

RESEARCH ARTICLE

EVALUATION OF THE G.R.E.A.T. PROGRAM

Short- and Long-Term Outcome Results from a Multisite Evaluation of the G.R.E.A.T. Program

Finn-Aage Esbensen

University of Missouri—St. Louis

D. Wayne Osgood

Pennsylvania State University

Dana Peterson

University at Albany

Terrance J. Taylor and Dena C. Carson

University of Missouri—St. Louis

Youth gangs continue to garner substantial attention from the media, public, and academic researchers as a result, in large part, of the violence attributed to gang members. Several prevention, intervention, and suppression programs have been introduced to address problems associated with youth gangs, but to date, relatively few have been deemed as promising, let alone as effective (e.g., Esbensen, Freng, Taylor, Peterson, and Osgood, 2002; Howell, 2012; Klein and Maxson, 2006; Maxson, Egley, Miller, and Klein, 2013; Reed and Decker, 2002).

Given the disruptive influence that gangs pose on school safety and academic performance (as well as on communities), gangs and associated violence are targets of prevention

This research was made possible, in part, by the support and participation of seven school districts, including the School District of Philadelphia. This project was supported by Award No. 2006-JV-FX-0011 from the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. The opinions, findings, and conclusions or recommendations expressed in this article are those of the authors and do not necessarily reflect the views of the Department of Justice or of the seven participating school districts. We would like to express our appreciation to the students who made this project possible by completing the student questionnaires. And this project would have been impossible without our team of colleagues and research assistants; special thanks to Adrienne Freng, Kristy Matsuda, J. Michael Vecchio, and Stephanie A. Wiley for their invaluable assistance. Direct correspondence to Finn-Aage Esbensen, Department of Criminology and Criminal Justice, University of Missouri—St. Louis, One University Blvd., St. Louis, MO 63121-4499 (e-mail: esbensen@umsl.edu).

and intervention efforts. Several programs have been developed and promoted as “effective,” and school administrators often are confronted with slick promotional materials advocating the “wonderfulness” of a wide array of programs claiming they will either reduce problem behaviors, increase social skills, promote positive youth behavior, or all of the above. Whenever possible, these school administrators should be encouraged to choose programs with a history of evaluation findings supporting program effectiveness. Although many programs exist, few have been subjected to rigorous program evaluations. Of particular importance is the lack of programs subjected to randomized control trials (RCTs). The current study presents one example of short- and long-term findings from a recent RCT assessing the effectiveness of a gang-prevention program—Gang Resistance Education and Training (G.R.E.A.T.). The findings from this study can aid recent efforts to provide empirically based information to school administrators and community leaders seeking to implement evidence-supported programs.

Despite the relative absence of the most rigorous evaluation designs (i.e., RCTs) assessing gang-prevention programs, an increasing number of agencies/organizations has developed criteria for classifying programs into various categories ranging from “not effective” to “effective” or “model” programs based on the findings of empirical evaluations. For example, the Blueprints Series (Mihalic, Fagan, Irwin, Ballard, and Elliott, 2002; Mihalic and Irwin, 2003) identifies model violence-prevention programs that have withstood rigorous scientific evaluations, and the Maryland Report (Sherman et al., 1997) assessed the effectiveness of a broad range of projects. In 2005, the Helping America’s Youth (HAY) Community Guide (Howell, 2009) rated programs identified by nonfederal agencies on three levels: Level 1 (exemplary or model programs based on evaluation designs of the “highest quality”), Level 2 (effective programs based on quasi-experimental research), and Level 3 (promising programs). Similarly, the Office of Juvenile Justice and Delinquency Prevention provides a listing of exemplary, effective, or promising programs (OJJDP, 2010), and in 2010, the National Institute of Justice introduced its “Crime Solutions” website, which identifies effective and promising programs (crimesolutions.gov).

Of particular relevance to the current study, the G.R.E.A.T. program is currently rated as “promising” by OJJDP and by Crime Solutions, and it is designated as Level 2 (effective) in the Helping America’s Youth rating scale (findyouthinfo.gov). Additionally, a recent systematic review found that the G.R.E.A.T. program was one of only a handful of gang-awareness programs meeting strict guidelines for determining program effectiveness (Gravel, Bouchard, Descormiers, Wong, and Morselli, 2013). These designations were initially based on findings from two multisite evaluations of the “original” program curriculum: one cross-sectional study conducted in 1995 (Esbensen and Osgood, 1999) and one longitudinal study conducted between 1995 and 1999 (Esbensen, Osgood, Taylor, Peterson, and Freng, 2001), but the current classifications are based on short-term findings from an evaluation of the revised G.R.E.A.T. program (Esbensen, Peterson, Taylor, and Osgood, 2012).

The G.R.E.A.T. program has been in existence since 1991 and has received some acclaim since its inception. Originally developed as a nine-lesson curriculum based on Drug Abuse Resistance Education (DARE), the program underwent a substantial curriculum revision after the findings of the two aforementioned studies. Once these revisions were made, there was considerable interest in determining whether the program would be found to be more effective at meeting program goals than was the case in the evaluations of the original G.R.E.A.T. program. In a recent publication, we reported on the 1-year posttreatment effects of the revised G.R.E.A.T. program (Esbensen et al., 2012). This article provides an overview of those results but focuses on the long-term program effects (up to 4 years posttreatment) while reporting additional analyses that examine (a) site-specific program outcomes and (b) the extent to which preexisting risk factors impact program effectiveness. Our findings contribute to the sparse body of knowledge about effective gang-prevention strategies.

We begin with a description of the G.R.E.A.T. program. Next, we turn to a recap of findings from previous evaluations, with a particular emphasis on critiques levied at both the program and the evaluation findings, and how the current program and evaluation overcome many of the limitations highlighted previously. We then describe the methodology employed and the results of the current evaluation of the revised G.R.E.A.T. program. We conclude with a discussion of how the current results fit with those of previous evaluations and what this means for gang-prevention programming.

G.R.E.A.T. Program¹

The G.R.E.A.T. program is a school-based gang- and violence-prevention program with three primary goals:

- (1) To teach youth to avoid gang membership
- (2) To prevent violence and criminal activity
- (3) To assist youth in developing positive relationships with law enforcement.

Developed as a universal prevention program targeting youth in early adolescence (i.e., 6th or 7th graders), the G.R.E.A.T. program was classified as a gang-awareness program in a recent review of gang programs (Gravel et al., 2013). The original G.R.E.A.T. program, developed by Phoenix-area police departments in 1991, was a cognitive-based program that taught students about crime and its effect on victims, cultural diversity, conflict resolution skills, meeting basic needs (without a gang), responsibility, and goal setting.^{2,3} Uniformed

1. This section describing the G.R.E.A.T. program is partially excerpted from Esbensen et al. (2012).

2. The core program component of G.R.E.A.T. is its middle-school curriculum, and often this is what is referred to with the term "G.R.E.A.T. program." Other optional components of G.R.E.A.T. are an elementary-school curriculum, a summer program, and G.R.E.A.T. Families.

3. For a detailed account of the political context surrounding the development of the original G.R.E.A.T. program, consult Winfree, Peterson Lynskey, and Maupin (1999).

law enforcement officers taught the curriculum in schools, and teachers were requested to complement the program content during regular classes.

The revised G.R.E.A.T. program contains much of the substance of the original program but, importantly, was also informed by the work of educators and prevention specialists and the growing body of risk factor research (see Esbensen et al., 2002; Esbensen, Peterson, et al., 2011, for a detailed account of the program review that informed the curriculum revision). As a result, the revised G.R.E.A.T. program was expanded from 9 to 13 lessons. It is still taught primarily by uniformed law enforcement officers (largely police officers and sheriff's deputies, but federal agents from the U.S. Marshals and the Bureau of Alcohol, Tobacco, and Firearms as well as District Attorneys also have been trained and certified to teach G.R.E.A.T.), and it incorporates classroom management training of officers and a focus on students' skill development through cooperative learning strategies—important pedagogical tools for educational settings (Gottfredson, 2001).⁴

Two school-based programs, the Seattle Social Development Model (SSDM) and Life Skills Training (LST), guided the revision of the G.R.E.A.T. program. LST is classified as a model program by the rigorous Blueprint standards, whereas the SSDM has received acclaim from a variety of sources. The SSDM is a comprehensive model that seeks to reduce delinquency and violence by building a positive learning environment incorporating several different classroom management components, such as cooperative learning, proactive classroom management, and interactive teaching (Catalano, Arthur, Hawkins, Berglund, and Olson, 1998). The LST program is a 3-year intervention in which two annual booster sessions supplement the initial program (Dusenbury and Botvin, 1992). LST consists of three components:

- (1) Self-management skills
- (2) Social skills
- (3) Information and skills directly related to the problem of drug abuse.

The revised G.R.E.A.T. program adopted some of the strategies from LST (in fact, some of the LST curriculum writers participated in the rewriting of the G.R.E.A.T. program), including an emphasis on the development of skills rather than on the assimilation of knowledge, and it incorporated problem-solving exercises and cooperative learning strategies.

During the revision of the G.R.E.A.T. program, incorporation of findings from research identifying risk factors for gang affiliation and violent offending was a primary enhancement to the program. While recognizing the importance of risk factors in all five domains (i.e., community, school, peer, family, and individual), the curriculum writers acknowledged that

4. Information about the G.R.E.A.T. program and an overview of the G.R.E.A.T. lessons included in the middle-school curriculum can be found at great-online.org/.

a school-based program could best address risk factors in the school, peer, and individual domains. As such, the revised curriculum addresses the following risk factor areas: school commitment, school performance, association with conventional or delinquent peers, susceptibility to peer influence, involvement in conventional activities, empathy, self-control (impulsivity, risk-seeking, self-centeredness, and anger control), perceived guilt, neutralization techniques (for lying, stealing, and hitting), and moral disengagement (e.g., Battin, Hill, Abbott, Catalano, and Hawkins, 1998; Esbensen and Deschenes, 1998; Esbensen, Huizinga, and Weiher, 1993; Esbensen, Peterson, Taylor, and Freng, 2010; Hill, Howell, Hawkins, Battin-Pearson, 1999; Howell and Egley, 2005; Klein and Maxson, 2006; Maxson and Whitlock, 2002; Maxson, Whitlock, and Klein, 1998; Pyrooz, Fox, and Decker, 2010; Thornberry, 1998; Thornberry, Krohn, Lizotte, Smith, and Tobin, 2003).

Research also has demonstrated the deleterious cumulative effects of risk exposure; the greater the number of risk factors or the greater the number of risk domains experienced, the greater the odds of youth gang and violence involvement, with these increases in risk associated with exponential increases in odds of becoming gang involved (Esbensen et al., 2010; Thornberry et al., 2003). This collective body of risk factor research suggests that prevention programs should attempt to address risk factors in multiple domains and to do so earlier, rather than later, in adolescence, both before the factors accumulate and before the typical age of onset for gang involvement—i.e., prior to the age of approximately 14 years of age (Esbensen and Huizinga, 1993; Hill et al., 1999; Thornberry et al., 2003). To this end, the revised G.R.E.A.T. curriculum addresses multiple risk factors across multiple domains and is taught in 6th or 7th grade, when students average 11–13 years of age.

Comparing Previous Evaluations with the Current Evaluation

Two previous multisite evaluations of the original G.R.E.A.T. program were conducted (Esbensen and Osgood, 1999; Esbensen, Osgood, et al., 2001). These evaluations found different degrees of “success” of the G.R.E.A.T. program at meeting its stated goals. A brief background on these studies provides context for the current study’s findings.

