Over the past decades, public concern over youth violence has led to a proliferation of prevention programs as well as a corresponding push to identify programs that “work.” A more accurate understanding of effectiveness as well as failure can be found by reframing the questions to ask what works, for whom, and under what conditions.

What Works (and What Does Not) in Youth Violence Prevention: Rethinking the Questions and Finding New Answers

Nancy G. Guerra, Paul Boxer, Clayton R. Cook

Dramatic rises in youth violence in the United States beginning in the 1980s coupled with high visibility acts such as school shootings have resulted in a corresponding proliferation of programs designed to prevent aggression and violence in children and youth. Parallel with this increasing programmatic expansion, there have been repeated calls for rigorous evaluations of programs and identification of “best practices” that merit dissemination and implementation.

Still, in spite of extensive efforts to document effective practices in youth violence prevention, there remains a somewhat confusing array of evidence for effectiveness and lack of effectiveness, as well as a broad range of interpretations of this evidence.

On the one hand, there are clear indications that individual programs can have an impact on preventing or reducing aggression and related behaviors, and positive evaluations of specific programs abound (Farrell, Meyer, and White, 2001; Henggeler and others, 1996). Looking at program reviews and meta-analysis, positive effects have been evident primarily with certain types of programs—most typically those derived from cognitive-behavioral principles (Guerra, Boxer, and Kim, forthcoming; Wilson, Lipsey, and Derzon, 2003). On the other hand, some reviews of program outcomes and attempts to
identify model programs have been less encouraging. For instance, a recent conference convened by the National Institute of Health (NIH) attempted to find a consensus on prevention programs for adolescents based on a commissioned meta-analysis of studies since 1995 (Chan and others, 2004). Only sixty-seven of more than sixteen hundred studies of youth violence prevention programs were included on the basis of rigorous scientific criteria, and evidence for effectiveness was noted in only fourteen studies. Similarly, a center at the University of Colorado has been reviewing youth violence prevention programs to designate “blueprint” programs based on rigorous standards in evaluation and the existence of at least one replication; only eleven programs out of hundreds examined have been selected as blueprint interventions (Mihalic and others, 2004).

From both a program and a policy perspective, it is at best difficult to sift through this seemingly contradictory assortment of evidence. Perhaps the key to unlocking this array of findings and interpretations is to rethink the very questions addressed. Asking “what works” in youth violence prevention suggests a simple, dichotomous answer (that is, works or does not work) about efforts to prevent or modify a problem behavior marked by its complexity and the multiplicity of associated risk and protective mechanisms (Boxer and Dubow, 2002; Cicchetti and Rogosch, 1996; Guerra and Huesmann, 2004). Even under the best circumstances it is unlikely that a violence prevention program “worked” for all: some participants may have improved, some may have stayed the same, and some may have increased their aggression, as has been noted in recent examinations of iatrogenic or negative effects interventions (Boxer, Guerra, Huesmann, and Morales, 2005; Dishion, McCord, and Poulin, 1999).

In the present chapter, we suggest that a more accurate understanding of effectiveness as well as failure can be found by reframing the questions and corresponding evaluation designs. Rather than asking, “What works in youth violence prevention?” we suggest that it is most important to ask: (a) what works and what does not work, (b) for whom, and (c) under what conditions? We discuss each component in more detail, providing illustrative examples from a large-scale prevention study, the Metropolitan Area Child Study (MACS). As we point out, this study provides an example of a multi-component, multi-context intervention that was more effective for some children and less effective for others.

Of course, the importance of assessing effectiveness for moderators of intervention is by no means a recent phenomenon, as several researchers have advocated for this type of investigation for quite some time (Frick, 2001; Kazdin, 2002). Such research has been seen as critical for providing valuable information that will lend itself to a better “fit” between the individual and the intervention. However, despite the recognized importance of examining moderators, it is also the case that beyond considerations of age and sex differences in outcomes, relatively few studies of youth violence prevention programs systematically address potential moderators of
outcomes. As we discuss in this chapter, an important next step is to define more carefully characteristics of individuals and contexts that can maximize program fit.

