Published in final edited form as:

Am J Prev Med. 2004 January ; 26(1 Suppl): 12–19.

The Study Designed by a Committee:

Design of the Multisite Violence Prevention Project

David B. Henry, PhD, Albert D. Farrell, PhD, and The Multisite Violence Prevention Project

University of Illinois at Chicago (Henry), Chicago, Illinois; and the Virginia Commonwealth University (Farrell)

Abstract

This article describes the research design of the Multisite Violence Prevention Project (MVPP), organized and funded by the National Center for Injury Prevention and Control (NCIPC) at the Centers for Disease Control and Prevention (CDC). CDC's objectives, refined in the course of collaboration among investigators, were to evaluate the efficacy of universal and targeted interventions designed to produce change at the school level. The project's design was developed collaboratively, and is a 2×2 cluster-randomized true experimental design in which schools within four separate sites were assigned randomly to four conditions: (1) no-intervention control group, (2) universal intervention, (3) targeted intervention, and (4) combined universal and targeted interventions. A total of 37 schools are participating in this study with 8–12 schools per site. The impact of the interventions on two successive cohorts of sixth-grade students will be assessed based on multiple waves of data from multiple sources of information, including teachers, students, parents, and archival data. The nesting of students within teachers, families, schools and sites created a number of challenges for designing and implementing the study. The final design represents both resolution and compromise on a number of creative tensions existing in large-scale prevention trials, including tensions between cost and statistical power, and between internal and external validity. Strengths and limitations of the final design are discussed.

“To quote an ancient proverb, ‘A camel is a horse designed by a committee.’ {This movie} must have been designed by a camel.” (Roger Ebert, *Chicago Sun Times*, April 26, 2002)

Proverbs abound concerning committees and the end products of their deliberations. Virtually all suggest that the designs that emerge from such efforts are overly cumbersome and lack integration and elegance. According to common wisdom, design is best left to individual effort. This paper reports the processes and end product of a committee of investigators who, in spite of common wisdom, set out to design a large-scale test of the efficacy and effectiveness of violence-prevention interventions among middle-school children.

Address correspondence to: David B. Henry, PhD, Department of Psychiatry, Institute for Juvenile Research, Families and Community Research Group, 840 S. Wood Street, Chicago IL 60612-7347. dhenry@psych.uic.edu.

Multisite Violence Prevention Project (corporate authors) includes: **Centers for Disease Control and Prevention, Atlanta GA:** Robin M. Ikeda, MD, MPH; Thomas R. Simon, PhD; Emilie Phillips Smith, PhD; LeRoy E. Reese, PhD (all Division of Violence Prevention, National Center for Injury Prevention and Control); **Duke University, Durham NC:** David L. Rabiner, PhD; Shari Miller-Johnson, PhD; Donna-Marie Winn, PhD; Steven R. Asher, PhD; Kenneth A. Dodge, PhD (all Center for Child and Family Policy except Asher [Department of Psychology]); **University of Georgia, Athens GA:** Arthur M. Home, PhD (Department of Counseling and Human Development Services); Pamela Orpinas, PhD (Department of Health Promotion); William H. Quinn, PhD (Department of Child and Family Development); Carl J. Huberty, PhD (Department of Educational Psychology); **University of Illinois at Chicago, Chicago IL:** Patrick H. Tolan, PhD; Deborah Gorman-Smith, PhD; David B. Henry, PhD; Franklin N. Gay, MPH (all Institute for Juvenile Research, Department of Psychiatry); **Virginia Commonwealth University, Richmond VA:** Albert D. Farrell, PhD; Aleta L. Meyer, PhD; Terri N. Sullivan, PhD; Kevin W. Allison, PhD (all Department of Psychology).

The impetus for this effort was Cooperative Agreement No. 99067 issued by the National Center for Injury Prevention and Control (NCIP) of the Centers for Disease Control and Prevention (CDC) in 1999. This announcement invited investigators to participate in a multisite collaboration to develop and test school- and community-based violence-prevention strategies. The goal of this initiative was to provide valid empirical evidence of the efficacy of universal and community-based interventions.¹ The application process was unusual in that it required applicants to provide support for their capacity to conduct research in this area rather than putting forth a research design. Participating schools were therefore selected on the basis of principals' and teachers' willingness and ability to participate in a research project. For purposes of recruitment all details of the project known to the investigators at the time were explained to principals and teachers. Specifically, the project was described as requiring random assignment of the school to a condition that might not include intervention, and a commitment not to introduce other violence-prevention programs during the intervention years of the study.

