The author(s) shown below used Federal funds provided by the U.S. Department of Justice and prepared the following final report:

Document Title: The Effectiveness of School-Based Violence Prevention Programs for Reducing Disruptive and Aggressive Behavior

Author(s): Sandra Jo Wilson, Mark W. Lipsey

Document No.: 211376

Date Received: September 2005

Award Number: Commissioned Paper

This report has not been published by the U.S. Department of Justice. To provide better customer service, NCJRS has made this Federally-funded grant final report available electronically in addition to traditional paper copies.

Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.
The Effectiveness of School-Based Violence Prevention Programs for Reducing Disruptive and Aggressive Behavior

Sandra Jo Wilson and Mark W. Lipsey
Center for Evaluation Research and Methodology
Institute for Public Policy Studies
Vanderbilt University

Revised Report for the National Institute of Justice
School Violence Prevention Research Planning Meeting
May 2005

---

1 This work was supported by grants from the National Institute of Mental Health (MH57766) and the William T. Grant Foundation.
The Effectiveness of School-Based Violence Prevention Programs for Reducing Disruptive and Aggressive Behavior

Executive Summary

- More than 200 research studies of the effectiveness of school-based programs for preventing or reducing aggressive behavior were located through vigorous search and synthesized using meta-analysis techniques.

- Most of the studies involved programs implemented primarily for research or demonstration purposes with a high degree of researcher involvement. Less than 15% involved evaluations of programs implemented under routine circumstances.

- Students from preschool to high school were included, but the literature emphasized programs for elementary and middle school students.

- School violence prevention programs fell into four broad categories: universal programs, selected/indicated programs, comprehensive programs, and special schools/classes. The mean effect sizes for each category were as follows:
  - Universal programs: ES=.18
  - Selected/Indicated programs: ES=.29
  - Comprehensive programs: ES=.06
  - Special schools/classes: ES=.06

- Treatment dose (in the form of treatment duration, frequency, or implementation quality) was uniformly influential. Programs with no or few implementation difficulties or with greater duration or frequency tended to produce larger reductions in aggressive behavior.

- The outcomes for different treatment modalities within the universal and selected/indicated formats (e.g., social skills training, cognitively-oriented programs, behavioral programs, counseling) were not significantly different; that is, the modalities appeared to be equally effective at reducing aggressive behavior.

- In general, larger program effects were achieved with relatively higher risk students.

- School violence programs were generally effective at reducing the more common types of aggressive behavior seen in schools, including fighting, name-calling, intimidation, and other negative interpersonal behaviors, especially among higher risk students. It is still unknown, however, whether the types of programs reviewed here (or any others) would be effective in preventing the rare but serious incidents of school violence perpetrated by very disturbed youth.
Introduction

Extremely violent events in schools draw our attention to programs designed to prevent and reduce school violence. Highly publicized school shootings, though tragic, are thankfully rare. In fact, less than 1% of the over 2,000 homicides of school-aged children in the 1999-2000 school year happened at school (DeVoe, et al., 2004). Further, over the past 10 years, victimization of school-aged children has declined both at and away from school. Nevertheless, some forms of antisocial behavior are common in schools. For example, according to principal reports from the 1999-2000 school year, 71% of public schools experienced a violent crime and over half took serious disciplinary action for some children (NCES, 2004).

Teachers and students also report negative behaviors in schools. According to Gottfredson, et al. (2000) teachers report that student misconduct is common and interferes with their efforts to teach. Among middle school students surveyed, 22% reported being threatened and 41% reported hitting or threatening to hit other students. Fewer high school students report such incidents (16% report being threatened and 32% report hitting or threatening to hit other students), but the numbers are not trivial. Minor behavior problems appear to be relatively common in schools. In addition, though serious violence is rare, some schools can be dangerous places for children.

There are a variety of approaches to school violence prevention and maintaining safe schools, including surveillance (e.g., metal detectors, security guards, etc.), deterrence (e.g., rules, regulations, zero tolerance policies), and psychosocial intervention. This report focuses on summarizing the effectiveness of programs of the latter sort. This paper presents an update of the meta-analysis reported in Wilson, Lipsey, and Derzon (2003). That article included research reported up to early 2000. This report includes research through 2004.

Because extreme violence is rare in schools, the psychosocial programs that are the subject of this report typically focus on more common forms of aggressive behavior such as fighting, bullying, verbal conflict, and disruptive behavior. These behaviors, even when not overtly violent, may inhibit learning and create interpersonal problems for those involved. In addition, minor forms of aggressive behavior can escalate (Garofalo, Siegel, & Laub, 1987) and schools that do not effectively counteract this progression may create an environment in which violence is normatively acceptable (Goldstein, Harootunian, & Conoley, 1994). Thus, it is appropriate for schools to attempt to reduce behaviors such as fighting, name-calling, bullying, and general intimidation that can create a negative school climate and lead to more serious violence. The issue addressed in this paper is the effectiveness of school-based programs for preventing or reducing these forms of aggressive and disruptive behavior.

Such programs, generally called violence prevention programs, are ubiquitous in U.S. schools (Gottfredson, et al., 2000). This terminology is somewhat misleading, however. Many of these programs do not directly target violence but, rather, seek to change such potential mediators as social and communication skills. Nor are they necessarily preventive — some serve children who already exhibit serious behavior problems,
including violence. Nonetheless, we will refer to these interventions as violence prevention programs because the term is familiar to most practitioners, the same strategies are applied for both preventive and treatment purposes, and the goal of these interventions, implicitly if not explicitly, is violence prevention.

Method

Eligibility Criteria

Studies were selected for this meta-analysis based on a set of detailed criteria, summarized as follows:

1. The study was reported in English no earlier than 1950 and involved a school-based program for children attending any grade, pre-K through 12th.
2. The study assessed intervention effects for at least one outcome variable that represented either (a) aggressive or violent behavior (e.g., fighting, bullying, person crimes); (b) disruptive behavior (e.g., behavior problems, conduct disorder, acting out); or (c) problem behavior (i.e., measures that include both internalizing and externalizing problems).
3. The study used an experimental or quasi-experimental design that compared students exposed to one or more identifiable interventions with one or more control or comparison conditions on at least one qualifying outcome variable. To qualify as an experimental/quasi-experimental design, a study was required to meet at least one of the following criteria:
   - Subjects were randomly assigned to conditions;
   - Subjects in the treatment and control conditions were matched and the matching variables included a pretest for at least one qualifying outcome variable or an acceptable proxy;
   - If subjects were not randomly assigned or matched, the study must have both a pretest and a posttest on at least one qualifying outcome variable or sufficient demographic information to describe the initial equivalence of the treatment and control groups.