The Metropolitan Area Child Study (MACS)

MACS was a longitudinal, quasi-experimental aggression prevention field study that included eight cohorts of urban and inner-city elementary school children (Metropolitan Area Child Study Research Group, 2002). Because of the multiply determined nature of children’s aggression, particularly in more disadvantaged urban settings characterized by high rates of crime and poverty and sparse availability of community resources, a central hypothesis of the study was that interventions under these more difficult circumstances needed to be multi-year and multi-component, and target cognition, skills, and behavior across multiple contexts. At the same time, limited funding for comprehensive prevention programming, even when found effective, may preclude adequate funding of such programs, particularly in poorer communities. To address both concerns, we designed the study to test multiple levels of intervention effort over a two-year period that involved individual children as well as teachers and classrooms, the peer group, and the family. The design included four experimental conditions, with each condition representing an increase in dosage and extension of contexts affected. Specifically, schools were randomly assigned to one of four conditions: level A, classroom-only general enhancement program (two-year social-cognitive curriculum and teacher training); level B, classroom enhancement plus small group intervention for high-risk children (two-year intensive social-cognitive training); level C, classroom enhancement plus small group plus family intervention for high-risk children (one-year family therapy); and level D, no treatment control.

We also were interested in examining moderators of intervention effectiveness at the individual and school or community level. At the individual level, we addressed the question of whether developmental timing within the elementary school years would affect intervention outcomes. Given the trend toward “earlier is better” in the field of prevention research and practice, we wanted to examine empirically whether outcomes varied for children who began the program in the early (Grades 2–3) versus late (Grades 5–6) elementary school years.

At the school or community level, many studies have reported that preventive interventions do not work equally well in all settings, with economic and social constraints often limiting the ability of interventions to effect change in aggression (Aber and others, 2002; Hughes and others, 2005). To test whether school or community resources moderated intervention outcomes, schools in our study were divided into low-resource schools and moderate-resource schools based on a broad array of school and community indicators, including neighborhood crime and violence rates,
funding allocations, number of children receiving free or reduced lunch, and number of rental housing units.

As we have reported previously, the intervention was effective in preventing aggression in some but not all children. A more complete description of the study design, interventions, and results can be found in the report of the Metropolitan Area Child Study Research Group (MACS) (2002). In some cases, iatrogenic effects were noted. As we will discuss in more detail in the following sections, effects on aggression were moderated by individual and school or community factors. Had we not examined these moderators and looked only at main effects across all children and schools, we would have concluded that the intervention did not work. However, a more accurate conclusion was that the intervention worked for some children under some conditions and did not work for other children under other conditions.

Rethinking the Question: What Works and What Does Not Work

The long-term goal of the MACS intervention was to prevent serious violent behavior among urban and inner-city youth. Because this behavior typically does not emerge until adolescence, the short-term goals for elementary school children were to reduce aggression, reduce associated risk factors (that is, poor social problem-solving skills), and increase associated protective factors (that is, school achievement). Risk and protective factors were considered secondary outcomes, and children’s aggression was considered a primary outcome. Whether the intervention “worked” or not hinged on whether we were able to demonstrate reductions in aggressive behavior relative to the control condition for the most aggressive children in our sample. Our conceptualization of effectiveness hinged on demonstrated behavior change in aggression as a proxy for later violence. However, across a range of prevention studies there is substantial variability in how effectiveness is conceptualized, suggesting a need for greater precision. Of particular importance are differences in specificity of outcomes and the clinical significance of findings.

Specificity of Outcomes Vis-à-vis Aggression and Violence

There is remarkably little consistency across aggression and violence prevention studies in specificity of outcomes. In many cases, the distinction between aggression and violence is blurred, although violence is usually considered a more extreme form of aggression. Most frequently, the term “aggression” is used to describe children’s behavior and the term “youth violence” is used to describe adolescent behavior (Guerra and Knox, 2002). Following this logic, we would expect programs focused on younger
children to include measures of their aggressive behavior and programs focused on adolescents to include measures of more serious violent behavior as evidence that the program works. In practice, even violence prevention programs for adolescents rarely measure actual effects on reducing violent behavior per se, in part due to the low base rates of serious violence, although many programs assess some type of aggressive behavior. Program evaluations also frequently include a range of other measures of associated risk factors, including individual attributes (self control), related behaviors (delinquent acts), and sanctions (school suspensions).

**Clinical Significance of Findings**

Along with the specific constructs and measures relied upon to assess the outcomes of youth violence prevention programming, of great importance in determining whether a program “works” is the particular criterion used to make that judgment. The common practice of reporting only statistical significance has clear limitations. Indeed, it is important to consider both statistical and clinical significance. For example, is the treated group now indistinguishable from a normative population, whereas the non-treated group is not (Kendall, Marrs-Garcia, Nath, and Sheldrick, 1999)? Has the quality of life of the treated group improved in ways that the quality of life of the non-treated group has not (Gladis, Gosch, Dishuk, and Crits-Christoph, 1999)?