A CDC review panel selected teams of investigators at four universities to develop and implement the Multisite Violence Prevention Project (MVPP). These were Duke University, The University of Georgia (UGA), the University of Illinois at Chicago (UIC), and Virginia Commonwealth University (VCU). Beginning in Fall of 1999, investigators from these sites and CDC began a series of monthly meetings and weekly conference calls. They reviewed the current state of knowledge in youth violence prevention and gaps in the literature, developed specific research questions, designed a study to address these questions, developed the interventions to be evaluated, and developed plans for implementation and coordination across sites. Consultants such as Gibbons et al.² and Gottfredson³ met with the investigative team to give advice on designs. The process of arriving at a research design was iterative, with designs proposed and evaluated according to criteria that emerged during these discussions. Details regarding the rationale for the specific research questions and some of the challenges faced in developing this collaboration are addressed elsewhere in this volume. This article describes the design of the MVPP study, with particular attention to how it addressed critical knowledge gaps; and discusses some of the primary challenges, strengths, and weaknesses of the design.

Considerations in Developing the Research Design

The investigative team assembled by CDC was challenged to design a study according to criteria that were at times mutually contradictory. First and foremost, the design had to be consistent with the project's research questions and theoretical assumptions about mechanisms of change. In order to have scientific credibility, it was essential that it possess high levels of internal and external validity.⁴ The design also had to be feasible given constraints related to budgets, school characteristics, and school district sizes. This section outlines some of the criteria that shaped the project's ultimate design.

One of the key decisions during the planning year was that the project focus on producing school-level change. Although school-level constructs such as norms and climate often are measured with individual-level indicators and may have parallels at the individual level of analysis, the appropriate focus for a study of school-level change would be school-level variables. This focus necessitated development of a between-school design in which schools, rather than individual students or classrooms of students were assigned randomly to conditions. This decision had important implications for statistical power, methods of data analysis, sampling, and methods of assessment.⁵ We consider each of these issues in this paper.

A second factor concerned the focus on comparing the effects of two different intervention strategies, both of which have shown promise in preventing aggressive behavior: (1) a universal intervention implemented with all sixth-grade students and teachers, and (2) a targeted intervention implemented with students selected on the basis of their levels of aggression and

their families. Although the original intention was to extend the intervention into multiple grade levels, it was soon apparent that even the substantial resources CDC had devoted to the project would not be sufficient to support such an effort. Limiting the number of grades was not optimal from a design point of view, but was weighed against even less-desirable options, such as removing the targeted intervention or reducing the number of schools. In the end, we chose to intervene directly with the sixth-grade students, their teachers, and their families⁶ and to determine the efficacy of this approach before extending it into other grade levels.

The design had to be capable of examining several possible change mechanisms through which the targeted intervention might produce school-level effects. Because most serious aggressive and violent behavior is committed by relatively few children in any school, the targeted intervention was expected to produce direct effects on school-level rates of aggression by lowering rates among these individuals. In addition, because high-risk children often have substantial influence on the learning, adoption, and perpetuation of aggressive and violent behavior by their peers, the targeted intervention also had the potential of producing indirect effects through its impact on school-level normative processes.⁷ The design thus had to be capable of assessing both the direct and indirect effects of the interventions at the school level of analysis. From the perspective of school-level change, it should be noted that sixth graders represented the youngest cohort of students within middle schools and may not have the same level of influence as their older peers. However, we designed the assessment schedule to include follow-up assessments of both universal and targeted samples as each cohort of sixth graders moved through middle school. By the final year of planned assessments, the cohort that originally received the interventions will be in eighth grade, and a second cohort who received the interventions the year following the first cohort will be in seventh grade.

A third consideration was the necessity of a true experimental design to assess efficacy. Other potentially efficacious interventions have been tested in quasi-experimental designs lacking true random assignment^{8,9} or in designs troubled by differential attrition.¹⁰ The absence of a true experimental design makes it difficult to conclude that preventive interventions are efficacious, particularly because the effects of such interventions are often modest in size.^{10–12} Such designs are critical for producing the scientifically valid knowledge needed to advance the field of violence prevention. More generally, as with all designs, the development of an appropriate design involved a careful balance between the often competing demands of internal and external validity,¹³ which are discussed below.

A fourth consideration was that the design have adequate statistical power to answer the research questions being posed. At the same time, the design had to be feasible given the constraints of budgets, school characteristics, and school district sizes. During the process of considering various designs, we consulted with experts including Gibbons et al.,² Hedeker et al.,¹⁴ Raudenbush,¹⁵ and Bryk and Raudenbush¹⁶ to render opinions regarding the statistical power of each alternative.