The first was a cross-sectional study of nearly 6,000 8th graders attending public schools in 11 U.S. cities conducted in 1995 (Esbensen and Osgood, 1999). The study found many results supportive of the original G.R.E.A.T. program’s effectiveness at reaching its goals. A variety of modeling strategies was employed, with three increasingly restrictive samples examined. Under the most restrictive analyses, G.R.E.A.T. students were found to be significantly “better” than non-G.R.E.A.T. students on 14 of 33 outcome measures examined. Program participants were consistently found to have lower levels of drug use and minor delinquent offending than nonparticipants. Examining attitudinal measures with consistent findings across modeling strategies, G.R.E.A.T. students had more negative attitudes about gangs, fewer delinquent friends, more friends involved in prosocial activities, greater commitment to peers promoting prosocial behaviors, less likelihood of acting impulsively, higher self-esteem, more commitment to success at school, and higher levels of attachment to both

mothers and fathers than their non-G.R.E.A.T. counterparts. Additionally, program effects on five outcome measures—peer delinquency, friends' involvement in prosocial activities, commitment to peers who promote prosocial activities, self-esteem, and commitment to success at school—were found to be stronger for males (relative to females), and the effects for two outcomes—commitment to and involvement with prosocial peers—were stronger for Black and Hispanic youth (relative to White youth).

The second evaluation was a prospective longitudinal study of more than 2,000 youth attending public schools in six U.S. school districts. The students were followed from 7th grade (6th in one site) until 11th grade (10th in one site). In 15 of the 22 schools that participated, the classrooms were randomly assigned to treatment and control conditions; in the remaining schools, because of constraints such as G.R.E.A.T. officers' schedules, the classrooms were assigned to condition based on matching procedures (e.g., one teacher's morning class was assigned to the treatment condition, whereas the same teacher's afternoon class was assigned to the control condition). The results of the longitudinal analyses were less supportive of the program than the cross-sectional results. Specifically, 5 of the 32 outcome measures were found to be consistent with beneficial program effects in preprogram versus postprogram (all 4 years combined) contrasts; G.R.E.A.T. students were found to have lower rates of victimization, more negative views of gangs, more favorable attitudes toward the police, more involvement with prosocial peers, and reduced levels of risk seeking. The results examining trends over time were less pronounced, with only three outcomes reaching statistical significance (victimization and involvement with and commitment to prosocial peers) and evidence that effects were delayed (rather than immediate). It is important to note, however, that 25 of the 32 outcome measures examined were in a direction consistent with positive program effects. Also, in contrast to the previous cross-sectional analyses, no significant differences in program effects were found across subgroups by sex or race/ethnicity.

Many of the accolades the G.R.E.A.T. program has received were based, in some part, on the relatively positive findings of the cross-sectional study and on the finding of small lagged effects on some program outcomes in the longitudinal evaluation. That is not to say that these studies were definitive "proof" that the original G.R.E.A.T. program was an undeniable "success." In fact, the results from the longitudinal evaluation were viewed as evidence of a lack of program effect and contributed to the comprehensive program review and revision. Some commentators were critical of the G.R.E.A.T. program and raised concerns about the previous evaluations. Klein and Maxson (2006), for example, noted that the most promising results were found employing the least rigorous methodological design: the cross-sectional study. The more rigorous longitudinal design found less support for the program, as demonstrated by the relative lack of significant differences between treatment and control groups after program exposure and only modest program effects when differences were found. They also highlight the lack of a significant program effect on gang membership, which is the key program outcome.

Klein and Maxson (2006) identified three factors that could account for the failure of the program to reduce the odds of gang membership. First, the original G.R.E.A.T. program was based on a “failed” program model: DARE. Second, the original G.R.E.A.T. program was not “gang specific”; rather, it was based on more general social skills targeted at delinquency prevention. Third, the program was aimed at a population with relatively low rates and probabilities of gang membership. Specifically, Klein and Maxson argued that this universal program focusing on all 7th-grade classrooms would be unlikely to reach the target group because few 7th graders attending schools are involved with gangs.

Ludwig (2005) presented additional concerns about the effectiveness of the original G.R.E.A.T. program. In addition to reinforcing the point that evaluations of the G.R.E.A.T. program found no effect on key dependent variables of gang involvement, drug use, or delinquency, Ludwig also noted that sample attrition throughout the study reduces the confidence that we should have about program effectiveness found in the longitudinal study.

Interest was renewed in the question of program effectiveness after the revised curriculum was fully implemented in 2003. In July 2006, the National Institute of Justice selected the University of Missouri—St. Louis to conduct a process and outcome evaluation of the revised G.R.E.A.T. program. The current program and evaluation address many limitations of the previous program and evaluation designs and build on the results of those previous studies. First, as described, the G.R.E.A.T. program underwent major changes after a substantial curriculum review based in large part on the findings of the previous evaluations. Of particular importance was an emphasis on linking specific program lessons with risk factors found to be important in gang joining and delinquency. In short, whereas the revised program still deals with general social skills and the prevention of delinquency, greater attention is now paid to the risk factors found to be associated with gangs. Second, the criticism that the original G.R.E.A.T. program was modeled after the DARE program was addressed during the curriculum review, with the revised G.R.E.A.T. program now modeled after two highly acclaimed school-based prevention programs (LST and SSDM). Third, Klein and Maxson’s (2006) critique of the universal targeted population raised the issue of efforts attempting to reduce statistically rare events. As many gang researchers have noted, gang membership is a rare event, even in the most at-risk neighborhoods or sub-populations. At the same time, the past 20+ years of gang research have demonstrated that gangs and gang-involved youth are found in communities not only across the United States but across the world (e.g., Covey, 2010; Esbensen and Maxson, 2012; Hagedorn, 2008). Although one can question the utility of trying to prevent a statistically rare event, it does not seem reasonable to abandon general prevention efforts, especially given researchers’ and practitioners’ inability to identify unique risk factors for gang membership and recent studies indicating a great deal of overlap in risk factors for gang membership and violence (Esbensen et al., 2010; Peterson and Morgan, 2013).⁵ Finally, with respect to

5. See Klein and Maxson’s (2006) review of the gang risk factor literature.

methodological issues raised by Ludwig and others, extensive efforts were made to increase both the active consent rates and the survey completion rates in the current evaluation. The results of these efforts are reported in the Methods section.

Although a previous study reporting short-term program effects of the revised G.R.E.A.T. program was published in 2012 (Esbensen et al., 2012), the current study focuses on long-term effects across 4 years posttreatment.⁶ This long-term emphasis is important not only to determine whether short-term effects are sustained over time but also because it captures youth at the ages of highest risk of gang joining (Klein and Maxson, 2006) and because delayed effects were detected in the previous longitudinal evaluation (Esbensen, Osgood, et al., 2001). Additionally, supplemental analyses reported in the current study (a) investigate the extent to which the overall results are replicated at each of the seven individual research sites and (b) control for preexisting risk factors. These important questions address the universality of program effects and introduce a more rigorous assessment than was possible in the previous study. As such, the current study goes well beyond the 1-year program effects reported in the 2012 study.

Methods

Site and School Selection

Seven cities (Albuquerque, NM; Chicago, IL; a Dallas-Fort Worth [DFW], TX, area district; Greeley, CO; Nashville, TN; Philadelphia, PA; and Portland, OR) were selected to provide a diverse sample of schools and students. Sample selection was guided by three main criteria:

- (1) Geographic and demographic diversity
- (2) A substantial number of officers delivering the program to some, but not all, students
- (3) Information provided by the National Gang Center about cities' level of gang activity.

The goal was to develop a sample that was geographically and demographically diverse across cities with varying degrees of gang activity. The student and school sample is representative of the students and schools in each of the seven cities' school districts. The final sample consists of 3,820 students (for whom active consent was obtained) nested within 195 classrooms (102 received G.R.E.A.T. and 93 did not receive the program) in 31 schools.

Active Parental Consent

Active parental consent was required for student participation (see Esbensen, Melde, Taylor, and Peterson, 2008, for a detailed description of the active consent process), and as stated previously, significant effort was made to improve these rates over what was achieved in the previous evaluation. Teachers were recruited to assist with the process, and the

6. In several sections of this article, we report long-term effects alongside the previously reported short-term effects for comparison purposes.

combined effort of teachers and evaluators produced a commendable active consent rate of 78%. Of the 4,905 students represented on the classroom rosters at the time of the consent process, 89.1% of youths ($n = 4,372$) returned a completed consent form, with 77.9% of parents/guardians ($n = 3,820$) allowing their child's participation and 11.3% ($n = 552$) declining.⁷

Research Design and Random Assignment of Classrooms

The outcome evaluation employs an experimental longitudinal panel design (a randomized control trial with long-term follow-up) in which classrooms in each of the participating schools were randomly assigned to the treatment (i.e., G.R.E.A.T.) or control condition.⁸ Once it was determined in which grade level (6th grade in 26 schools and 7th grade in 5 schools) and in which core subject area (commonly social studies but also in English and science classes) the program would be taught, we enumerated all the grade-level classrooms (ranging from 3 to 12). In situations with an odd number of classes, we made the a priori decision to oversample treatment classes (in partial recognition of the fact that many of the principals were reluctant to "deprive" any of their students of the program). The list of classes was then numbered from one through highest, and a table of random numbers was consulted to select the classrooms in which G.R.E.A.T. would be taught. Unselected classrooms comprised the control group.

All students in the treatment and control classrooms were eligible to participate in the evaluation, and those for whom active parental consent was obtained ($N = 3,820$) were then asked to participate in the evaluation by completing a confidential group-administered pretest questionnaire. After completion of the G.R.E.A.T. program in each school, students in both the experimental and control groups were then requested to complete posttests and four annual follow-up surveys. The retention rates across the six waves of data included in the outcome analyses reported in this article were 98.3%, 94.6%, 87.3%, 82.8%, 74.2%, and 71.9%, respectively, for Wave 1 (pretest) through Wave 6 (4 years posttreatment).⁹ These response rates reflect the diligent efforts of the research assistants working on this project. It is particularly challenging to track students through multiple schools and school districts, especially in a highly mobile sample: Although initially enrolled in 31 middle schools at pretest, students were surveyed in more than 200 different schools in Waves 5 and 6 when the students were in high school. We tracked students in each of the seven cities, identifying the schools (or cities) to which students had transferred. In several

7. This might be compared with an active consent rate of 57% of students being allowed to participate in the previous longitudinal evaluation of the original G.R.E.A.T. program (Esbensen, Osgood, et al., 2001).

8. This is an improvement over the previous longitudinal evaluation design, in which random assignment was possible in only 15 of 22 participating schools (Esbensen, Osgood, et al., 2001).

9. This compares with completion rates of 87%, 80%, 86%, 76%, 69%, and 67% in the previous longitudinal evaluation.

instances (especially for students who had moved outside of the district), this required soliciting information from school administrative assistants, teachers, or other students because, somewhat surprisingly, this information often was not available from the central district office or from computerized records. These efforts at locating students, combined with multiple visits to individual schools (in some instances more than 10 trips to survey chronically truant students), contributed to the fact that we could survey virtually all the students still enrolled in schools in the original districts. We obtained permission from principals at each of the new schools to survey the transfer students—clearly, a time- and labor-intensive effort, but one well worth achieving these high response rates.

Student Sample Characteristics

Based on responses provided at Wave 1, the sample is split evenly between males and females; most (55%) youth reside with both biological parents, and the majority (88%) was born in the United States (see Table 1). The sample is racially/ethnically diverse, with Hispanic youth (37%), White youth (27%), and African American youth (18%) accounting for 82% of the sample. Approximately two thirds of the youth (61%) were aged 11 years or younger at the pretest, representing the fact that 26 of the 31 schools delivered the G.R.E.A.T. program in 6th grade. Three of the six Chicago schools and two of four schools in Albuquerque taught G.R.E.A.T. in 7th grade; thus, students in these sites were somewhat older than students in the other sites.

Measurement

Program Goals

To assess program effectiveness, it was essential that measures of the three program goals be included in the student surveys. Additionally, the G.R.E.A.T. lessons targeted several secondary outcomes that sought to reduce known risk factors for delinquency and gang membership. We developed a student questionnaire that captured the essence of this skills building program, including many of the risk factors associated with gang membership as well as lesson-specific social skills (e.g., dealing with peer pressure and being able to say no). To reiterate, the G.R.E.A.T. program has three primary goals, as follows:

- (1) To help youth avoid gang membership
- (2) To reduce violence and criminal activity
- (3) To help youth develop a positive relationship with law enforcement.

