HOW GREAT IS G.R.E.A.T.? RESULTS FROM A LONGITUDINAL QUASI-EXPERIMENTAL DESIGN*

FINN-AAGE ESBENSEN
University of Missouri-St. Louis

D. WAYNE OSGOOD
Pennsylvania State University

TERRANCE J. TAYLOR
University of Nebraska at Omaha

DANA PETERSON
University of Nebraska at Omaha

ADRIENNE FRENG
University of Wyoming

Research Summary:
This manuscript presents results from the National Evaluation of the Gang Resistance Education and Training (G.R.E.A.T.) program, a school-based prevention program targeting middle-school students. A longitudinal quasi-experimental research design was conducted from 1995 through 1999. Beneficial program effects emerged gradually over time so that there was, on average, more pro-social change in the attitudes of G.R.E.A.T. students than the non-G.R.E.A.T. students four years following program exposure.

Policy Implications:
Two specific policy recommendations stem from this research. First, law enforcement officers can be effective providers of school-based prevention programs. Second, to better assess program effectiveness, evaluations should include design features that allow for assessment of long-term or delayed program effects.

* Earlier drafts of this paper have been presented at the 1998 and 1999 annual meetings of the American Society of Criminology. This research is supported under Award 94-IJ-CX-0058 from the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. Points of view in this document are those of the authors and do not necessarily represent the official position of the U.S. Department of Justice.

Our research evaluates the effectiveness of the Gang Resistance Education and Training (G.R.E.A.T.) program, which is illustrative of a prevention approach to the gang problem. The Phoenix Police Department, along with other "Valley of the Sun" police departments, introduced this school-based program in 1991 to provide "students with real tools to resist the lure and trap of gangs" (Humphrey and Baker, 1994:2). Modeled after the DARE (Drug Abuse Resistance Education) program, uniformed law enforcement officers introduce students to conflict resolution skills, cultural sensitivity, and the negative aspects of gang life during the nine-week G.R.E.A.T. program. In 1992, the Bureau of Alcohol, Tobacco, and Firearms (BATF) joined forces with the Phoenix Police Department by promoting and funding the program. (For a review of the historical development of the G.R.E.A.T. program, consult Winfree et al. 1999.) G.R.E.A.T. has proliferated throughout the country during its short history and by the end of 1997 had been incorporated into school curricula in all 50 states and the District of Columbia.

The stated objectives of the G.R.E.A.T. program are to reduce gang activity and to teach students about the negative consequences of gang involvement. The curriculum consists of eight lessons (taught in nine sessions) offered once a week to middle-school students, primarily seventh graders. Officers are provided with detailed lesson plans containing clearly stated purposes and objectives. In order to achieve the program's objectives, the eight lessons cover such topics as conflict resolution, goal setting, and resisting peer pressure. Discussion about gangs and their effects on the quality of people's lives are also included. The eight lessons are as follows:

1. Introduction—Acquaint students with the G.R.E.A.T. program and presenting officer.
2. Crime/Victims and Your Rights—Students learn about crimes, their victims, and their impact on school and neighborhood.
3. Cultural Sensitivity/Prejudice—Students learn how cultural differences impact their school and neighborhood.
4. Conflict Resolution (two sessions)—Students learn how to create an atmosphere of understanding that would enable all parties to better address problems and work on solutions together.
5. Meeting Basic Needs—Students learn how to meet their basic needs without joining a gang.
6. Drugs/Neighborhoods—Students learn how drugs affect their school and neighborhood.
7. Responsibility—Students learn about the diverse responsibilities of people in their school and neighborhood.
8. Goal Setting—Students learn the need for goal setting and how to establish short and long term goals.

As evidenced by the curriculum, the G.R.E.A.T. program is intended to provide life skills that empower adolescents with the ability to resist peer pressure to join gangs. The strategy is a cognitive approach that seeks to produce attitudinal and behavioral change through instruction, discussion, and role playing.

A notable feature of the program is its target population. In contrast to suppression and intervention programs, which are directed at youths who already are gang members, G.R.E.A.T. is intended for all youth. This is the classic, broad-based prevention strategy that is found in medical immunization programs: One intervenes broadly, with a simple and relatively unintrusive program, well before any problem is detectable and without any attempt to predict who is most likely to be affected by the problem.

To date, two published evaluations assessing program effectiveness have reported small but positive program effects on students' attitudes and behavior (Esbensen and Osgood, 1997 1999; Palumbo and Ferguson, 1995), while an earlier assessment of program effects two years after exposure failed to detect any significant differences between G.R.E.A.T. and non-G.R.E.A.T. students (Esbensen, 2001). Esbensen and Osgood reported findings from the cross-sectional component of the National Evaluation of the G.R.E.A.T. program. This study relied on surveys completed by 5,935 eighth-grade students in 11 cities across the continental United States one year after program delivery. Esbensen and Osgood (1999) found that students who had completed the G.R.E.A.T. program reported committing fewer delinquent acts and expressed more prosocial attitudes, including, among others, more favorable attitudes toward the police, higher levels of attachment to parents and self-esteem, and greater commitment to school. Using a multisite, pretest/post-test research design, Palumbo and Ferguson (1995) found the students had a "slightly increased ability" to resist the pressures to join gangs. The authors acknowledged, however, that "the lack of a control group prevents assessments of the internal validity. Therefore, it cannot be concluded that the results . . . were due to GREAT (sic) as opposed to other factors" (p. 600). The present study improves on this earlier research by combining their strengths, using both a control group and comparisons over time at multiple sites.

In addition to these outcome evaluations, the National Evaluation of G.R.E.A.T. included a number of other evaluation components. A process evaluation examining the fidelity of program implementation was completed during the early years of the project, concluding that the program integrity, both with regard to officer training and classroom instruction, was high (Esbensen et al., forthcoming; Sellers et al., 1998). In addition to the student surveys, which are the focus of the current

research, parents, school personnel, and G.R.E.A.T. officers also completed mail questionnaires inquiring about satisfaction and perceived effectiveness of the program. An overall high level of satisfaction with the G.R.E.A.T. program was expressed by these three populations (Freng, 2001; Peterson, 2001; Taylor, 2001).

EVALUATION RESEARCH DESIGN

Although the development of the G.R.E.A.T. curriculum was not theory driven, the design of the National Evaluation was (Winfree et al., 1996). The theories judged to be most relevant to the program were social learning theory (Akers, 1985) and self-control theory (Gottfredson and Hirschi, 1990). The identification of relevant theoretical constructs is critical to the short-term evaluation of prevention programs because prevention necessarily takes place well before the outcome of major concern (gang membership) is likely to occur. Thus, our evaluation placed considerable emphasis on theoretical constructs that are logically related to the program's curriculum and that are both theoretically and empirically linked to gang membership and delinquency (Grasmick et al., 1993; Hawkins and Catalano, 1993; Huizinga et al., 1994; Winfree et al., 1994).

Winfree et al. (1996) have elaborated on the relationship between the G.R.E.A.T. curriculum and the theoretical constructs included in this evaluation. For example, Lesson 4 of G.R.E.A.T. (conflict resolution) deals with concepts closely linked to self-control theory's anger and coping strategies. Lesson 5 (meeting basic needs) has conceptual ties to the risk-taking element of self-control theory. Lessons 6, 7, and 8 include elements addressing delayed gratification and impulsive behavior by attempting to teach responsibility and goal setting, including personal and career goals.

Elements of social learning theory appear in Lessons 1, 3, and 4. These lessons introduce definitions of laws, values, norms, and rules supportive of law-abiding behavior. Tolerance and acceptance (Lesson 3), for instance, are presented as values that reduce conflict and subsequently violence. Further, Lesson 4 addresses conflict resolution and steps students can take to ward off negative peer influences.

MEASURES

Measures included in the student questionnaires can be divided into two main categories: attitudinal and behavioral. As discussed above, the attitudinal measures included in these instruments can be classified as measures of distinct theoretical perspectives (e.g., social learning, social

control, social strain, and self-control).[1] For the current paper, those distinctions play little role—all are relevant both to the G.R.E.A.T. curriculum and to delinquency and gang membership. Thus, we refer to them only as attitudinal variables.

Of primary importance were measures of perceptions regarding the appropriateness of certain behaviors and measures of peer group conduct. Given the significant role of peers in gangs and delinquency, several different scales were used to tap the extent to which the youths felt committed to their peer group. Questions measuring students' involvement in school and community activities were also included in the questionnaires. Brief descriptions of attitudinal measures included in the student questionnaire are provided in Appendix I.