Designs Considered

Three different types of research designs, and several minor variations on each, were considered before arriving at a final design. The first was a *within-school cohort design*, in which different cohorts of students within each school would be assigned randomly to four conditions: (1) no treatment control, (2) universal intervention only, (3) targeted-intervention only, and (4) universal-plus-targeted interventions. This design would have substantially increased the statistical power available from a relatively small number of schools. Unfortunately, it had significant weaknesses. Chief among these were potential carryover effects within schools (e.g., high rates of student retention could result in cohorts not being independent of each other), the fact that not all orders were feasible (e.g., moving from the

universal intervention in which teachers participated in workshops on violence prevention to a control condition), and diffusion caused by interactions between cohorts (e.g., intervention teachers sharing curricular material with control teachers). It would have also taken 4 years for schools to complete all four conditions.

Several designs attempted to increase statistical power by reducing the number of experimental conditions. A *three-group between-school design* would have assigned schools randomly to three conditions: (1) no-treatment control, (2) universal intervention only, and (3) universal and targeted interventions. This stepped design was similar to that used in the Metropolitan Area Child Study.¹⁰ Although it did not risk interaction between conditions as in the first design, it would not have determined the effects of the targeted intervention alone. A *two-group between-school design* would have assigned schools randomly to either a no-treatment control condition or a condition that included both the universal and targeted interventions. This “all or none” design would have examined the combined effects of both intervention approaches, but would not have allowed us to determine the relative effects of each. Ultimately both designs were rejected because they did not provide a test of the relative impact of each intervention approach alone and in combination.

We also considered a *mixed between- and within-school design*. Within this design, half the schools would have been selected randomly for implementation of the universal interventions. Within each school, half the “high-risk” students within each grade would have been assigned randomly to the targeted intervention. In other words, the universal intervention would have been tested in a between-school design, and the targeted intervention in a within-school design. This design had the advantages of increased statistical power for detecting effects of the universal interventions at the school-level and the targeted intervention at the individual level. However, it would not have evaluated the impact of the targeted intervention at the school level, and did not provide for a true control condition, in that each school would have had a group of targeted children who received an intervention.

Features of the Final Design

The final design was a 2×2 cluster-randomized, true experimental design, as is illustrated in Figure 1. This design features true random assignment of schools to conditions representing all possible combinations of the universal and targeted interventions. As such, it provides a basis for testing the main effects of either intervention, collapsed across levels of the other intervention, as well as the theoretically interesting issue of the interaction between intervention conditions.

School Selection and Random Assignment

The design is termed “cluster-randomized” because individuals are assigned randomly to conditions in clusters (i.e., schools), within each site.^{2,15,17} The number of schools participating at each site varied because of differences in school size. Three schools were assigned to each condition in Chicago, which included K–8 schools with 61 to 120 sixth graders per school. Two to three schools were assigned to each condition at the other three sites that included mostly large middle schools with 200 to 400 sixth graders per school. Across the four sites this provided a total of approximately nine schools per condition for a total of 37 schools. (The Chicago site recruited 12 schools and the Georgia site recruited nine schools, and assigned three schools to one of the conditions and two schools each to the others. The remaining sites recruited eight schools each.)

At all sites, schools were randomized to conditions without matching or stratification. Three of the sites recruited virtually all large middle schools in their local school systems. The Chicago site had a large number of relatively small schools from which to select schools for

participation in the study. Schools in Chicago were selected to be approached for participation based on the grades served (K–8), enrollment (>1100 students), percentage of the school population designated as low income (>55%), residence of most students (>75%) within the school district boundaries, and location (travel time by public transportation from the university <1 hour). Schools that met these criteria were considered for participation if both principals and teachers agreed to participate in random assignment of schools to conditions, and not to implement any other violence-prevention program during the course of the study, even if assigned to the control condition. Of the 18 Chicago schools that met these criteria, the two largest schools were selected to participate in a pilot study in which students were assigned randomly to conditions by classroom. The final 16 participating schools were selected during randomization.

At each site, representatives from the schools, and other relevant community stakeholders were invited to a dinner meeting where the project was described and schools were randomized to conditions. For example, at the Richmond site, the dinner was attended by representatives of the eight participating schools; the school superintendent; representatives from CDC, the local community services board, and the university; and the area's congressman. We believed that explaining the importance of random assignment in a relaxed atmosphere and making the assignment process a public ritual would increase compliance in the ensuing years of the study. Representatives from each school in randomly selected order took turns selecting plain envelopes, each of which contained the description of a condition in the study.