Gang membership is measured by a single-item question that is part of a larger set of questions about youth gangs. Specifically, students answered the question, “Are you now in a gang?” This self-nomination approach has been found to be a valid and robust measure of gang affiliation (e.g., Esbensen, Winfree, He, and Taylor, 2001; Thornberry

TABLE 1

Sample Characteristics at Wave 1

| | Full Sample N = 3,820 % | ABQ n = 591 % | CHI n = 500 % | DFW area n = 614 % | GRE n = 582 % | NSH n = 590 % | PHL n = 457 % | POR n = 486 % |
|---------------------|-------------------------------|---------------------|---------------------|--------------------------|---------------------|---------------------|---------------------|---------------------|
| Sex | | | | | | | | |
| Male | 50 | 50 | 50 | 54 | 52 | 55 | 43 | 42 |
| Female | 50 | 50 | 50 | 46 | 48 | 46 | 57 | 58 |
| Race | | | | | | | | |
| White | 27 | 16 | 7 | 20 | 34 | 45 | 12 | 51 |
| African American | 18 | 4 | 29 | 21 | 2 | 23 | 44 | 7 |
| Hispanic | 37 | 49 | 56 | 46 | 50 | 17 | 20 | 13 |
| American Indian | 4 | 10 | 1 | 2 | 5 | 1 | 4 | 4 |
| Asian | 4 | 2 | 1 | 6 | 1 | 6 | 4 | 9 |
| Multiracial | 8 | 14 | 2 | 5 | 4 | 4 | 12 | 13 |
| Other | 4 | 5 | 2 | 1 | 5 | 5 | 5 | 3 |
| Age | | | | | | | | |
| 11 or younger | 61 | 35 | 18 | 74 | 77 | 80 | 61 | 79 |
| 12 | 29 | 43 | 44 | 25 | 22 | 19 | 35 | 20 |
| 13 or older | 10 | 23 | 38 | 2 | 2 | <1 | 4 | 1 |
| Mean Age | 11.48 | 11.87 | 12.22 | 11.27 | 11.23 | 11.19 | 11.42 | 11.21 |
| Living Arrangement | | | | | | | | |
| Both bioparents | 55 | 52 | 57 | 60 | 58 | 60 | 38 | 58 |
| Single parent | 20 | 20 | 19 | 15 | 14 | 18 | 24 | 15 |
| 1 Bio/1 stepparent | 13 | 15 | 12 | 14 | 15 | 12 | 18 | 13 |
| 1 Bio/1 other adult | 7 | 7 | 7 | 7 | 7 | 7 | 11 | 8 |
| Other relatives | 3 | 6 | 3 | 3 | 4 | 2 | 8 | 5 |
| Other arrangement | 2 | 1 | 1 | 1 | 3 | 2 | 2 | 1 |
| Immigration Status | | | | | | | | |
| Born outside U.S. | 12 | 10 | 9 | 13 | 11 | 15 | 11 | 15 |
| Born in U.S. | 88 | 90 | 91 | 87 | 89 | 85 | 89 | 85 |

et al., 2003). To measure *delinquency and violent offending*, students completed a 14-item self-reported delinquency inventory, including response categories that allowed for assessment of both ever and annual prevalence as well as frequency of offending during the past 6 months (past 3 months at Wave 2, the posttest). We treated this self-report inventory as a composite measure of general delinquency (examining both a variety and a frequency score) but also created a separate measure of violent offending consisting of three items (attacked someone with a weapon, used a weapon or force to get money or things from people, and been involved in gang fights). To measure the third specific program goal (improving relations with law enforcement), students answered six questions tapping general *attitudes toward the police* as well as two additional questions measuring students' attitudes about police officers as teachers.

Additional Program Objectives

In addition to these three program goals, the 13 G.R.E.A.T. lessons address risk factors for gang joining and life skills thought necessary to prevent involvement in gangs and delinquency (see, e.g., Hill et al., 1999; Klein and Maxson, 2006; Maxson and Whitlock, 2002; Maxson et al., 1998; Thornberry et al., 2003). These mediating variables are treated as implied program objectives and are included in our outcome analyses. We therefore examined the extent to which students exposed to G.R.E.A.T. had improved or enhanced skills that would enable them to resist the lures of gang membership and resist peer pressure to engage in illegal activities. The G.R.E.A.T. lessons encourage students to make healthy choices such as being involved in more prosocial activities and associating more with prosocial peers and less with delinquent peers. The lessons also teach students to improve their communication skills by being active listeners and being better able to interpret verbal and nonverbal communication, targeting these skills to improve students' empathy for others.

In all, 33 outcomes are assessed in these analyses, comprising five behavioral outcomes (gang affiliation, general delinquency, and violent offending—the latter two measured as both frequency and variety indices) and 28 attitudinal measures, including the two measures of attitudes to the police; guilt associated with norm violation; attitudes about gangs; refusal skills; collective efficacy; neutralizations for lying, stealing, and hitting; resistance to peer pressure; associations with delinquent and prosocial peers; prosocial involvement; commitment to negative and to positive peers; school commitment; guilt; empathy; self-centeredness; anger; impulsivity; risk seeking; conflict resolution; calming others; active listening; problem solving; self-efficacy; awareness of services; and altruism. (For a full listing of scales and scale characteristics, see the Appendix.)

Analysis Strategy

The *posttest-through-4-year posttreatment* analysis strategy is an elaboration of that used by Esbensen et al. (2012) for the first two posttreatment waves of outcome measures. These analyses, using MLwiN software (Rasbash, Steele, Browne, and Goldstein, 2009), include the outcome measures obtained on five occasions after treatment (Waves 2 through 6, Level 1) for a total of 15,693 observations nested within 3,739 individual students (Level 2) in 195 different classrooms (Level 3), in 31 schools (Level 4), in 7 cities (Level 5). The analyses allowed for residual mean differences for students, classrooms, and schools through random intercept terms at each level and for cities through dummy variable fixed effects (because of the small number of cities). By mean centering the treatment versus control explanatory variable within schools, we ensured that differences across schools in mean levels of outcomes did not bias the estimate of program effects inadvertently. The model also included a variance component to allow for the possibility that program impact varied across schools (i.e., a random coefficient for treatment versus control at the school level), which ensured an appropriately conservative significance test of program impact. The

variation in program impact across schools did not reach statistical significance at $p < .05$ for any outcomes. The analyses also controlled for the pretest measure of each outcome. We assessed the pretest comparability of treatment and comparison groups through a version of this model that omits time as a level of analysis.

The model allows for change over time through a quadratic function. We were careful to code this function so that the main effect for treatment would reflect mean differences across the entire posttreatment period. We accomplished this by capturing the function through orthogonal polynomials (coded across Waves 2–6 as linear = $-2, -1, 0, 1, 2$; quadratic = $2, -1, -2, -1, 2$). We then centered these terms within each person to adjust for any individual differences associated with attrition. Analyses included random variance components for the linear and squared terms at the individual, classroom, and school levels, thus allowing for the possibility of systematic differences in trajectories at each of those levels.

Our analytic model is designed so that the coefficient for treatment versus control provides an overall assessment of program impact, and the interactions between that term and the linear and squared terms for time reflect change over time in program impact (with significance assessed by a joint test of those two interaction terms). We applied a linear version of this model to most of the outcomes. The measure of gang membership is dichotomous and thus required a logistic version of the model. The self-report measures of general and violent delinquency were highly skewed integer variables, for which a negative binomial model was most appropriate. For the linear models, our tables show the magnitude of program effects in standard deviation units of difference between treatment and control groups (also known as Cohen's d), transformed so that positive values reflect beneficial program effects. For the logistic and negative binomial models, we report the percentage difference between treatment and control in odds (for logistic) or mean rate of offenses (negative binomial).

One objective of this multisite evaluation was to include students from diverse settings to allow us to address the issue of transferability of the program. The seven participating cities were selected to represent large and small cities, racially homogenous and racially heterogeneous populations, and cities across the geographical range of the United States. To examine the generalizability and transferability of the program, we implemented a version of the model that provides separate estimates of program effects and time trends for each city. We accomplished this by replacing all the fixed regression coefficients in the base model (except the pretest outcome measure) by their interactions with dummy variables for every site (leaving no reference site). The variance components remained the same.

A body of literature exists that has suggested that youth with greater preexisting risk might benefit more from some programs than youth at low risk (Andrews et al., 1990; Lipsey, 2009). Indeed, the cross-sectional results reported by Esbensen and Osgood (1999) found some evidence that the G.R.E.A.T. program was more effective for males (relative to females) and African American and Latino youth (relative to White youth)—groups

commonly found to be at higher risk of gang membership. To examine this issue, additional analyses were run to test whether the program impact differed between high- and low-risk youth.¹⁰ To measure risk, we first identified respondents who reported belonging to a gang in Waves 2 through 6. We then conducted a logistic regression analysis with that measure as the outcome and sex, race/ethnicity, and 35 Wave 1 measures (the 33 variables identified previously and 2 measures of school and community disorder) of all of the outcome variables as predictors. The fitted values from that analysis differentiate respondents for their probability of joining a gang by the end of the study. These fitted values were most strongly correlated with Wave 1 gang membership ($r = 0.80$), delinquency ($r = 0.74$), and peer delinquency ($r = 0.57$). We defined high-risk youth as the 25% of the sample with the greatest probability of joining a gang and low risk as the remaining 75% of the sample. We tested for differential program effects on high- versus low-risk youth by adding to the base model the two-way interaction of risk with classroom treatment assignment and the three-way interactions of risk and treatment assignment with linear and quadratic change. Finally, we also assessed the extent to which program effects differed by the subgroups (sex and race/ethnicity) compared in the previous evaluations, conducting sex (or race/ethnicity)-by-treatment interactions and examining group-by-treatment interaction over time. These analyses indicated that only for a few (1 or 2 out of 33) outcomes did program effects differ significantly by sex or race/ethnicity, certainly no more than by chance.

Results

Preliminary analyses examined the comparability of treatment and comparison groups on pretest measures. Across the entire set of 33 outcome measures, the differences tended to be small but slightly favored the treatment group, with the mean Cohen's $d = 0.017$ for the 28 measures to which it applies. The differences reached $p < .05$ for three measures and $p < .10$ for a total of seven, which is somewhat more than expected by chance, but not to a statistically significant degree. For instance, the binomial distribution indicates that $p = .23$ for obtaining three or more "significant" results by chance in 33 tests. Furthermore, the least probability any of these differences was $p = .02$, which is far greater than the Bonferroni standard of .0015 for 33 significance tests. To be cautious, we report results from analyses that control for pretest scores. That control had negligible consequence for the magnitude of estimated program effects, but it did increase their precision.

The results across the posttest-through-4-years posttreatment are consistent with those found for the 1-year posttreatment analyses (see Esbensen et al., 2002); the effect sizes, however, are somewhat smaller (see Table 2). In the 1-year posttreatment analyses, program impact was significant at the .05 level for a total of 11 of 33 outcomes, and an additional

10. In addition, this analysis addresses, in part, Klein and Maxson's (2006) critique that this universal program fails to target the most in-need youth and that effects from universal programs such as this might be diluted because of the large number of low-risk youth in the sample.

