullying. The first of these was conducted by Wilson, Gottfredson, and Najaka (2001), focusing on aggressive and delinquent behaviors broadly defined (including nonviolent antisocial behaviors). Overall, the effect size for the interventions was small but significant. Although a significant step forward in understanding the effect of school violence anti-bullying programs, the interpretability of the analysis is made somewhat difficult by the inclusion of nonviolent antisocial behavior in many of the outcomes, including numerous behaviors not related to bullying, such as vandalism.

Wilson, Lipsey, and Derzon (2003) expand on Wilson and colleagues' (2001) work by using a pre–posttest change design in their meta-analysis to allow for the incorporation of single group pre–posttest studies (e.g., no control group) that were not included in the previous analysis. Specifically, this latter meta-analysis considered the effect size for change for control groups and experimental groups separately, allowing for the inclusion of research reports for which there was no control group. It is not surprising that this design is somewhat controversial but does have the benefit of assessing the effect of nonexperimental findings such as those commonly implemented in school settings. In this analysis, small but significant effects are found for treatment groups, with little to no effect seen for control groups. The authors enthusiastically conclude that a broad range of anti-violence programs produces a significant reduction in violence in school settings.

Wilson et al. (2003) present a critical issue in the use of meta-analysis, namely, the interpretation of the effect size. As meta-analysis benefits from the compounded effects of combined sample size across multiple studies, it is arguable that the likelihood that the effect for any intervention (or variable) would not be significant is very slim. In other words, noting that the results from a meta-analysis are positive and significant is not, in and of itself, terribly meaningful or interpretable, as this "significance" may be due more to the vastly increased power of a combined sample size rather than any practical effect that is actually occurring in the population of interest. This perspective is an extension of Cohen's (1994) observation that null hypothesis testing is fundamentally flawed in the way that it is oftentimes interpreted and that positive findings are not necessarily more interpretable than null findings.

Cohen recommends the interpretation of effect sizes and the confidence interval around those effect sizes, rather than focusing on statistical significance, per se. It is this suggested methodology that will be adopted as part of the meta-analytic review for this study.

The issue of effect size was addressed by Wilson et al. (2003). The authors present the effect sizes in terms of d, which has some favorable statistical properties but is not easily understandable in the sense of producing a coefficient of determination (e.g., r^2) and tends to look more impressive (e.g., produces larger numbers) than does r. Recalculated in terms of the coefficient of determination r^2, it can be seen that whereas the effect sizes for the control groups are essentially .00 (or 0% effect), those for the experimental groups range from < .01 (or ≤ 1% effect) to a maximum of .02 (or 2% effect). Thus, although statistically significant, it is unclear that the effectiveness in producing change exhibited by Wilson et al.'s (2003) analysis necessarily warrants the enthusiastic support voiced by the authors. The possibility remains that this may be due to the inclusion of effectively nonexperimental designs lacking a control group in their analysis, although it was the intent of the authors to examine just these studies.

Publication Bias

Publication bias (or the "file drawer effect") occurs when articles with statistical significance are selected for publication more often than are articles that do not obtain significance (Rosenthal & Rosnow, 1991). When this occurs, peer-reviewed publications may provide a biased sample of all of the studies actually carried out, as nonsignificant studies are not included in the body of scientific knowledge. The file drawer effect may be passed on to meta-analyses given that unpublished nonsignificant studies would likely be difficult to locate. As noted by Rosenthal and Rosnow, there is no perfect method for identifying whether a body of knowledge is compromised by the file drawer effect. However, Rothstein, Sutton, and Borenstein (2005) note that there are several methods for estimating publication bias (also see Ferguson, 2007a, for a description). Similar to the way that concordance among multiple goodness-of-fit indices in confirmatory factory analysis can be used to demonstrate overall goodness of fit of a factor structure (see Lance & Vandenberg, 2002, for a discussion), concordance among these various indices of publication bias can provide consistent evidence of this concern in a meta-analysis.