Multisource, Multiwave Measurement

Another important feature of this design is its multi-source, multiwave assessment of key constructs. Data were collected on both a *cohortwide* sample and a *targeted* sample at each school. The cohortwide sample consisted of a random sample of approximately 80 sixth-grade students from each school (this sample includes all sixth graders at the smaller schools). The cohort-wide sample represents the student population at each school and provides estimates of change at the school level. The targeted sample included 15 to 20 students at each school selected on the basis of their levels of aggression and the degree to which they exerted influence over other students.^{18,19} Data from the targeted sample serve multiple purposes. They provide a basis for evaluating the impact of the targeted intervention on individual participants. They can also be used to determine the extent to which targeted students serve as mediators of change at the school level. Finally, it provides a basis for examining the relative impact of the universal and targeted interventions on students displaying higher initial levels of aggression. This is particularly important given previous studies of universal interventions that have found stronger intervention effects for such students.^{5,10,20,21}

For each sample, multiple sources of data are obtained across multiple waves of measurement.²² Sources include student reports, teacher reports, and data from school and court records. In addition, parent reports are obtained for targeted students, and teachers complete measures to assess the impact of the interventions on them and on their perceptions of the school. This measurement strategy is expected to produce highly accurate estimates of changes in aggression and key mediating variables.

Within the project's design, schools were assigned randomly to conditions that will remain in effect for two successive cohorts of students. The design thus provides an opportunity to examine the cumulative effect of 2 years of intervention on school-level changes. For several reasons, we anticipate that intervention effects may be stronger for students in Cohort 2 (i.e., there will be an Intervention $\times$ Cohort interaction). Cohort-2 students at schools assigned to the universal intervention will be entering schools in which most sixth-grade teachers and seventh-grade students will have participated in the universal interventions during the previous school year. During the second year of implementation, we were also able to improve

participation rates in the teacher, family, and student interventions, and to begin the interventions earlier in the school year. Although this design complicates the interpretation of findings, it provides a basis for evaluating the types of effects schools might expect as they implement programs across multiple years.

The design includes multiple sites representing somewhat different samples of students and schools, all of whom were considered at high ecologic and individual risk for aggression and violence. During the first year of implementation (i.e., 2001–2002), 2637 students were selected randomly and interviewed for the cohort-wide sample at pretest, and 522 students were selected for the targeted sample. The ethnic composition of the sample varied somewhat across sites.¹ At three sites, the sample consisted primarily of minority youth and their families, with one of the sites including a higher concentration of Hispanic youth and families. About half (50%) of the cohortwide sample reported living with two parents (including biological, step, or foster), compared with 35% of the targeted sample. Youth poverty rates in the school districts served by the project ranged from 17% to 37% with an average of 28% versus the national average of 16% (Table 1). Violent crime rates committed by juveniles in the counties served by the schools ranged from 47 to 74.1 per 100,000, with an average of 63 versus the national average of 43. These community-risk indices suggest that the initiative includes communities with substantial poverty and crime, which are likely to limit opportunities for positive involvements and to expose children to violent acts by their peers and to violent victimization by others.

School-Level Analyses

Because schools were assigned randomly to conditions, they will represent the units of analysis. Although assessment focuses on individual students, teachers, and families, these are regarded, metaphorically, as repeated observations of school levels of behavior. We will, therefore, analyze the data using a random-effects regression model approach that can model clustering of observations within higher levels of analysis as well as individual growth curves, which will include multiple waves of data as they become available.^{23–26} Random effects regression models assume that the available data at any given point of measurement estimate the group growth trend and each individual's data represents that individual's deviation from the group trend at that point of measurement. This assumption permits valid estimates of slopes and intercepts with cases that have missing waves of data, whether or not the data are missing at random (for reasons unrelated to the variables under study). Valid parameter and standard error estimates can be obtained even when there are varying numbers of measurements with individuals, and varying numbers of individuals within settings. This approach will also enable us to consider factors at both the individual level (e.g., gender, baseline aggression) and school-level (e.g., school size, school norms) that may influence outcomes.