TABLE 2

One-Year and Entire Four-Year Postprogram Effect Estimates for Attitudinal and Behavioral Measures Controlling for Between-City Differences, Overall Change over Time, and the Pretest Outcome Measure

| | 1 Year Postprogram | | | All 4 Years Postprogram | | |
|--|--------------------|----------|-----------|-------------------------|----------|-----------|
| | Program Effect | <i>b</i> | <i>SE</i> | Program Effect | <i>b</i> | <i>SE</i> |
| Attitudinal Measures | | | | | | |
| Impulsivity | 0.015 | -0.012 | 0.024 | 0.021 | -0.017 | 0.021 |
| Riskseeking | 0.041 | -0.041 | 0.030 | 0.053 | -0.051* | 0.025 |
| Anger | 0.057 | -0.056* | 0.026 | 0.049 | -0.049* | 0.023 |
| Self-centeredness | 0.054 | -0.046* | 0.022 | 0.038 | -0.031 | 0.025 |
| Attitudes toward the police (ATP) | 0.076 | 0.070* | 0.024 | 0.058 | 0.055* | 0.023 |
| G.R.E.A.T. ATP | 0.204 | 0.190* | 0.033 | 0.144 | 0.129* | 0.029 |
| Prosocial peers | 0.051 | 0.050† | 0.030 | 0.040 | 0.038 | 0.024 |
| Peer pressure | 0.079 | -0.050* | 0.020 | 0.044 | -0.031 | 0.019 |
| Negative peer commitment | 0.050 | -0.047 | 0.029 | -0.002 | 0.002 | 0.030 |
| Positive peer commitment | -0.010 | -0.011 | 0.037 | 0.007 | 0.008 | 0.032 |
| Delinquent peers | 0.083 | -0.051* | 0.021 | 0.025 | -0.017 | 0.018 |
| Lying neutralizations | 0.066 | -0.066† | 0.034 | 0.042 | -0.041 | 0.027 |
| Stealing neutralizations | 0.018 | -0.016 | 0.030 | 0.017 | -0.015 | 0.029 |
| Hitting neutralizations | 0.105 | -0.122* | 0.032 | 0.079 | -0.095* | 0.030 |
| School commitment | 0.020 | 0.015 | 0.021 | 0.031 | 0.023 | 0.017 |
| Guilt | 0.028 | 0.016 | 0.016 | 0.007 | 0.004 | 0.018 |
| Conflict resolution | -0.018 | -0.008 | 0.013 | -0.009 | -0.004 | 0.011 |
| Calming others | -0.004 | -0.002 | 0.014 | 0.010 | 0.005 | 0.012 |
| Refusal skills | 0.090 | 0.043* | 0.013 | 0.049 | 0.022* | 0.010 |
| Prosocial involvement index | 0.047 | 0.056† | 0.030 | 0.020 | 0.039 | 0.032 |
| Empathy | -0.008 | -0.005 | 0.022 | 0.012 | 0.008 | 0.018 |
| Active listening | 0.028 | 0.019 | 0.020 | 0.044 | 0.028 | 0.017 |
| Problem solving | 0.027 | 0.025 | 0.024 | -0.019 | -0.017 | 0.022 |
| Self-efficacy | -0.004 | -0.003 | 0.024 | 0.007 | 0.004 | 0.021 |
| Awareness of services | 0.015 | 0.012 | 0.021 | 0.016 | 0.012 | 0.018 |
| Collective efficacy | 0.125 | 0.075* | 0.021 | 0.096 | 0.055* | 0.015 |
| Attitudes about gangs | 0.114 | 0.102* | 0.031 | 0.094 | 0.079* | 0.024 |
| Altruism | 0.051 | 0.031 | 0.019 | 0.058 | 0.033* | 0.017 |
| Behavioral^a | | | | | | |
| Delinquency (frequency) ^b | 7.0% | -0.073 | 0.072 | 5.0% | -0.053 | 0.059 |
| Delinquency (variety) ^b | 7.0% | -0.072 | 0.048 | 5.0% | -0.052 | 0.039 |
| Violent offending (frequency) ^b | 10.0% | -0.107 | 0.179 | 11.0% | -0.106 | 0.122 |
| Violent offending (variety) ^b | -1.0% | 0.007 | 0.108 | 7.0% | -0.070 | 0.083 |
| Gang ^c | 39.2% | -0.498* | 0.162 | 24.0% | -0.271* | 0.135 |

Note. *SE* = standard error.

^aProgram effect as percent reduction.

^bNegative binomial model.

^cLogistic regression model.

† $p < .10$. * $p < .05$.

three were marginally significant at the .10 level (prosocial peers, prosocial involvement, and lying neutralizations). Combining the data for the entire 4 years (waves 2–6) post-treatment, we find 10 significant differences, including 8 of the same outcomes that were significant at 1 year posttreatment. The following list identifies the differences for post-test-through-4-years posttreatment; those identified with an asterisk also were noted in the 1-year posttreatment analyses. Three outcomes were significant at 1 year posttreatment but not for posttest-through-4-years posttreatment (self-centeredness, peer pressure, and delinquent associations).

- Lower rates of gang membership (24% reduction in odds)*
- More positive attitudes to police (ES = 0.058)*
- More positive attitudes about police in classrooms (ES = 0.144)*
- Less positive attitudes about gangs (ES = 0.094)*
- More use of refusal skills (ES = 0.049)*
- Higher collective efficacy (ES = 0.096)*
- Less use of hitting neutralizations (ES = 0.079)*
- Less anger (ES = 0.049)*
- Higher levels of altruism (ES = 0.058)
- Less risk seeking (ES = 0.053)

With respect to the three specific program goals, the odds of belonging to a gang during the posttest-through-4-years postprogram were 24% lower for the G.R.E.A.T. students, and they continued to have more positive attitudes toward the police in general and to officers in the classroom, compared with non-G.R.E.A.T. students. Estimates of program impact did not reach statistical significance, however, for delinquency (general or violent offending). Importantly, the treatment group continued to express less favorable attitudes about gangs, and several risk factors associated with gang membership also were found to be less pronounced among the G.R.E.A.T. students. Students who had participated in the program were more risk averse, expressed better anger control, and employed fewer neutralizations regarding the use of violence in response to different scenarios. Additionally, as described, several measures were developed and included in the analyses to assess skills taught in the G.R.E.A.T. lessons. For example, the curriculum teaches (through students' role-playing) strategies for students to use to avoid undesired activities in which their friends encourage them to participate. Students in the treatment group were more apt to report use of these refusal techniques. The G.R.E.A.T. students also reported higher levels of altruism and collective efficacy; that is, they indicated that they value doing things for others (e.g., "It feels good to do something without expecting anything in return") and that they can make a difference in their communities (e.g., "It is my responsibility to do something about problems in our community"). These values are reflected in a component of the G.R.E.A.T. program called the "Making My School a G.R.E.A.T. Place" project. This G.R.E.A.T. project provides students the opportunity to have an impact on their

environment by improving their school, surrounding area, or both. The project is intended to be an ongoing part of the program and to be completed by the end of the 13th lesson.

In contrast to these positive program effects, our long-term (posttest-through-4-years postprogram) analyses failed to discern a difference between the G.R.E.A.T. students and the control group on a range of peer-related factors: prosocial peers, peer pressure, negative peer commitment, positive peer commitment, and delinquent peers. Three of these potential outcomes were marginally significant ($p < .10$) in the 1-year post treatment analyses (prosocial peers, peer pressure, and delinquent peers), suggesting that the peer effect is muted over time. Also, the program did not produce statistically significant differences for several social skills or risk factors emphasized in one or more lessons: conflict resolution, calming others, active listening, problem solving, empathy, self-efficacy, awareness of services, prosocial involvement, neutralizations for lying and stealing, guilt, school commitment, self-centeredness, and impulsivity. The latter two outcomes are subcomponents of the larger self-control measure developed by Grasmick, Tittle, Bursik, and Arneklev (1993). The program impact for two other components of self-control (risk seeking and anger) did reach significance. These aspects of the program that did not differentiate the groups suggest that perhaps attitudes are more easily influenced than is behavior. A large proportion of these remaining nonsignificant factors consists of social skills variables representing program components that teach students factual information or how to modify their behavior (e.g., availability of services, active listening, calming others, and problem solving). That is, students are instructed on where to find assistance when needed and on the importance of listening to others when they speak, how to calm others who are upset, and constructive (and nonviolent) ways to solve problems that develop.

Site-Specific Analyses Posttest-Through-4-Years Posttreatment

One evaluation objective was to address the transportability of the program. That is, can G.R.E.A.T. be effectively taught in a variety of settings? To address this issue, we included seven diverse cities in the study, and in this set of analyses, we explore the extent to which the aggregate-level differences are replicated in the seven different cities. As shown in Table 3, the findings are mixed. At 1 year posttreatment (the first columns for each site), the overall findings are largely replicated in three sites (Albuquerque, the DFW area site, and Portland). A few program effects (including lower odds of gang membership) were noted in Philadelphia, but null findings were found in Greeley, Nashville, and Chicago (see Table 3).

It is important to consider whether these differences across sites in program impact reflect genuine differences in effectiveness or result from a combination of smaller sample sizes and chance variation inevitable among estimates of limited precision. Interaction tests give clear evidence that differences in impact across sites are statistically reliable for only G.R.E.A.T. attitudes toward police and negative peer commitment. For both, $p = .0011$, which surpasses the Bonferroni corrected value of $p < .0015$ (for $p < .05$, 33 tests). For the entire set of 33 outcomes, a total of 4 tests reached the nominal level of $p < .05$ and 6

T A B L E 3

One-Year and Entire Four-Year Site-Specific Program Effect Estimates for Attitudinal and Behavioral Measures Controlling for Between-City Differences, Overall Change over Time, and the Pretest Outcome Measure—Significant Effect Sizes Only

| | Albuquerque | | Chicago | | DFW Area | | Greeley | | Nashville | | Philadelphia | | Portland | |
|--------------------------|-------------|-------------------|---------|----------|-------------------|-------------------|---------|----------|-----------|----------|--------------|----------|-------------------|-------------------|
| | 1 Year | 1-4 Year | 1 Year | 1-4 Year | 1 Year | 1-4 Year | 1 Year | 1-4 Year | 1 Year | 1-4 Year | 1 Year | 1-4 Year | 1 Year | 1-4 Year |
| Impulsivity | | | | | | | | | | | | | | |
| Risk seeking | | | | | 0.13 [†] | | | | | | | | 0.19 | 0.17 |
| Anger | | | | | 0.11 [†] | | | | | | | | | |
| Self-centeredness | | | | | 0.20 | 0.10 [†] | | | | | | | | 0.18 [†] |
| ATP | 0.15 | | | | 0.17 | 0.14 [†] | | | | | | | 0.17 | 0.14 |
| G.R.E.A.T. ATP | 0.39 | 0.29 | | | 0.23 | 0.16 | | | 0.17 | | | | 0.43 | 0.36 |
| Prosocial peers | | 0.13 | | | | 0.16 | | | | | | | | 0.13 |
| Peer pressure | 0.18 | 0.12 [†] | | | 0.16 | | | | | -0.14 | | | | |
| Negative peer commit | 0.19 | 0.12 [†] | | | | | | | | -0.25 | | | 0.14 [†] | |
| Positive peer commit | | | | | | 0.19 | | | | | | | | |
| Peer delinquency | 0.32 | 0.18 | | | 0.12 [†] | | | | | | | | | |
| Lying neutralizations | | | | | | | | | | -0.13 | | 0.18 | | |
| Stealing neutralizations | | | | | | | | | | -0.17 | | | | |
| Hitting neutralizations | 0.16 | 0.13 | | | 0.13 [†] | | | | | | | | 0.17 | 0.16 |
| School commitment | | 0.10* | | | | | | | | -0.13 | | | | |
| Guilt | | | | | | 0.17 | | | | | | | | |
| Conflict resolution | | | | | | | | | | | | | | |
| Calming others | | | | | | 0.16 | | | | | | | | |
| Refusal skills | 0.15 | 0.11* | | | 0.15 | 0.11 | | | | | | | 0.17 | 0.11 [†] |

Continued

T A B L E 3

Continued

| | Albuquerque | | Chicago | | DFW Area | | Greeley | | Nashville | | Philadelphia | | Portland | |
|-----------------------|-------------|----------|---------|----------|----------|----------|---------|----------|-----------|----------|--------------|----------|----------|----------|
| | 1 Year | 1-4 Year | 1 Year | 1-4 Year | 1 Year | 1-4 Year | 1 Year | 1-4 Year | 1 Year | 1-4 Year | 1 Year | 1-4 Year | 1 Year | 1-4 Year |
| Prosocial involvement | | | | | | | | | | | | | | |
| Empathy | | | | | 0.13† | | | | | | | | | |
| Active listening | | | 0.17 | | 0.22 | | | | | | | | | |
| Problem solving | | | | | | | | | | | | | | |
| Self-efficacy | | | | | | | | | | | | | | |
| Awareness of services | | | | | | | | | | | | | | |
| Collective efficacy | 0.19* | 0.17 | | | 0.17 | | | | | | | | 0.19† | |
| Attitudes about gangs | 0.20 | | 0.24 | | 0.25 | | | | | | | | 0.20 | |
| Altruism | 0.14* | | 0.18 | | 0.19 | | | | | | | | 0.14† | |
| Delinquency (freq) | | | | | 31%† | | | | | | | | | |
| Delinquency (variety) | | | | | 23% | | | | | | | | | |
| Violent (freq) | | | | | | | | | | | | | | |
| Violent (variety) | | | | | | | | | | | | | | |
| Gang | 71% | 58% | | | | | | | | | | | 65% | 48% |
| | | | | | | | | | | | | | 61%† | |

Note: Negative estimates, such as those found in Nashville and Greeley, indicate a negative program effect.