In the MACS, the primary measure used to assess changes in aggression was a linear composite of ratings made by both teachers and peers. The teacher report measure was Achenbach’s (1991) Teacher Report Form (TRF) of the Child Behavior Checklist. As a clinical rating scale, the TRF produces T score equivalents of raw scores that indicate whether a target child’s problem behaviors are present to a degree similar to typical children referred for clinical psychological treatment. One critical finding in the evaluation of the MACS was that in the moderate resource schools, the full level C intervention for the younger age group produced general tendencies toward less aggression compared to control, whereas in the low-resource schools, the reverse was observed. These divergent results were produced through analyses of the composite (teacher report plus peer rating) aggression measure. However, when considered with respect to clinical status levels using just the TRF, a different pattern emerged. Here, in both the moderate- and low-resource schools, the full intervention actually produced beneficial effects on clinical status levels for the younger age group. In the moderate-resource schools, 18.3 percent of children in the full intervention moved from clinical status to sub-clinical status compared to 3.1 percent of the controls. In the low-resource schools, the proportions were 10.3 percent and 3.4 percent, respectively. This result, which demonstrates effects of the intervention on very high-risk youth, was masked by analyses examining more general trends in the different samples.
Rethinking the Question: For Whom

As we have discussed previously, it is likely that interventions have different effects for different individuals based on several key factors. Even absent specific hypotheses for differential effects, it is important to begin to enumerate potential individual level moderators of intervention effectiveness, the “for whom” in understanding outcomes of aggression and violence prevention programs. We propose three important potential individual moderators that have been examined to some degree in violence prevention research. These are developmental timing, predisposition to aggression, and peer social status.

**Developmental Timing.** From a developmental perspective, comprehensive youth development and violence prevention programming should begin before birth and continue into adolescence and young adulthood (Weissberg and Greenberg, 1998). From a policy and practice perspective, it is often necessary to select specific ages for priority funding in order to focus resources on age groups most likely to benefit from the intervention. This suggests the importance of identifying whether certain interventions are more effective during certain developmental periods. Depending on the type of intervention, it is possible to make specific hypotheses in relation to developmental timing. For example, we would expect interventions aimed at changing peer group norms to be more effective during adolescence when the influence of the peer group is most salient.

In some cases, a general type of intervention such as social-problem solving or social-skills training has been used across multiple age groups. There has been consensus that aggression prevention interventions should start early, in part due to the relative stability of children's aggression from the early school years (Huesmann, Eron, Lefkowitz, and Walder, 1984). This “earlier is better” standard has also been supported by studies suggesting that children who begin their aggression early are more likely to continue this behavior into adolescence and are more likely to develop entrenched negative behaviors that are difficult to change as they get older (Dishion and Patterson, 1992; Moffitt, 1993). This would suggest that early intervention is recommended, but particularly for those children already displaying aggressive behavioral tendencies (that is, secondary or selective prevention). In fact, the early starter-late starter model (Moffitt, 1993) also identifies a late starter group that begins antisocial behavior in adolescence, suggesting that, for these youth, intervening during the adolescent years is recommended.

Most studies that have considered the moderating influence of age have examined effects within a specific age period, typically based on school level (elementary, middle, high school). In the MACS study, we examined whether outcomes varied for children who began the program in the early (Grades 2–3) versus late (Grades 5–6) elementary school years. We found that the comprehensive program that included children, teachers, peers, and families was effective but only for the younger age group, supporting the
emphasis on early intervention. However, we did not include a middle school or high school sample in the intervention study.

In a recent meta-analysis of the effects of social skills training (SST) for students with externalizing behavior disorders, Gresham and others (forthcoming) reported a U-shaped curvilinear relation between the efficacy of SST and positive student behavior change as a function of age. Specifically, age moderated the effectiveness of SSTs, whereby SSTs were most effective during the early childhood and the adolescent years and less effective during the middle childhood years. As the authors discuss, it may be that SSTs during the early childhood years are affecting the “early starters” and SSTs during the adolescent years are affecting the “late starters,” while SSTs during middle school years are actually affecting the early starters whose aggression has become more chronic over time and hence more resistant.