Challenges Inherent in the Design

Although the project's 2×2 design appears straight-forward, it poses serious challenges that extend beyond those typically encountered in violence-prevention research. The source of many of these challenges is the nesting of participants within schools and sites. Although MVPP has one of the largest samples of schools in a violence-prevention efficacy study to date, school-level random assignment reduces the number of “subjects” to the number of schools and requires multilevel data analysis techniques to address this nesting. Other challenges relate to the often-conflicting goals of providing strong experimental evidence of efficacy while implementing the interventions under conditions that were close to the scale and circumstances under which violence-prevention programs are implemented in schools. For example, virtually all schools were already implementing some type of violence-prevention programming. We made no attempt to persuade schools to give up programs that they were

using because we wanted to evaluate the interventions under the same conditions they would face when implemented at scale. However, in order to achieve some level of statistical control over pre-existing violence-prevention programs in our analyses, we recorded other programs being used in each school.

Maintaining adequate statistical power presents a serious challenge. Assessing multiple persons in each school at multiple times produces highly precise and reliable measurements of school-level parameters. However, the statistical power of the design, and thus its sensitivity to true intervention effects, is more strongly related to the number of schools than to the total number of participants.^{14,26}

Our best estimate of the power of the final design is that it will be able to detect a modest mean difference at post-test of 0.21 standard deviations with power of 0.8. An obvious method for increasing power would be to increase the number of participating schools. However, the logistics of such an undertaking far exceed even the substantial level of funding currently committed to this project. In addition, some of the sites recruited virtually all of the schools in their surrounding areas. We have therefore attempted to maximize statistical power by other means. First, we have attempted to avoid interactions between sites and conditions by maintaining as much uniformity in assessment and delivery of interventions as possible. The presence of such an interaction would sap the power of the study to such an extent that even moderate effect sizes would be difficult to detect (R Gibbons, personal communication, Department of Psychiatry, UIC, January 9, 2000). Examining effects across two separate cohorts also provides an opportunity to replicate any trends that may be observed within the first cohort. Even though power increases most when additional schools are added, analyzing data from both cohorts through including cohort by condition interactions in the statistical models will increase the power of the design.

A second challenge relates to the *tension between internal and external validity*, or between efficacy and effectiveness.²⁷ This led to compromises on various aspects of the project's design. For example, although efforts were made to standardize implementation of the interventions through the use of manuals and close supervision of interventionists, modifications were sometimes needed to reflect intersite differences in variables such as school size, structure, and scheduling; school district policies; and geographic considerations (e.g., distance between schools). Choices of who (intervention staff or school staff) would implement the interventions also reflect the tension between internal and external validity. We chose project staff to implement the interventions in order to maximize the effect size (thus increasing internal validity) even though we don't believe that hiring outside staff would be feasible for most school districts. This tension was also reflected in the timing of intervention and assessment activities.

Completing the recruiting, consenting, and assessment process on the scale required by this study with participants who were often hard to locate was quite time consuming. Delaying the start of intervention activities until all pretest data had been collected was desirable in terms of internal validity. Unfortunately, because this would have delayed the start of the intervention until well into the school year, it would have impeded efforts to change school climate, and would have made it difficult to complete the interventions within the school year. As a compromise, the interventions were started after the bulk of the pretest data were collected, but some overlap was allowed between the initial intervention sessions and the wrapping up of the pretest data collection. Finally, this tension can be seen in the school recruitment and assignment. The sample of schools represents most of the types of school settings in which middle school-aged children are involved. Represented are urban and rural K–8 schools, small middle schools, and large consolidated middle schools where sixth graders have more limited contact with students in other grades. Thus, the study's results should be generalizable to a

variety of school settings. However, this variation in settings calls into question the equivalence of schools and raises the likelihood that there will be an interaction between the intervention and type of school, both of which have implications for internal validity.

A third challenge concerns the *interpretation of any differences in effects across cohorts*. One possible method of handling these data would be to pool data on individuals across cohorts, but such an approach would not allow for detecting differences in effects between cohorts. The second cohort provides an opportunity to assess the cumulative effect of two years of intervention at the school level. The impact of the interventions on Cohort 2 must be interpreted within the context of what occurred during each year of intervention. The effects of the interventions for the second cohort will depend on the effectiveness of the intervention in the prior year and on differences in implementation across cohorts. Intervention with Cohort 1 students and teachers may produce changes at the school level that may be reflected in pretest scores of Cohort-2 students entering these schools the following year. Including a cohort by condition interaction in analyses involving data from both cohorts will allow evaluation of differences in effects between cohorts and consistency of effects across cohorts.