Search and Retrieval of Studies

An attempt was made to identify and retrieve the entire population of school violence prevention studies, published and unpublished, that met the eligibility criteria specified above. The primary source was a comprehensive search of bibliographic databases, including Psychological Abstracts, Dissertation Abstracts International, ERIC (Educational Resources Information Center), U.S. Government Printing Office publications, National Criminal Justice Reference Service, and MedLine. Second, the bibliographies of previous meta-analyses and literature reviews (e.g., Durlak, 1995; 1997; Lösel & Beelmann, 2003; Wilson, Gottfredson, & Najaka, 2001) and the tables of contents of relevant journals were reviewed for eligible studies. Finally, the bibliographies of retrieved studies were themselves examined for candidate studies. Identified studies were retrieved from the library, obtained via interlibrary loan, or requested directly from the author. We obtained and screened more than 95% of the reports identified as potentially eligible through these sources.
We are currently in the process of updating the database to include literature published since the mid-2000 cutoff date for the original article. This update is not yet complete and, thus, the literature included herein that was published between 2001 and 2005 is not comprehensive.

**Coding of Study Reports**

Eligible studies were coded by trained research assistants familiar with social science research. Coding was done directly into a computer database using pre-formatted computer screens and supported by detailed computerized and paper coding manuals. The coding was reviewed by the first author and disagreements or questions were resolved through discussion with the coding team. In addition, to assess coder reliability, approximately ten percent of the studies were selected at random and recoded by a different coder.

**Effect Size Coding.** Study findings were coded to represent the mean difference in aggressive behavior between experimental conditions at the posttest measurement. The effect size statistic used for these purposes was the standardized mean difference (Cohen, 1988; Lipsey & Wilson, 2001), defined as the difference between the treatment and control group means on an outcome variable divided by their pooled standard deviation. This effect size statistic indexes the outcomes for the treatment group relative to the control group in standard deviation units. Thus ES=.50 indicates that the outcome for the treatment group was more favorable than that for the control group by an amount equivalent to half a standard deviation on the respective outcome measure.

**Study Descriptor Coding.** In addition to effect size values, information was coded for each study that described the methods and procedures, the intervention, and the subject samples. The items coded to describe study methods and procedure included details of the design, measures, and attrition. Those coded to describe the subject samples included age, gender, ethnicity, prior antisocial behavior, and risk for later antisocial behavior. The intervention was described by coding the type of program, duration, intensity, setting, and format of the program, delivery personnel, and other such characteristics.

The reliability of the coding for the study descriptors was generally high. For categorical items, intercoder agreement ranged from 73% to 100%. For continuous items, the intercoder correlations ranged from .76 to .99. A copy of the full coding protocol is available from the authors.

**Results**

**General Effects of School Violence Prevention Programs**

The literature retrieval and coding process yielded data from 372 school-based studies. The bibliographic references for these studies are included in Appendix A. The research studies included in this meta-analysis examined program effects on many different
outcomes, ranging from aggression and violence to social skills, academic performance, and self-esteem. This report will focus primarily on the outcomes most relevant to school violence prevention, namely aggressive behavior. However, we will first present the overall means and confidence intervals for each outcome category.

We have grouped each outcome measure from our eligible studies into one of fourteen categories, the weighted mean effect sizes and 95% confidence intervals for which are shown in Figure 1 below. The construct categories are defined as follows. Our main outcome of interest is aggressive and disruptive behavior, which involves a variety of negative interpersonal behaviors including fighting, hitting, bullying, verbal conflict, disruptiveness, acting out and the like. The most common type of measure in this category is a teacher-report survey. Next, there are three categories of behavior problems that are closely related to aggression. These are problem behavior (i.e., measures that include both internalizing and externalizing behaviors like the CBCL Total score), anger/hostility/rebelliousness, and activity level/attention problems. We have a small number of outcomes relating to children’s relationships with antisocial peers as well as a small number of substance use outcomes. Next, there are two categories relating to social adjustment. The first, and most common after aggression/disruption, includes measures of specific social skills like communication skills, social problem solving, conflict resolution skills, and the like. Social adjustment, on the other hand, focuses more on measures of how well children get along with their peers, i.e., do they have friends? are they well-liked or rejected? etc. The two categories of school outcomes are school performance (e.g., achievement tests, grades) and school participation (e.g., tardiness, truancy, dropout). The personal adjustment category includes measures of self-esteem, self-concept, and other measures of general well-being. Next, we have a category that includes various measures of students’ knowledge and attitudes about problem behavior. Internalizing includes measures of anxiety, depression, and the like. And, the family outcomes category includes a wide range of family functioning variables.

Figure 1 shows that the confidence intervals for all the variables (except three) do not cross the zero line – this means that the mean effect sizes are statistically significant. This suggests that, in general, school violence programs have positive effects on a wide variety of outcomes. The effects are particularly strong in the area of social skills, which is not surprising since the majority of programs specifically target aspects of social behavior. Substance use, antisocial peers, and family relations are not significantly different from zero, indicating that the programs in our database did not significantly impact these outcomes. But, these are also the three smallest categories of outcomes and were not generally the primary target of any of the programs we reviewed.

---

2 The problem behavior outcomes (which provide only a total score for internalizing and externalizing problems combined) are not included with the aggressive behavior outcomes that are the focus of this paper.

3 Ideally, we would have liked to examine program effects only on aggressive behavior. However, almost none of the measures that call themselves “aggressive” behavior measures focus solely on physically aggressive interpersonal behavior. Many include disruptiveness, acting out, and other forms of behavior problems that are negative, but not necessarily aggressive.
General Characteristics of Studies Measuring Program Effects on Aggressive and Disruptive Behavior

Of the 372 school-based studies, 236 studies included outcomes related to aggressive behavior. We excluded a small number of programs from that number (n=17) because they did not directly target aggressive behavior or behaviors closely related to aggression (e.g., social skills). These programs were either academically oriented or focused on a specific target population (e.g., kids from divorced families). Although these programs may have important influences on aggressive behavior in school, these benefits are secondary to the programs’ primary goals. In addition, most schools would not select programs of this type as their primary violence prevention strategy.
Therefore, this report focuses on 219 studies. These studies generated over 600 posttest group comparison effect sizes on some form of aggressive or disruptive behavior and represent nearly 50,000 individual subjects. The general characteristics of these studies are shown in Table 1. Ninety percent of the studies were conducted in the United States and nearly 75% were conducted by researchers in the fields of psychology or education. Fewer than 20% of the studies were conducted prior to 1980 and most were published in peer-reviewed journals (60%), with the remainder reported in unpublished media such as dissertations, theses, conference papers, and technical reports.

The subject samples had a range of demographic characteristics. Most samples were comprised of a mix of boys and girls, but some all boy samples (16%) and a few all girl samples (6%) were also present. Minority children were well represented with about a third of the studies having primarily minority youth. However, an additional third of the studies did not report ethnicity information on their subject samples. All school ages were included, from preschool through high school; the average age was around 10. A range of risk levels was also present, from general population students to students already exhibiting aggressive behavior. The modal risk designation was selected, that is, the sample was selected as “at risk” for aggressive or violent behavior based on individual characteristics, such as activity level. Students in the selected, environment category were those who were not selected individually for treatment, but their schools were selected because of some environmental or neighborhood risk, such as poverty or high crime rates. General population samples (i.e., not selected because of the risk status of the subjects) represented about one fourth of the groups, and indicated samples represented about twenty percent. Indicated youth were those selected for intervention because they exhibited serious behavior problems such as violence or aggression.