One of the more important objectives of the G.R.E.A.T. program is to reduce adolescent involvement in criminal behavior and gangs. We measured this involvement through self-reports of illegal activity by the respondents. This technique has been used widely during the past 30 years and provides a good measure of actual behavior rather than a reactive measure of police response to behavior (e.g., Hindelang et al., 1981; Huizinga, 1991; Huizinga and Elliott, 1986). Appendix II contains a list of behaviors included in the self-report inventory. Questions measuring self-reported drug use and victimization were also included in this section of the student questionnaire.

Gang membership was determined by students' responses to two filter questions: "Have you ever been a gang member?" and "Are you now in a gang?" Any student answering either of these questions in the affirmative responded to a series of questions requesting information about gang structure, gang activity, and attitudes about the gang. Prior to the "gang section" of the questionnaire, all students responded to perceptual questions about gangs, including the following: identification of good and bad things associated with gang membership and approval of gang membership.

LONGITUDINAL RESEARCH DESIGN

Two of the three previously published evaluations of the G.R.E.A.T. program contain methodological limitations. As stated, the Palumbo and Ferguson (1995) study did not include a comparison group, whereas the Esbensen and Osgood (1997, 1999) evaluation of G.R.E.A.T. utilized a cross-sectional design. This latter design lacks a pretest measure and

1. During the summer of 1995, members of the G.R.E.A.T. training staff were asked to critique the student questionnaire and to provide additional questions that would tap the program's message.

requires the ex post facto creation of a comparison group. Although statistical procedures can strengthen the validity of this method (Heinsman and Shadish, 1996), some consider it a weak design (e.g., Sherman et al., 1997).

The longitudinal research strategy implemented in the current evaluation, with a quasi-experimental research design and assignment of classrooms to treatment, serves two very important functions. First, this assignment process should create groups of G.R.E.A.T. and non-G.R.E.A.T. students that are at equal risk for future delinquency and gang involvement. Second, the longitudinal research design greatly increases statistical power for detecting program effects by controlling for previous individual differences and examining change over time.

Site Selection

Six cities were selected for inclusion in the longitudinal phase of the National Evaluation. The first criterion was the existence of a viable G.R.E.A.T. program. A second criterion was geographical location. A third criterion was the cooperation of the school districts and the police departments in each site. As such, we selected an East Coast city (Philadelphia, Pennsylvania), a West Coast location (Portland, Oregon), the site of the program's inception (Phoenix, Arizona), a Midwest city (Omaha, Nebraska), a "non-gang" city (Lincoln, Nebraska),[2] and a small "border town" with a chronic gang problem (Las Cruces, New Mexico).

Quasi-experimental Research Design

The longitudinal study included relatively equal sized groups of treatment (G.R.E.A.T.) and control (non-G.R.E.A.T.) students in the seventh grade (at the outset of the evaluation) at five of the sites and in sixth grade at the sixth site. Table 1 reports the number of students at each of the sites. Because G.R.E.A.T. is a classroom-based program, assignment was implemented for classrooms rather than for individual students. When data are pooled across sites, there will be a large enough sample of classrooms for confidence in our results, even when classrooms are used as the unit of analysis. The longitudinal sample consisted of 22 schools, 153 classrooms, and more than 3,500 students.[3]

2. Prior to the 1994 National Institute of Justice survey assessing the spread of gangs in America, Lincoln had reported no gangs (Curry and Decker, 1998:20–23). In 1994 and subsequent surveys conducted by the National Youth Gang Center, Lincoln has reported the presence of youth gangs.

3. At the beginning of the academic year, class lists from all of the participating classrooms were obtained. Each student appearing on these lists was assigned a uniquely identifiable number to be used throughout the longitudinal data collection. These lists contained names of individuals who had moved or failed to enroll for some

Table 1. Questionnaire Completion Rates

Site	Sample	Active Consent Sample	Pretest[a]	Post-Test[b]	1 Yr[c]	2 Yr	3 Yr	4 Yr
Las Cruces	626	301	518 (83)	519 (83)	275 (91)	242 (80)	228 (76)	225 (75)
Lincoln	653	425	595 (91)	351 (83)	388 (91)	366 (86)	329 (77)	331 (78)
Omaha	672	470	440 (94)	414 (88)	390 (83)	354 (75)	328 (70)	329 (70)
Philadelphia	465	228	388 (83)	317 (68)	174 (76)	147 (64)	159 (70)	125 (55)
Phoenix	569	300	493 (87)	434 (76)	250 (83)	195 (65)	191 (64)	163 (54)
Portland	583	321	502 (86)	468 (80)	284 (88)	246 (77)	186 (58)	204 (64)
TOTAL	3568	2045	2936 (87)	2503 (80)	1761 (86)	1550 (76)	1419 (69)	1377 (67)

[a] Completion rates based on initial sample size in all sites except Omaha where active parental consent was required for the pretest.
[b] Completion rates based on initial sample size in Las Cruces, Philadelphia, Phoenix, and Portland. Active parental consent was required for the post-test survey in Lincoln; so the active consent sample was used to calculate completion rates in Lincoln and Omaha.
[c] Completion rates for all four annual follow-up surveys were based on the active consent sample.

The assignment process was a critical feature of the research design. During late summer and early fall of 1995, procedures for assignment of classrooms to treatment and control conditions were developed at each of the 22 middle schools participating in the longitudinal study. Because the G.R.E.A.T. program was implemented differently at each site, unique solutions were required to implement random assignment at each site and, in some situations, at each school.[4] The exact nature of the process was dependent on what was possible at each site, but in all cases, the goal was to minimize the potential for differences between the sets of treatment and control classes.

In one site, for example, all seventh-grade students were required to complete a nine-week health class. This health class had been selected by the district as the logical placement of the G.R.E.A.T. program. This greatly facilitated the evaluation's assignment process in that no a priori selection factors were involved in the assignment of students to these classes. In two of the four schools participating in the evaluation, all health classes during the first quarter received G.R.E.A.T., whereas all health classes at the other two schools served as controls. During the second

other reason. Students whose names did not appear on the lists but who were in attendance were added to the list. We did not remove students from the list; so our initial sample and the pre-test and post-test completion rates provide a conservative estimate.

4. Although we had initially planned to implement random assignment, this was not feasible in every school. In the majority of schools (15 of the 22), school and law enforcement were amenable to random assignment and classrooms were assigned through a random process. In the remaining seven schools, assignment had to be more purposive, based on officer availability or limitations imposed by the school district.

quarter, this process was reversed. This arrangement allowed for the integrity of the assignment process to be met while minimizing the impact on school and police personnel.

In another site, three different procedures were used. In the first school, assignment was by classrooms (three to each condition). One week after the assignment, however, one of the control teachers was released due to a district-wide fiscal shortfall and students in that class were merged into the remaining five classes. This change resulted in a somewhat larger treatment group than control group, but this did not introduce any systematic bias. At two of the other schools, a literal flip of the coin determined which "house"—group of teachers and their classes—would receive the G.R.E.A.T. program and which would be withheld. At the fourth school, three teachers each taught two classes of the same subject. To reduce disruption to the school and to reduce teacher-induced bias into the research project, we selected one of each teacher's classes for G.R.E.A.T. instruction and the other for the control condition. Thus, to reiterate, the actual assignment process varied from city to city and even from school to school. Each modified process sought to ensure the integrity of the research objective while accommodating the class scheduling and structure of the schools.

Active Consent Procedures

The University of Nebraska Institutional Review Board approved a research design that allowed passive parental consent (the absence of a signed refusal implies consent) during the pretest and post-test data collection. These surveys were conducted two weeks before and two weeks after delivery of the G.R.E.A.T. program. Active parental consent (the absence of signed permission implies refusal) was planned for the subsequent annual surveys. These procedures were also approved by each of the participating school districts. During the spring of 1995, however, one school district (Omaha) enacted a new policy requiring active consent for all research conducted in the schools. Thus, in Omaha, active parental consent was obtained prior to the pretests. In the other five sites, the pretests include all students in attendance whose parents had not excluded them through the passive consent procedure (only 13 students).

A modified Dillman (1978) total design method was utilized to obtain the active consent forms, although the specific procedures varied slightly in terms of timing and sequencing across the six sites. The following serves as an "ideal type" of the procedures that were followed. During the spring and summer of 1996, three direct mailings were made to parents of survey participants. Included in the mailings were a cover letter, two copies of the parent consent form for student participation, and a business reply envelope. With substantial Spanish-speaking populations in Phoenix

and Las Cruces, mailings to parents in these cities included Spanish versions of the cover letter and consent form. In addition to the mailings, all parents not responding after the second mailing were contacted by telephone. School personnel also cooperated by distributing consent forms and cover letters at school. Teachers in all of the classrooms involved in the evaluation assisted with this process, rewarding students with a new pencil on return of the forms. Some teachers agreed to allow us to offer incentives such as pizza parties to classrooms in which a minimum of 70% of students returned a completed consent form. Other teachers offered incentives on their own, including earlier lunch passes and extra credit points.