Additional challenges involve *evaluating mediated and moderated relations*. The theoretical model of the GREAT Schools and Families intervention specifies several important theoretical mediators as targets of each intervention component.^{19,28,29} In addition to testing intervention effects on these mediating variables, it is important to test the relation of such changes to proximal and distal changes in violence and related behaviors. Change in school levels of aggression are predicted to be mediated by change in individual student, teacher, family, and school-level constructs. Somewhat different mediational processes are expected for the two interventions. Whereas the universal intervention is expected to directly affect all children, the targeted intervention is expected to produce its initial effects on the high-risk youth. Thus, efficacy of the targeted intervention on targeted children is, itself, hypothesized to be a mediator of the effects of the targeted intervention on school levels of aggression. The complexity of these hypothesized processes will require careful modeling of mediation across multiple levels of analysis. We believe that this can be done while still following the criteria for detecting mediation proposed by Kenny et al.³¹ and others.^{30,32} Although cross-level mediation has not been explored in the prevention literature, it has received some attention in the organizational literature³³ and in the violence-risk literature.³⁴

Few prevention studies have systematically undertaken detection of moderators, even though it is recognized that moderators are important considerations in prevention science.^{20,35} This may be due to design features that were not intended to examine variations in effects by subgroups. Because of its size and design, MVPP has a particular advantage for investigating moderator effects. Once again, the multi-level data and school-level assignment will present potential complications for these analyses. This issue of multi-level moderation is another issue that has received some attention in studies of organizational psychology,³⁶ which suggest methods for evaluating moderated relations that cut across levels of analysis.

The foregoing challenges exist because of the “school as subject” metaphor inherent in the final design. Because schools are the units that were randomized, they become the “subject” units to be used in analysis of the data.¹⁶ Individuals and points of measurement within individuals are expected to be clustered within schools. Evidence from the pilot study suggests that such clustering does indeed exist within the MVPP sample. Random effects models of teacher reports on the BASG (Behavioral Assessment System for Children) aggression, conduct problems, and learning problems subscales revealed significant ($p < 0.05$) intraclass correlations of approximately 0.07 on all three subscales. By contrast, the intraclass correlations for site were not significant in any analysis. These intraclass correlations suggest that approximately 7% of the variance in teacher reports of aggression is due to clustering

within schools. Thus, there appears to be empirical support for the notion that, in this sample, “school level of aggression” is a meaningful characterization.

Like all metaphors, the school-as-subject metaphor is only useful when it is not taken too literally. Aggression is an attribute that characterizes the behavior of individuals. Schools are behavior settings³⁷ in which individuals temporarily participate. To speak of reducing school levels of aggression is actually to speak of reducing schools' contributions to individual levels of aggression. The degree to which individual observations are clustered within schools can provide an estimate of the likely upper limit of the effectiveness of the aspects of the interventions that seek to change school climate or norms.

Conclusion

Although it is early to comment definitively on the strengths and weaknesses of the design, a few observations appear in order. The use of a between-school design will, very likely, prevent infiltration of effects between conditions. Student transfers between conditions are expected to be few, as are the opportunities for teachers to share what they have learned with teachers in schools assigned to other conditions. The design also appears to have reasonable internal validity. Although random assignment of the schools to conditions is a prime factor in maximizing internal validity, the relatively small number of units randomized within each site may not eliminate pretest differences across conditions.³⁸ Fortunately, with the initial cohort in the present study, early indications suggest that randomization of the schools succeeded in producing near equivalence across conditions.

Among other potential weaknesses of the design is the need for high levels of participation and faithful implementation of the interventions in order to maintain internal validity. At present, implementation levels have been at acceptable levels for the universal intervention, and at less-than-desirable levels for the targeted intervention. We believe, however, that the challenges we have faced in achieving high participation levels in the targeted intervention are not at all unlike those faced by the schools themselves when attempting to motivate parents and families to participate in aspects of their children's education. In this case, finding effects for the intervention under the conditions of low participation rates could be considered evidence for the effectiveness of the intervention under conditions close to what might be expected if the interventions were implemented at scale. The implementation of the intervention in a single grade is also a weakness of the design, albeit one necessitated by power and cost considerations. Developing interventions directed at each grade level simultaneously may ultimately prove to be the most appropriate strategy for producing cohort-wide effects.²¹ In the meantime, the present study's focus on multiple cohorts within the same grade level provides a basis for the initiation of such an effort.

The proverb quoted at the beginning of this paper impugns committee designs for combining poorly matched and coordinated parts in an attempt to honor disparate viewpoints of committee members. At the same time, it must be noted that the ungainly camel is perfectly designed to meet the extreme demands of life in the desert. The field of violence prevention, and indeed that of behavioral interventions generally, suffer from an absence of scientifically rigorous evidence gathered in real-world settings.³⁹ The design of the MVPP Study includes many seemingly mismatched and contradictory elements. Nonetheless, we believe it is well suited to the needs of the field at this time.