Research Goals

This meta-analysis seeks to expand on the laudable work of the previous studies. First, this analysis is designed to improve on the precision of the previous studies, in particular through moderator analyses. To examine the possibility that Wilson et al.'s (2001) analysis may have been influenced by the inclusion of nonviolent antisocial acts in their analysis, outcome variables that are specifically related to violent behavior and those related to more general externalizing or antisocial behavior will be considered for differential impact. Similarly, the differential impact of prevention programs on elementary school, middle school, and high school youth will be considered, as well as whether the children were identified as being at risk prior to the program or were part of the general school population. These moderator analyses should help in understanding specifically to which populations of children and under what circumstances violence-prevention programs are most effective.

Second, this analysis will specifically address several related issues. First, out of concern that Wilson and colleagues' (2003) analysis may have been influenced by the inclusion of nonexperimental studies, only clearly randomized experimental studies will be included in these analyses. Thus, this analysis takes the opposite approach of Wilson et al. by focusing on the science rather than practice of violence prevention programs. Second, unlike previous analyses, the focus will be less on the achievement of statistical significance and instead on effect size and the confidence interval around the effect size to ascertain an estimate of the "true" impact of violence prevention programs in the school-based population. Last, this analysis will include a rigorous analysis of publication bias that is typically not a part of most meta-analyses to ascertain whether the effect of school-based prevention programs may be inflated by the file drawer effect.

One further issue of previous meta-analyses bears mentioning: The previous meta-analyses have involved an all-inclusive approach, meaning that they have included in their analyses all potential studies independent of quality. As Bobko and Stone-Romero (1998) argue, meta-analyses are typically ill equipped to adequately address internal and external validity issues related to individual studies. This can be put more bluntly as the "junk in, junk out" problem, meaning that meta-analyses traditionally assume that all included studies are internally and externally valid. Results of such meta-analyses may be compromised by the poor methodology of some of the studies included. In response, this meta-analysis will use a modified "best practice" methodology (Ferguson, 2007a) with the intent to specifically include studies that meet certain criteria related to higher quality methodology.

Method

It is the purpose of this section to describe three elements of the meta-analytic process: (a) the method for selecting and categorizing studies included in the meta-analysis, (b) the method for calculating effect size estimates from the studies, and (c) the procedure for statistical and publication bias analysis.

Study Selection and Categorization

PsycINFO was searched for all articles published between the years of 1995 and 2006 (this criterion is discussed below) that included the following search terms: (school or educat*) and (interv* or prevent* or treat* or therap*) and (attack* or fight* or aggress* or violen* or externalizing or bully*). The references of primary sources revealed in this search were also examined for studies that were not discovered during this initial search.

Articles were judged relevant if they met the following criteria:

1. Articles had to have been published between the years of 1995 and 2006 (effectively, a 10-year publication period). Primarily, this was done to examine trends in recent research on school anti-bullying programs and to avoid potential contamination effects from earlier efforts that may have been more preliminary. Examining recent years of research only is a common element of many meta-analyses for interventions as intervention strategies tend to change over time.

2. Outcome variables had to clearly measure some element of bullying behavior or aggression toward peers, including direct aggressive behavior toward children in a school setting. Articles that measured only learning the details of a program (e.g., program mastery) or included only measures of behavior disorders such as ADHD, but not necessarily bullying, were not included. Similarly, programs that specifically targeted other forms of violent behavior unrelated to bullying behavior, such as gang activity, were not included.

3. Articles included had to involve some form of control or contrast group to test program effectiveness. Pre–post only designs were not included. Randomized assignment to control and contrast groups was required for inclusion, although this may have occurred at the individual, classroom, or school level.

4. Intervention programs had to be school based.

5. As this analysis is concerned with the potential for publication bias in peer-reviewed journals, only manuscripts published in peer-reviewed journals were included in the analysis. Book chapters, dissertation manuscripts, and unpublished manuscripts were not included in the analysis. Although it would be interesting and valuable to consider publication status (published or unpublished) as a moderator variable in the analysis, there was no evident method for assuring that all relevant unpublished manuscripts could be obtained (including those from unknown authors, or those intentionally or unintentionally suppressed by the authors).