A wide range of program and methodological characteristics is evident from Table 1. Most notable among these is that most studies were conducted mainly for research purposes with high levels of researcher involvement (research and demonstration programs), that nearly two-thirds of the programs were less than 20 weeks in length, and that almost 40% suffered from implementation problems.

---

4 The original meta-analysis published in the Journal of Consulting and Clinical Psychology included 221 studies, but only 172 of these involved treatment-control designs; thus the sample of studies presented here includes an additional 47 studies.
Table 1. Characteristics of the Studies in the Meta-Analysis

<table>
<thead>
<tr>
<th>Variable</th>
<th>N</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Subject Characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gender mix</td>
<td></td>
<td></td>
</tr>
<tr>
<td>All males (&gt; 95%)</td>
<td>36</td>
<td>16</td>
</tr>
<tr>
<td>&gt; 60% males</td>
<td>59</td>
<td>27</td>
</tr>
<tr>
<td>50% - 60% males</td>
<td>73</td>
<td>33</td>
</tr>
<tr>
<td>&lt; 50% males</td>
<td>25</td>
<td>11</td>
</tr>
<tr>
<td>No males (&lt; 5%)</td>
<td>14</td>
<td>6</td>
</tr>
<tr>
<td>Missing</td>
<td>12</td>
<td>6</td>
</tr>
<tr>
<td>Age of group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preschool &amp; kindergarten</td>
<td>19</td>
<td>8</td>
</tr>
<tr>
<td>6 through 10</td>
<td>96</td>
<td>44</td>
</tr>
<tr>
<td>11 through 13</td>
<td>60</td>
<td>27</td>
</tr>
<tr>
<td>14 and up</td>
<td>44</td>
<td>20</td>
</tr>
<tr>
<td>Predominant ethnicity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>64</td>
<td>29</td>
</tr>
<tr>
<td>Black</td>
<td>52</td>
<td>24</td>
</tr>
<tr>
<td>Hispanic</td>
<td>16</td>
<td>7</td>
</tr>
<tr>
<td>Other minority</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Mixed ethnicity</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>Missing</td>
<td>75</td>
<td>34</td>
</tr>
<tr>
<td>Risk level</td>
<td></td>
<td></td>
</tr>
<tr>
<td>General population</td>
<td>52</td>
<td>24</td>
</tr>
<tr>
<td>Selected, environment</td>
<td>25</td>
<td>11</td>
</tr>
<tr>
<td>Selected, individual characteristics</td>
<td>103</td>
<td>47</td>
</tr>
<tr>
<td>Indicated</td>
<td>39</td>
<td>18</td>
</tr>
<tr>
<td><strong>Program Characteristics</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Demonstration vs. routine practice</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Research programs</td>
<td>107</td>
<td>49</td>
</tr>
<tr>
<td>Demonstration programs</td>
<td>83</td>
<td>38</td>
</tr>
<tr>
<td>Routine practice</td>
<td>29</td>
<td>13</td>
</tr>
<tr>
<td>Program duration (weeks)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 to 6 weeks</td>
<td>44</td>
<td>20</td>
</tr>
<tr>
<td>7 to 19 weeks</td>
<td>93</td>
<td>43</td>
</tr>
<tr>
<td>20 to 37 weeks</td>
<td>48</td>
<td>22</td>
</tr>
<tr>
<td>38 and up</td>
<td>34</td>
<td>16</td>
</tr>
<tr>
<td>Frequency of service</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Less than weekly</td>
<td>21</td>
<td>10</td>
</tr>
<tr>
<td>1 to 2x per week</td>
<td>114</td>
<td>52</td>
</tr>
<tr>
<td>3 to 4x per week</td>
<td>23</td>
<td>11</td>
</tr>
<tr>
<td>Daily</td>
<td>61</td>
<td>28</td>
</tr>
<tr>
<td>Implementation quality</td>
<td></td>
<td></td>
</tr>
<tr>
<td>No problems or none mentioned</td>
<td>137</td>
<td>63</td>
</tr>
<tr>
<td>Possible implementation problems</td>
<td>37</td>
<td>17</td>
</tr>
<tr>
<td>Explicit implementation problems</td>
<td>45</td>
<td>21</td>
</tr>
</tbody>
</table>

*Percentages may not add up to 100 because of rounding.

bIt was often impossible to distinguish between a study with no attrition between pretest and posttest and a study that reported only the number of subjects available at posttest. Thus, although no attrition and unreported attrition are clearly different, they are, of necessity, combined in the same category.
**Effect Size Adjustments**

Effect sizes based on small samples are known to be biased; to adjust for this, all effect sizes were multiplied by the small sample correction factor, \(1 - \left(\frac{3}{4n} - 9\right)\), where \(n\) is the sample size for the study (Lipsey & Wilson, 2001). Also, each effect size was weighted by its inverse variance in all computations so that its contribution was proportionate to its reliability (Hedges & Olkin, 1985). Examination of the effect size distribution identified a small number of outliers with potential to distort the analysis; these were recoded (i.e., Windsorized) to less extreme values (Hedges & Olkin, 1985; Lipsey & Wilson, 2001). In addition, several studies used unusually large samples. Because the inverse variance weights chiefly reflect sample size, those few studies would dominate any analysis in which they were included. Therefore, the extreme tail of the sample size distribution was recoded to a maximum of 250 subjects per intervention or control group for the computation of weights. These adjustments to outliers allow us to retain them in the analysis with high-end values, but make those values less extreme so that they do not exercise highly disproportionate influence on the analysis results.

To create sets of independent effect size estimates for analysis, only one effect size from each subject sample was used in any analysis. When more than one was available, the effect size from the measurement source (or informant) most frequently represented across all studies (e.g., teachers’ reports, self-reports) was selected. We wanted to retain informant as a variable for analysis, so did not elect to average across effect sizes from different informants when more than one was reported. If there was more than one effect size from the same informant or source, however, their mean value was used.

Finally, many studies provided data sufficient for calculating mean difference effect sizes on the outcome variables at the pretest. In cases where pretest effect sizes were available, we adjusted the posttest effect sizes for pretest differences by subtracting the pretest value from the posttest value. In the regression models presented below, we tested whether there were systematic differences between effect sizes that were adjusted and those that were not by including dummy codes for adjustment in the regression models.