Based on previous experience with obtaining active consent from parents of middle-school students, we knew that we needed at least four weeks of intensive effort at each site to reach an acceptable response rate (Esbensen et al., 1996). Due to a combination of staffing patterns and the logistics of coordinating consent processes at five national sites, we staggered the consent process throughout the spring of 1996.[5] This allowed for several mailings and coordination with school personnel prior to the end of the school year, while also allowing for an additional mailing during the summer (prior to the first-year follow-up survey administered during the fall of 1996). With this staggered approach, we experimented with slightly different procedures at the six sites.

The results of the active consent process led to an overall retention of 57% of the initial sample (see Table 1), although the sample retention varied from a low of 48% at one site to a high of 70% at another (for a more detailed discussion of the active consent process and examination of the effects of active consent procedures on the representativeness of the sample, consult Esbensen et al., 1999). This variation in return rates is more likely due to demographic differences among the sites than to the particular sequence of methods used at each site. Lincoln, for example, is

5. In Lincoln, it became necessary to obtain active consent prior to the post-test. This imposed a serious time constraint on the research; the end of the semester was only four weeks away when this request was made. At the conclusion of the semester, student schedules were changed and it would no longer be possible to survey the students in their original classrooms. The questionnaires had to be completed prior to the end of the first semester. We, thus, had less than three weeks to actually implement the active consent process and still have time to administer the questionnaires. In this site, we implemented simultaneous mail and classroom distribution of forms and visited the classrooms every other day to collect returned forms. In spite of the time limitation imposed for the active consent procedure, we were successful in obtaining a 70% response rate prior to the post-test administration. A subsequent mailing to nonrespondents after the post-test resulted the return of additional consent forms, raising the final response rate to 76%. These additional students were retained in the study, however, without data from the immediate post-test.

a relatively stable community with a high percentage of intact families, predominantly white, and with the majority of adults having more than a high-school education, all characteristics found to be associated with high response rates (Ellickson and Hawes, 1989). On the other hand, Philadelphia is predominantly African-American and has fewer parents with more than a high-school diploma than the other five sites. It would appear that community demographic characteristics play an important role in response rates. In spite of such differences, however, we can conclude that through diligence and the use of multiple methods, researchers can obtain participation rates in excess of 50% under active consent procedures.

To conclude, at a minimum, the following general procedures were used in all six sites: at least three mailings, follow-up phone calls after the second mailing, collaboration of school teachers, and the offering of incentives to those students returning completed forms. These procedures resulted in an overall response rate of 67% (57% providing affirmative consent and 10% withholding consent), whereas 33% of parents failed to return the consent forms.

Questionnaire Completion Rates

The completion rates for the student survey were excellent. Of the 2,045 students for whom active parental consent was obtained, 1,761 (86%) students completed surveys during the one-year follow-up. Annual retention rates for years two, three, and four were 76%, 69% and 67% respectively (see Table 1). Given the multisite, multischool sample, combined with the fact that respondents at five of the six sites made the transition from middle school to high school between the year 1 and year 2 surveys, this completion rate is commendable. Hansen et al. (1990) examined attrition in a meta-analysis of 85 longitudinal studies and reported an average completion rate of 72% for the 19 studies with a 24-month follow-up period. Few of these 19 studies included multisite samples. Tebes et al. (1992) reported on the attrition rates from middle school to high school. In their study examining differential attrition for different age groups, they report losing 41.3% of their sample between eighth and ninth grade!

For the second-, third-, and fourth-year follow-up surveys, considerable difficulty was introduced into the retention of the student sample. As the cohort moved from middle school to high school, combined with normal mobility patterns, students were enrolled in more than 10 different high schools in each of four sites (Omaha, Phoenix, Portland, and Philadelphia), and by the last data collection effort, participating students were enrolled in more than 100 different schools. It was necessary to contact school officials at these schools, whether fewer than 10 respondents or more than 100 were enrolled at the school. In some instances, these new

schools were in different districts, which required approval from the necessary authorities to survey their students. In spite of these logistical concerns, we successfully obtained completed questionnaires from 67% of the sample in the fourth-year follow-up survey.

DESIGN OF THE LONGITUDINAL ANALYSIS OF PROGRAM IMPACT

Our research design involves four nested levels of analysis: Waves of data collection are nested within individual students who are followed over time, those students are nested within classrooms in which the program was (or was not) delivered, and the classrooms are nested within schools. It is important that our analysis takes into account the nature of our research design at each of these levels of analysis.

Correctly specifying our statistical model avoids two potential problems. The first problem would be violating the assumption of independence among observations, which would lead to erroneous tests of statistical significance, potentially making chance fluctuations appear significant. Dependence occurs when there are systematic patterns of similarity among observations, and those similarities are not fully explained by variables included in the analysis. Because there is no way of guaranteeing that all sources of similarity among nested observations can be identified and measured, it is prudent to allow for the possibility of dependence among them. Modern multilevel regression models, such as Bryk and Raudenbush's (1992) hierarchical linear modeling (HLM) and Goldstein's (1995) (MLwiN), permit us to do so by adding residual variance components that reflect systematic variations in means and in relationships across higher level units of analysis (i.e., individuals, classrooms, and schools). We address this concern through a four-level hierarchical model, which we estimate using Goldstein's (1995) MLwiN program. MLwiN is comparable to Bryk and Raudenbush's (1992) HLM, but MLwiN is better suited to our purposes because HLM is limited to three levels of analysis.

The second potential problem for the analysis is that our estimate of program effect may be biased by extraneous features of our research design. This bias would arise if our comparison between the treatment and control groups was influenced by pre-existing differences between individuals, classrooms, or schools. We have been careful to structure our statistical model to avoid such biases and to create comparisons that provide the most meaningful assessment of the impact of G.R.E.A.T. Below, we explain how we coded our data to accomplish this, and we also specify the variance components included in our model.

WITHIN-INDIVIDUAL CHANGE OVER TIME

Our interest is in within-individual change over time, as would be reflected in differences between an individual's scores across waves of data. With several waves of data, at least some degree of sample attrition is inevitable; so we must be careful that the changing sample composition does not bias estimates of change. We avoid that bias by studying change through contrasts between waves and including in our model individual means (across waves) on those contrasts. That statistical control ensures that the contrasts become strictly within-individual comparisons, not influenced by any stable individual differences (Bryk and Raudenbush, 1992:121–123). We also include in our model a variance component for individual means across time, which allows for dependence among observations from the same person.

The general form of our analysis follows Osgood and Smith's (1995) strategy for program evaluations with extended longitudinal designs. In order to focus attention on change attributable to participation in the program, the most important element of the model is on a contrast between the pretest measure (coded as 0) and the five later waves of data (all coded as 1). The coefficient for this contrast will equal the mean change between the pretest and all five later waves of data combined. Pooling the post-program observations in this fashion yields an especially powerful test for any change attributable to the program.

Individual differences in the mean value of this contrast arise because of missing data. Respondents who contribute data for all waves have a mean of five-sixths, whereas those who complete only the pretest and immediate post-test have a mean of one-half. If we did not control for individual means on this contrast, its regression coefficient would be an undifferentiated amalgam of genuine change over time and pre-existing individual differences associated with attrition. Estimating a separate coefficient for the individual means separates these two relationships so that the contrast between waves is limited to within-individual change.

Only respondents who completed both the preprogram questionnaire and at least one later survey contribute to the results for this contrast. All other respondents were eliminated from the analysis. Of the 2,714 students who completed the pretest questionnaire, 2,293 students (84.5%) met this criterion. These students represent 145 classrooms from 20 schools at the six research sites.[6]

Osgood and Smith (1995) recommend using a polynomial function of time to capture the pattern of change during the post-test period. With five waves of post-test data in our study, a quadratic function is sufficient

6. Two schools in Philadelphia were eliminated from the analysis due to an inadequate number of either treatment of control classrooms.

to capture any systematic trend. To maintain a straightforward interpretation of the pre-post contrast, we coded the linear and quadratic terms to be orthogonal to it [i.e., values of 0, −2, −1, 0, 1, 2 and 0, 2, −1, −2, −1, 2 across waves (Judd and McClelland, 1989)]. In the presence of a large positive value for the pre-post contrast, a moderate positive value for the linear term would indicate continuing increase over time, whereas a moderate negative value would indicate decline after initial positive change. The quadratic term reflects curvilinear change during the post-program period. As with the first contrast, our model includes individual means on the linear and quadratic terms in order to capture any association with attrition and thereby to limit results to within-individual change.