References

1. Multisite Violence Prevention Project. The Multisite Violence Prevention Project: background and overview. *Am J Prev Med* 2004;26(suppl):3–11. [PubMed: 14732182]

2. Gibbons RD, Hedeker D, Elkin I. Some conceptual and statistical issues in analysis of longitudinal psychiatric data. *Arch Gen Psychiatry* 1993;50:739–50. [PubMed: 8357299]
3. Gottfredson, DC. School-based crime prevention. In: Sherman, LW.; Gottfredson, DC.; Mackenzie, D.; Eck, J.; Reuter, P.; Bushway, S., editors. *Preventing crime: what works, what doesn't, what's promising: a report to the United States Congress* (NCJ 171676). Washington DC: U.S. Department of Justice, Office of Justice Programs; 1997. p. 125-82.
4. Campbell, DT.; Stanley, JC. *Experimental and quasi-experimental Designs for research*. New York: Houghton-Mifflin; 1963.
5. Farrell AD, Meyer AL, White KS. Evaluation of Responding in Peaceful and Positive Ways (RIPP): a school-based prevention program for reducing violence among urban adolescents. *J Clin Child Psychol* 2001;30:451–63. [PubMed: 11708233]
6. Meyer AL, Farrell AD. Social skills training to promote resilience and reduce violence in African American Middle School Students. *Educ Treatment Child* 1998;21:461–88.
7. Henry D, Guerra NG, Huesmann LR, Tolan PH, VanAcker R, Eron LD. Normative influences on aggression in urban elementary school classrooms. *Am J Community Psychol* 2000;28:59–81. [PubMed: 10824274]
8. Cunningham PB, Henggeler SW. Implementation of an empirically based drug and violence prevention and intervention program in public school settings. *J Clin Child Adolesc Psychol* 2001;30:221–32.
9. Farrell AD, Valois RF, Meyer AL, Tidwell R. Impact of the RIPP violence prevention program on rural middle school students: a between-school study. *J Prim Prev* 2003;24:143–67.
10. Metropolitan Area Child Study Research Group. A cognitive-ecological approach to preventing aggression in urban settings: initial outcomes for high risk children. *J Consult Clin Psychol* 2002;70:179–94. [PubMed: 11860044]
11. Conduct Problems Prevention Research Group. Initial impact of the fast track prevention trial for conduct problems: I. The high-risk sample. *J Consult Clin Psychol* 2001;67:631–47.
12. Conduct Problems Prevention Research Group. Initial impact of the fast track prevention trial for conduct problems: II. Classroom effects. *J Consult Clin Psychol* 1996;67:648–57.
13. Kazdin, AE. *Research design in clinical psychology*. Vol. 4th. Boston MA: Allyn & Bacon; 2003.
14. Hedeker D, Gibbons RD, Waternaux C. Sample size estimation for longitudinal designs with attrition: comparing time-related contrasts between two groups. *J Educ Behav Stats* 1999;24:70–93.
15. Raudenbush SW, Liu X. Statistical power and optimal design for multisite randomized trials. *Psychol Methods* 2000;5:199–213. [PubMed: 10937329]
16. Bryk, AS.; Raudenbush, SW. *Hierarchical Linear Models: applications and data analysis methods*. Newbury Park CA: Sage; 1992.
17. Boruch, RF. Randomized controlled experiments for evaluation and planning. In: Bickman, L.; Rog, DJ.; Debra, J., editors. *Handbook of applied social research methods*. 1998. p. 161-91.
18. Henry, DB. *Validity of Teacher Nominations and Ratings in Selecting Influential Aggressive Students for the MVPP Project*. Chicago IL: Institute for Juvenile Research; Jun. 2001 Technical Report
19. Smith EP, Gorman-Smith D, Quinn WH, Rabiner DL, Tolan PH, Winn DM. Multisite Violence Prevention Project. Community-based multiple family groups to prevent and reduce violent and aggressive behavior: the GREAT Families Program. *Am J Prev Med* 2004;26(suppl):39–47. [PubMed: 14732186]
20. Farrell AD, Meyer AL, Kung EM, Sullivan TN. Development and evaluation of school-based violence prevention programs. *J Clin Child Psy* 2001;30:207–20.
21. Farrell AD, Meyer AL, Sullivan TN, Kung EM. Evaluation of the Responding in Peaceful and Positive Ways (RIPP) Seventh Grade Violence Prevention Curriculum. *J Child Fam Stud* 2003;12:101–20.
22. Miller-Johnson S, Sullivan TN, Simon TR. Multisite Violence Prevention Project. Evaluating the impact of interventions in The Multisite Violence Prevention study: Samples, procedures, and measures. *Am J Prev Med* 2004;26(suppl):48–61. [PubMed: 14732187]
23. Murray D, McKinlay S, Martin D, et al. Design and analysis issues in community trials. *Evalu Rev* 2000;24:493–514.