A total of 42 published studies made up of 45 separate observations was found that met the above criteria. Three of the included studies included two separate observations of different programs at different times. The combined sample size for the included articles was 34,713, which represents a pool of the sample groups from the individual studies. Articles in this study were coded for the presence of several potential moderator variables, namely, (a) grade level of the program implementation (elementary age, middle school age, or high school age), (b) whether the samples were drawn from populations identified as at risk for violence or were part of the general school population, and (c) whether the outcome variables clearly used measures of violent behavior or focused on broader measures of bullying behaviors such as teasing.

Calculating Effect Size Estimates

Pearson's r, a flexible and easily interpreted index of effect size, was used as the effect size estimate in this study. Correlation coefficients were transformed to Fisher's z, weighted, averaged, and transformed back to a pooled r, denoted r_+. This provides for a reasonably conservative estimate of the actual effect noted in the study, as it does not include estimates of effect size that may be due solely to chance.

Another issue that arises is that of multiple measures for the same construct occurring within a study (multiple dependent measures). Ideally, if the reliability between the measures were known or reported, it would be possible to calculate composite score correlation (Hunter & Schmidt, 2004). However, as this information is often not reported, simple mean correlations were computed. In studies in which both univariate (e.g., bivariate correlations) and multivariate (e.g., partial correlations or correlations adjusted by beta-weights in a multiple regression) were available, only the latter were included in the meta-analysis, as this provided better indices of the unique shared variance between intervention and outcome.

Table 1
Meta-Analytic Results for All Studies and Moderator Subgroups

Group	k	N	r_+	95% CI	Homogeneity Test
All studies	45	34713	.12(.09)	(.08, .17)	$X^2(44) = 800.1, p \leq .001$
Nonviolent bullying	23	14597	.12(.04)	(.06, .17)	$X^2(22) = 196.1, p \leq .001$
Violence	22	20116	.13(.13)	(.05, .20)	$X^2(21) = 534.3, p \leq .001$
Elementary school	26	19112	.15(.13)	(.08, .22)	$X^2(25) = 582.5, p \leq .001$
Middle school	14	14933	.08(.04)	(.02, .14)	$X^2(13) = 136.1, p \leq .001$
High school	5	668	.13(.14)	(.03, .24)	$X^2(4) = 7.1, p \geq .10$
Low risk	26	28185	.09(.09)	(.02, .15)	$X^2(25) = 621.9, p \leq .001$
High risk	19	6528	.19(.08)	(.11, .27)	$X^2(18) = 176.8, p \leq .001$

Note: k = number of independent studies; N = number of participants; r_+ = pooled correlation coefficient random effects model (fixed effects model is in parentheses); CI = confidence interval; violence = violent behaviors outcome variables.

Statistical and Publication Bias Analyses

The Comprehensive Meta-Analysis (CMA) software program was used to fit both random and fixed effects models. Although both fixed and random effects models are presented as part of the analysis, random effects models allow for generalization to a broader population of studies than do fixed effects models. Hunter and Schmidt (2004) also argue that random effects models are appropriate when population parameters may vary across studies, as is likely here. Confidence intervals reported will be based on the random effects model. Publication bias was assessed using the six methods described above. General agreement between the six measures was considered to be evidence for or against publication bias.

Findings

Table 1 presents the results from the meta-analysis on the entire group of studies as well as for subgroups of studies broken down by levels of potential moderators (outcome variable type, grade level, and risk status).

In all cases, the effect of anti-bullying programs on violent or broader bullying outcomes was positive and significant with an overall r equal to .12. However, as meta-analyses capitalize on pooled sample size, statistical significance in and of itself is not meaningful. Of greater concern is the interpretation of the effect size of the outcome (Cohen, 1994). As can be seen from Table 1, none of the effect sizes cross the zero point. We interpret this as indicating that this effect is "real" and positive in the population. However, the size of the effect is very small. Specifically, converting r to the coefficient of determination r^2 (for ease of understanding, we will denote this in terms of % age $r^2 \times 100$) reveals that the impact of anti-bullying programs ranges from less than 1% impact (for low-risk children) to 3.6% (for high-risk children). Thus, it can be said that although anti-bullying programs produce a small amount of positive change, it is likely that this change is too small to be practically significant or noticeable. Results were best for programs that specifically targeted high-risk youth, although even here, the overall effect size was small.