PROGRAM IMPACT

G.R.E.A.T. is a school-based program, delivered simultaneously to entire classrooms rather than separately to individual students. Participation in the program was captured by a contrast assigned a value of +.5 for students in classrooms that received the program and a value of −.5 for students in classrooms that did not. For the assumption of independence to hold, all similarity within classes would have to be explained by the treatment effect and by any control variables. As with all classroom-based studies, there are many other sources of similarity as well, such as the teacher's classroom management style, which police officer delivers the program, and all the factors that determine which students end up in which classrooms. We therefore include in our model classroom-level variance components for both mean level on the outcome measure and for the contrasts over time.

There is also the potential for systematic differences between the schools participating in the evaluation, and in the assignment of classrooms to treatment and comparison groups within schools. Any variation across schools in the proportion of students in the two groups would produce between-school variation on this contrast, confounding the comparison between treatment and control groups with pre-existing differences between schools. Most schools had equal numbers of classrooms in the two groups; so differences in rates of program participation were not substantial, ranging from 42% to 69% of students participating in G.R.E.A.T. As with the contrasts over time, we eliminate this possible bias by controlling for school means on the contrast of treatment versus control. This procedure has the same effect as using dummy variables to control for all differences between schools, as we did in our cross-sectional evaluation of G.R.E.A.T. (Esbensen and Osgood, 1999).

Success of the G.R.E.A.T. program would be evidenced by more favorable change over time in the treatment group than in the control

group. Therefore, our estimate of program impact takes the form of interactions between the within-individual contrasts over time and the within-school contrast between G.R.E.A.T. and comparison classrooms. There is one such interaction term for each between-wave contrast, with the interaction for the first over-time contrast (preprogram versus all post-program waves) most directly reflecting any overall beneficial impact of the program.

To allow for the possibility of variation across schools in the comparability of G.R.E.A.T. and control classrooms and in treatment effects, our model included school-level variance components for three terms: the overall mean, the treatment versus control pretest difference, and the overall treatment effect (i.e., the interaction of preprogram versus post-program with treatment versus control). Additional variance components are not feasible with this number of schools, and we judged these three to be most relevant to a sound test of program effectiveness. These variance components rarely proved statistically significant, but they were retained in the model except when one or more terms had to be eliminated for MLwiN to converge on satisfactory estimates.

RESULTS

COMPARABILITY OF TREATMENT AND CONTROL GROUPS

We begin by examining whether the treatment and control groups were comparable at the beginning of the study. Although our analysis controls for any such differences (by focusing on within-individual change), the evaluation was designed to produce comparable groups, and the study is much stronger if the design succeeded than if we artificially approximate comparability through statistical controls. We make the pretest comparison of groups as part of the full multilevel model of program impact, in which the coefficient for G.R.E.A.T. versus control reflects differences at the pretest (given the dummy coding of pre-post contrast and its interaction with the treatment-control contrast). This estimate controls for mean differences between schools and takes into account variability across schools and across classrooms within schools.

As can be seen in Table 2, the two groups are, on the whole, very similar, with only very small differences between the groups on all variables. Even so, there are two variables for which the difference between groups is great enough to be significant at the .05 level (victimization and negative beliefs about gangs) and three more that surpass the .10 level (self-reported status offenses, peer delinquency, and prosocial peers). This number is somewhat more than would be expected by chance for these 32 significance tests (1.6 at .05 and 3.2 at .10). The differences do not appear problematic, however. The smallest probability is not beyond chance for

this number of comparisons (i.e., 1 out of 32 equals .031 and the smallest p value is .034), and with over 2,000 cases, we have abundant statistical power to detect even very slight differences. Furthermore, all of the pretest differences that approach statistical significance indicate less favorable means in the treatment group than in the control group. Thus, if there is any genuine difference, it is slight and favors the control group.

Table 2. Pretest Comparison of G.R.E.A.T. Treatment Group and Control Group, From Full Multilevel Model

	Difference	S. E.	S. D.	p
Gang Membership[a]	–.210	.323	.034	.543
Drug Use	.034	.027	–.059	.199
Total Self-Report Delinquency	.020	.021	–.050	.331
Minor Self-Report Delinquency	.028	.025	–.052	.266
Person Self-Report Delinquency	.010	.021	–.023	.645
Property Self-Report Delinquency	.011	.018	–.028	.524
Status Self-Report Delinquency	.051	.031	–.077	.098
Victimization	.056*	.026	–.116	.034
Gang, Negative	–.034*	.017	–.105	.046
Gang, Positive	–.003	.016	.011	.837
Attitudes about Gangs	.015	.048	–.018	.756
Attitudes toward Police	–.062	.043	–.079	.149
Dangerous School Environment	.026	.028	–.045	.347
Peer Delinquency	.071	.037	–.105	.058
Prosocial Peers	–.078	.044	–.102	.076
Peer Commitment, Negative	.035	.055	–.032	.525
Peer Commitment, Positive	–.069	.047	–.067	.140
Guilt	–.005	.026	–.009	.850
Neutralization Total	.041	.045	–.051	.361
Neutralization for Fighting	.091	.061	–.083	.136
Neutralization for Lying	.020	.049	–.021	.680
Neutralization for Stealing	.005	.051	–.006	.922
Impulsiveness	.000	.034	.000	.994
Risk Seeking	.068	.044	–.073	.122
Self-Esteem	–.025	.032	–.035	.438
Social Isolation	–.020	.048	.021	.676
School Commitment	–.020	.035	–.027	.572
Limited Educational Opportunity	.005	.033	–.007	.884
Maternal Attachment	.047	.062	.036	.448
Paternal Attachment	–.011	.067	–.008	.869
Parental Monitoring	.000	.036	.000	.997
Cultural Identification	.042	.040	.061	.300

* $p < .05$.
[a] Gang membership is a dichotomous variable; so its regression coefficients are logistic. The standardized difference is based on translating the logistic coefficients to probabilities and dividing the probability difference by the observed standard deviation.

OVERALL PROGRAM IMPACT

Tables 3 and 4 present the results for our analysis of the impact of the

G.R.E.A.T. program on our set of outcome measures. Our primary interest is in Table 3, which reflects program effects on the overall change from the preprogram to post-program periods. Here, we see evidence of a small but systematic beneficial effect of the program. The program effect is statistically significant for five of the outcome measures: victimization, negative views about gangs, attitudes toward police, prosocial peers, and risk seeking. Although these five outcomes are a fraction of the total set of 32 outcomes measures, they are a greater number than would be expected by chance (1.6 at $p < .05$). More importantly, in all cases, the direction of the difference favors participants in G.R.E.A.T. (e.g., less victimization and more prosocial peers), which is also true of peer delinquency and commitment to negative peers, for which $p < .10$. Indeed, all but four of the 32 estimates of program impact are in the direction favorable to G.R.E.A.T., and this preponderance is far greater than would be expected by chance (sign test: $Z = 4.24, p < .001$). Thus, we conclude that the beneficial direction of the program impact is statistically reliable. At this same time, it is important to realize that the magnitude of this positive impact is small. The average standardized program effect for the five significant outcomes is only .11, and the average across all measures is only .04.

There is less evidence of program effects on trends during the post-program period (Table 4). Program effects on the linear trend reach statistical significance for three outcome measures, and one for the quadratic trend. This is no more than would be expected by chance. It is interesting to note, however, that for 25 of the 32 outcome measures, the linear trend is in the direction of a greater movement toward positive adjustment for G.R.E.A.T. participants than for nonparticipants (sign test: $Z = 3.18, p < .001$). Thus, program benefits are more likely increasing over time than fading away.