**Program Format and Treatment Modality**

The collection of school-based programs analyzed here represents a wide range of intervention programming and illustrates the variety of strategies available for school-based intervention. We have divided these programs into four groups based on the general format of the programs and (in some cases) by treatment modality within format. The different formats tend to differ on a number of methodological, participant, and intervention characteristics that make it unwise to combine them into a single analysis. The four intervention formats are as follows:

- **Universal programs**: these programs are delivered in classroom settings to the entire classroom; children are generally not selected individually for treatment but receive treatment simply because they are students in a program classroom. However, schools are frequently selected because they are in low socioeconomic
status and/or high crime neighborhoods. The children in these universal programs may be considered at risk by virtue of their socioeconomic background or neighborhood risk.

- **Selected/Indicated programs**: these programs are delivered to students who are selected especially to receive treatment by virtue of the presence of some risk factor, including disruptiveness, aggressive behavior, activity level, etc. Most of these programs are delivered to the selected children outside of their regular classrooms (and may use either group or individual formats), although some programs are delivered in the regular classrooms but are targeted for the selected children.

- **Special schools or classes**: these programs involve special schools or classrooms that (for the students involved) serve as a usual classroom or school. Children are placed in these special schools or classrooms because of some behavioral or school difficulty that is judged to warrant their placement outside of mainstream classrooms. The programs in this category include special education classrooms for behavior disordered children, alternative high schools, and schools within schools programs.

- **Comprehensive/Multimodal programs**: these programs generally involve multiple modalities and multiple formats, including both classroom-based and pull-out programs. They may also involve programs for parents and capacity building components for school administrators and teachers in addition to the programming provided for the students. The defining characteristic of these programs is that they include multiple treatment elements and formats.

The weighted mean effect sizes and 95% confidence intervals for each program format are shown in Figure 2. The confidence intervals for the selected/indicated, universal, and comprehensive programs are beyond the zero line, indicating that the means for these programs are statistically significant. In addition, the means for selected/indicated and universal programs are significantly different from the mean for comprehensive programs. Does this mean that selected/indicated or universal programs are “better” than comprehensive programs? We don’t believe so. There are important differences in study and subject characteristics across the different formats. Because of these differences, it makes little sense to compare outcomes across formats. For example, the selected/indicated programs are delivered to specially identified and selected children; it is not surprising that this format has the largest overall mean since the children in these programs likely have significant room for improvement. Although some comprehensive programs have components for specially selected children, we would not expect the outcomes on a whole school to be similar to the kinds of outcomes you might expect with a targeted programs on a specialized samples. Thus, we do not believe we can compare results across formats and will therefore present the results separately by format. This separation by format has the added benefit of allowing us to better identify important study and subject characteristics that are associated with positive outcomes for particular program formats. We will now turn to an examination of study results for each of the four program formats.
Results for Universal Programs

There were sixty-one universal programs in the database. All of these were delivered in classroom settings to entire classes of students.\(^5\) Four treatment modalities were used, as shown in Table 2. The overall weighted mean posttest effect size for the universal programs was .18 and was significantly different from zero.

<table>
<thead>
<tr>
<th>Modality</th>
<th>Description</th>
<th>Prototypical Program</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavioral strategies</td>
<td>Interventions involve the use of various behavioral techniques, such as rewards, token economies, contingency contracts, and the like to modify or reduce inappropriate behavior.</td>
<td>Good Behavior Game</td>
<td>5</td>
</tr>
<tr>
<td>Cognitively-oriented programs</td>
<td>Interventions focus on changing thinking processes or cognitive skills; programs focus on solving social problems, controlling anger, inhibiting hostile attributions, etc.</td>
<td>I Can Problem Solve</td>
<td>37</td>
</tr>
</tbody>
</table>

\(^5\)There were three universal programs that were delivered to entire classrooms, but certain children (those at risk) were selected for analysis. These were retained in the universal format category because the experiences of these children were more similar to the universal programs than the selected/indicated programs.
Social skills programs

Interventions are designed to help youth better understand social behavior and learn appropriate social skills. Children learn communication skills, fighting avoidance skills, group entry skills, eye contact, “I” statements, etc.

Counseling, Talk therapy

These programs utilized traditional group therapy techniques in classroom settings.

<table>
<thead>
<tr>
<th>Social skills programs</th>
<th>Interventions are designed to help youth better understand social behavior and learn appropriate social skills. Children learn communication skills, fighting avoidance skills, group entry skills, eye contact, “I” statements, etc.</th>
<th>Skillstreaming</th>
<th>17</th>
</tr>
</thead>
<tbody>
<tr>
<td>Counseling, Talk therapy</td>
<td>These programs utilized traditional group therapy techniques in classroom settings.</td>
<td>Reality Therapy</td>
<td>2</td>
</tr>
</tbody>
</table>

General Moderators of Observed Effects on Aggressive and Disruptive Behavior for Universal Programs

Tests of the homogeneity of the effect sizes using the Q-statistic (Hedges & Olkin, 1985) showed more variability across studies than expected from subject-level sampling error ($Q_{60} = 87$). That is, some studies produced effect sizes that were larger than the corresponding mean across studies while others produced effects that were smaller. This variation was expected to be associated with the nature of the interventions, subjects, and methods in the studies of universal programs. Our next step, therefore, was to identify the study characteristics most strongly associated with effect size. An inverse variance weighted multiple regression analysis was conducted using mixed effects models (Raudenbush, 1994). The dependent variable in these analyses was the effect size for aggressive behavior.

In the first stage of this analysis, different clusters of variables were entered into the model one at a time with the general features of the study methods and procedures entered first, then those representing subject characteristics and nonspecific program characteristics, and, finally, specific intervention categories. Variables included in the study method cluster included the amount of attrition, research design (random vs. nonrandom), an assessment of pretest similarity between treatment and comparison groups (based on the research design, pretest information, demographic comparisons across groups and author acknowledgment), and a dummy code representing archival and observational measures (the excluded category was self- and teacher-reported paper-and-pencil measures). Also entered at this stage were two variables assessing the role of researchers in the delivery of the intervention. One variable indicated whether the program was a research project, a demonstration project, or a routine practice program. The second variable involved the role of the researcher/evaluator in the delivery of the program and ranged from the researcher or the research staff actually delivering the services to the children to the researcher having little or no involvement in the delivery of services.

Variables included in the subject cluster included the age of the sample, gender, and risk level. The ethnic mix of the sample could not be included because of excessive missing data. The cluster of nonspecific program attributes included treatment dosage variables, such as program duration, number of sessions per week, and implementation quality, and program format variables such as service personnel (teachers, laypersons,
Finally, dummy codes representing the two most common intervention strategies were entered (i.e., cognitively-oriented programs, and social skills programs). The results of this analysis (not shown) demonstrated that each cluster, in turn, added significantly to the prediction of effect size. Overall, this model accounted for 32% of the effect size variance. Of that amount, 31% was associated with the method cluster, 46% with the subject cluster, and the remainder (23%) with the two clusters of intervention characteristics.

In the next step of the analysis, a reduced regression model that included only the empirically or conceptually most important variables was fit to the data. This procedure involved dropping, in stepwise fashion, individual variables from each cluster that did not make a significant independent contribution. However, not every nonsignificant predictor variable was eliminated during this process. Some key variables relating to method and treatment dosage were retained in the model despite their lack of significance in order to make their weak relationship explicit.