With regard to publication bias assessment, results are presented for the full selection of studies in Table 2.

Table 2
Publication Bias Results for Main Meta-Analytic Results

	FSN	OFSN	RCT	RT	DTTF	95% CI	Bias?
All studies	2992	0	$p \leq .001$	$p \geq .05$	.10	(.05, .15)	Yes

Note: FSN = Fail-Safe *N*; OFSN = Orwin's (1983) Fail-Safe *N* (based on fixed effects model for $r \geq .10$); RCT = significance of Begg and Mazumdar's (1994) rank correlation test; RT = significance of Egger's Regression (Egger, Davey-Smith, Schneider, & Minder, 1997); DTTF = Corrected r_+ point value for publication bias from Duval and Tweedie's (2000) trim and fill; CI = confidence interval for Duval and Tweedie's trim and fill.

The Fail-Safe *N* procedure suggested that 2,992 unknown studies with zero percent effect size would be necessary to render the results of the meta-analysis nonsignificant. As each additional study would also contribute to the sample size, this is not surprising and this result presents an unrealistically comforting view of the potential for publication bias. Orwin's (1983) Fail-Safe *N* (based on the fixed effects model) suggests that, effectively, zero studies would be necessary to bring the overall effect below the level of $r \leq .10$, which is the point that falls below Cohen's (1992) minimum recommendations for an interpretable effect size. This is also below the $r \geq .20$ that we had adopted as the practical significance cutoff based on Lipsey's (1998) suggestions for intervention research. These results argue against the practical significance of the effectiveness of school-based anti-bullying programs, in particular given that Duval and Tweedie's (2000) correction for publication bias estimates that the true effect is closer to $r = .10$ than $r = .12$.

Tests of homogeneity were positive, with the exception of studies targeting high school students. This suggests the presence of potential moderator effects.

Evaluation of Moderator Effects

In this analysis, outcome variable (violence or broader bullying behaviors), grade level of treatment delivery (elementary school, middle school, high school), and risk status (general school population or children identified as at risk) were examined for potential moderator effects. This was done by correlating each of these variables with the effect size of the individual study. Results of bivariate correlations suggested that outcome effect size is associated with at-risk status ($r = .37$, $p \leq .05$) but not grade level of program implementation ($r = -.15$, $p \geq .05$) or outcome measure ($r = -.09$, $p \geq .05$). This was further examined using a stepwise multiple regression with grade level, outcome variable, and at-risk status as predictor variables and effect size as the outcome variable. Results yielded a regression coefficient, $R = .37$ ($R^2 = .14$), which was statistically significant, $F(1, 43) = 6.7$, $p \leq .01$. In the regression equation, only at-risk status was a significant predictor of effect size ($\beta = .12$).

Discussion

The purpose of this analysis was to examine the effectiveness of school-based anti-bullying programs. Given the expense and effort expended in implementing these programs (Sherman, 2000), it would be useful for policy makers to understand the effect on bullying and even more serious violent behavior. Given that meta-analytic reviews tend to produce significant

results due to pooled sample size, irrespective of the effect size of the results, this study sought to alleviate some of the flaws of meta-analytic interpretation by focusing on the strength of effect size rather than statistical significance, as statistical significance can be misleading (Bobko & Stone-Romero, 1998). Specifically, examining the overall effect size as well as the confidence interval around the effect size can help us understand whether an effect appears to be a true effect in the population as well as whether this effect has practical significance. Lipsey (1998) suggested $r \geq .20$ as a cutoff for practical significance and Cohen (1992) suggested a cutoff of $r \geq .10$ for a small effect. The overall effect size in this analysis was (using the random effects model) $r = .12$, which falls below Lipsey's suggestion but is slightly above Cohen's cutoff. Examining the coefficient of determination, the proportion of variance in violent behavior positively affected by anti-bullying programs is 1.4%. Examining the confidence interval around the effect size, the confidence interval never crosses the zero point, which lends support to the conclusion that school-based anti-bullying programs produce a true effect in the population. However, adjusting for observed publication bias, the adjusted effect size appears closer to $r = .10$. Furthermore, Orwin's Fail-Safe N (which uses the fixed effects model) suggested that effectively zero null-results studies would be necessary to bring the overall effect size below either Cohen's or Lipsey's recommendations for trivial results. Thus, it is not possible to conclude that school-based anti-bullying programs produce a meaningful or practically significant effect on bullying or violent behavior among schoolchildren.