Figure 1 illustrates the pattern of positive program effects that held for all the variables, with significant program effects on change from the preprogram to post-program periods. In the first of these examples, the overall rate of victimization declined throughout the study. With respect to prosocial peers, a different trend is observed; our respondents first suffered an overall loss of prosocial peers, followed by an increase in the final two years. For both outcomes, participants in G.R.E.A.T. had somewhat less favorable adjustment prior to the start of the program, meaning a higher rate of victimization and fewer prosocial peers. With each assessment after the completion of the program, the participants' adjustment improved relative to the control group, until the initial difference was reversed in the final two years. Although the change is small, note that it is in a pattern generally considered to be interpretable as a program effect, even if the initial difference was genuine. As Cook and Campbell note

Table 3. Analysis of Program Impact: Preprogram Versus Post-Program Contrast

	Difference	S. E.	S. D.	*p*
Gang Membership[a]	–.102	.350	.016	.771
Drug Use	–.043	.030	.074	.143
Total Self-Report Delinquency	–.010	.018	.025	.576
Minor Self-Report Delinquency	–.017	.026	.032	.519
Person Self-Report Delinquency	–.005	.018	.011	.794
Property Self-Report Delinquency	.003	.019	–.007	.891
Status Self-Report Delinquency	–.040	.029	.061	.162
Victimization	–.065*	.023	.135	.004
Gang, Negative	.033*	.015	.102	.028
Gang, Positive	–.014	.016	.049	.368
Attitudes about Gangs	–.029	.039	.035	.454
Attitudes toward Police	.073*	.037	.092	.049
Dangerous School Environment	–.032	.032	.055	.319
Peer Delinquency	–.057	.032	.085	.070
Prosocial Peers	.085*	.034	.111	.011
Peer Commitment, Negative	–.093	.053	.083	.081
Peer Commitment, Positive	.048	.049	.047	.322
Guilt	.001	.024	.003	.954
Neutralization Total	–.021	.039	.026	.584
Neutralization for Fighting	–.071	.050	.065	.160
Neutralization for Lying	–.017	.049	.017	.733
Neutralization for Stealing	.022	.046	–.024	.633
Impulsiveness	–.009	.033	.012	.781
Risk Seeking	–.093*	.039	.100	.016
Self-Esteem	.044	.031	.064	.153
Social Isolation	–.009	.038	.009	.822
School Commitment	.013	.033	.017	.700
Limited Educational Opportunity	.003	.029	–.004	.919
Maternal Attachment	–.015	.053	–.012	.777
Paternal Attachment	.086	.056	.059	.128
Parental Monitoring	.020	.035	.026	.579
Cultural Identification	.008	.041	.011	.853

* $p < .05$

[a] Gang membership is a dichotomous variable, so its regression coefficients are logistic. The standardized difference is based on translating the logistic coefficients to probabilities and dividing the probability difference by the observed standard deviation.

(1979:111), the reversal of an initial group difference cannot result from artifacts such as regression to the mean, ceiling effects, or maturation.

VARIATION IN PROGRAM EFFECTIVENESS BY PRIOR RISK

We also explored the possibility that the impact of the G.R.E.A.T. program may depend on students' levels of risk for delinquency and gang membership. Our earlier cross-sectional analyses of program impact had examined the consistency of program effects across demographic groupings (Esbensen and Osgood, 1997; 1999). There, we found evidence that G.R.E.A.T. was more effective with groups that are at higher risk for

Table 4. Analysis of Program Impact: Post-Program Trends

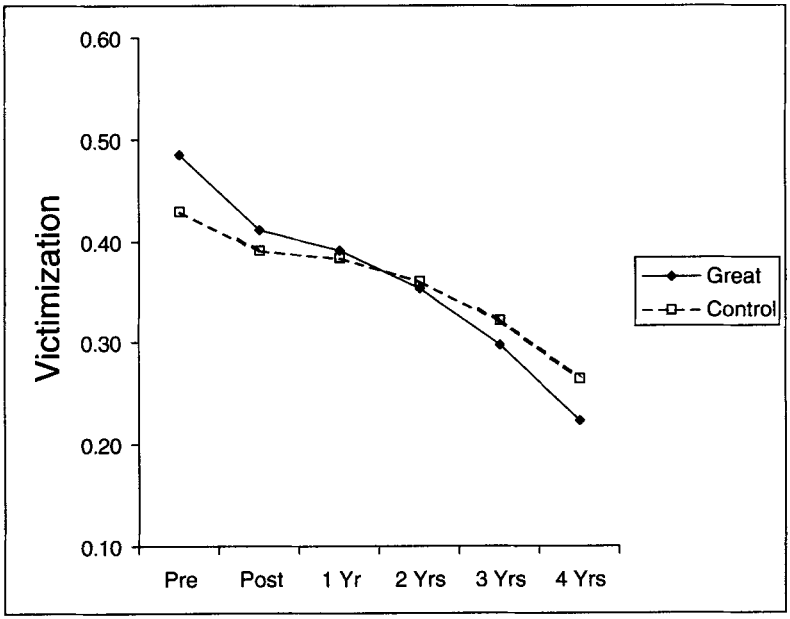
	Linear Trend			Quadratic Trend		
	Coef.	S. E.	*p*	Coef.	S. E.	*p*
Gang Membership[a]	–.206	.145	.154	–.072	.109	.507
Drug Use	–.010	.010	.283	.002	.007	.755
Total Self-Report Delinquency	–.004	.006	.540	–.009	.005	.060
Minor Self-Report Delinquency	–.012	.009	.214	–.011	.006	.074
Person Self-Report Delinquency	–.014*	.006	.020	–.001	.005	.788
Property Self-Report Delinquency	–.002	.007	.764	–.011*	.005	.032
Status Self-Report Delinquency	.006	.013	.631	–.002	.009	.798
Victimization	–.015*	.007	.031	–.001	.005	.882
Gang, Negative	.003	.005	.608	.003	.004	.413
Gang, Positive	–.007	.005	.117	.002	.004	.563
Attitudes about Gangs	–.010	.014	.494	.007	.009	.422
Attitudes toward Police	.022	.013	.094	.011	.009	.219
Dangerous School Environment	–.009	.014	.483	–.009	.007	.223
Peer Delinquency	–.019	.011	.080	–.007	.009	.406
Prosocial Peers	.022	.012	.071	–.005	.009	.626
Peer Commitment, Negative	–.032	.018	.074	.013	.014	.339
Peer Commitment, Positive	–.003	.017	.843	–.004	.013	.785
Guilt	–.006	.010	.575	.001	.006	.919
Neutralization Total	–.015	.013	.247	–.010	.009	.224
Neutralization for Fighting	–.020	.016	.208	–.012	.012	.307
Neutralization for Lying	–.024	.016	.132	–.011	.011	.309
Neutralization for Stealing	–.003	.015	.853	–.007	.011	.507
Impulsiveness	.007	.013	.586	–.006	.009	.475
Risk Seeking	–.002	.014	.871	.002	.010	.871
Self Esteem	.011	.012	.336	–.005	.008	.487
Social Isolation	.005	.014	.693	–.001	.010	.911
School Commitment	.005	.012	.692	.000	.008	.991
Limited Educational Opportunity	–.003	.011	.770	.005	.007	.521
Maternal Attachment	–.003	.019	.889	–.014	.012	.274
Paternal Attachment	–.007	.021	.740	–.005	.015	.746
Parental Monitoring	.023*	.011	.040	–.004	.008	.589
Cultural Identification	.016	.013	.205	.012	.008	.142

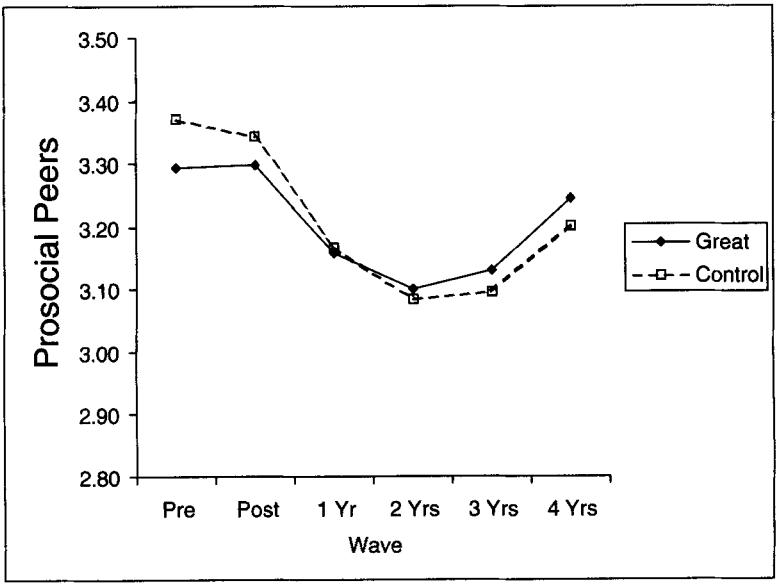
* $p < .05$.
[a] Gang membership is a dichotomous variable; so its regression coefficients are logistic. The standardized difference is based on translating the logistic coefficients to probabilities and dividing the probability difference by the observed standard deviation.

delinquency, specifically, males and minority group members. In the present longitudinal analysis, the data from the pretest measure allowed us to measure risk of future delinquency and gang membership directly, rather than inferring risk indirectly from demographic proxies.