The final reduced model accounted for 36% of the variance in effect sizes (Table 3). Among the variables relating to study method and procedure, two variables representing study attrition and the form of measurement were retained in the model, although attrition was not significant. Archival and observational measures both produced larger effects than measures reported by teachers or by the subjects themselves.

Table 3. Weighted Mixed Effects Multiple Regression: Universal Programs

<table>
<thead>
<tr>
<th>Variable</th>
<th>β</th>
<th>B</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Objective measures</td>
<td>.43</td>
<td>.26</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Attrition (% loss)</td>
<td>-.04</td>
<td>-.06</td>
<td>ns</td>
</tr>
<tr>
<td>Higher risk subjects</td>
<td>.14</td>
<td>.07</td>
<td>ns</td>
</tr>
<tr>
<td>Age</td>
<td>-.46</td>
<td>-.04</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Risk by age interaction</td>
<td>-.29</td>
<td>-.05</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Routine practice program</td>
<td>-.09</td>
<td>-.03</td>
<td>ns</td>
</tr>
<tr>
<td>Implementation quality</td>
<td>.06</td>
<td>.18</td>
<td>ns</td>
</tr>
<tr>
<td>Program duration (weeks)</td>
<td>-.33</td>
<td>-.18</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Cognitively-oriented programs</td>
<td>-.25</td>
<td>-.12</td>
<td>ns</td>
</tr>
<tr>
<td>Regression constant</td>
<td></td>
<td></td>
<td>.56</td>
</tr>
<tr>
<td>Overall model</td>
<td>Q (9) = 31.42, p &lt; .01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>Q (51) = 55.15, ns</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R² = .36</td>
<td>N=61</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Next, a variable associated with the operation and delivery of the programs was retained in the model, though it was not significant. This variable indicates whether the program was a routine program (versus one delivered primarily by researchers for research or demonstration purposes). With the other variables controlled in the model, routine practice programs were not significantly different from researcher-involved programs.
Of the three subject characteristics included in the model, only two were significant – age and the risk by age interaction. Overall, programs delivered to groups of younger children produced larger effects than those for older children. The interaction effect indicates even larger effects for younger, higher risk children. Recall, however, that the student participants for universal programs are not individually selected for treatment. Thus the risk categories for this set of programs generally distinguish between children in schools or neighborhoods that are high risk and children from low crime neighborhoods with higher socioeconomic status. The significant interaction suggests that universal programs may be particularly beneficial for schools in troubled areas.

Three attributes of the programs themselves were retained in the model. Two of these variables relate to the strength or dose of the intervention. Implementation quality was not significant in the model, but it was retained because we thought it important to make this finding explicit. The other dosage-related variable in the model (treatment duration) was significant and indicated that longer programs were associated with smaller effects. Although the longer programs tended to have more implementation difficulties, the lower effectiveness of the longer programs cannot be explained completely by implementation failures because the implementation variable is controlled in the model. It is possible that when delivered over a longer period of time, children may receive less intense treatment contact, resulting in reduced program effectiveness.

The final program attribute in the model is a dummy code for the cognitively-oriented programs (the most common in this format category). Though the effect is negative, it is not significant. In general, this suggests that when other variables are controlled, there are no significant differences in effectiveness across the different treatment modalities used in universal programs. We also tested the other treatment modalities and none were significant.

In the original article (Wilson, Lipsey, & Derzon, 2003), we did not separate the program formats (i.e., universal, selected/indicated, comprehensive, special). We did, however, present the results for the routine practice programs separately from the results of the research and demonstration programs. While both routine programs and implementation quality were important in our original analysis, they were notably not significant in our current analysis of the universal programs. This has mainly to do with our separation of the program formats. In the case of the universal programs, there were only seven routine practice programs. The weighted mean for these seven routine programs was smaller than the mean for the research and demonstration programs, but with only seven programs, the difference was not significant. The implementation quality variable also trends in the expected direction, but was not significant for the universal programs. As we shall see below, implementation quality is more important with some of the other program formats. It thus appears that the children’s age and risk level were the stronger influences on the outcomes of the universal programs.

To illustrate the significant variables from the final reduced regression model, we have produced a series of tables that show the average effect sizes broken down by the key variables. With small numbers of effect sizes in some of the cells, it is difficult to draw
strong conclusions, but the tables show interesting patterns that can inform future research.

Table 4. Mean Effect Sizes by Treatment Modality and Form of Outcome Measure

<table>
<thead>
<tr>
<th>Treatment Modality</th>
<th>Objective measures (archives, observations)</th>
<th>Self-reports &amp; teacher-reports</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean ES</td>
<td>N</td>
</tr>
<tr>
<td>Behavioral programs</td>
<td>--</td>
<td>0</td>
</tr>
<tr>
<td>Cognitively-oriented programs</td>
<td>.33</td>
<td>1</td>
</tr>
<tr>
<td>Social skills programs</td>
<td>.30</td>
<td>9</td>
</tr>
<tr>
<td>Counseling programs</td>
<td>.16</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 4 presents the weighted mean effect sizes for each treatment modality for the two forms of outcome measurement. For the largest two treatment modalities, social skills and cognitively-oriented programs, the archival and observational measures produce significantly larger effects, although the number of effect sizes in the individual cells is small. This suggests that the common self- and teacher-reported measures of aggressive behavior (e.g., Child Behavior Checklist, Teacher-Child Rating Scale) may not be sensitive to changes induced by typical universal violence prevention programs. This finding illustrates that different types of measurement instruments can produce different results and that multiple measures should be used to assess program results. In addition, since the paper-and-pencil teacher surveys are used so frequently in this kind of research, their link to actual classroom behavior should be established clearly.

Table 5. Mean Effect Sizes by Age Group and Risk

<table>
<thead>
<tr>
<th>Age Group</th>
<th>Low Risk Subjects</th>
<th>Higher Risk Subjects (low SES)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean ES</td>
<td>N</td>
</tr>
<tr>
<td>K-Elementary</td>
<td>.13</td>
<td>20</td>
</tr>
<tr>
<td>Middle school</td>
<td>.14</td>
<td>15</td>
</tr>
<tr>
<td>High school</td>
<td>.21</td>
<td>3</td>
</tr>
</tbody>
</table>

Table 5 illustrates the significant interaction between age and risk that we introduced in the regression model in Table 3. The younger higher risk children appear to achieve the greatest benefit from universal programming. Recall that these high risk children are at risk because their families are low socioeconomic status or they live in high crime or high poverty neighborhoods. The schools these children attend are also likely to be less affluent and may be more disorganized than the schools attended by the low risk children. Thus, perhaps universal programs are more beneficial for the higher risk students (especially those who are in elementary school), because their schools are so disorganized and that universal prevention activities provide a boost to these young children. Note, however, that there are only three universal programs for high school children, so it is difficult to evaluate the effects of universal programs on the full spectrum of ages.
Though implementation quality was not a significant contributor to the regression model, the overall pattern of results shown in Table 6 is as expected, with smaller program effects for programs that had difficulties with implementation. For each treatment modality for which a comparison can be made, the average effect size for the programs without implementation difficulties is greater than that for similar programs with implementation problems. For the behavioral and social skills programs, the effect sizes for the well-implemented programs were twice the size of those for the problem programs.