† $p < .10$. * $p < .05$.

reached the nominal level of $p < .10$, which is somewhat more than chance but not notably so. Also recall that we did not find significant school-level variance in program impact for any outcomes. Whether the differences among sites reflect chance fluctuations or genuine differences in effectiveness, the results of Table 3 make clear that any given implementation of the program might or might not achieve results consistent with the overall average.

The results for site-specific program impact *across all 4 years posttreatment* (the second columns for each site in Table 3) are similar to those found at 1 year posttreatment. Once again, the results in Albuquerque, Portland, and the Texas site resemble the aggregate results. Philadelphia experienced a few positive outcomes, whereas Chicago and Greeley once again had null findings. For posttest-through-4-years posttreatment, however, the G.R.E.A.T. students in Nashville reported five negative program effects (more susceptibility to peer pressure, more commitment to negative peers, less school commitment, and greater neutralizations for lying and stealing). Overall, the site-specific results are robust with the posttest-through-4-years posttreatment results similar overall to those found for 1 year posttreatment with the caveat that the 1-year posttreatment effect sizes, as is the case with the full-sample results, are somewhat larger.

Preexisting Risk Analyses Posttest-through-4-Years Posttreatment

To test for the possibility that the G.R.E.A.T. program might be more suitable for high-risk youth, we used Wave 1 data to identify students at risk for gang membership. Specifically, we used sex, race/ethnicity, and 35 attitudinal and behavioral measures (the 33 outcome measures plus school and community disorder) from Wave 1 as predictors of being a gang member in any subsequent wave (i.e., Waves 2 through 6). Then, we saved the predicted probabilities as the risk measure. Although there is no set standard for classifying risk, we dichotomized the risk measure and identified the top 25% as at risk (a method used, e.g., by Farrington and Loeber, 2000; Hill et al., 1999). To minimize missing data, we substituted scale means for any missing Wave 1 predictors when computing the risk score. None of the treatment-by-risk interactions is significant, but to test for the possibility that effects might change over time, we examined also risk-by-treatment-by-time interactions. Several significant three-way interactions emerged, and the pattern is consistent. The three-way interactions suggest that most of the beneficial impact is associated with the high-risk students in the early waves and that the treatment/control difference for high-risk youth fades over time. Some evidence shows that the treatment is increasingly beneficial for low-risk youths over time, but that pattern is far from consistent.

Table 4 provides a summary of the analyses of differential impact in relation to risk. The variables are coded so the main effects retain their original meaning.¹¹ Four of the 33

11. The overall impact effects reported are similar to, but not exactly the same as, those reported for the aggregate-level analyses in Table 2 because this model adds risk level as a predictor and all its interactions with treatment condition and time (both linear and squared).

TABLE 4

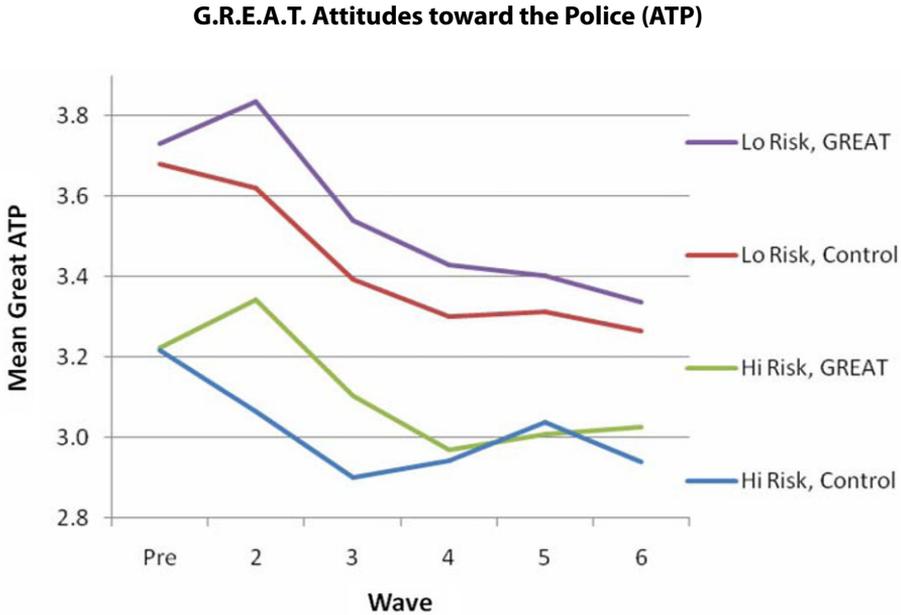
Interaction Effects of Risk by Impact and Risk by Impact by Time

| | Risk × Impact | | | Risk × Impact × Time |
|-------------------------------|---|------------------------|-----------|-----------------------------------|
| | Program Effect Difference (<i>d</i>) | <i>b</i> Difference | <i>SE</i> | Wald Chi Square (2 <i>df</i>) |
| Attitudinal | | | | |
| Impulsivity | 0.036 | -0.029 | 0.044 | 5.95 [†] |
| Risk seeking | 0.042 | -0.041 | 0.051 | 9.36* |
| Anger | 0.042 | -0.042 | 0.052 | 4.35 |
| Self-centeredness (ATP) | 0.062 | -0.051 | 0.043 | 4.48 |
| G.R.E.A.T. ATP | -0.053 | -0.050 | 0.048 | 4.52 |
| Prosocial peers | 0.017 | 0.015 | 0.047 | 4.87 [†] |
| Peer pressure | 0.077 | 0.073 | 0.048 | 0.22 |
| Negative peer commitment | 0.095 | -0.066 [†] | 0.036 | 15.96* |
| Positive peer commitment | -0.032 | 0.032 | 0.054 | 6.04* |
| Delinquent peers | 0.055 | 0.064 | 0.058 | 3.95 |
| Lying neutralizations | 0.098 | -0.067 [†] | 0.036 | 17.64* |
| Lying neutralizations | 0.099 | -0.096 [†] | 0.050 | 15.51* |
| Stealing neutralizations | 0.013 | -0.012 | 0.047 | 9.62* |
| Hitting neutralizations | -0.035 | 0.041 | 0.059 | 4.43 |
| School commitment | 0.015 | 0.011 | 0.039 | 7.36* |
| Guilt | 0.066 | 0.041 | 0.032 | 10.27* |
| Conflict resolution | 0.013 | 0.006 | 0.022 | 8.39* |
| Calming others | -0.004 | -0.002 | 0.024 | 4.23 |
| Refusal skills | 0.051 | 0.024 | 0.024 | 3.22 |
| Prosocial involvement index | -0.030 | -0.059 | 0.073 | 0.76 |
| Empathy | -0.053 | -0.036 | 0.037 | 4.36 |
| Active listening | 0.003 | 0.002 | 0.033 | 1.33 |
| Problem solving | 0.004 | 0.004 | 0.048 | 5.03 [†] |
| Self-efficacy | 0.019 | 0.013 | 0.037 | 0.61 |
| Awareness of services | -0.056 | -0.043 | 0.042 | 2.08 |
| Collective efficacy | 0.056 | 0.032 | 0.032 | 0.13 |
| Attitudes about gangs | 0.088 | 0.074 [†] | 0.043 | 2.66 |
| Altruism | -0.022 | -0.013 | 0.031 | 4.06 |
| Behavioral | | | | |
| Delinquency (frequency) | 11.1% | -0.105 | 0.127 | 9.58* |
| Delinquency (variety) | 10.5% | -0.100 | 0.079 | 6.21* |
| Violent offending (frequency) | 40.8% | -0.342 | 0.302 | 2.54 |
| Violent offending (variety) | 26.7% | -0.236 | 0.181 | 5.46 [†] |
| Gang | -16.8% | 0.184 | 0.255 | 7.84* |

Notes. *df* = degrees of freedom; *SE* = standard error.

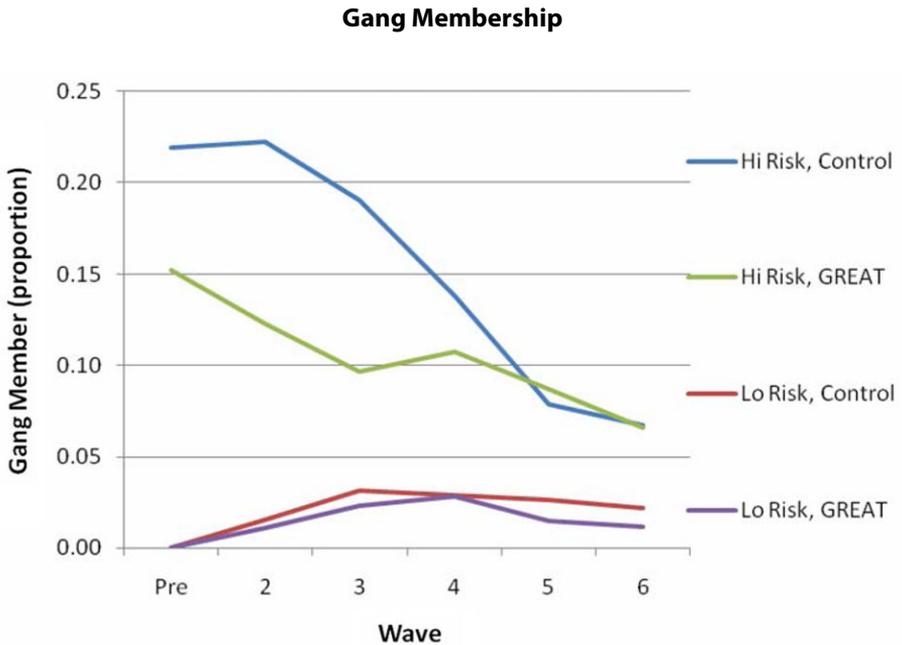
[†]*p* < .10. **p* < .05.

FIGURE 1



risk-by-treatment interactions reached the .10 level of significance, but none reached the .05 level, a pattern that could easily arise by chance. Twelve of the three-way interactions (risk by treatment by time) were significant at the .05 level and four more reached the .10 level. Furthermore, the significance levels for three outcomes surpassed the Bonferroni correction criterion of $p < .0015$ and a total of seven reached $p < .01$, giving strong evidence of genuine rather than chance effects for the dataset as a whole. Figures 1 and 2 provide examples of the three-way interactions for the four combinations of high versus low risk and G.R.E.A.T. versus control. Figure 1 shows that for G.R.E.A.T. attitudes toward police, the treatment and control groups are comparable at the pretest for both high-risk and low-risk youth. In Waves 2 and 3, the treatment group shifts toward more favorable attitudes than the control group, and the resulting difference is more pronounced among high-risk youth. Across Waves 4 through 6, the treatment versus control difference largely disappears for the high-risk youth, whereas a moderate difference remains for the low-risk youth. For gang membership, Figure 2 shows that among the high-risk youth, a somewhat larger proportion of control rather than treatment youth were gang members, and that the G.R.E.A.T. program led to greater reductions in membership for the treatment group than controls through Waves 2 and 3. By Wave 6, however, this treatment effect was no longer apparent. The rates of gang membership were much lower in the low-risk group, of course, but we observe suggestions of a beneficial program effect gradually emerging so that at Wave 6, the rate of gang membership was only half as high in the treatment group as the control group.