We defined risk empirically by the relationship of pretest measures to anti-social outcomes at the one-year follow-up interview. Specifically, we conducted a regression analysis using as an outcome measure the mean of standardized scores on self-reported delinquency, self-reported drug use, ever having been in a gang, peer delinquency, expected guilt for deviant

Figure 1. Program Impact for Two Outcome Measures: Victimization and Prosocial Peers





acts (reversed), and favorable attitudes about gangs. The predictor variables from the pretest were age, sex, race, parents' education, and four factor scores that capture 51% of the variance of the pretest values for the entire set of outcome measures. This regression succeeded in explaining 34% of the variance in the generalized anti-social outcome measure. We defined the high-risk group as respondents in the top 25% of predicted anti-social outcomes based on this regression.

To determine whether program impact depended on a student's risk of anti-social outcome, we added this measure to our analysis of overall program impact as an explanatory variable. The statistical model included the main effect for risk and all two- and three-way interactions of risk with program participation (G.R.E.A.T. versus control group) and time (the three contrasts across waves). The three-way interaction of risk, G.R.E.A.T., and the pre-post program contrast would most directly reflect whether the program was more beneficial for high-risk youth than for low-risk youth. We conducted this analysis for 16 of the principal outcome measures (excluding subscales and including all variables with significant program impact). These analyses revealed no more than chance level differences between high- and low-risk youth in the impact of G.R.E.A.T. Also, there was no consistent pattern in the direction of the findings, so that students with a high risk of negative outcomes were no more or less likely to benefit from the program than were low-risk students.

TESTING ALTERNATIVE EXPLANATIONS OF THE PROGRAM EFFECT

We conducted three additional analyses to test whether our findings of program benefits may be attributable to weaknesses in the research design rather than to genuine effects of G.R.E.A.T. We were especially wary that the variables that showed the greatest program impact also tended to be variables on which there were larger pretest differences between treatment and control groups. Although our research design minimizes the possibility of genuine differences between the classrooms assigned to G.R.E.A.T. versus control classrooms, the data available for analysis is limited by attrition over time, which is also a function of the consent procedures. Furthermore, consent rates were higher in the treatment group than in the control group at three of the six research sites, and in a fourth, there was a higher rate of attrition in the control group in years 3 and 4. It is conceivable that these differential rates of participation would introduce differences between treatment and control groups that would be the source of the apparent program benefits. The reader should remember that our focus on within-individual change already dampens the impact of any such differences on our results. The purpose of the additional analyses was to further ensure that the apparent program effects were genuine.

Our first method to test the robustness of the program impact findings was to control for preprogram risk for negative outcomes. We did so by adding the continuous version of our risk measure (see preceding section) to our impact analysis. Our second approach to bolstering the comparability of the treatment and control groups was based on propensity scores (Rosenbaum and Rubin, 1983; Winship and Morgan, 1999). Specifically, we used logistic regression to estimate the probability (or propensity) for attrition, using as predictors the same set of pretest scores used to define the risk groups. We then compared the distributions of the treatment and control groups on these propensity scores, and probabilistically eliminated cases to produce matching distributions for the two groups. This propensity matching equates the two groups on the complex of factors associated with attrition. Because the consent process and choice of schools was distinct at each site, the propensity analysis and matching was site specific. We used the propensity matching to address the participation differences we had identified, thus, matching on propensity for consent at three sites and for participation in the fifth and sixth waves at one site. Matching resulted in the loss of about 8% of the total sample. Finally, pretest and attrition differences between treatment and control groups were most pronounced at one site (Philadelphia); so the risk of invalid results is greatest there. Therefore, our third approach to addressing the potential difference between groups was to repeat the analyses using only the other five sites.

For expediency, we limited these three sets of analyses to the seven outcome measures for which the overall program effect reached or approached statistical significance (i.e., $p < .10$). We reasoned that if these effects maintained their magnitude and statistical significance, then confidence in the overall set of findings was justified.

Each of these three approaches reduced the pretest differences between G.R.E.A.T. participants and the control groups, although none eliminated them. Using risk scores as a control variable was most effective in this regard. Despite the greater pretest comparability, the magnitudes of the estimated program effects were essentially unchanged. Indeed, the very small changes that occurred favored the treatment group at least as often as the control group. Furthermore, although two of these strategies sacrificed sample size, there were virtually no meaningful changes in statistical significance. For instance, in the analysis controlling for risk, the probability level for attitudes toward police rose from .049 to .052, whereas the probability level for peer delinquency fell from .070 to .031. Thus, it appears unlikely that our findings showing a positive impact of participating in G.R.E.A.T. could be due to pre-existing differences between the treatment and control groups. This increases our confidence that the G.R.E.A.T. program does have modest beneficial effects.

DISCUSSION

The G.R.E.A.T. program is a school-based prevention program that seeks to reduce adolescent involvement in gangs. Uniformed law enforcement officers teach this primarily cognitive-based program to middle-school students. Of primary interest in this evaluation was the question: Can a cognitive-based prevention program produce a measurable treatment effect? A related issue of considerable policy interest concerns the role of law enforcement in such programs; that is, are officers suitable deliverers of prevention programs in schools? Previous evaluations of similar law enforcement prevention efforts have provided mixed results. For example, the DARE program has been the object of numerous evaluations with what can be described, at best, as mixed results (e.g., Lynam et al., 1999; Rosenbaum and Hanson, 1998). Another program, law-related education (LRE), has not been as widely adopted nor evaluated as DARE, but the extant evaluations have generally been positive (see Gottfredson (2001) for a discussion). Contrary to the mixed reviews of DARE, the previously published outcome evaluations of the G.R.E.A.T. program have been modestly positive (Esbensen and Osgood, 1999; Palumbo and Ferguson, 1995). The four-year results reported here are consistent with those of the previous G.R.E.A.T. studies.

The consistency of these findings is important given that different research designs and slightly different methods were utilized in the three studies. Here, the focus is on the two studies that were conducted as part of the National Evaluation. The cross-sectional evaluation of the G.R.E.A.T. program was completed in 1995 in 11 cities using *anonymous* questionnaires completed by students under *passive parental consent* procedures. The longitudinal evaluation was conducted in six cities (four that were included in the cross-sectional study) from 1995 to 1999 using *confidential* questionnaires restricted to those students for whom *active parental consent* had been obtained. In spite of these methodological differences, results from the two evaluations were remarkably similar. Those students participating in the G.R.E.A.T. program expressed more prosocial attitudes after program completion than did those students who had not been exposed to the G.R.E.A.T. curriculum.

In spite of these consistent, yet modest, positive effects of the G.R.E.A.T. program, two issues need our attention. First, the program's primary stated objective is to reduce gang activity. Although the cross-sectional evaluation did find slightly lower rates of gang membership and self-reported delinquency, this was not the case in the longitudinal study. Second, although the cross-sectional findings reflected a difference between groups one year after program completion, the longitudinal design did not produce any significant group differences until three to four

years after program exposure. Had the evaluation been concluded after a one- or two-year follow-up period, our conclusions would have been different. For a discussion of the two year results, consult Esbensen (2001). The two graphs presented in Figure 1 reveal that it was not until the third year that the groups began to diverge. Each of these issues will be discussed more fully below.

The dual goals of the G.R.E.A.T. program (as stated in the G.R.E.A.T. Officer Instructor's Manual) are "to reduce gang activity, and to educate a population of young people as to the consequences of gang involvement." Additional inquiries of the G.R.E.A.T. management staff determined that another objective of the program was to develop positive relations with law enforcement. These three objectives are addressed through the eight-lesson curriculum that targets both attitudinal and behavioral change. At the outset of the National Evaluation, the research team developed a questionnaire that would allow examination of the effectiveness of program content as well as the stated objectives. Two of these program objectives appear to have been met through the curriculum; the G.R.E.A.T. students reported more favorable attitudes toward the police and more negative attitudes about gangs than did the non-G.R.E.A.T. students. Of the five evaluation outcomes that achieved statistical significance, it is worth noting that neither gang membership nor rates of self-reported delinquency were lower for the G.R.E.A.T. students than for the control group. All but one of the seven behavioral measures, however, were in the direction suggesting a program effect.

The finding that the benefit of G.R.E.A.T. became evident only gradually over many years can be considered curious and unexpected. For a short-term program such as this, many would expect any impact to be strongest immediately and to be subject to decay over time. However, other evaluations have reported similar lagged or long-term effects (e.g., Berrueta-Clement et al., 1984; Hawkins et al., 2000; Olds et al., 2001; Tremblay et al., 2001). Why this delayed effect occurs is less clear than the fact that it does occur. Several possible explanations come to mind. First, young adolescence is a stressful and anxiety-filled stage of life during which most adolescents experience considerable ambiguity with regard to appropriate attitudes and behaviors. Second, the organizational structure of American schools may contribute to this stress; at ages 11 or 12, children move from the comfort of relatively small and stable elementary schools to larger, more diversified middle or junior high schools, and then at ages 14 or 15, the young adolescent is forced to make another transition to an even larger, more diverse high-school setting. As the child reaches the age of 16 and 17, some of the angst of adolescence is resolved and the child has adapted to the high-school setting (tenth or eleventh grade).