**Predisposition to Aggression.** An important consideration for prevention programming is the extent to which change occurs among the more aggressive youth. Indeed, many large-scale evaluation studies have reported significant effects for all youth (as part of primary or universal preventive efforts) as well as for subgroups of more aggressive youth participating in more intensive secondary or selective prevention programming (Conduct Problems Prevention Research Group, 2002). In the MACS study, significant reductions in aggression were only found for the high-risk more aggressive children who participated in the most intensive intervention programming, although both low- and high-risk children participated in the universal classroom enhancement program offered at all levels of intervention.

In addition to examining whether program outcomes vary as a function of an individual’s baseline level of aggression, it is also likely that outcomes may vary within the more aggressive group. In other words, there may be specific subgroups of aggressive children who respond differentially to a particular intervention. For example, one subgroup may be related to whether or not other problem behaviors or disorders are present (that is, comorbidity). On the one hand, we might expect that aggressive children with comorbid problems or disorders would be particularly resistant to intervention efforts, given the complexity and extent of their problems. However, intervention effectiveness may also depend on the nature of the comorbid disorders. For example, some studies have found that children with comorbid conduct problems and depression were actually more responsive to interventions than children with conduct problems alone (Beauchaine, Gartner, and Hagen, 2000). It may be that comorbidity is more problematic when aggression or conduct problems are linked with problems that might increase the explosive or chronic nature of aggression and violence (Lynam, 1998).

**Peer Social Status.** One of the most robust findings in the literature is the link between peer rejection and aggression: rejection and aggression are highly correlated and rejection exacerbates aggression among children.
initially disposed to behave aggressively (Dodge and Pettit, 2003; Newcomb, Bukowski, and Pattee, 1993). The negative consequences of aggression within a child’s peer group presumably should provide a source of motivation for children to decrease their aggression. However, over the past decade, several studies have identified a subgroup of aggressive children who are able to maintain a popular status among their peers (for example, Farmer and others, 2003; Rodkin, Farmer, Pearl, and Van Acker, 2000).

Analyses of data from the MACS illustrate this possibility. Boxer, Huesmann, Hanish, and Guerra (forthcoming) examined pre-test and post-test peer-rated indicators of aggression and popularity in the late intervention. Across conditions, the correlation between aggression and popularity became more positive from pre-test to post-test, consistent with earlier studies. Interestingly, however, the amount of change in the correlation was statistically significant only in the intervention conditions and not in the control condition. Follow-up analyses suggested that youth who began the MACS late intervention program at relatively high levels of popularity (greater than 0.5 SD above the mean) were more likely to increase in their aggressiveness in active treatment than in the control condition. In contrast, low-popularity youth showed decreases in their aggression in the context of intervention.

Rethinking the Question: Under What Conditions
Implementing violence prevention programs in real-world settings also requires attention to contextual conditions that may facilitate or interfere with prevention outcomes. Although studies examining the moderating role of contextual factors are limited, previous research suggests the importance of two broad categories of contextual moderators. These are conditions that support the learning and maintenance of aggression and conditions that interfere with high-quality implementation of the intervention. Because many aggression and violence prevention programs (including the MACS program) are implemented in school settings, we use school context to illustrate how these moderators may affect outcomes.

Conditions That Support the Learning and Maintenance of Aggression. Some school-based preventive interventions focus on individual skill building or promoting competencies through classroom-level curricula. Other school-based programs emphasize changing contextual conditions that support aggression, such as teacher reinforcement for aggression and school-wide norms supporting aggression. More recently, the emphasis on comprehensive programming has led to programs that seek to change both individual and contextual factors, as illustrated in the MACS study. However, regardless of the extent to which contextual conditions are addressed, it is unlikely that all conditions that support the learning and maintenance of aggression can be considered in any single program—in many cases, these conditions are simply beyond the scope of school-based prevention.
The learning and maintenance of aggression is more likely under conditions where aggression is seen as “normative” (that is, the extent to which it is seen as acceptable in a given setting) and where there are ample opportunities for reinforcement. This was supported in the MACS study. As Henry and others (2000) report, children's aggressive behavior was more likely to improve when they were in classrooms with lower levels of normative beliefs supporting aggression, and their aggressive behavior was more likely to escalate when they were in classrooms with higher levels of normative support. This finding is also consistent with a recent study by Aber and others (2002). In their evaluation of the Resolving Conflict Creatively program, a classroom-based elementary school program, they found positive effects on aggression only in classrooms where children's normative beliefs supporting aggression were low.