Examining moderator effects, it appears that the effectiveness of school-based anti-bullying programs for at-risk youth may bear somewhat more fruit, with the coefficient of determination closer to 3.6%. Thus, to the extent that anti-bullying programs are used in the schools, it may be more effective to direct them specifically at youth who are identified as at risk. It should be noted, however, that even this effect size is fairly small and does not meet Lipsey's (1998) recommended minimum effect size for interventions.

It is worth noting that one relevant issue not examined in this analysis that is likely to have a significant effect on programmatic success is that of program fidelity. In other words, program staff who are more reliable in following the standardized intervention may have a moderating effect on the success of the program. Examining programmatic fidelity as a moderator variable in a meta-analysis such as this one is difficult to achieve, however, as most published programs imply or claim to have program fidelity or do not provide data on program fidelity. However, this issue is certainly one for further research.

Conclusion

It is concluded that school-based anti-bullying programs are not practically effective in reducing bullying or violent behaviors in the schools. This conclusion is likely to be disappointing for policy makers and the general populace given the increased interest in targeting bullying and other violence in the schools. Given the laudable goal of reducing violence in the schools and the enormous financial and research effort devoted to developing anti-bullying programs, it would be useful to consider why these programs demonstrate such limited effectiveness. Several explanations are offered here, although further research would be necessary to truly understand this phenomenon.

The first suggestion is that bullying, in particular (as opposed to more serious violence), is more advantageous to bullies than is nonbullying. Although bullies are sometimes thought of as being insecure and, in effect, victims of their own aggression, research has suggested the opposite. Aside from aggressive and dominant personality traits, bullies have high self-esteem and are effectively normal children (Olweus, 1995). Bullying and violent behavior may, in many cases, be an effective strategy in climbing the social dominance hierarchy among children at the expense of other children. As anti-bullying programs may encourage equality in problem solving that would involve bullies reducing their social dominance, there may simply be no incentive offered by these programs to entice bullies or violent children to follow these strategies. Put another way, anti-bullying programs may be of benefit to victims of bullying at the potential expense of the social dominance that bullies enjoy. As bullies or violent children see no personal benefit in following the program recommendations, they simply reject them. Nonviolent children may attempt to follow the program suggestions but may ultimately become frustrated when violent children do not. Thus, in the absence of effective incentives for the bullies themselves, there is little reason to believe that these programs would succeed.

A second suggestion is that the behaviors in question, namely, antisocial and violent behavior, may involve both genetic as well as nongenetic causal influences. It has become common to suggest that violent behavior is learned (e.g., American Psychological Association, 2006), which may be a politically expedient position rather than one based on science. Indeed, behavioral genetics studies of antisocial behavior and violence suggest that genetic inheritance predicts at least 50% of these behaviors, ranging from bullying and school violence to criminally violent and antisocial acts (Ferguson, 2007b; Larsson, Andershed, & Lichtenstein, 2006; Rhee & Waldman, 2002). Research on specific genes has identified the 5-HTT serotonin transporter gene (Retz, Retz-Junginger, Supprian, Thome, & Rosler, 2004), the valine/methionine variant in the catechol O-methyltransferase (COMT) gene (Thapar et al., 2005), and the MAOA gene located on the X-chromosome (Caspi et al., 2002) as being involved in violent behavior (although Huizinga et al., 2006, could not replicate the finding for MAOA). Although such genes certainly interact with the environment to produce behavior, treatments that remain uninformed of the genetic basis of violent behavior may be at a disadvantage. Identifying violence as genetically based does not mean that it could not be malleable to behaviorally based intervention strategies. However, a biologically based phenomenon may prove somewhat more resistant to behavioral interventions, and any such intervention may need to be aware of the biological basis of the phenomenon if that intervention is to succeed.