Thus, prior prevention or intervention experiences may, at this time, begin to manifest themselves.

The lagged effects found in this research, when considered in the context of other similar program effects, suggest that program development and associated evaluations may well benefit from adopting a developmental or life-course perspective. Outcomes related to attitudinal and behavioral change may not be discernable in the short-run and may well be mediated by specific developmental issues, such as peer group affiliation, growing independence from family of origin, and school transitions. As noted by Tremblay et al. (2001:335):

> If a preschool intervention aims at preventing delinquency, the impact of the intervention must obviously be measured when the delinquent behavior usually appears, that is, no earlier than preadolescence. Clearly, we must expect interventions that aim to change the course of human development will have long-term effects. In fact, there may be more long-term effects than short-term effects.

From this perspective, these lagged effects do not appear either curious or unexpected.

To conclude, we would like to pose three summary statements and recommendations. First, to test for program effectiveness, a developmental perspective may be beneficial. This would entail longitudinal research designs that allow for examination of both short- and long-term program effects.

Second, evaluations need to consider not only stated program goals, but implied goals as well. The G.R.E.A.T. program, for instance, is generally described as a gang prevention program, with stated objectives of reducing gang membership and teaching students about the negative aspects of gangs. As mentioned above, a third goal of the program was the development of positive relationships with law enforcement. Had we assessed program effectiveness from the restricted perspective of reducing gang membership, we would have concluded the program to be ineffective. Including measures of attitudes toward the police and attitudes about gangs allowed for a more comprehensive assessment of program effectiveness.

Third, with respect to the issue of whether law enforcement officers can be effective deliverers of prevention curricular in schools, results of this evaluation suggest that they can be effective teachers. With the current proactive, community-oriented policing emphasis of law enforcement, prevention or school resource officers may play a beneficial role in crime prevention. Our findings of a positive program effect indicate that evaluations of officer-taught, school-based, cognitive prevention programs

need not be restricted to "feel-good" or satisfaction measures; such programs can have a measurable impact on students attitudes and limitedly on behaviors.

REFERENCES

Akers, Ronald L.
1985 Deviant Behavior: A Social Learning Approach, 3d ed. Belmont, Calif.: Wadsworth.

Berrueta-Clement, John R., Lawrence J. Schweinhart, W. Steven Barnett, Ann S. Epstein, and David P. Weikart
1984 Changed Lives: The Effects of the Perry Preschool Program on Youths through Age 19. Ypsilanti, Mich.: High/Scope Press.

Bryk, Anthony S. and Stephen W. Raudenbush
1992 Hierarchical Linear Models: Applications and Data Analysis Methods. Newbury Park, Calif.: Sage.

Cook, Thomas D. and Donald T. Campbell
1979 Quasi-Experimentation: Design & Analysis Issues for Field Settings. Boston, Mass.: Houghton Mifflin Company.

Curry, G. David and Scott H. Decker
1998 Confronting Gangs: Crime and Community. Los Angeles, Calif.: Roxbury Press.

Dillman, Don A.
1978 Mail and Telephone Surveys: The Total Design Method. New York: Wiley.

Ellickson, Phyllis L. and Jennifer A. Hawes
1989 An assessment of active versus passive methods of obtaining parental consent. Evaluation Review 13:45–55.

Elliott, Delbert S., David Huizinga, and Suzanne S. Ageton
1985 Explaining Delinquency and Drug Use. Beverly Hills, Calif.: Sage.

Esbensen, Finn-Aage
2001 The National Evaluation of the Gang Resistance Education and Training (G.R.E.A.T.) Program. In Jody Miller, Cheryl L. Maxson, and Malcolm W. Klein (eds.), The Modern Gang Reader, 2nd Edition. Los Angeles, CA: Roxbury.

Esbensen, Finn-Aage and D. Wayne Osgood
1997 Research in Brief: National Evaluation of G.R.E.A.T. Washington, D.C.: U.S. Department of Justice.
1999 Gang Resistance Education and Training (G.R.E.A.T.): Results from the national evaluation. Journal of Research in Crime and Delinquency 36:194–225.

Esbensen, Finn-Aage, Elizabeth P. Deschenes, Ronald E. Vogel, Jennifer West, Karen Arboit, and Lesley Harris
1996 Active parental consent in school-based research: An examination of ethical and methodological issues. Evaluation Review 20:737–753.

Esbensen, Finn-Aage, Adrienne Freng, Terrance J. Taylor, Dana Peterson, and D. Wayne Osgood
forthcoming 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 (eds.), Responding to Gangs: Evaluation and Research. Washington, D.C.: U.S. Department of Justice. In Press.

Esbensen, Finn-Aage, Michelle H. Miller, Terrance J. Taylor, Ni He, and Adrienne Freng
1999 Differential attrition rates and active parental consent. Evaluation Review 23:316–335.

Freng, Adrienne
2001 Parents speak out: Parent questionnaires from the Gang Resistance Education and Training (G.R.E.A.T.) Evaluation. Paper presented at the Annual Meeting of the Academy of Criminal Justice Sciences, Washington, D.C., April.

Goldstein, Harvey
1995 Multilevel Statistical Models. London: Arnold.

Gottfredson, Denise C.
2001 Schools and Delinquency. New York: Cambridge University Press.

Gottfredson, Michael R. and Travis Hirschi
1990 A General Theory of Crime. Stanford, Calif.: Stanford University Press.

Grasmick, Harold G., Charles R. Tittle, Robert J. Bursik, Jr., 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.

Hansen, William B., Nancy S. Tobler, and John W. Graham
1990 Attrition in substance abuse prevention research. Evaluation Review 14:677–685.

Hawkins, J. David and Richard F. Catalano
1993 Communities that Care: Risk-Focused Prevention Using the Social Developmental Model. Seattle, Wash.: Developmental Research and Programs, Inc.

Hawkins, J. David, Jie Guo, Karl Hill, Sara Battin-Pearson, and Robert Abbott
2000 Long Term Effects of the Seattle Social Development Intervention on School Bonding Trajectories. Seattle, Wash.: University of Washington, Social Development Research Group.

Heinsman, Donna T. and William R. Shadish
1996 Assignment methods in experimentation; when do nonrandomized experiments approximate answers from randomized experiments? Psychological Methods 1:154–69.

Hindelang, Michael J., Travis Hirschi, and Joseph G. Weis
1981 Measuring Delinquency. Beverly Hills, Calif.: Sage.

Huizinga, David
1991 Assessing violent behavior with self-reports. In Joel Milner (ed.), Neuropsychology of Aggression. Boston, Mass.: Kluwer.

Huizinga, David and Delbert S. Elliott
1986 Reassessing the reliability and validity of self-report delinquency measures. Journal of Quantitative Criminology 2:293–327.

Huizinga, David, Finn-Aage Esbensen, and Anne W. Weiher
1991 Are there multiple paths to delinquency? Journal of Criminal Law and Criminology 82:83–118.

Huizinga, David, Rolf Loeber, and Terence P. Thornberry
1994 Urban Delinquency and Substance Abuse. Washington, D.C.: U.S. Department of Justice.

Humphrey, Kim R. and Peter R. Baker
1994 The GREAT Program: Gang resistance education and training. FBI Law Enforcement Bulletin, Sept. 1–4.

Judd, Charles M. and Gary H. McClelland
1989 Data Analysis: A Model-Comparison Approach. New York: Harcourt Brace Jovanovich, Publishers.

Lynam, Donald R., Richard Milich, Rick Zimmerman, Scott P. Novak, T.HK. Logan, Catherine Martin, Carl Leukefeld, and Richard Clayton
1999 Project DARE: No effects at 10-year follow-up. Journal of Consulting and Clinical Psychology 67:1–4.

Olds, David, Charles R. Henderson, Jr., Robert Cole, John Eckenrode, Harriet Kitzman, Dennis Luckey, Lisa Pettitt, Kimberly Sidora, Pamela Morris, and Jane Powers
2001 Long-term effects of nurse home visitation on children's criminal and antisocial behavior. In Alex Piquero and Paul Mazerolle (eds.), Life-Course Criminology. Belmont, Calif.: Wadsworth.

Osgood, D. Wayne and Gail Smith
1995 Applying hierarchical linear modeling to extended longitudinal evaluations: The Boys Town Follow-Up Study. Evaluation Review 19:3–38.