FIGURE 2



Discussion

Schools are a desirable location to offer universal programs with an emphasis on preventing an array of adolescent problem behaviors including bullying, drug use, dating violence, gang affiliation, and others (Jimerson, Nickerson, Mayer, and Furlong, 2012). Although school-based violence-prevention/intervention programs are widespread, knowledge of their effectiveness is often lacking (Alford and Derzon, 2012; Gottfredson, 2001). Given teacher and administrator concern about the “loss of instructional time” to nonacademic activities, school administrators increasingly rely on “evidence-based practices” when making decisions about which, if any, programs to allow into their schools. The G.R.E.A.T. program is one primary prevention program that, based on our evaluation, holds promise.

In addition to increased placement of prevention programming in schools, the past 20 years have observed an increase in the presence of police officers on school campuses, as both School Resource Officers (e.g., Finn and McDevitt, 2005; Na and Gottfredson, 2013; Petteruti, 2011) and prevention program providers (e.g., DARE and G.R.E.A.T.). The research reported in this article addresses the efficacy of a program that uses law enforcement officers to deliver a gang-prevention and violence-reduction program.

The G.R.E.A.T. program is one choice that school administrators have when selecting from a vast list of prevention programs. G.R.E.A.T. is currently rated as “promising” by OJJDP and by Crime Solutions, and it is designated as “Level 2” (effective) in the Helping

America's Youth rating scale (findyouthinfo.gov). These designations, although initially based on findings from two previously published evaluations of the original G.R.E.A.T. program, have incorporated and are now based on the short-term results reported from the current evaluation of the revised G.R.E.A.T. program (Esbensen et al., 2012). To recap those previous studies, a cross-sectional study conducted in 1995 found that G.R.E.A.T. students were substantially "better" than non-G.R.E.A.T. students on a variety of attitudinal and behavioral outcomes (Esbensen and Osgood, 1999). A more rigorous longitudinal evaluation conducted between 1995 and 1999 found less support for the program (in terms of the number of significant differences, effect sizes, and the presence of delayed—rather than immediate—effects) (Esbensen, Osgood, et al., 2001). Still, because most of the results were in the direction of positive programmatic effects, G.R.E.A.T. was deemed by raters as a program holding "promise." This was particularly true given the relatively short program dosage (i.e., nine 1-hour lessons delivered over a span of 9 weeks).

Previous critiques of the original program and earlier evaluations raised several concerns. First, some commentators labeled G.R.E.A.T. a "failed program" based on a lack of significant effects on delinquency or gang membership (Ludwig, 2005; Klein and Maxson, 2006). Additionally, when positive programmatic effects were found between G.R.E.A.T. and non-G.R.E.A.T. students, effect sizes were modest (Klein and Maxson, 2006). Third, two well-known gang researchers suggested that the lack of significant program effects were not surprising, given the program's emphasis on factors related to general delinquency (as opposed to gang-specific issues), its modeling after the failed DARE program, and the fact that it targeted a population at low risk of gang involvement (Klein and Maxson, 2006). Finally, the previous longitudinal evaluation was criticized for the extent of sample attrition occurring during the examination period (Ludwig, 2005).

The G.R.E.A.T. program underwent a substantial overhaul after a curriculum review (see Esbensen et al., 2002; Esbensen, Peterson, et al., 2011, for a detailed account of the program review). Many changes were sparked by findings from these early evaluations. The program was expanded from 9 to 13 lessons, and substantial effort was made to link specific program lessons to evidence-based risk factors for gang joining and delinquency found in prior research. Practitioners and researchers versed in gangs and school-based prevention were brought together to offer suggestions for program modifications. Then, professional curriculum writers were employed to develop the specific program lessons. This effort led to the "revised" G.R.E.A.T. program that is the focus of the current study.

After the revised program was fully implemented in 2003, interest in assessing the effectiveness of the G.R.E.A.T. program was renewed. Our most recent work (Esbensen et al., 2012), based on a longitudinal evaluation design that included full random assignment and improved active consent and retention rates, reported that a relatively low dosage (13 lessons) primary prevention program can have measurable effects on a diverse sample of students 1 year posttreatment. This article extends that research by reporting the results of treatment effects up to 4 years posttreatment. We also address two additional questions:

- (1) Were the aggregate results replicated in each of the seven study sites?
- (2) Did the results vary based on youths' preexisting levels of risk?

G.R.E.A.T. Goals and Objectives

The posttest-to-4-year postprogram analyses examined the direct effects of G.R.E.A.T. on the three main program goals (preventing gang involvement, reducing delinquency and violence, and improving views of law enforcement) as well as on several risk factors associated with gang affiliation that were targeted in the curriculum. The results identify positive program effects on a number (10 of 33) of these program objectives. Compared with students in the control classrooms, students in G.R.E.A.T. classrooms expressed more positive attitudes to the police and lower odds of gang membership. They reported also more use of refusal skills, lower support for neutralizations regarding violence, less favorable attitudes about gangs, lower levels of risk seeking and anger, higher levels of altruism, and a higher degree of collective efficacy. It is important to highlight that eight of the ten differences found across 4 years posttreatment also were evident among the 11 differences 1-year post program delivery, indicating a sustained, long-term program effect on those outcomes.

The effect sizes are small (and program rating schemes weight this important aspect of program impact), which remains a criticism lodged by reviewers and by rating schemes. The Blueprints program, for example, declined to classify the revised program as “promising” largely because of the small effect sizes (S. Mihalic, personal communication, April 8, 2013). In our view, that assessment fails to take into account the limited scope and cost of the G.R.E.A.T. program. It is important to note that we are independent evaluators not program developers, and we have no stake in this program's success, financially or otherwise. From our first introduction to the G.R.E.A.T. program in the early 1990s, our shared sentiment was and remains skeptical that there would be a measurable effect of a 9- or 13-lesson program with the average lesson being less than 40 minutes, which is diluted even more by absenteeism and scheduling issues. Finding such beneficial program effects across multiple studies has surprised us, and their consistency forces us to take them seriously. We ask not only ourselves but also the critics what effect size is reasonable to expect given the low dosage and the general audience targeted by this program, and how large must the effects be to justify the use of a program requiring such limited investment?

The revised program and the most recent evaluation design overcome many of the limitations critics noted for the original program and evaluations of it. The program itself is now more “evidence-based,” focused on key risk factors found to be important for gang joining. Additionally, more pedagogically sound strategies (such as active learning as opposed to didactic lecture) comprise a bulk of the program lessons. These two factors provide reason for optimism that the revised program should be more effective at preventing gang membership than its original configuration. These program revisions might be responsible for the divergence in findings related to gang membership between the current evaluation

and its previous counterparts. Specifically, the increased focus of the revised curriculum on theoretically and empirically based risk factors for gang joining, coupled with a more effective “skills-based” programmatic structure, might be the primary reasons why the current study finds G.R.E.A.T. participants report reduced odds of gang joining, relative to their non-G.R.E.A.T. counterparts. Conversely, programmatic effects might be more easily uncovered based on the higher rates of study participation relative to the previous longitudinal study.

All this being said and despite many significant effects in favor of the G.R.E.A.T. program, our current results also include several effects that failed to reach statistical significance. Next, we focus specifically on some of the findings among the social skills and peer-related measures. We focus on these two areas because of consistent (non) effects. Before discussing them, however, we remind readers that chance might be the source of the weaker results for these outcomes. The lack of significance is definitely not proof of “no effect,” and differences in program impact between these outcomes and the others are rarely if ever statistically significant (judging from their standard errors and implied confidence intervals).

Social skills. Our overall lack of findings with regard to several social skills might engender disappointment. In discussing the lack of change in several skills among G.R.E.A.T. students, we speculated previously that effecting attitude change might be easier than stimulating behavioral change. That is, a greater proportion of attitudinal than skills-based behavioral changes existed among the significant differences found between G.R.E.A.T. and control students, and a greater proportion of skills-based factors among the nonsignificant differences. One skill for which we did find a significant difference, however, was G.R.E.A.T. students’ greater use of refusal skills. In our classroom observations of lesson delivery, we noted that this component, more than other social skills components, used role plays between students and the officer, with the officer attempting to lure the student into deviant behavior and the student practicing a host of methods to resist involvement. Students relished this exercise, actively paying attention and participating. We suggest it is possible that students’ greater interest in and ability to practice this skill might have produced the positive results and that offering students more opportunity to rehearse the other social skills might yield the intended programmatic effect.

Peer effects. Two of the three program effects (resistance to peer pressure and association with delinquent peers) that were found 1 year posttreatment but not for the full 4 years posttreatment are related to the role of the peer group, and one additional peer outcome that reached marginal significance ($p < .10$) at 1-year posttreatment also failed to reach significance across the entire 4 years posttreatment. Two other peer-related variables (commitment to positive and to negative peers) also failed to reach significance at both time periods. These results raise two issues, as follows:

- (1) Can an individual-targeted program impact peer factors?
- (2) If yes, then can these results be sustained over time?

The answer to the first question is mixed; modest differences were found between treatment and control students on the peer-related outcomes and risk factors at 1 year posttreatment. The answer to the second question seems to be no; for the 4 years posttreatment as a whole, peer-related differences for the full sample were no longer statistically significant. These results, although disappointing, might be expected: Peers play a major role in the lives of adolescents, and a few brief lessons encouraging youth to avoid negative peer influences might not be sufficient to overcome these influences to achieve the intended outcome.¹²

Program Effects by Site

Some questions are raised by the site-specific results regarding the utility of the G.R.E.A.T. program as a general gang-prevention program applicable in a variety of settings. Three diverse cities (Albuquerque, a DFW suburb, and Portland) experienced program results similar to the larger sample. These sites represent cities with a large Hispanic population (Albuquerque), a city that has the largest percentage of White residents in the United States (Portland), and a city that is part of a large megalopolis (the DFW area site). One city has a long history of gangs (Albuquerque), whereas the other two have relatively new gang problems. The cities with null findings are also diverse—one is among the largest cities in the nation (Chicago) with pockets of extreme disadvantage and high rates of violent crime, whereas the other city (Greeley) is the smallest in the sample (less than 100,000 inhabitants) but with a pronounced gang problem that emerged in the past two decades. A few program effects (notably, lower odds of gang involvement and less positive attitudes about gangs) were found in Philadelphia, a city similar to Chicago in many ways, being large and having neighborhoods facing long-standing poverty, violence, and gang activity.

These findings highlight the importance of conducting multisite evaluations not only to assess the transportability of the program or policy but also to allow for the possibility that contextual effects in some sites might not allow for the detection of program effects (Type II error). For example, whereas one of our considerations in selecting the final sites for the evaluation was program saturation (i.e., we excluded sites in which the G.R.E.A.T. program had a long history, thereby introducing the possibility of program contamination in the control group), it was only after agreements had been obtained that we learned that the Nashville Police Department had an extensive involvement in the schools, teaching the G.R.E.A.T. elementary-level component in 3rd or 4th grade, DARE in 5th grade, and then the G.R.E.A.T. middle-school component in 6th grade, as well as a DARE booster session in 9th grade. Thus, the absence of a positive program effect in Nashville might be an artifact of this police saturation in the schools.

Trying to make sense of these site-specific differences led us to consider several potential explanations. First, it is essential to keep in mind that differences of this magnitude are

12. We temper this with the reminder that program effects seem to vary by site, and at least in one site, the program does produce significant and lasting differences on peer-related variables.

little more than would be expected by chance alone. Next, we revisited the results and considered several potential school factors (e.g., school size, school characteristics, and student demographics) but could not isolate factors that shed light on the findings. As part of another project, we revisited all the cities, schools, and neighborhoods in the hope that we could observe neighborhood characteristics that could help explain the disparate results, but again, we gained no satisfactory insights.

We also examined the possibility that the site differences reflect differential program implementation fidelity.¹³ Fortunately our research design allowed us to examine this possibility as we went to great lengths to assess officer implementation fidelity by observing 492 unique G.R.E.A.T. classroom deliveries and assigning a fidelity score (ranging from 1 to 5) to each classroom (for more information on the assessment of implementation fidelity, see Esbensen, Matsuda, Taylor, and Peterson, 2011). Analyses failed to identify significant differential program effects associated with program quality; only 1 of the 33 potential outcomes (attitudes toward officers in the classroom) showed a more favorable outcome for students in classrooms in which officers implemented the program with increased fidelity ($p < .05$). One possibility for the overall null finding is that 27 of the 33 officers implemented the program with good-to-excellent fidelity. Only three officers were deemed to have not implemented the program (one each in Albuquerque, Greeley, and Philadelphia) and three (one each in the DFW site, Nashville, and Chicago) to have marginal implementation. A seventh officer was deemed to have implemented the program in three classrooms, but because of classroom management issues, the officer failed to implement the program in two other classrooms. Given the overall program fidelity, there might have been insufficient statistical power to detect implementation effect.