A third suggestion may be that these programs are being implemented during a time in which school violence has already decreased significantly (for reasons that are poorly understood) and that children's violent behavior has already hit a "floor effect." As noted earlier, significant violent events in the schools have declined significantly over the past decade (National Center for Education Statistics, 2005). Although it is unclear that bullying behaviors have necessarily followed suit, it is possible that this is the case as well. In particular, when the majority of school-based anti-bullying programs are targeted at low-risk general population youth, the implementation of these programs in terms of both population and timing may be flawed. Programs targeted at seriously at-risk youth may be more effective and a better expenditure.

With regard to policy implications, this meta-analysis provides information that may be of significant value in making decisions about whether or how to implement school-based anti-bullying programs. In particular, this study highlights how focusing on statistical significance may provide misleading information and lead to flawed policy. Ultimately, it will be the responsibility of policy makers to decide whether the cost of implementation of these programs is worth a relatively small payoff (it is unlikely that a 1.4% difference in the behavior of individuals would be noticeable). Policy makers are cautioned to examine the performance of specific anti-bullying programs, as several anti-violence programs (e.g., Metropolitan Area Child Study Research Group, 2002) actually increased aggression under some circumstances. Nonetheless, it is important that policy be informed by the actual effect that these programs have rather than simply by examining statistical significance.

In sum, this study sought to examine the overall effect of school-based anti-bullying programs on bullying and other serious violent behaviors. Results of this study suggest that anti-bullying programs produce an effect that is positive and statistically significant but practically negligible. It is hoped that this study will stimulate discussion in this area and provide an impetus for improving school-based anti-bullying programs.

References

Ahmed, E., & Braithwaite, V. (2006). Forgiveness, reconciliation, and shame: Three key variables in reducing school bullying. *Journal of Social Issues*, *62*(2), 347-370.

Ambert, A. (1994). A qualitative study of peer abuse and its effects: Theoretical and empirical implications. *Journal of Marriage and Family*, *56*, 119-130.

American Psychological Association. (2006). *An APA brochure on youth violence*. Retrieved September 2, 2006, from http://www.apa.org/ppo/issues/pbviolence.html

Begg, C., & Mazumdar, G. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics*, *50*, 1088-1101.

Bobko, P., & Stone-Romero, E. (1998). Meta-analysis may be another useful research tool but it is not a panacea. In G. Ferris (Ed.), *Research in personnel and human resources management, Vol. 16* (pp. 359-397). Greenwich, CT: JAI Press.

Bollmer, J. M., Harris, M. J., & Milich, R. (2006). Reactions to bullying and peer victimization: Narratives, physiological arousal, and personality. *Journal of Research in Personality*, *40*, 803-828.

Braithwaite, J. (1989). *Crime, shame, and integration*. Cambridge, UK: Cambridge University Press.

Caspi, A., McClay, J., Moffitt, T., Mill, J., Martin, J., Craig, I., et al. (2002). Role of genotype in the cycle of violence in maltreated children. *Science*, *297*, 851-854.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*, 155-159.

Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, *49*, 997-1003.

DeVoe, E., Dean, K., Traube, D., & McKay, M. (2005). The SURVIVE community project: A family-based intervention to reduce the impact of violence exposures in urban youth. *Journal of Aggression, Maltreatment and Trauma*, *11*, 95-116.

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*, 89-99.

Egger, M., Davey-Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple graphical test. *British Medical Journal*, *315*, 629-634.

Farrell, A., Meyer, A., Kung, E., & Sullivan, T. (2001). Development and evaluation of school-based violence prevention programs. *Journal of Clinical Child Psychology*, *30*, 207-220.

Farrell, A., Meyer, A., Sullivan, T., & Kung, E. (2003). Evaluation of the Responding in Peaceful and Positive Ways (RIPP) seventh grade violence prevention curriculum. *Journal of Child and Family Studies*, *12*(1), 101-120.

Farrell, A., Meyer, A., & White, K. (2001). Evaluation of Responding in Peaceful and Positive Ways (RIPP): A school-based prevention program for reducing violence among urban adolescents. *Journal of Clinical Child Psychology*, *30*(4), 451-463.