24. Bock, RD. Within-subject experimentation in psychiatric research. In: Gibbons, RD.; Dysken, MW., editors. Statistical and methodological advances in psychiatric research. New York: Spectrum Books; 1983. p. 59-90.
25. Bock, RD. Measurement of human variation: a two-stage model. In: Bock, RD., editor. Multi-Analysis of educational data. Orlando FL: Academic Press; 1989. p. 319-42.
26. Raudenbush SW. Statistical analysis and optimal design for cluster randomized trials. *Psychol Methods* 1997;173-85.
27. Dodge KA. The science of youth violence prevention: progressing from developmental epidemiology to efficacy to effectiveness to public policy. *Am J Prev Med* 2001;20(suppl):63-70. [PubMed: 11146262]
28. Meyer AL, Allison KW, Reese LE, Gay F. Multisite Violence Prevention Project. Choosing to be violence free in middle school: the student component of the GREAT Schools and Families universal program. *Am J Prev Med* 2004;26(suppl):20-28. [PubMed: 14732184]
29. Orpinas P, Horne AM. Multisite Violence Prevention Project. A teacher-focused approach to prevent and reduce students' aggressive behavior: the GREAT Teacher Program. *Am J Prev Med* 2004;26(suppl):29-38. [PubMed: 14732185]
30. Baron RM, Kenny D. The moderator-mediator variable distinction in social psychological research: Conceptual, strategic and statistical considerations. *J Pers Soc Psychology* 1986;1173-82.
31. Kenny, DA.; Kashy, DA.; Bolger, N. Data analysis in social psychology. In: Gilbert, DT.; Fiske, ST.; Lindzey, G., editors. The handbook of social psychology, volume 1. Oxford UK: Blackwell Publishers; 1998. p. 233-65.
32. MacKinnon D, Krull JL, Lockwood CM. Equivalence of the mediation, confounding, and suppression effects. *Prevention Science* 2000;1:173-212. [PubMed: 11523746]
33. Griffin MA, Mathieu JE, Jacobs RR. Perceptions of work contexts: Disentangling influences at multiple levels of analysis. *J Occup Organ Psychol* 2001:563-79.
34. Tolan PH, Gorman-Smith D, Henry DB. The Developmental Ecology of Urban Males' Youth Violence. *Dev Psychol* 2003;39:274-91. [PubMed: 12661886]
35. Tolan, PH.; Guerra, NG. What works in reducing adolescent violence: an empirical review of the field. Center for the Study and Prevention of Youth Violence Monograph. Boulder CO: University of Colorado; 1994.
36. Schriesheim CA, Coglisier CC, Neider LL. Is it "trustworthy?": a multiple-levels-of-analysis reexamination of an Ohio state leadership study, with implications for future research. *Leadersh Q* 1995;6:111-45.
37. Barker, RG. Ecological psychology: concepts and methods for studying the environment of human behavior. Stanford CA: Stanford University Press; 1968.
38. Hsu LM. Random sampling, randomization and equivalence of contrasted groups in psychotherapy outcome research. *J Consult Clin Psychol* 1989;131-7. [PubMed: 2647799]
39. Aral SO, Peterman TA. Do we know the effectiveness of behavioural interventions? *Lancet* 1998;351(suppl III):33-6. [PubMed: 9652720]

Figure 1.
Graphic depiction of the Multisite Violence Prevention Project (MVPP) study research design.

Table 1

Community poverty and crime characteristics by site

Variable	U.S. Total	Site				Overall
		UIC	Duke	UGA	VCU	
Poverty (% of children under 18) ^a	15.7%	30.6%	17.0%	26.1%	36.6%	27.6% ^c
Juvenile arrests for violent crime/ 100,000 population ^b	42.7	62.6	47.0	74.1	69.2	63.2 ^c

^a School district youth poverty data for 1997, Source: U.S. Census Bureau.

^b County juvenile violent crime statistics. Source: FBI Uniform Crime Reports, 1997.

^c Average percentage or rate/100,000.

UIC, University of Illinois at Chicago; UGA, The University of Georgia; VCU, Virginia Commonwealth University.