Palumbo, Dennis J. and Jennifer L. Ferguson
1995 Evaluating Gang Resistance Education and Training (GREAT): Is the impact the same as that of Drug Abuse Resistance Education (DARE)? Evaluation Review 19:591–619.

Peterson, Dana
2001 G.R.E.A.T. prevention: School personnel perspectives. Paper presented at the Annual Meeting of the Academy of Criminal Justice Sciences, Washington, D.C., April.

Rosenbaum, Dennis P. and Gordon S. Hanson
1998 Assessing the effects of school-based drug education: A six-year multilevel analysis of Project D.A.R.E. Journal of Research in Crime and Delinquency 35:381–412.

Rosenbaum, Paul R. and Donald B. Rubin
1983 The central role of the propensity score in observational studies for causal effects. Biometrika 70:41–55.

Sellers, Christine S., Terrance J. Taylor, and Finn-Aage Esbensen
1998 Reality check: Evaluating a school-based gang prevention model. Evaluation Review 22:590–608.

Sherman, Lawrence W., Denise Gottfredson, Doris MacKenzie, John Eck, Peter Reuter, and Shawn Bushway
1997 Preventing Crime: What Works, What Doesn't, What's Promising. Washington, D.C.: National Institute of Justice.

Taylor, Ronald D., Ronin Casten, Susanne M. Flickinger, Debra Roberts, and Cecil D. Fulmore
1994 Explaining the school performance of African-American adolescents. Journal of Research on Adolescence 4:21–44.

Taylor, Terrance J.
2001 What makes G.R.E.A.T. officers think G.R.E.A.T.'s great? Paper presented at the Annual Meeting of the Academy of Criminal Justice Sciences, Washington, D.C., April.

Tebes, Jacob K., Davis L. Snow, and Michael W. Arthur
1992 Panel attrition and external validity in the short-term follow-up study of adolescent substance use. Evaluation Review 16:151–170.

Tremblay, Richard E., Frank Vitaro, Lucie Bertrand, Marc LeBlanc, Helene Beauchesne, Helene Boileau, and Lucille David
2001 Parent and child training to prevent early onset of delinquency: The Montreal Experimental Study. In Alex Piquero and Paul Mazerolle (eds.), Life-Course Criminology. Belmont, Calif.: Wadsworth.

Webb, Vince and Chris Marshall
1995 The relative importance of race and ethnicity on citizen attitudes to the police. American Journal of Police 14:45–66.

Winfree, L. Thomas, Jr., Teresa Vigil-Backsrom, and G. Larry Mays
1994 Social learning theory, self-reported delinquency, and youth gangs: A new twist on a general theory of crime and delinquency. Youth and Society 26:147–177.

Winfree, L. Thomas, Jr., Finn-Aage Esbensen, and D. Wayne Osgood
1996 Evaluating a school-based gang prevention program: A theoretical perspective. Evaluation Review 20:181–203.

Winfree, L. Thomas, Jr., Dana Peterson Lynskey, and James R. Maupin
1999 Developing a local police and federal law enforcement partnership: The sometime tortuous journey from a great idea to implementation of G.R.E.A.T. Criminal Justice Review 24:145–168.

Winship, Christopher and Stephen L. Morgan
1999 The estimation of causal effects from observational data. Annual Review of Sociology 25:659–707.

Finn-Aage Esbensen is the E. Desmond Lee Professor of Youth Crime and Violence in the Department of Criminology and Criminal Justice at the University of Missouri-St. Louis. He is currently involved in several research projects with a special focus on gang activity. He was Principal Investigator of the National Evaluation of the Gang Resistance Education and Training program.

D. Wayne Osgood is a Professor in the Crime, Law, and Juvenile Program of the Department of Sociology at Pennsylvania State University, University Park. His research interests include a variety of topics concerning crime, delinquency, and other deviant behaviors; prevention and intervention programs for juvenile delinquency; and statistical methods.

Terrance J. Taylor is currently a Ph.D. student in the Department of Criminal Justice at the University of Nebraska at Omaha. His research interests include criminological theory, juvenile delinquency and victimization, research methodology, and policing.

Dana Peterson is a doctoral student in Criminal Justice at the University of Nebraska at Omaha. Her present research interests include gangs and youth violence, families and delinquency, and sex and gender issues in crime.

Adrienne Freng is an Assistant Professor in the Department of Criminal Justice at the University of Wyoming. Her current research interests include: race and crime issues, specifically dealing with Native Americans, gangs, and juvenile delinquency. Currently, she is completing research examining the role of race in gang membership.

APPENDIX I: SUMMARY OF SCALES

Unless otherwise indicated, these measures were adopted from the National Youth Survey (Elliott et al., 1985) or the Denver Youth Survey (Huizinga et al., 1991).

1. School Environment: nine items measuring safety in the schools, e.g., "There are gang fights at my school."
2. Parental Monitoring: four items measuring communication with parents about activities, e.g., "My parents know who I am with if I am not at home."
3. Maternal Attachment: six semantic differential items tapping emotional attachment to the mother or mother-figure.
4. Paternal Attachment: six semantic differential items tapping emotional attachment to the father or father-figure.
5. Self-control (Grasmick et al., 1993): eight items about impulsive and risk-taking behavior, e.g., "Sometimes I will take a risk just for the fun of it."
6. Ethnic Identity (Taylor et al., 1994): four items measuring students' ethnic identity, e.g., "I feel good about my cultural or ethnic background."
7. Attitudes to Police (Webb and Marshall, 1995): seven items such as "Police officers are honest."
8. Peer Involvement: six items about the extent to which respondent spends time with the peer group.
9. Commitment to Negative Peers: three questions such as "If your friends were getting you in trouble at home, how likely is it that you would still hang out with them?"
10. Commitment to Positive Peers: two questions such as "If your friends told you not to do something because it was against the law, how likely is it that you would listen to them?"
11. Limited Opportunity: four items measuring perceived limited educational opportunities. A representative question is "You'll never have enough money to go to college."
12. Neutralization: nine items tapping the respondent's belief that it is okay to engage in some deviant behaviors if extenuating factors are present. For instance, "It's okay to tell a small lie if it doesn't hurt anyone."
13. Guilt: 16 questions asking how guilty the youths would feel if they did such things as "hit someone with the idea of hurting them" or "used alcohol."
14. Self-Concept: a six-item scale consisting of statements such as "I am a useful person to have around."

15. School Commitment: seven items tapping the youth's desire to succeed in school, e.g., "I try hard in school."
16. Positive Peer Behavior: eight items about the kinds of prosocial things in which friends have been involved.
17. Negative Peer Behavior: 16 items about illegal activities in which the friends have been involved.
18. Positive Reinforcements for Gang Membership (Winfree et al., 1994): identification of six possible benefits that may accrue to gang members (e.g., protection, money, excitement).
19. Negative Punishers for Gang Membership (Winfree et al., 1994): seven items tapping negative consequences of gang affiliation (e.g., trouble with police, guilt, getting hurt).

APPENDIX II: SUMMARY OF SELF-REPORT MEASURES

The self-report questions were preceded by the following introduction: "Studies have found that everyone breaks the rules and laws sometimes. Please indicate how many times in the past____ months you have done each thing. If you have not done these things, enter '0'." (Respondents were also asked an "ever prevalence" question in the pretest and post-test questionnaires. In order to prevent overlap in the reporting period, the number of months included in the recall period were of varying lengths—12 months for the pretest, 3 months for the post-test, and 6 months for the annual follow-up surveys.)

Self-Reported Delinquency:

1. Skipped classes without an excuse.
2. Lied about your age to get into some place or to buy something.
3. Avoided paying for things such as movies, bus or subway rides.
4. Purposely damaged or destroyed property that did not belong to you.
5. Carried a hidden weapon for protection.
6. Illegally spray painted a wall or a building.
7. Stolen or tried to steal something worth less than $50.
8. Stolen or tried to steal something worth more than $50.
9. Gone into or tried to go into a building to steal something.
10. Stolen or tried to steal a motor vehicle.
11. Hit someone with the idea of hurting them.
12. Attacked someone with a weapon.
13. Used a weapon or force to get money or things from people.
14. Been involved in gang fights.
15. Shot at someone because you were told to by someone else.
16. Sold marijuana.
17. Sold other illegal drugs such as heroin, cocaine, crack, or LSD.

Alcohol and Other Drug Use:

1. Tobacco products.
2. Alcohol.
3. Marijuana.
4. Paint, glue or other things you inhale to get high.
5. Other illegal drugs.

Victimization:

1. Been hit by someone trying to hurt you.
2. Had someone use a weapon or force to get money or things from you.
3. Been attacked by someone with a weapon or by someone trying to seriously hurt or kill you.
4. Had some of your things stolen from you.