Ferguson, C. J. (2007a). Evidence for publication bias in video game violence effects literature: A meta-analytic review. *Aggression and Violent Behavior*, *12*, 470-482.

Ferguson, C. J. (2007b). *Genetic contributions to antisocial personality and behavior: A meta-analytic review from an evolutionary framework*. Unpublished manuscript.

Gini, G. (2006). Social cognition and moral cognition in bullying: What's wrong? *Aggressive Behavior*, *32*, 528-539.

Haynie, D. L., Nansel, T., & Eitel, P. (2001). Bullies, victims, and bully/victims: Distinct groups of at-risk youth. *Journal of Early Adolescence*, *21*, 29-49.

Hazler, R. J., Carney, J. V., & Granger, D. A. (2006). Integrating biological measures into the study of bullying. *Journal of Counseling and Development*, *84*, 298-307.

Heydenberk, R. A., Heydenberk, W. R., & Tzenova, V. (2006). Conflict resolution and bully prevention: Skills for school success. *Conflict Resolution Quarterly*, *24*(1), 55-69.

Huizinga, D., Haberstick, B., Smolen, A., Menard, S., Young, S., Corley, R., et al. (2006). Childhood maltreatment, subsequent antisocial behavior, and the role of monoamine oxidase A genotype. *Biological Psychiatry*, *60*, 677-683.

Hunter, J., & Schmidt, F. (2004). *Methods of meta-analysis: Correcting error and bias in research findings*. Thousand Oaks, CA: Sage.

Jolliffe, D., & Farrington, D. P. (2006). Examining the relationship between low empathy and bullying. *Aggressive Behavior*, *32*, 540-550.

Kyriakides, L., Kaloyirou, C., & Lindsay, G. (2006). An analysis of the revised Olweus Bully/Victim Questionnaire using the Rasch measurement model. *British Journal of Educational Psychology*, *76*, 781-801.

Lance, C., & Vandenberg, R. (2002). Confirmatory factor analysis. In F. Drascow & N. Schmitt (Eds.), *Measuring and analyzing behavior in organizations: Advances in measurement and data analysis* (pp. 221–224). New York: Jossey-Bass/Pfeiffer.

Larsson, H., Andershed, A., & Lichtenstein, P. (2006). A genetic factor explains most of the variation in the psychopathic personality. *Journal of Abnormal Psychology*, *115*, 221-230.

Lipsey, M. W. (1998). Design sensitivity: Statistical power for applied experimental research. In L. Bickman & D. J. Rog (Eds.), *Handbook of applied social research methods* (pp. 39-68). Thousand Oaks, CA: Sage.

Marini, Z. A., Dane, A. V., Bosacki, S. L., & YLC-CURA. (2006). Direct and indirect bully-victims: Differential psychosocial risk factors associated with adolescents involved in bullying and victimization. *Aggressive Behavior*, *32*, 551-569.

Metropolitan Area Child Study Research Group. (2002). A cognitive-ecological approach to preventing aggression in urban settings: Initial outcomes for high-risk children. *Journal of Consulting and Clinical Psychology*, *70*, 179-194.

Nansel, T. R., Overpeck, M., Pilla, R. S., Ruan, W. J., Simons-Morton, B., & Scheidt, P. (2001). Bullying behaviors among U.S. youth: Prevalence and association with psychosocial adjustment. *Journal of the American Medical Association*, *285*(16), 2094-2100.

National Center for Education Statistics. (2005). *Indicators of school crime and safety: 2005*. Washington, DC: Author.

Olweus, D. (1995). Bullying or peer abuse at school: Facts and interventions. *Current Directions in Psychology*, *4*, 196-200.

Olweus, D., Limber, S., & Mihalic, S. (1999). *The bullying prevention program: Blueprints for violence prevention*. Boulder, CO: Center for the Study and Prevention of Violence. Retrieved October 24, 2006, from http://www.colorado.edu/cspv/safeschools/bullying/bullying_bppsummary.html

Orwin, R. (1983). A Fail-Safe *N* for effect size in meta-analysis. *Journal of Educational Statistics*, *8*, 157-159.