Table 6. Mean Effect Sizes by Treatment Modality and Implementation Quality

<table>
<thead>
<tr>
<th>Treatment Modality</th>
<th>No Implementation problems</th>
<th>Implementation problems</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean ES</td>
<td>N</td>
</tr>
<tr>
<td>Behavioral programs</td>
<td>.22</td>
<td>2</td>
</tr>
<tr>
<td>Cognitively-oriented programs</td>
<td>.31</td>
<td>10</td>
</tr>
<tr>
<td>Social skills programs</td>
<td>.25</td>
<td>19</td>
</tr>
<tr>
<td>Counseling programs</td>
<td>.30</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 7. Mean Effect Sizes by Treatment Modality and Subject Risk Level

<table>
<thead>
<tr>
<th>Treatment Modality</th>
<th>Low Risk Subjects</th>
<th>Higher Risk Subjects (low SES)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean ES</td>
<td>N</td>
</tr>
<tr>
<td>Behavioral programs</td>
<td>.24</td>
<td>2</td>
</tr>
<tr>
<td>Cognitively-oriented programs</td>
<td>.09</td>
<td>27</td>
</tr>
<tr>
<td>Social skills programs</td>
<td>.19</td>
<td>7</td>
</tr>
<tr>
<td>Counseling programs</td>
<td>.26</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 7 shows the mean effect sizes for the treatment modalities by the risk status of the students. The pattern of means suggests that the cognitively-oriented programs may be particularly well-suited for students in troubled schools or neighborhoods, while behavioral programs may be better suited for higher socioeconomic status students (though the numbers are small for the behavioral programs). The social skills programs appear to work about equally well across the full spectrum of students.
**Results for Selected/Indicated Programs**

There were 103 programs of this format, distinguished by their selective targeting of interventions to individually selected children. Nearly all of these programs were delivered outside of the classroom to small groups or to individual students. The overall weighted mean effect size for the selected/indicated programs was .29 and was significantly different from zero. Five treatment modalities were identified. Three of the five modalities, social skills training, counseling, and cognitively-oriented programs, had generally the same features as those used in the universal format. The behavioral programs, however, were generally implemented with small groups of children or with individual children rather than on a class-wide basis. The peer mediation programs were typical; the subjects in studies of mediation programs were those students who experience an interpersonal conflict and received mediation services from their peers.\(^6\)

<table>
<thead>
<tr>
<th>Modality</th>
<th>Description</th>
<th>Prototypical Program</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td>Behavioral strategies</td>
<td>Interventions involve the use of various behavioral techniques, such as rewards, token economies, contingency contracts, and the like to modify or reduce inappropriate behavior.</td>
<td>Behavioral contracts</td>
<td>16</td>
</tr>
<tr>
<td>Cognitively-oriented programs</td>
<td>Interventions focus on changing thinking processes or cognitive skills; programs focus on solving social problems, controlling anger, inhibiting hostile attributions, etc.</td>
<td>I Can Problem Solve, Think Aloud</td>
<td>38</td>
</tr>
<tr>
<td>Social skills programs</td>
<td>Interventions are designed to help youth better understand social behavior and learn appropriate social skills. Children learn communication skills, fighting avoidance skills, group entry skills, eye contact, “I” statements, etc.</td>
<td>Skillstreaming, conflict management training</td>
<td>17</td>
</tr>
<tr>
<td>Counseling, Talk therapy</td>
<td>These programs utilized traditional group therapy techniques in classroom settings.</td>
<td>Reality Therapy, Rational Behavior Therapy</td>
<td>25</td>
</tr>
<tr>
<td>Peer Mediation</td>
<td>Student mediators are trained to offer mediation services for peers who experience interpersonal conflicts. Training generally focuses on a series of conflict resolution steps.</td>
<td>Community Board Program</td>
<td>5</td>
</tr>
</tbody>
</table>

Tests of the homogeneity of effect sizes for the selected/indicated programs showed significant variability across studies ($Q_{102}=157$). As with the universal programs, this variability was expected to be associated with methodological and substantive characteristics of the studies of selected/indicated programs. Our analysis for these programs proceeded much the same as the analysis for universal programs. We first entered clusters of method, subject, and program characteristics into a full model, then

---

\(^6\) There are other peer mediation programs in the database, but the outcomes are measured either on the student mediators themselves or the whole school. There are two of these and they are included in the social skills category with the universal programs.
selectively removed some nonsignificant variables to arrive at the reduced model presented in Table 9. This model was statistically significant and accounted for 36% of the variance in aggressive/disruptive outcomes for selected/indicated programs. As above, we have retained some nonsignificant method and study variables in the model to make explicit their weak relationship with effect size.

The final regression model includes twelve variables relating to study method, subject characteristics, implementation, service personnel, and treatment modality. Regarding study method, effect sizes that were adjusted for pretest differences between the treatment and comparison groups were smaller than effect sizes that were not adjusted. Studies that experienced more attrition had smaller program effects than those with less attrition. The method of assignment to treatment and comparison groups was not significant in the model, indicating that with the other variables in the model held constant, experimental and quasi-experimental designs did not produce appreciably different results. Note that randomized designs made up the majority (70%) of programs in the selected/indicated format, while the universal format programs were evaluated predominantly with nonrandom designs.

Table 9. Weighted Mixed Effects Multiple Regression: Selected/Indicated Programs

<table>
<thead>
<tr>
<th>Variable</th>
<th>β</th>
<th>B</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pretest adjusted ES</td>
<td>-.25</td>
<td>-.32</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Random assignment</td>
<td>.06</td>
<td>.07</td>
<td>ns</td>
</tr>
<tr>
<td>Attrition (% loss)</td>
<td>-.23</td>
<td>-.98</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Implementation quality</td>
<td>.22</td>
<td>.13</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Higher risk subjects</td>
<td>.25</td>
<td>.20</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Age</td>
<td>.16</td>
<td>.02</td>
<td>&lt;.10</td>
</tr>
<tr>
<td>Researcher delivery</td>
<td>-.10</td>
<td>-.10</td>
<td>ns</td>
</tr>
<tr>
<td>Service professional delivery</td>
<td>-.09</td>
<td>-.11</td>
<td>ns</td>
</tr>
<tr>
<td>Individual one-on-one format</td>
<td>.07</td>
<td>.08</td>
<td>ns</td>
</tr>
<tr>
<td>Cognitively-oriented programs</td>
<td>-.10</td>
<td>-.11</td>
<td>ns</td>
</tr>
<tr>
<td>Counseling programs</td>
<td>-.02</td>
<td>-.02</td>
<td>ns</td>
</tr>
<tr>
<td>Behavioral programs</td>
<td>.13</td>
<td>.18</td>
<td>ns</td>
</tr>
<tr>
<td>Regression constant</td>
<td>-.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall model</td>
<td></td>
<td>Q (12) = 57.13, p &lt; .01</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>Q (90) = 99.69, ns</td>
<td></td>
</tr>
<tr>
<td>R² = .36</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N = 103</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Similar to universal programs, higher risk subjects achieved greater benefits from violence prevention programs than lower risk subjects, although the students in these programs were generally at much higher risk than the students in universal programs. With the selected/indicated programs, a few lower risk children were involved, but the distinction here is mainly between students who are already exhibiting serious behavior problems and those who have risk factors that may lead to later problems. The age variable was marginally significant (and also positively correlated with subject risk...
programs with older students tended to show larger program effects than those with younger children.