Program Effects by Preexisting Risk

The findings for preexisting risk are complex but straightforward. Although we did not find any risk-by-treatment interaction effects, we did uncover a pattern of three-way interaction of risk by treatment by time. The three-way interactions suggest that most of the beneficial impact is associated with the high-risk students in the early waves and that the treatment/control difference for high-risk youth fades over time. Some evidence indicates that the treatment is increasingly beneficial for low-risk youths over time, but that pattern is far from consistent. What these findings mean for universal versus targeted gang-prevention programming is therefore somewhat ambiguous, although the suggestion might be that

13. Given the literature regarding the importance of implementation fidelity, we investigated the relationship of program impact to the quality of G.R.E.A.T. program delivery. Each officer was observed an average of 15 times by trained research assistants. To address this question, we added to the base model the two-way interaction of the officer rating with classroom treatment assignment and the three-way interactions of officer rating and treatment assignment with linear and quadratic change. We avoid confounding these interactions with overall treatment effects by grand mean centering the officer rating and assigning all control classrooms the mean officer rating.

high-risk students (as demonstrated in prior research) have greater gains than do low-risk students, especially in the short term, but that low-risk students also receive program benefits.

Conclusions

The research team responsible for the current evaluation conducted the original G.R.E.A.T. studies in the 1990s (an 11-city cross-sectional study and a 6-city longitudinal quasi-experimental study). Our familiarity with the original program and the evaluation designs and subsequent results facilitate our ability to place the current results within the larger context of school-based gang-prevention programs. Although we have familiarity with the program as evaluators and did provide recommendations regarding program content and delivery based on findings from our first evaluation, it is important to emphasize that we have not been involved in program development; our sole role has been as program evaluators. We note that findings of positive program effects are unfortunately rare in independent prevention trials (Eisner, 2009). Our previous studies of the original G.R.E.A.T. curriculum found a 1-year posttreatment program effect in the cross-sectional study (see Esbensen and Osgood, 1999), but no effect was observed at that time period in the longitudinal quasi-experimental design (Esbensen, Osgood, et al., 2001). In that latter study, we did find a sleeper or lagged effect (3 and 4 years posttreatment) for five outcomes: more favorable attitudes to police, lower victimization, more negative attitudes about gangs, more prosocial peers, and less risk-seeking behavior. Contrary to that previous longitudinal study, the current longitudinal experimental study of the revised G.R.E.A.T. curriculum did find a positive program effect 1 year posttreatment. Importantly, three of the lagged program effects found across 4 years posttreatment in the previous study were replicated in this study for effects across the 4 years (more favorable attitudes to the police, more negative attitudes about gangs, and less risk seeking). Whereas the original program had no appreciable short- or long-term effect on gang involvement, an evaluation of the revised program found reduced odds of gang membership (39% for the first 12 months and 24% across the entire 48 months postprogram). Given the results of the current evaluation, it is important to restate that the G.R.E.A.T. program underwent a major review and revision subsequent to the previous evaluation results. The original G.R.E.A.T. program was a “canned” nine-lesson program with an emphasis on didactic teaching methods. The current 13-lesson G.R.E.A.T. curriculum emphasizes skills building and the use of cooperative learning strategies—both strategies borrowed from other school-based “model” or acclaimed programs.

The fact that both evaluations (of the original and revised program) found decidedly more favorable attitudes toward the police among the G.R.E.A.T. students suggests that this kind of law-enforcement-based prevention program can have a positive impact on youth-police relations. This is particularly important given recent findings that perceptions of police legitimacy often are muted among gang members (particularly those embedded in criminal networks), a factor associated with their increased involvement in crime (Papachristos,

Meares, and Fagan, 2012). Also, it is important to note that studies of both the original and the revised curriculum produced evidence that the G.R.E.A.T. program is associated with more negative views of gangs. We view these similarities in findings as suggestive of an overall consistency in the program and further speculate that the additional program effects of the revised G.R.E.A.T. program are likely a result of the revised and enhanced curriculum.

The current study is not without limitations. Study participants were enrolled in public schools in seven U.S. cities. Students who attended private schools, other districts, those whose parents declined participation, and those who were absent during survey administration periods were not included. We attempted to survey as many eligible students as possible, making more than 10 trips to schools to try to reach those who were habitually truant or otherwise unavailable. We also attempted to survey students who transferred schools within the original and adjacent districts; those who moved to districts outside of the original metro areas were typically lost. Consequently, we might have lost a disproportionate share of gang members and other “at-risk” youth. Additionally, we have no alternative measures of delinquency or gang membership other than the students’ self-reports. Future studies might find it useful to collect measures of school disciplinary reports, police reports, and other indicators.

The G.R.E.A.T. program is no panacea for the gang problems confronting many schools and neighborhoods. However, our findings suggest that G.R.E.A.T. holds promise as a primary gang-prevention program, overall and in several of our seven individual research sites. Although it is important to note that the effect sizes are small (ranging from 0.05 to 0.14 over 48 months posttreatment), it is equally important to emphasize that this is a low-dosage program. The curriculum consists of 13 lessons, generally delivered once a week in less than 40 minutes. Furthermore, the realities of program delivery such as student absenteeism, teacher announcements, fire drills, snow days, officer illness, and shortened day schedules mean that most of the G.R.E.A.T. students do not receive the full recommended dosage. That statistically significant differences were found for 11 outcome measures (and another 3 with marginal significance) 12 months posttreatment and for 10 measures across 4 years posttreatment we find surprising and certainly promising.

References

- Alford, Aaron A. and James Derzon. 2012. Meta-analysis and systematic review of the effectiveness of school-based programs to reduce multiple violent and antisocial behavioral outcomes. In (Shane R. Jimerson, Amanda B. Nickerson, Matthew J. Mayer, and Michael J. Furlong, eds.), *Handbook of School Violence and School Safety: International Research and Practice*. New York: Routledge.
- Andrews, D.A., Ivan Zinger, Robert D. Hoge, James Bonta, Paul Gendreau, and Francis T. Cullen. 1990. Does correctional treatment work? A clinically relevant and psychologically informed meta-analysis. *Criminology*, 38: 369–404.

- Battin, Sara R., Karl G. Hill, Robert D. Abbott, Richard F. Catalano, and J. David Hawkins. 1998. The contribution of gang membership to delinquency beyond delinquent friends. *Criminology*, 36: 93–115.
- Catalano, Richard F., Michael W. Arthur, J. David Hawkins, Lisa Berglund, and Jeffrey J. Olson. 1998. Comprehensive community- and school-based interventions to prevent antisocial behavior. In (Rolf Loeber and David P. Farrington, eds.), *Serious and Violent Juvenile Offenders: Risk Factors and Successful Interventions*. Thousand Oaks, CA: Sage.
- Covey, Herbert C. 2010. *Street Gangs Throughout the World*, 2nd Edition. Springfield, IL: Charles C. Thomas.
- Dusenbury, Linda and Gilbert J. Botvin. 1992. Competence enhancement and the development of positive life options. *Journal of Addictive Diseases*, 11: 29–45.
- Eisner, Manuel. 2009. No effects in independent prevention trials: Can we reject the cynical view? *Journal of Experimental Criminology*, 5: 163–183.
- Esbensen, Finn-Aage and Elizabeth P. Deschenes. 1998. A multisite examination of youth gang membership: Does gender matter? *Criminology*, 36: 799–828.
- Esbensen, Finn-Aage, Adrienne Freng, Terrance J. Taylor, Dana Peterson, and D. Wayne Osgood. 2002. Putting research into practice: The national evaluation of the Gang Resistance Education and Training (G.R.E.A.T.) program. In (Winifred L. Reed and Scott H. Decker), *Responding to Gangs: Evaluation and Research*. Washington, DC: U.S. Department of Justice, National Institute of Justice.
- Esbensen, Finn-Aage and David Huizinga. 1993. Gangs, drugs, and delinquency in a survey of urban youth. *Criminology*, 31: 565–589.
- Esbensen, Finn-Aage, David Huizinga, and Anne W. Weiher. 1993. Gang and non-gang youth: Differences in explanatory factors. *Journal of Contemporary Criminal Justice*, 9: 94–116.
- Esbensen, Finn-Aage, Kristy N. Matsuda, Terrance J. Taylor, and Dana Peterson. 2011. Multi-method strategy for assessing program fidelity: The national evaluation of the revised G.R.E.A.T. program. *Evaluation Review*, 35: 14–39.
- Esbensen, Finn-Aage and Cheryl L. Maxson. 2012. *Youth Gangs in International Perspective: Results from the Eurogang Program of Research*. New York: Springer.
- Esbensen, Finn-Aage, Chris Melde, Terrance J. Taylor, and Dana Peterson. 2008. Active parental consent in school-based research: How much is enough and how do we get it? *Evaluation Review*, 32: 335–362.
- Esbensen, Finn-Aage and D. Wayne Osgood. 1999. Gang resistance education and training (GREAT): Results from the national evaluation. *Journal of Research in Crime and Delinquency*, 36: 194–225.
- Esbensen, Finn-Aage, D. Wayne Osgood, Terrance J. Taylor, Dana Peterson, and Adrienne Freng. 2001. How great is G.R.E.A.T.? Results from a quasi-experimental design. *Criminology & Public Policy*, 1: 87–118.
- Esbensen, Finn-Aage, Dana Peterson, Terrance J. Taylor, and Adrienne Freng. 2010. *Youth Violence: Sex and Race Differences in Offending, Victimization, and Gang Membership*. Philadelphia, PA: Temple University Press.

- Esbensen, Finn-Aage, Dana Peterson, Terrance J. Taylor, Adrienne Freng, D. Wayne Osgood, Dena C. Carson, and Kristy N. Matsuda. 2011. Evaluation and evolution of the Gang Resistance Education and Training (G.R.E.A.T.) Program. *Journal of School Violence*, 10: 53–70.
- Esbensen, Finn-Aage, Dana Peterson, Terrance J. Taylor, and D. Wayne Osgood. 2012. Results from a multi-site evaluation of the G.R.E.A.T. program. *Justice Quarterly*, 29: 125–151.
- Esbensen, Finn-Aage, L. Thomas Winfree, Jr., Ni He, and Terrance J. Taylor. 2001. Young gangs and definitional issues: When is a gang a gang and why does it matter? *Crime & Delinquency*, 47: 105–130.
- Farrington, David P. and Rolf Loeber. 2000. Some benefits of dichotomization in psychiatric and criminological research. *Criminal Behaviour and Mental Health*, 10: 100–122.
- Finn, Peter and Jack McDevitt. 2005. *National Assessment of School Resource Officer Programs: Final Project Report*. Cambridge, MA: Abt Associates. Retrieved from ncjrs.gov/pdffiles1/nij/grants/209273.pdf.
- Gottfredson, Denise C. 2001. *Schools and Delinquency*. New York: Cambridge University Press.
- Grasmick, Harold G., Charles R. Tittle, Robert J. Bursik, and Bruce J. Arneklev. 1993. Testing the core empirical implications of Gottfredson and Hirschi's General Theory of Crime. *Journal of Research in Crime and Delinquency*, 30: 5–29.
- Gravel, Jason, Martin Bouchard, Karine Descormiers, Jennifer S. Wong, and Carlo Morselli. 2013. Keeping promises: A systematic review and new classification of gang control strategies. *Journal of Criminal Justice*, 41: 228–241.
- Hagedorn, John M. 2008. *A World of Gangs: Armed Young Men and Gangsta Culture*. Minneapolis: University of Minnesota Press.
- Hill, Karl G., James C. Howell, J. David Hawkins, and Sara R. Battin-Pearson. 1999. Childhood risk factors for adolescent gang membership: Results from the Seattle Social Development Project. *Journal of Research in Crime and Delinquency*, 36: 300–322.
- Howell, James C. 2009. *Preventing and Reducing Juvenile Delinquency: A Comprehensive Framework*. Thousand Oaks, CA: Sage.
- Howell, James C. 2012. *Gangs in America's Communities*. Thousand Oaks, CA: Sage.
- Howell, James C. and Arlen Egle, Jr. 2005. *Gangs in Small Towns and Rural Areas*. NYGC Bulletin. Edited by The Office of Juvenile Justice and Delinquency Prevention. Washington, DC: U.S. Department of Justice.
- Jimerson, Shane R., Amanda B. Nickerson, Matthew J. Mayer, and Michael J. Furlong (eds.). 2012. *Handbook of School Violence and School Safety: International Research and Practice*, 2nd Edition. New York: Routledge.
- Klein, Malcolm W. and Cheryl L. Maxson. 2006. *Street Gang Patterns and Policies*. New York: Oxford University Press.
- Lipsey, Mark W. 2009. The primary factors that characterize effective interventions with juvenile offenders: A meta-analysis review. *Victims & Offenders*, 4: 124–147.