Parkins, I. S., Fishbein, H. D., & Ritchey, P. N. (2006). The influence of personality on workplace bullying and discrimination. *Journal of Applied Social Psychology*, *36*(10), 2554-2577.

Phillips, D. A. (2007). Punking and bullying: Strategies in middle school, high school, and beyond. *Journal of Interpersonal Violence*, *22*(2), 158-178.

Retz, W., Retz-Junginger, P., Supprian, T., Thome, J., & Rosler, M. (2004). Association of serotonin transporter promoter gene polymorphism with violence: Relation with personality disorders, impulsivity and childhood ADHD psychopathology. *Behavioral Sciences and the Law*, *22*, 415-425.

Rhee, S., & Waldman, I. (2002). Genetic and environmental influences on antisocial behavior: A meta-analysis of twin and adoption studies. *Psychological Bulletin*, *128*, 490-529.

Rosenthal, R., & Rosnow, R. (1991). *Essentials of behavioral research: Methods and data analysis*. New York: McGraw-Hill.

Rothstein, H., Sutton, A., & Borenstein, M. (2005). *Publication bias in meta-analysis: Prevention, assessment, and adjustments*. Hoboken, NJ: Wiley.

Sharp, S., & Smith, P. (1994). *Tackling bullying in your school: A practical handbook for teachers*. London: Routledge.

Sherman, L. W. (2000). The safe and drug-free schools act. *Brookings Papers on Education Policy*, pp. 125-156.

Taub, J. (2001). Evaluation of the Second Step Violence Prevention Program at a rural elementary school. *School Psychology Review*, *31*, 186-200.

Thapar, A., Langley, K., Fowler, T., Rice, F., Turic, D., Whittinger, N., et al. (2005). Catechol O-methyltransferase gene variant and birth weight predict early-onset antisocial behavior in children with attention-deficit/hyperactivity disorder. *Archives of General Psychiatry*, *62*, 1275-1278.

Van Schoiack-Edstrom, L., Frey, K., & Beland, K. (2002). Changing adolescents' attitudes about relational and physical aggression: An early evaluation of a school-based intervention. *School Psychology Review*, *31*, 201-216.

Wilson, D., Gottfredson, D., & Najaka, S. (2001). School-based prevention of problem behavior: A meta-analysis. *Journal of Quantitative Criminology*, *17*, 247-272.

Wilson, S., Lipsey, M., & Derzon, J. (2003). The effects of school-based intervention programs on aggressive behavior: A meta-analysis. *Journal of Consulting and Clinical Psychology*, *71*, 136-149.

Christopher J. Ferguson, PhD, is an assistant professor of clinical and forensic psychology within the Department of Behavioral, Applied Sciences and Criminal Justice at Texas A&M International University in Laredo. He holds a degree in clinical psychology from the University of Central Florida in Orlando. His research interests have focused on violent behavior, positive and negative effects of violent video games, and evolutionary psychology.

Claudia San Miguel, PhD, is an assistant professor of criminal justice within the Department of Behavioral, Applied Sciences and Criminal Justice at Texas A&M International University in Laredo. She is the coauthor of *The Death Penalty: Constitutional Issues, Commentaries, and Cases*. She is currently working on research involving border violence between the United States and Mexico and intelligence-led policing. She is also working on research involving the trafficking of women and children. In addition to teaching courses in policing, legal issues, and criminology, she has also taught at the International Law Enforcement Academy located in Roswell, New Mexico. The academy is sponsored by the U.S. Department of State.

John C. Kilburn, Jr., PhD, is the department chair of Behavioral, Applied Sciences and Criminal Justice at Texas A&M International University in Laredo. His research interests include community security and reactions to crime. His previous work has been published in journals such as *Criminal Justice Review*, *Urban Affairs Review*, and *Social Forces*.

Patricia Sanchez, MA, is a graduate of the Master's of Arts in Counseling Psychology program at Texas A&M International University in Laredo. Currently, she provides family counseling at the STCADA clinic. She is studying for her LPC and LCDC licenses in Texas.