We also would expect school-based interventions to be less effective in classrooms where teachers reinforce children's aggression (for example, through increased attention) and in schools with inconsistent sanctions for aggression and higher levels of tolerance for such behavior (if these conditions are not addressed directly by the intervention). Studies have shown that classroom and school climates that reinforce aggression serve to foster and sustain more severe forms of aggression (Greene, 2005; Kallestad and Olweus, 2003). Supports and sanctions for aggression may also contribute to an overall school climate that fosters or inhibits the development of aggression (McEvoy and Welker, 2002).

Conditions That Interfere with High-Quality Implementation of the Intervention. Perhaps the most important finding in the MACS intervention, to date, is that intervention outcomes were moderated by level of school adversity. Positive effects were found only in schools with at least moderate levels of resources, and negative effects were found under some conditions but only in schools with low levels of resources (although positive effects for clinical significance were noted across both moderate and low-resource schools). The finding that school adversity moderated outcomes is consistent with a number of aggression and violence prevention efforts that have demonstrated poorer outcomes in more distressed and disadvantaged settings (Aber and others, 2002; Hughes and others, 2005). Ironically, although interventions may be most needed in low-resource schools, the realities of these low-resource schools may also make these interventions less effective.

It is possible that these interventions are less effective because of the multiplicity of risk factors that exist in more disadvantaged settings: even the most comprehensive interventions are unlikely to address this array of factors. However, it is also important to consider the ways in which adversity may interfere with high quality implementation of interventions. For instance, teachers and administrators in low-resource schools are likely to
struggle to “survive” rather than thrive. Children may come to school ill-prepared for the day, teachers may feel overburdened with challenges of teaching in a low-resource environment, and schools are likely to be understaffed and have relatively high rates of turnover. The addition of a new violence prevention curriculum or program may tax an already overburdened system (Gresham, 1989; Telzrow, McNamara, and Hollinger, 2000).

Conclusions and Future Directions

In spite of numerous evaluations of youth violence prevention programs and the urgent need for evidence-based practices, there are still relatively few documented “best practices” to guide policies and programs. Even considering interventions with evidence of effectiveness in more than one setting, it is still unlikely that such programs work equally well for all youth under all conditions. Indeed, most “best practice” guidelines suggested by agencies such as the CDC (Thornton and others, 2000) as well as various researchers (for example, Boxer and Dubow, 2002; Tolan and Guerra, 1994) typically are main effects recommendations. They do not take into account the critical issues of “what works and what does not, for whom, and under what conditions.”

To provide greater clarity and direction for practice in youth violence prevention, we suggest that program evaluations focus attention not only on outcomes, but also on moderators of outcomes. As we have illustrated with findings from the Metropolitan Area Child Study, evaluations that do not consider specific moderators may deem a program to be ineffective when, in fact, it was effective, but only for some children. In the MACS example, the full intervention was effective only for younger more aggressive children in schools with greater resources. However, when clinical significance was considered, the program was effective in reducing the aggression levels of the most extremely aggressive youth to normative levels regardless of school resources. It may be that school resources are particularly important for implementation overall, but within schools a teacher with one or two highly aggressive children might also be motivated to implement an intervention beyond limitations of school resources.

Because there are numerous violence prevention programs that have shown efficacy with distinct populations, such as young children at risk for serious conduct problems (Webster-Stratton and Reid, 2003) or older adolescents already engaged in delinquent behavior (Henggeler and others, 1996), we propose that researchers focus on second-level evaluation studies to assess the effectiveness of those interventions across different settings and for different intervention groups. For example, does multi-systemic therapy work equally well for adolescents living in the inner-city as well as in rural settings? Does it matter whether services are provided by front-line direct care workers, by seasoned or less experienced teachers, or by clinical psychologists? Certainly these types of questions are being addressed by the
groups who have developed efficacious interventions. The thrust of our recommendation is that it might be most useful for others interested in ascertaining program effects to focus on the portability of existing strong programs and to identify target groups for whom existing interventions are not effective prior to developing new programs focused on groups not well served by current interventions. Not only is this an important next step for youth violence prevention programs, but it should add to literature on program effectiveness across multiple interventions that target a range of problems.

References


NANCY G. GUERRA is professor of psychology at the University of California Riverside and Principal Investigator of the Academic Center of Excellence on Youth Violence Prevention funded by the Centers for Disease Control and Prevention.

PAUL BOXER is assistant professor of psychology at the University of New Orleans and adjunct faculty associate in the Research Center for Group Dynamics at the University of Michigan.

CLAYTON R. COOK is a doctoral student in the school psychology program at the University of California Riverside.