- Ludwig, Jens. 2005. Better gun enforcement, less crime. *Criminology & Public Policy*, 4: 677–716.
- Maxson, Cheryl L., Arlen Egley, Jr., Jody Miller, and Malcolm W. Klein (eds.). 2013. *The Modern Gang Reader*, 4th Edition. New York: Oxford University Press.
- Maxson, Cheryl L., and Monica L. Whitlock. 2002. Joining the gang: Gender differences in risk factors of gang membership. In (C. Ronald Huff, ed.), *Gangs in America III*. Thousand Oaks, CA: Sage.
- Maxson, Cheryl, Monica L. Whitlock, and Malcolm W. Klein. 1998. Vulnerability to street gang membership: Implications for practice. *Social Service Review*, March: 70–91.
- Mihalic, Sharon F, Abigail A. Fagan, Katherine Irwin, Diane Ballard, and Delbert S. Elliott. 2002. *Blueprints for Violence Prevention Replications: Factors for Implementation Success*. Boulder: University of Colorado, Center for the Study and Prevention of Violence.
- Mihalic, Sharon F and Katherine Irwin. 2003. Blueprints for violence prevention: From research to real-world settings—factors influencing the successful replication of model programs. *Youth Violence and Juvenile Justice*, 1: 307–329.
- Na, Chongmin and Denise C. Gottfredson. 2013. Police officers in schools: Effects of school crime and the processing of offending behaviors. *Justice Quarterly*, 30: 619–650.
- Office of Juvenile Justice and Delinquency Prevention (OJJDP). 2010. *Model Programs Guide*. Retrieved from ojjdp.gov/mpg/.
- Papachristos, Andrew V., Tracy L. Meares, and Jeffrey Fagan. 2012. Why do criminals obey the law? The influence of legitimacy and social networks on active gun offenders. *Journal of Criminal Law & Criminology*, 102: 397–440.
- Peterson, Dana and Kirstin A. Morgan. 2013. Sex differences and the overlap in risk factors for gang membership and violence. *Journal of Crime and Justice*. DOI:10.1080/0735648X.2013.830393
- Petteruti, Amanda. 2011. *Education Under Arrest: The Case Against Police in Schools*. Washington, DC: Justice Policy Institute.
- Pyrooz, David C., Andrew M. Fox, and Scott H. Decker. 2010. Racial and ethnic heterogeneity, economic disadvantage, and gangs: A macro-level study of gang membership in urban America. *Justice Quarterly*, 27: 867–892.
- Rasbash, Jon, Fiona Steele, William J. Browne, and Harvey Goldstein. 2009. *A User's Guide to MLwiN, Version 2.10*. Bristol, U.K.: Bristol University Press.
- Reed, Winifred L., and Scott H. Decker. 2002. *Responding to Gangs: Evaluation and Research*. Washington, DC: U.S. Department of Justice, National Institute of Justice.
- Sherman, Lawrence W., Denise C. Gottfredson, Doris MacKenzie, John Eck, Peter Reuter, and Shawn D. Bushway. 1997. *Preventing Crime: What Works, What Doesn't, What's Promising*. Washington, DC: Department of Justice, Office of Justice Programs.
- Thornberry, Terence P. 1998. Membership in youth gangs and involvement in serious and violent offending. In (Rolf Loeber and David P. Farrington, eds.), *Serious and Violent Juvenile Offenders: Risk Factors and Successful Interventions*. Thousand Oaks, CA: Sage.

- Thornberry, Terence P., Marvin D. Krohn, Alan J. Lizotte, C. A. Smith, and Kimberly Tobin. 2003. *Gangs and Delinquency in Developmental Perspective*. New York: Cambridge University Press.
- Winfree, L. Thomas, Jr., Dana Peterson Lynskey, and James R. Maupin. 1999. Developing local police and federal law enforcement partnerships: G.R.E.A.T. as a case study of policy implementation. *Criminal Justice Review*, 24: 145–168.

Appendix

Scale Characteristics of Outcome Measures (Wave 1)

Impulsivity: Four items such as: I often act without stopping to think.

Scale mean = 2.97 (0.81); $\alpha = 0.59$

Response categories: 1) strongly disagree to 5) strongly agree

Risk-Seeking: Four items including: I like to test myself every now and then by doing something a little risky.

Scale mean = 2.60 (0.95); $\alpha = 0.77$

Response categories: 1) strongly disagree to 5) strongly agree

Anger: Four items including: I lose my temper pretty easily.

Scale mean = 3.08 (0.96); $\alpha = 0.74$

Response categories: 1) strongly disagree to 5) strongly agree

Self-Centeredness: Four items such as: If things I do upset people, it's their problem not mine.

Scale mean = 2.50 (0.82); $\alpha = 0.69$

Response categories: 1) strongly disagree to 5) strongly agree

Attitudes Toward Police: Six items such as: Police officers are honest.

Scale mean = 3.81 (0.82); $\alpha = 0.86$

Response categories: 1) strongly disagree to 5) strongly agree

G.R.E.A.T. ATP: Two items such as: Police officers make good teachers.

Mean = 3.58 (0.95)

Response categories: 1) strongly disagree to 5) strongly agree

Prosocial Peers: Four items including: During the last year, how many of your current friends have been generally honest and told the truth?

Scale mean = 3.42 (0.97); $\alpha = 0.83$

Response categories: 1) none of them, 2) few of them, 3) half of them, 4) most of them, and 5) all of them

Peer Pressure: Seven items such as: How likely is it that you would go along with your current friends if they wanted you to bully another student at school?

Scale mean = 1.27 (0.51); $\alpha = 0.82$

Response categories: 1) not at all likely to 5) very likely

Negative Peer Commitment: Three items including: If your group of friends was getting you into trouble at home, how likely is it that you would still hang out with them?

Scale mean = 1.68 (0.85); $\alpha = 0.81$

Response categories: 1) not at all likely to 5) very likely

Positive Peer Commitment: Two items including: If your group of friends told you *not* to do something because it was wrong, how likely is it that you would listen to them?

Scale mean = 4.19 (1.17); $\alpha = 0.80$

Response categories: 1) not at all likely to 5) very likely

Delinquent Peers: Seven items including: During the last year, how many of your current friends have attacked someone with a weapon?

Scale mean = 1.30 (0.54); $\alpha = 0.86$

Response categories: 1) none of them, 2) few of them, 3) half of them, 4) most of them, and 5) all of them

Lying Neutralizations: Three items such as: It's okay to tell a small lie if it doesn't hurt anyone.

Scale mean = 2.60 (0.98); $\alpha = 0.76$

Response categories: 1) strongly disagree to 5) strongly agree

Stealing Neutralizations: Three items such as: It's okay to steal something if that's the only way you could ever get it.

Scale mean = 1.64 (0.80); $\alpha = 0.83$

Response categories: 1) strongly disagree to 5) strongly agree

Hitting Neutralizations: Three items such as: It's okay to beat up someone if they hit you first.

Scale mean = 3.32 (1.11); $\alpha = 0.80$

Response categories: 1) strongly disagree to 5) strongly agree

School Commitment: Seven items such as: I try hard in school.

Scale mean = 3.92 (0.70); $\alpha = 0.77$

Response categories: 1) strongly disagree to 5) strongly agree

Guilt: Seven items including: How guilty or how badly would you feel if you stole something worth *less* than \$50?

Scale mean = 2.66 (0.55); $\alpha = 0.93$

Response categories: 1) not very guilty/badly, 2) somewhat guilty/badly, and 3) very guilty/badly

Conflict Resolution: Five items including: During the past year when you've gotten upset with someone, how often have you talked to the person about why you were upset?

Scale mean = 2.17 (0.46); $\alpha = 0.66$

Response categories: 1) never, 2) sometimes, and 3) often

Calming Others: Three items including: When someone was upset, how often have you asked the person why he/she was upset during the past year?

Scale mean = 2.41(0.51); $\alpha = 0.71$

Response categories: 1) never, 2) sometimes, and 3) often

Refusal Skills: Four items including: During the past year when you have tried to avoid doing something your friends tried to get you to do, how often have you told the person that I can't do it because my parents will get upset with me.

Scale mean = 2.33 (0.51); $\alpha = 0.70$

Response categories: 1) never, 2) sometimes, and 3) often

Prosocial Involvement Index: Four items including: During the past year, were you involved in school activities, or athletics?

Scale mean: 2.38 (1.14); $\alpha = 0.47$

Response categories: 1) yes and 2) no

Empathy: Five item including: I would feel sorry for a lonely stranger in a group.

Scale mean = 3.63 (0.65); $\alpha = 0.59$

Response categories: 1) strongly disagree to 5) strongly agree

Active Listening: Three items such as: I look at the person talking to me.

Scale mean = 3.66 (0.72); $\alpha = 0.60$

Response categories: 1) strongly disagree to 5) strongly agree

Problem Solving: Two items including: I talk to my friends about my problems.

Scale mean = 3.57 (0.91); $\alpha = 0.45$

Response categories: 1) strongly disagree to 5) strongly agree

Awareness of Services: Four items including: You know where a person can go for help if he/she is victimized.

Scale mean = 3.76 (0.65); $\alpha = 0.72$

Response categories: 1) strongly disagree to 5) strongly agree

Collective Efficacy: Three items including: It is my responsibility to do something about problems in our community.

Scale mean = 3.25 (0.77); $\alpha = 0.62$

Response categories: 1) strongly disagree to 5) strongly agree

Attitudes about Gangs: Two items: Getting involved with gangs will interfere with reaching my goals.

Scale mean = 3.72 (1.12); $\alpha = 0.71$

Response categories: 1) strongly disagree to 5) strongly agree

Altruism: Three items including: It feels good to do something without expecting anything in return.

Scale mean = 3.60 (0.83); $\alpha = 0.66$

Response categories: 1) strongly disagree to 5) strongly agree

Finn-Aage Esbensen is the E. Desmond Lee Professor of Youth Crime and Violence and also serves as Chair of the Department of Criminology and Criminal Justice at the University of Missouri—St. Louis. He also serves on the Steering Committee of the Eurogang Program of Research.

D. Wayne Osgood is a professor of criminology and sociology at Pennsylvania State University and lead editor of *Criminology*. His current research interests include relationships between adolescent friendship networks and delinquency, connections between time use and problem behavior, and effects of prevention programs.

Dana Peterson is an associate professor in the School of Criminal Justice, University at Albany (New York) and conducts research primarily in the areas of youth gangs and violence, with particular interest in how these are structured by sex and gender.

Terrance J. Taylor is an associate professor in the Department of Criminology and Criminal Justice at the University of Missouri—St. Louis. His primary research interest involves youth crime and violence. He received his Ph.D. in criminal justice from the University of Nebraska in 2002.

Dena C. Carson is an assistant research professor in the Department of Criminology and Criminal Justice at the University of Missouri—St. Louis. Her general research interests include youth violence, victimization, gangs, and delinquent peer groups. Her recent publications have appeared in *Youth & Society*, *Journal of Criminal Justice*, and *Youth Violence & Juvenile Justice*.