Regarding the characteristics of treatment delivery, programs that did not experience implementation difficulties tended to produce larger effect sizes than programs that had problems with implementation. Beyond the implementation variable, delivery personnel and session format (one-on-one vs. group treatment) did not significantly influence program effects. Finally, the three dummy codes for cognitively-oriented programs, social skills programs, and counseling are not significant, indicating that differences in effectiveness across different treatment modalities were small and nonsignificant.

The finding here that implementation quality is important for the selected/indicated programs parallels the similar finding in our original publication. Also borne out in this update is the influence of subject risk level on program outcomes. However, there were no differences between routine practice and the research/demonstration programs. Overall, the mean for routine programs is smaller than that for research/demonstration programs, but the outcomes for the few routine practice programs we have are quite variable and spread over the four formats. By separating the studies into the four format categories, we have created more circumscribed sets of programs, but we cannot readily examine differences between routine practice and research/demonstration programs.

Results for Comprehensive or Multimodal Programs

There were 17 comprehensive school violence programs in our database, distinguished by their multiple treatment components and formats. The average number of distinct treatment components for comprehensive programs was four, whereas the universal and selected/indicated programs typically have one treatment component (and occasionally two or three). The studies of comprehensive programs tended to involve larger samples of students, with two exceptions (see the two cases with broad confidence intervals in Figure 2). In addition, comprehensive programs were generally longer than the universal and selected/indicated programs. The modal program covered an entire school year and almost half of the programs were longer than one year. In contrast, the average program length for universal and selected/indicated programs was about 20 weeks.

The overall mean effect size for comprehensive programs was significant, though small (ES=.06). The homogeneity test showed that there was significant heterogeneity in the set of 17 comprehensive programs, indicating that some programs produced larger effect sizes than others (Q_{16}=34). We have produced the forest plot shown in Figure 2 to illustrate this variability. The forest plot shows the individual effect sizes and confidence intervals for each comprehensive program.
Figure 2. Effect Sizes and Confidence Intervals for Comprehensive Programs

Although the overall mean effect size for the comprehensive programs was small at .06, most of the individual program effects shown in Figure 2 are greater than zero. Identifying the characteristics that are associated with the most effective programs can help practitioners select appropriate comprehensive programming and provide insights into what aspects of program delivery might be important to emphasize.

Our analysis of the comprehensive programs proceeded similar to the analyses of heterogeneity for the universal and selected/indicated formats. Weighted multiple regression analysis was conducted to identify influential method, subject, and/or treatment characteristics. The results of that analysis are shown in Table 10.

Table 10. Weighted Fixed Effects Multiple Regression: Comprehensive Programs

<table>
<thead>
<tr>
<th>Variable</th>
<th>β</th>
<th>B</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Program duration (weeks)</td>
<td>.20</td>
<td>.0008</td>
<td>ns</td>
</tr>
<tr>
<td># sessions per week</td>
<td>.60</td>
<td>.06</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Higher risk subjects</td>
<td>-.41</td>
<td>-.11</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Age</td>
<td>-.49</td>
<td>-.02</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Regression constant</td>
<td>.59</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Overall model: Q (4) = 15.73, p < .01
Residual: Q (12) = 18.39, ns
R² = .46
N=17
Four variables were retained in the final model. For the comprehensive programs, more sessions per week and younger age were associated with larger treatment effects. Note that the direction of effect for the risk variable is negative. This suggests that lower risk subjects show larger treatment outcomes from comprehensive treatments than do the higher risk subjects, which is counter to what we found for universal and selected/indicated programs. However, we believe the risk and age variables are confounded in this analysis. Figure 3 below shows a scatterplot of weighted effect sizes by age, with the lower and higher risk samples designated by stars and pentagons, respectively. This figure shows that all of the programs for higher risk children were delivered to samples whose average age was under 9, while nearly all of the lower risk samples were over the age of 9. Thus the larger effects for younger children might be just as easily explained by their higher risk status. According to the Figure, there appears to be no major differences in outcome across age for the lower risk children. For the higher risk children, the studies with the youngest children achieved the largest outcomes. However, there were no programs for higher risk subjects older than age 9 so understanding the age differences in outcome for the higher risk children is difficult.
Results for Special Schools or Classes

There were 37 programs delivered in special schools or classrooms. Programs generally involved an academic curriculum plus programming that targeted social or aggressive behavior. The students in these programs typically had serious behavioral (and often academic) difficulties that resulted in their placement outside of mainstream classrooms. The weighted mean posttest effect sizes for the special programs was .07 and was not statistically significant.
Although the special programs were not significant overall, the Q test was significant, indicating that the distribution of effect sizes was heterogeneous. Therefore, we performed a weighted mixed effects multiple regression analysis to identify critical study variables. The final model is shown in Table 11.

Table 11. Weighted Mixed Effects Multiple Regression: Special Classes and Schools

<table>
<thead>
<tr>
<th>Variable</th>
<th>( \beta )</th>
<th>( B )</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonrandom assignment</td>
<td>.38</td>
<td>.28</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Implementation problems</td>
<td>-.39</td>
<td>-.31</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Higher risk subjects</td>
<td>.25</td>
<td>.25</td>
<td>&lt;.05</td>
</tr>
<tr>
<td>Regression constant</td>
<td>- .72</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Overall model: \( Q(3) = 30.32, p < .01 \)
Residual: \( Q(33) = 46.86, \text{ns} \)
R\(^2\) = .39
N = 37

Three variables showed up as important moderators of effect size: method of group assignment, level of risk of students, and implementation quality. Nonrandom assignment methods resulted in larger effects, studies with implementation problems found smaller effects, and programs with higher risk samples showed more positive results. The implementation variable is the largest in the model, indicating that fidelity to program components may be particularly critical for the special programs.

**Summary and Conclusions**

The issue addressed in this paper was the effectiveness of school-based programs for preventing or reducing aggressive and disruptive behavior. The results of 219 experimental and quasi-experimental studies of school-based psychosocial programs were synthesized using meta-analysis techniques. We divided the school violence prevention programs into four groups that represented distinct program formats: universal, selected/indicated, special classes/schools, and comprehensive. Overall, the universal, selected/indicated, and comprehensive programs were generally effective at reducing the more common types of aggressive behavior seen in schools, including fighting, name-calling, intimidation, and other negative interpersonal behaviors, especially among higher risk students. The special programs did not significantly impact aggressive and disruptive behavior.

The mean effect size for selected/indicated programs was .29 for aggressive behavior. We can translate this into terms that are more concrete by converting it into typical levels of aggressive behavior in schools. According to the 1999 Youth Risk Behavior Survey, 14.2% of students reported being in a physical fight on school grounds in the year prior to the survey. For 1995 and 1997, 15.5% and 14.8% of students reported being in physical fights (Centers for Disease Control and Prevention, 2002). If we use these figures to estimate that about 15% of untreated school children will get into a fight during a school year, the overall effect size of .29 for selected/indicated programs
translates into about a nine percentage point reduction in fighting. That is, if 15% of students who received no violence prevention programming were getting into fights before intervention, only about 6% of children in selected/indicated programs were getting into fights, less than half of the baseline rate. The most effective programs produced larger effects than this and, thus, would reduce rates of aggressive behavior even more. In addition, since many of the children in the selected/indicated programs were already exhibiting some problem behavior, it is likely that their baseline level of fighting behavior was higher than the general estimate of 15%. Thus, the reduction in aggressive behavior would be even greater.

For the universal programs, our mean effect size of .18 translates into about a five percentage point reduction in fighting over the 15% baseline for untreated children (from 15% to 10%). Though this effect size is smaller than that for the selected/indicated programs, this effect is not trivial. And, as with the selected/indicated programs, the most effective universal programs produced effects larger than this.

Not all programs were equally successful in reducing aggressive behavior. Treatment dose (in the form of treatment duration, frequency, or implementation quality) was uniformly influential. Programs with no or few implementation difficulties tended to produce greater reductions in aggressive behavior. In addition, comprehensive programs with greater session frequency per week were more effective than programs with fewer sessions. For universal programs, shorter programs appeared to be more effective than longer programs. We hypothesize that longer universal programs may be less intense than shorter ones and thus may not have the salience to influence student behavior.

In general, larger treatment effects were achieved with higher risk students. For the universal programs, students from high poverty, disadvantaged neighborhoods achieved the greatest benefits from violence prevention programming; this was especially true for children in early elementary grades. For the selected/indicated programs whose students were generally already exhibiting potentially problematic behavior, programs delivered to youth with more serious problems tended to have larger treatment effects.

Overall, the different treatment modalities within the universal and selected/indicated formats (e.g., social skills training, cognitively-oriented programs, behavioral programs, counseling) were not significantly different from each other; that is, the modalities appeared to be equally effective at reducing aggressive behavior. Though the psychosocial programs reviewed here were generally effective for reducing aggressive student behaviors, it is still unknown whether the types of programs reviewed here (or any others) would be effective in preventing the rare but serious incidents of school violence perpetrated by very disturbed youth.
Bibliography


Appendix A

Bibliography of Studies Included in the Meta-analysis


1 NOTE: Numbers in brackets at the end of each citation are the study identification numbers.
effects of the Perry Preschool Program on youths through age 19 (High/Scope Educational Research Foundation,
intervention on crime and delinquency in adolescence and early adulthood. In J. D. Burchard & S. N. Burchard
(Eds.), Prevention of delinquent behavior. Vermont conference on the primary prevention of psychopathology, Vol. 10
student conduct and satisfaction. Behavior Therapy, 8, 193-204. [1478]
(University Microfilms No. AAT68-00477) [645]
have adjustment problems (Doctoral dissertation, Rutgers University). Dissertation Abstracts International, B-
36/05, 2458. [879]
program for adolescents with adjustment problems. Journal of Community Psychology, 8, 110-116. [879]
with parent training and family-focused support to prevent conduct disorder in high-risk populations: the Fast
Track Multisite Demonstration Project. In C. F. Ferris & T. Grisso (Eds.), Understanding aggressive behavior in
children (pp. 256-264). New York: Academy Science. [445]
Using the Fast Track randomized prevention trial to test the early-starter model of the development of serious
Fast Track program. In R. D. Peters & R. J. McMahon (Eds.), Preventing childhood disorders, substance abuse, and
Applied & Preventive Psychology, 5, 19-35. [175, 499]
Block, J. (1978). Effects of a rational-emotive mental health program on poorly achieving, disruptive high school
Booth, B. A. (1995). A cognitively based anger control training program with aggressive adolescents in the school setting
(University Microfilms No. AAT95-40344) [498]
resolution skills to young adolescents. American Journal of Preventive Medicine, 12(5, Suppl), 65-74. [204]
in definitions of and attitudes towards bullying. School Psychology International, 17, 331-345. [614]
Microfilms No. AAT 8302209 ; DA-I A 43/09) [2142, 3238]
and Allied Disciplines, 40, 1051-1060. [926]


Burcham, B. G. (2002). Impact of school-based social problem-solving training on middle school students with disruptive behavior (Doctoral dissertation, University of Kentucky). Dissertation Abstracts International, 63(2-A), 507. (University Microfilms No. 0424543) [5030]


Burcham, B. G. (2002). Impact of school-based social problem-solving training on middle school students with disruptive behavior (Doctoral dissertation, University of Kentucky). Dissertation Abstracts International, 63(2-A), 507. (University Microfilms No. 0424543) [5030]


Conduct Problems Prevention Research Group (2002). The implementation of the Fast Track Program: An example of a large-scale prevention science efficacy trial. Journal of Abnormal Child Psychology, 30(1), 1-17. [4001]


Friedman, R. M., Filipczak, J., & Fiordaliso, R. (1977). Within-school generalization of the Preparation through Responsive Educational Programs (PREP) academic project. Behavior Therapy, 8, 986-995. [484]


Wilson Lipsey References
Page 7 of 22


Gottfredson, D. C. (1986). An empirical test of school-based environmental and individual interventions to reduce the risk of delinquent behavior. Criminology, 24, 705-731. [404, 408]


Hains, A. A. (1994). The effectiveness of a school-based, cognitive-behavioral stress management program with adolescents reporting high and low levels of emotional arousal. The School Counselor, 42(2), 114-124. [2631]


Huesmann, L. R., Spindler, A., McElwain, N., & Guerra, N. G. (no date). The roles of cognitive processes in moderating and mediating the development and prevention of childhood aggression. Ann Arbor, MI: University of Michigan, Institute for Social Research. [432]


Wilson Lipsey References
Page 12 of 22


Searey, S. (1986). The behavioral effects of teaching supportive and task-related behaviors to behaviorally disordered students in combination with cooperative learning strategies (Doctoral dissertation, University of Kansas). Dissertation Abstracts International, 47/6, 2124. (University Microfilms No. AAT86-19944) [2081]


[588]


Wilson Lipsey References


